The John von Neumann Institute for Computing (NIC) was established in 1998 by Forschungszentrum Jülich and Deutsches Elektronen-Synchrotron DESY to support the supercomputer-oriented simulation sciences. In 2006, GSI Helmholtzzentrum für Schwerionenforschung joined NIC as a contract partner.

The core task of NIC is the peer-reviewed allocation of supercomputing resources to computational science projects in Germany and Europe. The NIC partners also support supercomputer-aided research in science and engineering through a three-way strategy:

- Provision of supercomputing resources for projects in science, research, and industry.
- Supercomputer-oriented research and development by research groups in selected fields of physics and natural sciences.
- Education and training in all areas of supercomputing by symposia, workshops, summer schools, seminars, courses, and guest programmes for scientists and students.

The NIC Symposium is held biennially to give an overview on activities and results obtained by the NIC projects in the last two years. The contributions for this seventh NIC Symposium are from projects that have been supported by the supercomputers JUROPA, JUGENE, and JUQUEEN in Jülich. They cover selected topics in the fields of Astrophysics, Biophysics, Chemistry, Elementary Particle Physics, Materials Science, Condensed Matter, Computational Soft Matter Science, Earth and Environmental Research, Computer Science, Fluid Mechanics, and Plasma Physics.
Bibliographic information published by the Deutsche Nationalbibliothek. The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Publisher and Forschungszentrum Jülich GmbH
Distributor: Zentralbibliothek
52425 Jülich
Phone +49 (0) 24 61 61-53 68 · Fax +49 (0) 24 61 61-61 03
e-mail: zb-publikation@fz-juelich.de
Internet: http://www.fz-juelich.de/zb

Cover Design: Jülich Supercomputing Centre, Forschungszentrum Jülich GmbH

Printer: Grafische Medien, Forschungszentrum Jülich GmbH,
Porschen & Bergsch Mediendienstleistungen GbR (Cover)

Copyright: Forschungszentrum Jülich 2014

We thank A. Kurzmann (Karlsruher Institut für Technologie) and G. Gassner (Universität Stuttgart) for the images used in the cover design. For the image of the ship see http://www.pgs.com

Publication Series of the John von Neumann Institute for Computing (NIC)
NIC Series Volume 47

ISBN 978-3-89336-933-1

The complete volume is freely available on the Internet on the Jülicher Open Access Server (JUWEL) at http://www.fz-juelich.de/zb/juwel

Neither this book nor any part of it may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage and retrieval system, without permission in writing from the publisher.
Preface

Kurt Binder
Institut für Physik, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany
E-mail: kurt.binder@uni-mainz.de

Gernot Münster
Institut für Theoretische Physik, Universität Münster, 48149 Münster, Germany
E-mail: munsteg@uni-muenster.de

Manfred Kremer
Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: m.kremer@fz-juelich.de

As part of a well-established and very useful tradition, the John von Neumann Institute for Computing (NIC) has been holding NIC Symposia biennially since 2001 in order to present highlights of the research that is only possible by using extensively the resources provided by the NIC and the Jülich Supercomputing Centre (JSC). In June 2013, the massively parallel machine JUQUEEN of the JSC, which contains almost half a million processors, reached the status of “No. 1 in Europe” on the “TOP 500” list of supercomputers. However, the main achievement of the JSC is not just an excellent score in this type of ranking, but the fact that the excellent hardware performance actually does enable really outstanding research of the highest quality in research fields encompassing astrophysics, computational biology and biophysics, chemistry, elementary particle physics, materials science, condensed matter, computational soft matter science, earth and environment, computer science and numerical mathematics, fluid mechanics and plasma physics.

There are several ingredients which all play a pivotal role to continuously maintain the top standard of supercomputer based research at the NIC/JSC. In particular, the user-friendly and very helpful staff of the JSC provides fast and easy access to the machines and creates ample opportunities for well-targeted training. One must keep in mind that many of the computational codes (which often are based on code developments extending over decades and hence sometimes are very complex and sophisticated) are not immediately suited for a massively parallel computer architecture. In some cases the training and advice provided by the JSC staff allows an adaptation of these codes. In many cases the scientific problem is intrinsically in conflict with a massively parallel execution of the code, however: for these problems it is of crucial importance that the JSC has a dual hardware strategy, providing with the system JUROPA (which has a cluster architecture with faster, but less energy efficient, processors and only moderate parallelism) a “general purpose supercomputer”. This indeed has served the needs of about 90% of all the (several hundred!) research projects, that were carried out at the JSC in the two years since the last NIC symposium.
The fact that the access to supercomputer resources at the JSC is very valuable and urgently
needed by many researchers (who come from all over Germany, and via collaborations
also from other countries, mostly in Europe) can already be seen from the observation that
at the calls for proposals there is a demand which for JUROPA means an overbooking
by typically a factor of four; also JUQUEEN is overbooked, though by a factor of (on
average) about two only. So in almost all cases the allocations of resources are less than
the demands, and for some projects no resources can be given at all.

These decisions are made by a well-developed peer review process. There are two calls
for proposals of projects (of a duration of one year, which can be renewed) every year,
with granting periods starting at May 1st and Nov. 1st, respectively. The closing date
for the call is two months earlier in each case. The NIC/JSC has a huge data base on
expert referees, both from Germany and from abroad, which are selected according to
their proper expertise, and asked to send anonymous reports. About two weeks before the
granting date the NIC peer review board meets at Jülich, and reviews the outcome of the
refereeing process, adjusting the allocation of computing time in order to ensure both that
no resources would be wasted and that projects are supported according to their scientific
excellence. This well-established peer review procedure meeting the highest standards is
also a key factor for the longstanding success of the NIC.

Very similar procedures to deal with the allocation of computer resources have meanwhile
been implemented also by the two other German national supercomputing centres at
Garching (LRZ) and at Stuttgart (HLRS), as well as by the supercomputers at France, Italy
and Spain that have joined with the German supercomputing centres in the framework
“Partnership for Advanced Computing in Europe” (PRACE). PRACE is open for scientists
from most European Countries, and hence PRACE has played a key role in developing
computational science in countries that cannot afford to run their own supercomputing
centres. Since about two thirds of the available resources in PRACE are provided by
the German centres (JSC, HLRS and LRZ), Germany has taken the lead to disseminate
the know-how in the use of supercomputers throughout Europe, thus helping to make
Europe worldwide competitive in this key technology. In fact, in the last two periods
about as many “cycles” from JUGENE/JUQUEEN were given to PRACE as were
available for distribution through the NIC. By providing these resources in the past,
the JSC has already fulfilled basically all its duties from the PRACE contract, and it is
hoped that the JSC in the future can provide a larger share of its resources to the NIC again.

It is also a pleasure to mention that in the period covered by these proceedings the
collaboration between the JSC, the HLRS and the LRZ in the framework of the Gauss
Centre for Supercomputing (GCS) has been intensified. A steering committee (“GCS-
Lenkungsausschuss”) has been formed, which jointly advises these three supercomputing
centres, and organises a detailed comparative review process of the so-called “Gauss
Large Scale Projects”. Currently a project falls in this class if it demands resources of
at least 35 million core hours. It is clear that such prestigious large scale projects are
extremely costly, and hence the review process is even more stringent than for the “normal
projects”. The steering committee ensures that comparable standards are applied at all
three centres, and also takes care that each project uses the best suited hardware platform
(i.e., occasionally projects are transferred from one centre to another one). The NIC peer review board, combined with its counterparts from the HLRS and LRZ form a GCS review panel, which makes the decisions about the allocation of computing time to the Gauss Large Scale Projects. These projects come from all fields of science, from astrophysics to biophysics to engineering.

Apart from collaborating in the framework of GCS and PRACE with other supercomputer centres, the John von Neumann Institute (NIC) also carries out its own research on selected research fields, focused in the NIC research groups. While only one of these groups (on “materials physics”, headed by Prof. M. Müser) is located at the JSC, another group (working on simulations in elementary particle physics, within the framework of quantum chromodynamics (QCD) on the lattice, headed by Dr. R. Sommer) is located at DESY/Zeuthen, and two smaller groups (working also on lattice QCD, headed by Prof. O. Philipsen, and on relativistic transport simulations, headed by Prof. M. Bleicher) are both situated at the University of Frankfurt but partially sponsored by the GSI Darmstadt. It should be recalled that the NIC operates in the framework of a cooperation agreement between these research centres of the Helmholtz association, namely the Forschungszentrum Jülich (FZJ), the Deutsches Elektronen Synchrotron (DESY) at Hamburg, and the Gesellschaft für Schwerionenforschung (GSI) at Darmstadt. The research topics of the NIC research groups on the one hand make key contributions to the research programmes of these Helmholtz centres, and on the other hand they maintain and develop further the specific know-how on highly advanced supercomputer codes in these specific areas. The present volume contains short reports on research highlights from these groups.

The total number of research projects carried out at the NIC in the last two years is by far too large to cover all of them in a single volume. Therefore this volume contains a selection of only 41 reviews on projects that were selected, taking into account both the previously obtained referee reports on the project applications and the publication records that resulted. A complete account of the publications from all the projects can be found on the NIC website. There also particular scientific highlights are displayed in more detail, namely the “NIC excellence projects”. The NIC peer review board decides at each of its meetings, which of the many excellent projects deserves this prestigious award most (thus there can normally be only two NIC excellence projects per year). In the period covered by this book, the NIC excellence projects were

“The Small Scale Structure of the Universe” (S. Gottlöber, Potsdam)

“Numerical simulations of strongly correlated electron systems” (F. Assaad, Würzburg)

“Non-linear single particle response of glassforming systems in external fields” (J. Horbach, Düsseldorf)

“First Principles Modelling of minerals, melts and fluids at high pressures and high temperatures” (S. Jahn, Potsdam)

“Direct Numerical Simulation of turbulent mixing processes in the planetary boundary layer” (J. P. Mellado, Hamburg)

The present volume also contains reports from all these projects.
We thank all the authors who have contributed to this book for their valuable efforts, which provide a beautiful cross section through the wide range of research topics that is covered by the NIC projects. Special thanks go to the staff of the JSC for providing the high quality supercomputing infrastructure and excellent user service. This support clearly was crucial for the successful execution of the projects described here. We also thank Ms Martina Kamps, who compiled the contributions and produced this high quality book. Already here we want to thank Walter Nadler, Ms Britta Hoßfeld, and Ms Helga Offergeld for their help in organising the 7th NIC-Symposium.

Jülich, February 2014

K. Binder  G. Münster  M. Kremer
Contents

Paving the Road towards Pre-Exascale Supercomputing
D. Brömmel, U. Detert, S. Graf, T. Lippert, B. Orth, D. Pleiter, M. Stephan, E. Suarez 1

The NIC Research Groups

Relativistic Transport Simulations and the Search for the Quark-Gluon-Plasma
M. Bleicher, H. Petersen 17

Atomistic Modelling of Redox Reactions in Non-Equilibrium
W. B. Dapp, M. H. Müser 25

The Thermal Quark Hadron Transition in Lattice QCD with two Quark Flavours
O. Philipsen, M. Bach, V. Lindenstruth, C. Pinke 33

Tests of the Standard Model and Lattice Simulations
K. Jansen, H. Simma, R. Sommer 41

Astrophysics

Introduction
P. L. Biermann 51

Dynamics of Binary Neutron Star Mergers

The Plasma Environment of Comet 67P/Churyumov-Gerasimenko
C. Koenders, K.-H. Glassmeier, I. Richter 61

A Coherent Hubble Volume Simulation for All-Sky ISW Predictions and Large Scale Surveys
S. Gottlöber, J. M. Diego, W. Watson, I. Iliev, G. Yepes 69

The Circumstellar Medium of Massive Stars in Motion
J. Mackey, N. Langer, D. M.-A. Meyer, V. V. Gvaramadze, S. Mohamed, H. R. Neilson, A. Mignone 77
Computational Biology and Biophysics

Introduction
H. Grubmüller 89

Molecular Simulation-Based Structural Prediction of a Rhodamine-Labelled Biosensor
J. Dreyer, M. Brown Gonçalves, E. Ippoliti, P. Carloni 91

How Tertiary Interactions Between the L2 and L3 Loops Affect the Dynamics of the Distant Ligand Binding Site in the Guanine Sensing Riboswitch
C. A. Hanke, H. Gohlke 99

Automatic NOESY Assignment Using an Iterative CS-Rosetta Scheme
O. F. Lange 107

The Flexibility of Fibrinogen and its Initial Adsorption Stages at the Graphite and Mica Surface
S. Köhler, F. Schmid, G. Settanni 117

Chemistry

Introduction
W. Klopper 129

Force-Activated Reactivity of Disulfides or the “Strange Case of Dr. Jekyll and Mr. Hyde”
P. Dopieralski, J. Ribas-Arino, D. Marx 131

Ab initio Molecular Dynamics Simulations of Heterocyclic Chromophores in Aqueous Solutions
G. Bekçioğlu, C. Allolio, D. Sebastiani 139

Ab initio Molecular Dynamics Simulation of the Interaction of Liquid Water with Sub-Picosecond High-Intensity THz Pulses
P. Kr. Mishra, O. Vendrell, R. Santra 147
Elementary Particle Physics

Introduction
G. Münster 157

Isospin Splittings in the Light Baryon Octet from Lattice QCD + QED at the Physical Mass Point
Z. Fodor, S. Krieg, T. Lippert 159

Numerical Simulation of Supersymmetric Yang-Mills Theory
G. Bergner, I. Montvay, G. Münster, U. D. Özugurel, D. Sandbrink 169

Flavour Physics of Up, Down and Strange Quarks from Dynamical QCD × QED
G. Schierholz 177

More Results on Finite Temperature QCD with Wilson Fermions

Low-Energy Precision Physics and Lattice QCD

Materials Science

Introduction
R. O. Jones 205

Ab initio Modelling of Growth of Graphene for Silicon-Compatible Microelectronics
J. Dąbrowski, A. Fleszar, G. Lippert, G. Łupina 207

Nanometre-Scale ab initio Investigations of Functional Magnetic Materials
M. E. Gruner, P. Entel 215

First-Principle Calculations on the Phase Evolution of Si Electrodes in Li-Ion Batteries
J. Rohrer, K. Albe 223

Direct Excitation of Charge-Transfer States in the ZnPc-C_{60} Donor-Acceptor Complex
O. Brügner, M. Walter 231
Condensed Matter

Introduction
K. Binder, A. Muramatsu 241

Influence of Vibrational Modes on the Quantum Transport through a Nano-Device
F. Anders, A. Jovchev 243

Correlated Topological Insulators and Semi-Metals

Admolecule Structures at the Surface of Ice
M. Bockstedte, A. Michl, M. Kolb 259

Atomistic Modelling of Zr-Segregation in Symmetric Grain Boundaries of Molybdenum
O. Lenchuk, J. Rohrer, K. Albe 267

Computational Soft Matter Science

Introduction
K. Kremer 277

Dynamic Phase Transitions and Time Dependent Structuring Effects in Particle-Stabilised Emulsions
S. Frijters, F. Günther, J. Harting 279

Non-Linear Response of Glass-Forming Systems to External Forces
D. Winter, P. Chaudhuri, J. Horbach 287

Directing the Assembly of Block Copolymers: A Single-Chain-in-Mean-Field Simulation Study
M. Müller, D. Sun, U. Welling 295

Earth and Environment

Introduction
U. Hansen 305

3D Acoustic Full Waveform Tomography
First-Principles View on Element and Isotope Cycles in the Earth’s Interior
S. Jahn, V. Haigis, P. M. Kowalski, D. Künzel, G. Spiekermann, J. Wagner 317

Direct Numerical Simulation of Turbulent Mixing in the Planetary Boundary Layer
J. P. Mellado, C. Ansorge, A. de Lózar, J. R. Garcia, T. Keitzl, C. van Heerwaarden 325

Cutting the Edge of Regional Climate Models: Highly Resolved Climate Simulations in the Alpine Region
H. Truhetz, A. Prein, A. Csáki, A. Gobiet 333

Computer Science and Numerical Mathematics

Introduction
D. Kröner 343

Evaluation of a Multigrid Solver for 3-Level Toeplitz and Circulant Matrices on Blue Gene/Q
M. Bolten 345

Solving the Reactive Compressible Navier-Stokes Equations in a Moving Domain
R. Klöfkorn, M. Nolte 353

Scalability Tuning of the Load Balancing and Coupling Framework FD4
M. Lieber, W. E. Nagel, H. Mix 363

Fluid Mechanics

Introduction
N. A. Adams 373

Transitional and Turbulent Magnetohydrodynamic Flows in Uniform and Non-Uniform Magnetic Fields
T. Boeck, D. Krasnov, S. Tympel, O. Zikanov 375

Highly-Resolved Numerical Simulations of Bed Load Transport in a Turbulent Open Channel Flow
B. Vowinckel, T. Kempe, J. Fröhlich 383

Comparison of Navier-Stokes-Fourier Equation and Grad’s Moment Equation Solutions for Turbulence
G. Gassner, M. Torrilhon, A. Beck, S. Knechtel, T. Bolemann 391
Numerical Investigation of Reacting and Non-Reacting Flows in Gas Turbine Related Configurations
A. Fiolitakis, C. Eberle, A. Filosa, J. Prause, B. Noll

Computer Simulation of Flow Past Superhydrophobic Striped Surfaces
J. Zhou, A. V. Belyaev, E. S. Asmolov, O. I. Vinogradova, F. Schmid

Computational Plasma Physics

Introduction
K.-H. Spatschek

Laser Pulse Amplification by Raman and Brillouin Scattering towards Multi-Petawatt Level
G. Lehmann

Radiation Generation in Plasma-Based Accelerators with Controlled Electron Injection
Paving the Road towards Pre-Exascale Supercomputing

Dirk Brömmel, Ulrich Detert, Stephan Graf, Thomas Lippert, Boris Orth, Dirk Pleiter, Michael Stephan, and Estela Suarez

Institute for Advanced Simulation, Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: th.lippert@fz-juelich.de

Supercomputing at scale has become the decisive challenge for users, providers and vendors of leading supercomputer systems. On next-generation systems, approaching exascale by the end of the decade, we will be confronted with millions of cores, and the need of massive parallelism. Beyond aggregating ever larger compute performance also the ability to hold and efficiently process drastically increasing amounts of data will be key to enable future leading research facilities for computational science. We report in this article on the evolving supercomputing infrastructure at Jülich Supercomputing Centre (JSC), research and development activities on future HPC technologies and architectures as well as on the computational science research and collaboration with science areas which will require exascale supercomputing in the future.

1 Introduction

In 2005 the Jülich Supercomputing Centre (JSC) started it’s dual system strategy to most efficiently serve the application portfolio of the users of the Jülich Research Centre, the John von Neumann Institute for Computing in Germany and since mid of 2010 the Partnership for Advanced Computing in Europe (PRACE). Via the German Gauß Centre for Supercomputing in 2009 a first milestone was reached with the installation of the IBM Blue Gene/P system named JUGENE as highly scaling system (294,912 cores) and JUROPA (25,000 Intel Nehalem CPU cores) as highly flexible, general-purpose cluster system.

This dual system strategy has been carried forward end of 2012 with the installation of a new highly scaling, 28 rack IBM Blue Gene/Q system named JUQUEEN entering the TOP500 list at rank 7 world wide and as #1 in Europe (see Sec. 2). With it’s 458,752 cores more than 1.5 million hardware threads can be executed concurrently by a single application. Several applications have been proven to scale to this extent and are now members in the “JSC High-Q Club” (see Sec. 2.1). The compute system is flanked by a new GPFS storage system providing for the first time full end-to-end data integrity and a maximum I/O bandwidth of 200 GByte/s. To enable future architectures where non-volatile storage devices are integrated into the supercomputer to provide even higher bandwidth and significantly higher access rates, JSC collaborated with IBM on an active storage subsystem attached to JUQUEEN (see Sec. 3).

To prepare the next step of replacing the general-purpose system JUROPA by a system approach of 2 PFlop/s peak performance, a new test cluster called JUROPA-3 has been installed. The work on future architectures where such clusters are coupled to a booster comprising tightly coupled many-core devices is continued in the new DEEP-ER project (see Sec. 5) which extends the ongoing Dynamical Exascale Entry Platform Extended Reach (DEEP) project. A new exascale lab, the NVIDIA Application Lab, focusses on another type of many-core architectures, namely GPUs (see Sec. 6).
Figure 1. Timeline of the transition from JUGENE to JUQUEEN in 2012/13.

In October 2013 the Human Brain Project (HBP) – one of two large-scale initiatives selected out of six candidates to receive significant funding through the EU’s new Future and Emerging Technologies (FET) Flagship Programme – started a 2.5-year ramp-up phase. The task of the JSC in this project will be to develop, deploy and operate the main HBP supercomputer in Jülich. Sec. 7 gives an overview of the project structure and its goals.

2 JUQUEEN

When IBM first announced Blue Gene/Q as the third generation of their Blue Gene supercomputing family in 2011, JSC decided to enter a collaboration with IBM which resulted in the installation of a prototype system. This allowed exploring the new architecture including its scalability before the development of the system was completed. The first racks of a Blue Gene/Q production system arrived at Jülich in April 2012. Only a few weeks later the new system, called JUQUEEN, consisting of 8 racks could be made accessible to users for early production runs delivering a peak performance of 1.6 PFlop/s, 60% more than JUGENE. The first 8 racks were fulfilled through the Helmholtz programme “Supercomputing”.

JUGENE was an at that time still operational 72 racks Blue Gene/P system with 1 PFlop/s peak performance. The system had been very reliable and stable over several years such that a user job utilisation of over 90% could be achieved. The system was nevertheless shut down at the end of July 2012 and dismounted to free space for the final JUQUEEN system consisting of 28 racks. To minimise the impact on the users and to limit the downtime to a minimum, the transition (see Fig. 1) took place in several steps, starting with 8 racks, then switching to a 16 rack system, 24 racks, and finally 28 racks in February 2013. After the system went into production the job utilisation immediately reached 70% and is now at a level >90%. The additional 20 racks were fulfilled through the project.
“Peta GCS” of the German Gauß Centre for Supercomputing.
This translates into a more than five-fold increase of compute performance on Blue Gene available for scientific computing in Jülich, Germany and Europe. A large fraction of it is now available for users of the Gauss Centre for Supercomputing and PRACE.

2.1 The High-Q Club

As scaling up processing pipeline performance by increasing the clock frequency has reached its limits, performance boosts can only be achieved by scaling-out and increasing parallelism at different levels. From the Blue Gene/P to Blue Gene/Q the width of the SIMD vector instruction pipelines doubled and the number of threads per core as well as the number of cores per node increased by a factor 4.

To help our users in migrating their codes and leveraging the performance on the new architecture, we organised a first porting and tuning workshop in Spring 2013. Following this workshop and to promote the idea of exascale capability computing, we have established the High-Q Club, a showcase for codes able to utilise the entire 28-rack Blue Gene/Q machine at JSC. The club members comprise a collection of the highest scaling codes on JUQUEEN, through which we intend to encourage other developers to invest in tuning and scaling their codes. We want our users to show that they are capable of using all 458,752 cores, and for example more than 1 million concurrent threads on JUQUEEN.

The diversity of members of the High-Q Club establishes that it is possible to scale real applications to the complete JUQUEEN using a variety of programming languages and parallelisation models, demonstrating individual approaches to reach that goal. High-Q status thus marks an important milestone in application development towards future HPC systems that envisage even higher core counts.

To qualify for membership, developers should submit evidence of the scalability of their codes across all available cores. While we currently do not set a strict minimum efficiency, we do expect the codes to profit from additional cores with an increase in speed. The benchmark used should also be as close as possible to a production scenario: trivial kernels or libraries will not be accepted.

At the time of writing this article, the members of the High-Q Club were:

dynQCD
dynQCD is a code for simulations in the field of Lattice Quantum Chromodynamics and can be used for different fermion actions. The code is developed at the University of Wuppertal and the Simulation Laboratory for Nuclear and Particle Physics at JSC and is written in C. The code features hybrid-parallelisation using POSIX threads. Inter-process communication on Blue Gene/Q is done via the low-level System’s Programming Interface (SPI) by-passing higher-level communication libraries like MPI.

Gysela
Gysela is a GYrokineSic SEmi-LAgrangian code for plasma turbulence simulations developed at CEA Cadarache. It is, e.g., used in simulations of the electrostatic branch of the ion temperature gradient turbulence in tokamak plasmas. Gysela is written in Fortran90 and C and uses MPI, OpenMP and POSIX threads for parallelisation.
JuSPIC
JuSPIC\(^4\) is the Jülich Scalable Particle-in-Cell code for fully relativistic plasma simulations in electromagnetic fields. The non-linear interaction between the fields and the plasma is described by the relativistic Vlasov equation and Maxwell’s equations. JuSPIC is developed by the Simulation Laboratory for Plasma Physics at JSC in Fortran using MPI and OpenMP. Since it is also used to test new architectures and programming models, there also is a version that uses the StarSs model.

MP2C
The Massively Parallel Multi-Particle Collision Dynamics code (MP2C)\(^5\) implements a hybrid representation of solvated particles in a fluid to simulate soft matter physics and mesoscopic hydrodynamics. It is developed by the Simulation Laboratory for Molecular Systems at JSC and is written in Fortran. The parallelisation is based on MPI while scalable I/O on the full JUQUEEN is achieved by using SIONlib\(^6\).

µϕ (muPhi)
µϕ (muPhi)\(^7\) combines two packages to model and simulate water flow and solute transport in porous media. It can be used for the prediction and control of groundwater production, the assessment of water contamination and becomes more and more important for flood and climate prediction. Among other parts, the code make use the iterative solver template library (ISTL) developed at the University of Heidelberg within framework of the DUNE project. µϕ is written in C++ using MPI for parallelisation. Like MP2C it relies on SIONlib for scalable parallel I/O.

NEST
NEST\(^8\) is a simulator for spiking neural network models that focus on the dynamics, size and structure of neural systems rather than on the exact morphology of individual neurons. It includes over 25 neuron and 10 synapse models and uses a hybrid parallelisation scheme to perform the computations of the neuron and synapse dynamics. The code is a development by the NEST initiative in C++, combining MPI and OpenMP.

PEPC
The Pretty Efficient Parallel Coulomb\(^9\) solver is used for N-body simulations and was developed within the Simulation Laboratory for Plasma Physics at JSC. PEPC is not restricted to a specific force law or physical problem and can thus be applied to different problems, e.g., beam-plasma interaction, vortex dynamics, gravitational interaction or molecular dynamics simulations. The code is written in Fortran 2003 and C and is parallelised using MPI as well as OpenMP or POSIX threads.

PMG+PFASST
PMG+PFASST\(^10\) combines a parallel multigrid solver with a time parallel approximation scheme to solve ODEs with linear stiff terms. The two parts have been developed at the Lawrence Berkeley National Lab (PFASST) and the University of Wuppertal (PMG) and have been coupled to one application by developers from the cross-sectional team Mathematical Methods and Algorithms at JSC and Universita della Svizzera Italiana.
PMG+PFASST is written in Fortran 2003 and C. Parallelisation is implemented using MPI and POSIX threads.

**Terra-Neo**

Terra-Neo\textsuperscript{11} is used for modelling earth mantle dynamics. The development team is built from members of Ludwig-Maximilians-Universität München, Universität Erlangen-Nürnberg, Regionales Rechenzentrum Erlangen and Technische Universität München. The interdisciplinary team involves geophysics and algorithmic experts as well as software and hardware experts to enable the high performance and scalability on future HPC architectures. The Terra-Neo framework is implemented in C++ and Fortran using MPI and OpenMP for parallelisation.

**waLBerla**

waLBerla\textsuperscript{12} (widely applicable Lattice Boltzmann solver from Erlangen) is a computational fluid dynamics application. Originally, the waLBerla framework has been centred around the Lattice Boltzmann method for the simulation of fluid scenarios but in the meantime evolved to a code that is also suitable for a wide range of applications based on structured grids. It is developed in C++ and uses MPI and OpenMP. Accelerator devices are supported using CUDA and OpenCL.

2.2 **GPFS Storage Cluster – JUST**

For the storage back-end it is hard to keep pace with the performance increase of the compute systems. In parallel to the upgrade of the Blue Gene/P system JUGENE to the Blue Gene/Q system JUQUEEN the storage cluster JUST3 has been replaced by a new storage system JUST4. The main goal was to achieve a significant bandwidth improvement from about 60 GByte/s to approximately 200 GByte/s. This could only be achieved by a significant increase of the number of disks which was not straightforward. With both the number of disks and the capacity per disk increasing standard RAID technology would not have provided sufficient protection against failures resulting in data loss. With the old storage systems we observed a disk failure rate of 2-3 disks per week and rebuild times in the order of 12 hours during which performance could degrade significantly. For this reason, JSC was the first to install IBM’s GPFS Storage Servers (GSS).

Instead of RAID controllers GSS uses GPFS Native RAID (GNR), which is a software implementation of storage RAID technologies within GPFS. One key feature of GNR is the Declustered RAID concept where user data is stored in strips which are grouped in RAID arrays. The strips are distributed over multiple disks such that even if multiple disks breaks no data loss occurs. GNR supports 2- and 3-fault-tolerant Reed-Solomon codes and 3-way and 4-way replication. Whenever any of the RAID arrays becomes critical, i.e. when no further disk failure could be tolerated, GPFS will maximise rebuild priority accepting performance degradations. Once this critical state is left, rebuild priorities are lowered to improve performance. Tests have shown that the critical phase of the rebuild lasts only for a couple of minutes in contrast to 12 hours for classical RAID controllers.

Another key feature of GNR is the support of end-to-end checksums. These checksums are created and verified at the client which is writing and reading the data, respectively, and sent over the network together with the data. This significantly reduces the risk of silent
errors, i.e. errors which are not detected by the storage system itself.

Since September 2013 the GSS storage cluster JUST4 is in production and used to host two file systems, one scratch file system with 3.2 PBytes and one data file system with 1.9 PBytes dedicated to special, data intense projects.

3  Blue Gene Active Storage

Typical HPC architectures continue to move away from what is called Amdahl’s rule for a balanced I/O performance: one bit of I/O per second for each instruction per second. This growing performance gap is becoming more critical as there is a growing number of applications processing big data volumes. Higher performance in terms of bandwidth could be achieved by scaling-out to an even larger number of spinning disks. This strategy is not only limited by cost but also by power and would not be affordable in an exascale context. Current projections aim for an I/O bandwidth of up to 60 TBytes/s

There are several opportunities which will allow to mitigate and overcome this problem. Firstly, new non-volatile memory technologies while limited in capacity (within an affordable budget) allow to realise much higher bandwidth plus dramatically higher access rates. The latter is an important feature for a large set of data-intensive applications. The advantages of high-capacity spinning disk storage systems and high-performance non-volatile memory technologies can be combined in a hierarchical storage architecture as it was explored by a PRACE prototype at JSC. Secondly, active storage concepts provide an opportunity to reduce performance and energy costs of data movement. Active storage is an architectural concept where processing capabilities and storage are integrated.

Such a concept has been realised for JUQUEEN by Blue Gene Active Storage (BGAS). A BGAS node comprises a standard Blue Gene/Q I/O node as well as a PCIe card which integrates 2 TBytes of storage plus network ports towards external, large-capacity storage systems. Each of these nodes is connected by 2 network links to the Blue Gene/Q compute nodes, 6 network links to the neighbouring BGAS nodes within a 3-dimensional torus network and 2 10-GbE links to the external large capacity storage cluster JUST. The high bi-sectional bandwidth provided by this torus network enables fast movement of the data within the active storage system.

The various ways on how applications can efficiently use such a system are still to be explored. The architecture can, e.g., be used for post-processing data generated during large-scale simulations. In this case, the high bandwidth is exploited to write out huge amounts of data which is post-processed within the active storage system. Only a resulting, significantly reduced set of data is finally written to the external storage. Another option is to use such a system for multi-pass analysis, where massively parallel applications perform a large number of random read accesses to data sets loaded into the active storage system. Yet another option is the use of the fast storage to temporarily hold data which does not fit into the compute nodes main memory.

The BGAS system attached to JUQUEEN is the result of a cooperation with IBM in the framework of the Exascale Innovation Centre (EIC).
JUROPA-3

In order to prepare for a future replacement of the JUROPA HPC production system, a prototype system called JUROPA-3 has been installed at JSC in April 2013. JUROPA-3 is the outcome of a cooperation of JSC, ParTec Cluster Competence Center and T-Platforms, aiming at the development of solutions for fundamental questions in large-scale cluster computing. Topics like application check-point/restart, end-to-end data integrity, network topology, failure prediction and energy efficiency are addressed in this cooperation.

JUROPA-3 comprises 60 compute nodes, each equipped with 2 Intel Xeon E5-2650 CPUs (Sandy Bridge-EP) providing a total of 960 processor cores with a peak performance of 15.3 TFlops. In addition, 4 compute nodes are each enhanced with 2 NVIDIA Tesla K20X GPUs and another 4 nodes with 2 Intel Xeon Phi 5110P co-processors each. The co-processors amount to an extra performance of 18.5 TFlops peak. The compute nodes, the server nodes and the storage components of the cluster are connected by a 56 Gbit/s Infiniband network (FDR) with fat-tree topology providing high communication bandwidth for parallel applications and I/O (see Fig. 2).

JUROPA-3 uses the same hierarchical system structure with master nodes, administrative nodes and compute nodes like the current JUROPA production cluster. Several modifications and enhancements in JUROPA-3, however, are subject to further development and testing:

- Scientific Linux is used as the basic operating system combined with ParTec’s ParaStation software for cluster management.
- The majority of compute nodes run in diskless operation mode.
- 8 compute nodes include additional local disks for checkpoint/restart development.
• 16 fat nodes possess an increased amount of main memory (256 GB and 128 GB, respectively) which allow for production runs of structural mechanics applications.

• A small storage area (2 server nodes, 60 TB disk space) is dedicated to the development and testing of future Lustre parallel file system versions and end-to-end data integrity enhancements.

• A GPFS gateway node with Infiniband HCA on one side and 4x10 GbE on the other side provides for connectivity between JUROPA-3 and JSC’s file server JUST4. GPFS is mounted in addition to Lustre on all cluster nodes and serves as the main file system for home and scratch data.

• The SLURM resource manager will be integrated with ParaStation and be used for job management and control.

5 DEEP – Extended Reach (DEEP-ER)

Multiple challenges have to be overcome to make exascale computing possible by the end of the decade. Simply scaling up today’s HPC concepts and technology will not be sufficient: the overall power consumption must be drastically reduced; applications will have to be modified to scale up and extract performance from systems with millions of cores; resiliency methods must be developed to deal with the reduced mean time to failure (MTTF); and additional layers in the memory hierarchy are needed to reduce the increasing gap between the growing compute performance and the limited bandwidth of both memory and storage.

Two of the above mentioned exascale challenges (concurrency and power consumption) are addressed in the DEEP project\textsuperscript{16,17}. The novel DEEP architecture combines the high scalability of dedicated massively parallel systems with the ubiquity and cost effectiveness of commercial off-the-shelf HPC clusters. Its straightforward, scalable programming model allows applications to run their code parts with different scalability characteristics on the part of the system best suited to them. Furthermore, DEEP’s direct warm water-cooling concept and the use of many-core processors reduces drastically the overall energy consumption.

The DEEP architecture will be substantially extended with resiliency, memory hierarchies and I/O functionalities in the new project "DEEP – Extended Reach" (DEEP-ER)\textsuperscript{18}. The improvement in performance and power efficiency when using emerging memory technologies such as non-volatile memory (NVM) and memory attached to the network (NAM) will be explored. A prototype will be built and scientific applications will be ported to demonstrate the achieved improvements in scalability, parallel I/O efficiency, and system reliability.

In the DEEP-ER prototype (see Fig. 3) the computing power at the Booster Nodes (BN) will be provided by the second generation of Intel Xeon Phi. At the BNs, NVM devices represent one further level of memory hierarchy additional to the existent processor’s main memory. NVM allows for much larger memory sizes than DRAM and can be globally accessible from any part of the DEEP-ER prototype. A uniform high-speed interconnect will run across Cluster and Booster. NAM nodes attached to it will provide access to a
large shared memory pool for all Cluster and Booster nodes. The memory will be part of the node address spaces and support fine-grain access.

The simplification of the hardware architecture and the newly introduced components open the door for new functionality in the software environment, in particular regarding I/O and resiliency (see Fig. 4 and 5).

5.1 I/O

For highly efficient access to the I/O devices, the Fraunhofer parallel file system (FhGFS)\textsuperscript{19} will be extended and optimised, to allow for system-wide access to fast storage class mem-

---

**Figure 3.** Hardware architecture of the DEEP-ER prototype.

**Figure 4.** DEEP-ER I/O architecture.

**Figure 5.** DEEP-ER Resiliency architecture.
ory devices, such as NVM or NAM. FhGFS will use these devices to build independent units of cache groups for subsets of nodes, allowing the nodes in one group to perform certain I/O operations without being influenced by ongoing I/O operations in other groups.

Additionally, efficient access to the storage subsystem at different levels of abstraction will be provided by APIs originating from FhGFS itself, the parallel I/O library SIONlib\textsuperscript{6}, and the software developed by Exascale I/O Workgroup EIW\textsuperscript{20}, respectively. DEEP-ER will investigate their respective performance in comparative benchmarks.

Finally, to provide application developers convenient means to identify potential bottlenecks in the application I/O path, new storage management extensions will be integrated with the corresponding cluster management tools. New monitoring extensions (such as detailed live statistics or profiling capabilities) will be also added to the system. In this way the full potential of the DEEP-ER concept will be leveraged to match the storage access performance with the available compute power.

5.2 Resiliency

Employing the capabilities provided by the described I/O architecture and the characteristics of DEEP’s programming model, a dual-approach resiliency concept (see Fig. 5) will be developed. It combines an application-based multi-level checkpoint/restart mechanism with a less intrusive scheme for task-controlled recovering from component failures.

The first approach will provide a common framework to store, identify, validate and reload checkpoints from multiple levels of the memory hierarchy. The performance of local NVM memory will be exploited for frequent (cheap/local) checkpoints and NAM and permanent parallel I/O storage for less frequent, (expensive/global) ones. A common resiliency abstraction layer will implement all common mechanisms required to write, locate, test and read checkpoints to the multi-levels of the checkpoint/restart subsystem, allowing for a transparent use of its functionality.

The second approach is based on OmpSs\textsuperscript{21}. The OmpSs task-based programming model decomposes an application into stateless tasks arranged in a directed acyclic graph (DAG) representing the dependencies between these tasks. The OmpSs runtime system can detect a single task or a set of tasks failing because of a system fault, and it can then transparently re-execute these tasks from their respective beginnings. OmpSs’ concept of stateless tasks was already extended in DEEP by identifying highly scalable code parts to be offloaded to the Booster as tasks, too. These coarse-grain tasks potentially run on many processors for a considerable amount of time. Re-executing them from the beginning for fault recovery would be a waste of resources.

Therefore, DEEP-ER combines the resiliency enhancements achieved by restarting single tasks with the possibility to write checkpoints within such parallel tasks. In case a coarse-grain task is being restarted by the task-controlled mechanism, checkpoints stored in DEEP-ER’s multi-level system can be loaded to recover the most recent task state. Additional OmpSs annotations can be used by the application developer to identify the data required to be written at the checkpoint in a straightforward way. Furthermore, OmpSs might be used to enable a new class of checkpoints on a runtime-system-level, using the DAG to identify a synchronised state of the overall application.

Systems using the DEEP-ER developments will be able to run more applications increasing scientific throughput, and the loss of computational work through system failures
will be substantially reduced. Within the project, seven grand-challenge HPC applications will be optimised demonstrating the usability, performance and resiliency of the DEEP-ER Prototype. They will set the scale for the checkpoint/restart and I/O capabilities developed within the project. The DEEP-ER project started in October 2013 and will last three years. It is led by the Jülich Supercomputing Centre and brings together a total of 14 partners coming from both research and industry, including four PRACE supercomputing centres.

6 NVIDIA Application Lab at Jülich

Exploiting the performance of thousands of simple cores running at a low clock speed has the potential to significantly improve energy efficiency. This is, e.g., the case for applications where dense matrix operations dominate, like in the Linpack benchmark. This helped GPU accelerated systems to reach top positions in the Green500 list\textsuperscript{22}, the list of the most power-efficient HPC systems. To enable more applications to use GPUs is one of the key goals of the NVIDIA Application Lab in Jülich. This lab has been established in summer 2012. Since then it has built-up a broad application portfolio encompassing computational neuroscience, high-energy physics, radio astronomy, data analytics and others.

Another focus of the Lab is the parallelisation of applications on multiple GPUs. Not only aggregation of more compute performance but also the need for more memory capacity are reasons for using multiple GPUs. For the application developers this means that beyond the significant amount of parallelism at device level, an additional level of parallelism needs to be managed. For an application of the Jülich Institute for Neuroscience and Medicine (INM-1) called JuBrain, several parallelisation strategies have been investigated and tested\textsuperscript{23}. As part of an attempt to assemble a realistic, 3-dimensional model of the human brain, images of brain cuts have to be mapped onto each other. This image registration task involves repeated executions of a compute intensive kernel to calculate the mutual information metric. While low-resolution images can be processed using a single GPU, the available device memory is too small to hold high-resolution images.

JuBrain is only one of several data-intensive applications explored within the Lab. It could also be shown that density cluster analysis can be performed very efficiently on recent generations of GPUs\textsuperscript{24}. Cluster analysis aims on the identification of regions of similar objects in a multi-dimensional data set. It is a standard method of data analytics which can, e.g., be applied to protein folding simulation data. A significant speed-up of this analysis enables interactive processing of large data sets.

7 Human Brain Project

In January 2013, the European Commission selected the Human Brain Project (HBP) as one of two large-scale initiatives out of six candidates to receive significant funding through the EU’s new Future and Emerging Technologies (FET) Flagship Programme, starting October 2013. The 2.5-year ramp-up phase of the project (until March 2016), which is funded by the EU’s 7th Framework Programme, will be followed by a – partially overlapping – operational phase under the upcoming next framework programme, Horizon 2020. Federating more than 80 European and international research institutions (with more partners joining the consortium later on through a Competitive Call Programme) under the lead of
Henry Markram from the Swiss Ecole Polytechnique Fédérale de Lausanne (EPFL), the HBP as a whole is planned to last ten years and estimated to cost one billion Euros.

The goal of the HBP is to collect all existing knowledge about the human brain and to reconstruct the brain, piece by piece, in multi-scale models and supercomputer-based simulations of these models. The resulting “virtual human brain” offers the prospect of a fundamentally new understanding of the brain and its diseases and of novel, brain-like computing technologies.

To reach this goal, the HBP will build a European research infrastructure consisting of six ICT (Information & Communication Technology) Platforms, dedicated respectively to Neuroinformatics, Medical Informatics, Brain Simulation, Neuromorphic Computing, Neurorobotics, and High-Performance Computing. Together, these platforms will make it possible to federate neuroscience data from around the world, to integrate the data in unifying models and simulations of the brain, to validate the results against empirical data, and to make them available to the scientific community. The resulting knowledge on the structure and connectivity of the brain will open up new perspectives for the development of “neuromorphic” computing systems incorporating unique characteristics of the brain such as energy-efficiency, fault-tolerance and the ability to learn. The HBP’s models and simulations will enable neuroscientists to carry out in silico experiments on the virtual human brain that cannot be done in vivo for practical or ethical reasons.

Jülich plays a key role in the HBP as it contributes to the project a unique combination of expertise and infrastructure in the two relevant fields neuroscience and supercomputing. The increasingly close cooperation of the two research areas in Jülich becomes manifest in the joint activities of the Institute of Neuroscience and Medicine and the JSC within the Helmholtz Portfolio Theme Supercomputing and Modelling for the Human Brain and, in particular, within the new Simulation Lab Neuroscience. In fact, in the HBP, Jülich leads important subprojects in both neuroscience (“Strategic Human Brain Data”) and supercomputing (“The High-Performance Computing Platform”). In addition, Jülich researchers are work package or task leaders in other subprojects (Brain Simulation, Neuroinformatics, Theory).

The task of the HBP’s HPC Platform subproject is to build the supercomputing and data hard- and software infrastructure required to run cellular brain model simulations of the size of a full human brain, and to make this infrastructure available to the consortium and the wider community. Central element of the HPC Platform is the HBP Supercomputer, the project’s main production system, which will be built in stages to arrive at the exascale capability needed for cellular simulations of the complete human brain towards the end of the decade. It will be the task of the JSC to develop, deploy and operate the HBP Supercomputer in Jülich as the future European HPC Facility for brain research. Jülich will work with HPC industry in the ramp-up phase to arrive at suitable, innovative HPC solutions meeting the specific requirements of the HBP (such as large memory and interactivity), and to lay the technological foundation for the subsequent procurement of a pre-exascale machine in the next phase of the project. The interactive supercomputing capabilities that will be developed for the HBP will be invaluable not only for neuroscience but also for a broad range of other applications in the life sciences and elsewhere (e.g., in civil security research). While the HBP is poised to become the main driver for the future development of high-performance computing at JSC, the breadth of applications involved and their requirements will warrant the usability of the HBP Supercomputer for many other fields, too.
Besides the main HBP Supercomputer at Jülich the HBP’s HPC Platform will consist of a smaller software development system at CSCS (Lugano, Switzerland), a system for molecular-level simulations at BSC (Barcelona, Spain), and a system for massive data analytics at CINECA (Bologna, Italy). During the ramp-up phase of the project, the HBP will negotiate with further PRACE Tier-0 institutions that have expressed their interest in adding in-kind support to the Platform. A high priority goal is to establish a PRACE community access programme, also to be negotiated in the ramp-up phase. This would allow access to the Tier-0 capability of the HPC Platform, reviewed by the HBP’s International Access Board, via PRACE services.

Acknowledgments

The DEEP, DEEP-ER and Human Brain Project are partially funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement N° 287530, 610476 and 284941. The Exascale Innovation Center is supported by the State of Nordrhein-Westfalen.

References

3. High-Q Club website: http://www.fz-juelich.de/ias/jsc/high-q-club
17. DEEP website: http://www.deep-project.eu
18. DEEP-ER website: http://www.deep-er.eu
22. Green500 website: http://green500.org/
The NIC Research Groups
Relativistic Transport Simulations and the Search for the Quark-Gluon-Plasma

Marcus Bleicher and Hannah Petersen
Frankfurt Institute for Advanced Studies and Institut für Theoretische Physik,
Johann Wolfgang Goethe-Universität,
Ruth-Moufang-Straße 1, 60438 Frankfurt am Main, Germany
and
GSI Helmholtzzentrum für Schwerionenforschung, Planckstraße 1, 64291 Darmstadt, Germany
E-mail: {bleicher, petersen}@th.physik.uni-frankfurt.de

The hybrid version of the Ultra-relativistic Quantum Molecular Dynamics approach (UrQMD) is presented. This model allows to combine the relativistic Boltzmann-equation dynamically with relativistic fluid dynamics to allow for the exploration of the ultra hot and super dense matter created in collisions of heavy nuclei in todays largest particle accelerators. The conditions reached in the experiments performed at the CERN-LHC, BNL-RHIC and in the future at FAIR near Darmstadt will allow to gain insights into the properties of matter as it existed micro-seconds after the Big Bang or nowadays in the interior of Neutron Stars.

1 Introduction

Among the four fundamental forces of nature, i.e. gravity, electro-magnetism, the weak interaction and the strong interaction, the latter one is currently under detailed investigation at the worlds largest and most energetic particle accelerators. The strong interaction, also known as Quantum Chromo-Dynamics (QCD) governs the interaction of elementary coloured particles, i.e. quarks and gluons and their composites, e.g. the protons and neutrons in the atomic nucleus. To gain further insights into the interactions between the quarks and gluons and the properties of QCD, heavy ions, e.g. gold or lead are collided at very high energies in accelerators at the CERN-LHC, BNL-RHIC or in the future at the FAIR facility. These collisions generate temperatures on the order of 10 Billion degrees and densities of up to $10^{17}$ kg/m$^3$ and allow to explore the phase diagram of QCD. On the experimental side, the output of such a collision are reconstructed tracks in the detectors that allow to obtain information on the final particle momenta and masses. A direct observation of the collision zone is not possible due to its small size of only $10^{-14}$ m and its short life time of $10^{-22}$ seconds. Therefore, theoretical modelling is necessary to connect the final state observations with the properties of the matter during the collision. On the theoretical side, the non-abelian nature of QCD makes first principle calculations of the evolution dynamics essentially impossible and one has to rely on transport simulations based on phenomenological approaches to explore the dynamics. A sketch of the QCD phase diagram is depicted in Fig. 1. The horizontal axis denotes the density in units of the density of an atomic nucleus, while the vertical axis shows the temperature. The arrows indicate the paths at different beam energies through the phase diagram. At high temperatures and/or high net baryon densities the formation of a new state of matter, the Quark Gluon Plasma (QGP), is expected. In contrast to usual nuclear matter where the quarks and gluons are confined in bound states such as protons and neutrons, the QGP is a deconfined
state which contains freely moving quarks and gluons. To study the transition to this new phase of QCD is one of the main motivations for heavy ion research.

2 The Approach

Ideally, one would try to solve QCD directly as given by the Lagrangian. Lattice QCD simulations take this route and provide first principle information in the regime of small baryon densities and under the condition of a static and thermally equilibrated system by direct evaluation of the path integrals. This allows to extract fundamental properties of QCD, among them, e.g. the transport coefficients (viscosities and heat conductivity) and the equation of state, i.e. the relation between pressure, energy density and temperature of the matter. Unfortunately dynamical out-of-equilibrium calculations are not feasible in this approach. A less fundamental, but more suited approach is to employ hybrid transport simulations that use lattice QCD information (and phenomenological models) as input to describe the dynamics encountered in a heavy ion reaction. Here, one approximates Greens functions with a phase space interpretation leading to a system of coupled equations of Boltzmann-type or after a further averaging to (viscous) relativistic hydrodynamics.

Hybrid models, combining both, a Boltzmann-type dynamics, colour flux tubes or a colour glass picture for the initial off-equilibrium dynamics, followed by a fluid dynamical evolution for the intermediate part, coupled again to a Boltzmann dynamics for the dilute final phase have become the state-of-the art for the interpretation of heavy ion data\textsuperscript{3-9}. Fig. 2 exemplifies the different stages of a heavy ion reaction as given by the simulations. The time direction is from left to right. The points denote particles, while the surfaces denotes matter in the hydrodynamic stage.
Figure 2. Time evolution of a heavy ion reaction. The snapshots show the hybrid simulation results at characteristic stages of the reaction. The time direction is from left to right. The points denote particles, while the surfaces denote matter in the hydrodynamic stage. Taken from Ref. 2.

With this toolbox at hand, one is now able to relate fundamental properties of QCD to finally observable particle spectra. In the following we will demonstrate this with recent examples before we come to some technical aspects.

3 Measuring the Temperature: Photons

To answer the question, if one reaches the conditions to form Quark-Gluon-Plasma a measurement of the temperature is necessary. Out of the many possible observables,
Electromagnetic probes have the key advantage to leave the hot and dense region undisturbed: once they are created, they escape freely from the reaction zone, due to their negligible rescattering cross-sections. Besides dileptons, direct photon emission is therefore of greatest interest to gain an undisturbed sight into the reaction zone. Fig. 3 shows a comparison of the direct photon spectra from different model scenarios\textsuperscript{10} to the experimental data by the WA98-collaboration\textsuperscript{11}. Calculations without intermediate hydrodynamic stage (pure cascade mode) are shown as red crosses, hybrid calculations with hadron gas EoS as red solid line and Bag model calculations (i.e. an equation of state with a large latent heat) are depicted by the dark-green dotted line. The conclusion of this analysis is two-fold. Firstly, the photons transverse momentum distribution is related to the temperature of the source (in the simplest picture \( \frac{E}{dN/d^3p} \approx \exp(-p_\perp/T) \)). Secondly, the necessary increased photon yield around \( p_\perp \approx 2 - 3 \) GeV is only achievable, if the equation of state has a (phase) transition to the Quark-Gluon-Plasma.

4 Measuring the Diffusion and Drag Coefficients: Elliptic Flow and Quenching of Heavy Quarks

Elliptic flow is a space-momentum correlation that develops in the early stage of the reaction. Fig. 4 sketches the idea: The left figure shows the initially created almond shape in the overlap zone of the two colliding nuclei in space. Due to the asymmetric pressure gradients the spatial anisotropy translate into a momentum space anisotropy (right figure) that can be measured in experiment by the second Fourier coefficient \( v_2 \) (higher flow harmonics, \( v_n \), can also be extracted). The efficiency of the transformation from space to momentum space depends on the interaction strengths of the matter. If the matter has a high viscosity (low interaction rate, large mean free path), the transformation is inefficient and the resulting elliptic flow is small, however, if the viscosity is low, the elliptic flow is large. This method has been used to obtain experimental estimates for the shear viscosity.

![Figure 4](image-url)
over entropy ratio ($\eta/s \approx 1 - 2/(4\pi)$), see e.g. Ref. 12. Here we want to go a step further and estimate also the interaction strength of heavy Quarks with the medium\(^\text{13}\). Heavy quarks are an ideal probe for the QGP, because they are produced in the primordial hard collisions of the nuclear reaction and therefore probe the created medium during its entire evolution process. Most interesting here is the relation between the above discussed elliptic flow, $v_2$, and the nuclear modification (think of damping) factor, $R_{AA}$. While both are directly sensitive to the viscosity/interaction strength, the $R_{AA}$ decreases strongly with an increase of interactions strength, while $v_2$ increases. Therefore, a simultaneous description of both quantities provides strong constraints on the interaction between heavy quarks and the medium. The latest results are shown in Fig. 5 for calculations and data for lead+lead collisions at the LHC (lines show the calculation, symbols show the experimental data).

5 Event-by-Event Fluctuations

One of the proposed signatures for a phase transition to the Quark Gluon Plasma or a critical endpoint are event-by-event fluctuation observables. These observables require high statistics on the experimental but also on the theoretical side. Critical phenomena are not the only source of fluctuations in heavy ion collisions. Besides controllable differences like different beam energies at which the ions are brought to collision or different impact parameters (head-on collisions versus more peripheral touching), each event differs in its initial energy density distribution. Due to the finite number of nucleons in each nucleus and their fluctuating positions and interaction points the initial geometry in each event is irregular and has hot and cold spots (see Fig. 6 left).

Along similar lines as the elliptic flow explained above there are higher flow coefficients like triangular flow, $v_3$, that can be measured\(^\text{18,19}\). As shown in Fig. 6 (right) the value of triangular flow is very sensitive to the amount of fluctuations in the initial configuration. Comparisons to the experimental data indicate that full event-by-event simulations are necessary to reproduce the correct value of this observable. In addition, triangular flow and higher flow coefficients that are sensitive to even finer structures in the initial state deliver important constraints on the shear viscosity over entropy ratio of the Quark Gluon Plasma.
6 How to Improve the Speed of the Simulations?

The simulations mentioned above require a substantial amount of computational resources. Typically 100'000 simulations have to be performed with each hybrid calculation running for approx. 2 hours for a given parameter set and centrality. If one wants to explore event-by-event fluctuations up to 10 Million events will be needed. Therefore, to improve the speed of the simulation graphic cards (GPGPUs) provide an alternative to a pure CPU based code. We have explored the potential of GPGPUs to improve the speed of the relativistic fluid dynamical simulation as shown in Fig. 7. Here, the OpenCL-SHASTA consists of a C++-part, managing the memory allocation and enqueuing of the kernels. The kernels are routines written in OpenCL and completely run on the GPU or multi-core CPU. Kernels are executed in a parallel manner, and each singular instance of a kernel
is called a work-item. These work-items are mapped, in hardware, to the stream cores of GPUs and CPUs. The mapping occurs in small groups, whose size depends on the hardware used. The smallest possible group is called wavefront. Within each wavefront the execution flow must be uniform, i.e. when branching occurs within a wavefront, all branches are computed serially. The result of this re-write is a tremendous gain in speed up to factors of 500 on the newest GPGPUs (figures are done with the previous GPGPU version).

7 Concluding Remarks

Transport simulations have become a valuable tool to explore the properties of QCD-matter under extreme conditions. In the ideal case, they allow to link fundamental properties of QCD, e.g. extracted from lattice QCD, to observables measured by the experiments. However, this comes at the price of certain simplifications, e.g. restrictions to phase space distributions in contrast to Greens functions or additional phenomenological assumptions. In these proceedings, we have shown examples how final state particle spectra can shed light on the equation of state of QCD-matter, and on its transport properties. Finally we have discussed possible improvements to the simulation packages to allow for increased simulation speed. Nevertheless, the use of high performance computers is inevitable to keep up with the high quality and precision of the data provided by the experiments.

Acknowledgements

We would like to thank the John von Neumann Institute for Computing and the Jülich Supercomputing Centre for support. This work as also been supported by the Hessian LOEWE-Initiative through HIC for FAIR and the LOEWE-CSC. H. P. is supported by the Helmholtz association VH-NG-822.

References

1. www.gsi.de/start/fair/forschung_an_fair/kernmateriephysik.htm

23
11. M. M. Aggarwal et al. [WA98 Collaboration].
Atomistic Modelling of Redox Reactions in Non-Equilibrium

Wolf B. Dapp\textsuperscript{1} and Martin H. Müser\textsuperscript{1,2}

\textsuperscript{1} John von Neumann Institute for Computing (NIC), Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: \{w.dapp, m.mueser\}@fz-juelich.de

\textsuperscript{2} Department of Materials Science and Engineering, Universität des Saarlandes, 66123 Saarbrücken, Germany

We developed a new atomistic method to model (non-equilibrium) redox reaction using empirical force fields for use in MD simulations. To this end, we added the (formal) ionisation state as a discrete variable into the “split charge equilibration” method (SQE). This extension allows atoms to swap integer charges across bonds, in addition to exchanging fractional charges. We call this method “redoxSQE”, and, in first steps, used it to study contact electrification and to set up a model rechargeable nano-battery that reproduces the generic features of the discharge of a macroscopic battery qualitatively. Other popular charge-transfer force fields fundamentally cannot describe any history-dependent effect because they calculate the charge distribution as a unique function of atomic positions. For similar reasons, state-of-the-art DFT-based methods fail to describe redox reactions in non-equilibrium.

1 Introduction

Redox reactions involve a change of oxidation state, most commonly by exchanging an electron between a donor (which is oxidised) and an acceptor (which is reduced). Such reactions are of fundamental importance for instance in many biological processes (such as cellular respiration), but also in electrochemical cells, e.g., rechargeable batteries. Redox reactions are also at the heart of one of the oldest scientific experiments: Thales of Miletus rubbed cat-fur against a piece of amber and found that the latter becomes electrically attractive to things like wool. He sought rational explanations for this effect which we now know as triboelectricity and contact electrification. In fact, the word “electron” is the Greek word for amber. Our research group takes a top-down approach to introducing the concept of oxidation states into an atomistic empirical force-field description, by attempting to reproduce qualitatively (as an initial step) results for instance relating to battery discharge\textsuperscript{1}. Other groups\textsuperscript{2} are working on deriving SQE (see Sec. 2) plus oxidation states from the bottom up, from DFT (density functional theory).

In many cases, the molecular systems of interest either contain so many atoms or evolve on such long time scales that they are beyond the reach of DFT or \textit{ab initio}-based MD (molecular dynamics), and reasonably accurate empirical force fields are needed. Those have indeed been derived for a variety of systems, but a weak point remains the realistic handling of electrostatic interactions. Fixed-charge models are unsuitable, for example, at interfaces (e.g., a silicon dioxide layer on bulk silicon), or indeed for any other chemically heterogeneous systems, where the charge on each atom strongly depends on its chemical environment. Likewise, whenever a redox reaction occurs, the effective (fractional) charge of the participating atoms will be changed, in a quasi-discontinuous fashion. Here, we
will summarise selected recent results of our research group regarding the simulation of (non-equilibrium) redox reactions\textsuperscript{1,3,4}.

\section{SQE and redoxSQE}

The general strategy in methods that assign effective atomic charges on the fly is to minimise an energy expression (typically a quadratic expansion of the interaction between atoms) with respect to the charges. In the most popular \textit{charge equilibration} (QE) model, also used in the so-called “reactive force field” ReaxFF\textsuperscript{5}, the linear term involves the electronegativity, $\chi$, while the quadratic term is proportional to the chemical hardness of atoms, $\kappa$, which reflects its charge self-interaction. In principle, atomic hardesses and electronegativities are element-specific properties that can be obtained through finite differences from the free-atom electron affinity and the first ionisation energy. For a formal, DFT-based justification of the originally \textit{ad hoc} introduced parameters see work by Verstraeten \textit{et al.}\textsuperscript{2,6}. An alternative to QE approach is the \textit{atom-atom charge transfer} (AACT) model\textsuperscript{7}, which is based on the idea that charge is transferred through chemical bonds. Here, the quadratic term penalises charge transfer via a so-called bond hardness, $\kappa(b)$, which is essentially inversely proportional to the polarisability of the chemical bond. The \textit{split-charge equilibration} (SQE) model is a hybrid of these two approaches\textsuperscript{8}. Let $q_{ij}$ be the partial charge split between atom $i$ and atom $j$, with the symmetry relation $q_{ij} \equiv -q_{ji}$. The total effective charge on atom $i$ is given by

$$Q_i = \sum_{ij} q_{ij}, \quad (1)$$

and the expression to minimise becomes

$$V = V_C(\{R, Q\}) + \sum_i \left( \frac{\kappa_i}{2} Q_i^2 + \chi_i Q_i \right) + \sum_{i,j>i} \frac{1}{2} \kappa_{ij}^{(b)} q_{ij}^2, \quad (2)$$

where $V_C(\{R, Q\})$ is the Coulomb energy, which depends both on the atomic and the charge configuration. The bond hardness $\kappa_{ij}^{(b)}$ depends on the types of atoms forming the bond and on their separation. The QE model arises in the limit where all $\kappa_{ij}^{(b)}$ are set to zero (i.e., the system is infinitely polarisable), while AACT is equivalent to neglecting the atomic $\kappa$.

Both QE and AACT fail to reproduce certain generic trends. For example, AACT does not show the proper scaling of the polarisability for oligomers\textsuperscript{9} in the limit of small degrees of polymerisation $P$, and the skin depth for external fields in solids is always less than an atomic spacing\textsuperscript{10}. Furthermore, it fails to describe metals, or any material with a dielectric constant not $\approx 1$. Conversely, dielectric behaviour is ruled out in QE, and solids always behave as ideal metals\textsuperscript{10}. Moreover, polar molecules show the wrong scaling of the polarisability\textsuperscript{9} and dipole\textsuperscript{11} for long-chained polymers. Finally, because of its polarisability, QE also produces the wrong dissociation limit of molecules. For example, if a CH\textsubscript{4} molecule were separated from an H\textsubscript{2}O molecule, the molecules would carry nonphysical charges of about $\pm 0.4 \, e$, where $e$ is the elementary charge, unless artificial constraints were imposed. None of these artifacts arise with SQE.

However, like its competitors, the original formulation of SQE still fails to describe true ions. In a realistic parameterisation the bond hardness diverges once orbitals between
neighbored atoms no longer overlap at large interatomic distances. Then electropositive atoms automatically cease to donate their electrons to electronegative atoms, and both partners are neutral. This shortcoming is remedied with the introduction of a (formal) oxidation state as a variable. We modify Eq. 1, and now calculate the charge according to

$$Q_i = \sum q_{ij} + n_i e,$$

where $n_i$ is an integer number. The oxidation state can be thought of as excess integer charges on atom $i$. Note that no bond-related energy penalty applies when integer charges are moved about the system.

The concept of formal (integer) oxidation numbers makes it possible to simulate charging or discharging of Galvanic cells, or, more generally, any type of processes involving redox reactions, as well as electrostatic fields of zwitterionic molecules, which violate the principle of local charge neutrality. Moreover, the partial charge of an atom can now be made history dependent because, for any given atomic configuration, the system can assume a number of different minima of $V$, one for each unique $\{n_i\}$ configuration, akin to Landau-Zener levels.

The oxidation state dynamics can be parameterised so that it mimics a Landau-Zener process. The dynamics can involve “radiation”, i.e., a discontinuous (energy) change of the system. This could happen because the new oxidation state of the system necessitates a reoptimisation of all split charges. Alternatively, radiation-free redox reactions are possible either if the excess energy is supplied to the system as kinetic energy, or if the redox reaction happens exactly at $q_{ij} = e/2$. Radiation-free redox reactions play an important role in Marcus theory, which was the first generally accepted theory of electron transfer. Fig. 1 shows for the dissociation of a generic diatomic molecule energy levels and charge evolution as a function of separation.

In our implementation, a stochastic process determines for a specific bond whether we
attempt an integer charge transfer move. A Metropolis-type criterion on the energy decides whether such a trial move across a dielectric bond (i.e., one having non-zero bond hardness) is accepted or rejected. Similar to Tully surface hopping\textsuperscript{13}, an integer charge transfer takes the system to a different Landau-Zener level on which it will evolve subsequently. An integer charge transfer across a metallic bond is always accepted, because the bond hardness and therefore the change in energy during a trial move is zero – the integer charge transfer is perfectly balanced by the transfer of partial charge between the participating atoms.

We calculate the bond hardness as a divergent rational function according to the (heuristic) formula

\[
\kappa_{ij}^{(1)} = \begin{cases} 
\kappa_{ij}^{(p)} & r_{ij} \leq r_s, \\
\kappa_{ij}^{(p)} + \kappa_{ij}^{(0)} \frac{r_s^2}{r_s^2 - (r_{ij} - r_s)^2} & r_s < r_{ij} < r_1, \\
\infty & r_1 \leq r_{ij},
\end{cases}
\]

where \(\kappa_{ij}^{(p)}\) is a plateau value (\(\equiv 0\) for metals), \(\kappa_{ij}^{(0)}\) is a bond-specific parameter, and \(r_s\) and \(r_1\) are short- and long-separation cutoffs, respectively. This expression has been shown\textsuperscript{14} to work well for fractional charges in the homolysis of a variety of organic molecules, even for some radicals and transition states although the calibration was done on equilibrated structures satisfying the octet rule. The long-range cutoff is convenient for computational reasons.

In the following, we briefly discuss two applications of the \textit{redoxSQE} method: (i) to contact electrification, and (ii) to battery discharge. For more details, we refer the reader to the original literature\textsuperscript{1,3,4}.

\section{Contact Electrification}

If two neutral solids (e.g., gold and sodium) are brought into contact, charge transfers between them, and, upon separation, they retain some of that charge. This means that after contact formation there is an electrostatic attraction between the clusters that was not there before – even if one assumes an identical atomic configuration before and after. Such history dependence is not captured by most charge-transfer force fields. They determine fractional charges as a unique function of the instantaneous atomic positions just like conventional DFT computes a unique charge density for a given atomic configuration.

For metals, the mechanism of contact electrification is well established\textsuperscript{15}. Electrons transfer from the metal having the smaller work function to that with the larger one. The precise amount of transferred charge is affected by electrostatics, e.g., by the total capacitance of the metals and by the rate at which the two solids are pulled apart. Nevertheless, the direction of charge flow between initially neutral metals is entirely determined by their work functions.

The rubbing-induced charge transfer between dielectrics is much less well understood. Electron transfer\textsuperscript{16}, proton transfer\textsuperscript{17}, or the exchange of hydroxide or other ions\textsuperscript{18} have all been suggested as the possible origin of contact-dynamics-induced charging. Unlike metals, dielectrics cannot be arranged into a linear triboelectric series in an unambiguous fashion, sometimes the series is even cyclic. For example, during rubbing with the material
Figure 2. From Ref. 3. **Left:** visualisation of the contact electrification of a metal tip and a metal substrate. Red and blue indicate, respectively, the amount of negative and positive charge on an atom. Initially neither solid is charged. When they are brought to close proximity, charge can pass between them such that negative charge flows from the metal with the smaller work function to that with the larger work function. The charge distribution mainly lives on the surface of the clusters, as expected for metals, because this minimises the (repulsive) electrostatic energy. After the solids have been separated, no more charge can flow. With redoxSQE, both solids retain a constant, integer charge, while AACT predicts that both parts will be neutral again after separation. With DFT or QE, neither parts can be neutral at any point (not even during the approach), as charge is transferred non-locally and over arbitrary distances with those schemes. **Right:** Similar, except that here the bonds within each solid are modelled as dielectric. This prevents the charge from spreading across the solids’ surfaces. Also, the total transferred (integer) charge is smaller than for metals.

Fig. 2 demonstrates that our model reproduces the qualitative features of contact electrification for both cases. In the original publication\(^3\), we show a number of model systems, including the contact between two metal clusters, two dielectric solids, and the dissociation of an NaCl dimer. For the latter case, we also performed quantum-chemical calculations to show that the method can be adapted to fit a real system. In this case, the first of two possible outcomes is that the dimer breaks up to form two neutral products (as in fact all diatomic molecules do with the exception of FrCl). This happens in an inert environment, such as an argon atmosphere. QE – and likewise currently used approximations to the exact DFT functionals – incorrectly predict a remnant charge of \(\pm 0.5\) e on the atoms at infinite distance in this case (see also Fig. 1). Conversely, in a sufficiently polar environment (e.g., an NaCl dimer surrounded by \(\geq 4\) water molecules), the dimer dissociates as the ions Na\(^+\) and Cl\(^-\). RedoxSQE can reproduce both cases, as well as the energy curves and the ESP (electrostatic potential) partial charges with respect to the separation\(^3\).

The physical concept behind redoxSQE is that the oxidation state is largely independent of the instantaneous atomic configuration. This allows for history dependence in that the system can evolve on, and switch between, different, independent Landau-Zener levels. The oxidation state is a discrete quantity and a change involves a redox reaction implicating two or more nearby atoms. The effective total charges are determined anew after each transfer of integer charge.

### 4 Towards Modelling a Battery Atomistically

Battery research and development has been experiencing a tremendous surge in recent years, as it is crucial to overcome technological challenges related to energy production (e.g., buffering output peaks of renewable energy power plants), storage (e.g., for use in portable electronic devices) and usage (e.g., for automotive purposes). In a model system more complex than contact electrification, we took first steps towards modelling a battery...
Figure 3. From Ref. 1. **Left:** illustration of a battery setup with 1194 atoms. Charge is encoded in the colouring: blue means positive, red negative charge. For visual distinction, (fixed-charge) electrolyte particles are chosen smallest, independent of their LJ radii, and their charge colouring is halved. The medium-sized particles are cations, while the largest particles are metallic atoms. The separator (salt bridge) keeps non-electrolyte atoms from moving between the two half-cells but lets electrolyte atoms pass. A resistive external load $R$ (following Ohm’s law) completes the circuit. **Right:** Discharge curves of our battery demonstrator with different external resistors. The data represents an average over 4 independent runs, and each point is averaged over many MD time steps. The solid lines are inserted to guide the eye. The higher the external resistance, the closer the battery’s behaviour approaches an ideal discharge curve, the lower the load, the more the internal resistance dominates. In this property, and the shape of the discharge curve, our nano-battery resembles a macroscopic battery. Initially, the voltage declines sharply, as the electrodes are charged before ions are dissolved or adsorbed, respectively. This is followed by an extended plateau when the voltage stays constant as the charge transfer through the external resistance is balanced by an equal amount of ion transfer in the electrolyte. At the same time, additional charge on the electrodes is now compensated by dissolving and adsorbing ions. Finally, another steep decline concludes the discharge, as the electrodes are consumed and their surfaces passivated.

as a whole\(^1\), atomistically. This attempts to fill a gap between mesoscopic porous electrode models\(^{20,21}\) commonly used for commercial battery applications on the one hand, and DFT- or MD-based simulations of isolated and specialised aspects of individual processes\(^{22}\) happening in a half-cell on the other hand.

We set up a system that resembles a traditional Voltaic wet-cell, containing two metal electrodes, a liquid electrolyte (which we model as purely ionic), an adjustable Ohmic external load, and a salt bridge that keeps the electrodes from touching (which would create a short circuit), but allows electrolyte ions to pass. The short-range interaction is calculated with a Lennard-Jones (LJ) potential, and the electrolyte is modelled as a Kob-Andersen glass in order to avoid freeze-out. Fig. 3 (left panel) illustrates the setup. When the switch is closed, the difference in chemical potential between the two electrodes drives a current through the external load. The circuit is completed by electrolyte ions streaming from one half-cell to the other. In contrast to other charge transfer force fields, redoxSQE does not *equalise* the chemical potential between the electrodes. Instead, the method maintains a differential in chemical potential by allowing anode metal ions to be reduced and dissolve as cations. At the same time, cations adsorb to the cathode surface, are oxidised and take on the excess electrons as the second half-reaction to a full redox reaction. The process halts once the redox-active material is exhausted, i.e., the anode is dissolved and all cations have been oxidised at the cathode.

RedoxSQE also reproduces the polarisation charges naturally without having to intro-
duce explicit mirror charges. A Helmholtz double layer of electrolyte ions forms at both electrode interfaces where the potential drop occurs. In fact, one of the advantages of the method is that it allows to study the processes occurring at the electrolyte-electrode interface, also including clustering and dendrite formation.

Fig. 3 (right panel) shows the discharge behaviour for different external loads. Despite not having parameterised the simulation for any particular material but having chosen typical generic values for the free parameters, the results resemble the discharge curves of commercial batteries. Other macroscopic properties of battery discharge are also reproduced, such as a strong temperature dependence of capacity and available voltage, voltage recovery if the battery is allowed to relax (as happens in pulsed-discharge applications), as well as a voltage overshoot when the battery is charged. These generic features are largely independent of the parameterisation details, which hints at the transferability of the method.

In future work we will parameterise the model for specific materials of practical relevance, and improve the efficiency of the implementation so that we can model sufficiently large ensembles. However, the initial results are already quite encouraging first steps towards an atomistic model of a battery as a whole.

5 Summary

RedoxSQE can be made part of empirical force fields to model (non-equilibrium) redox reactions as they occur in problems of contact electrification and tribo-electricity as well as in electrochemical processes such as the discharge of a battery. It outperforms other charge-transfer force fields in that it does not suffer from their shortcomings, and is transferable.

We showed that redoxSQE is able to simulate processes that occur in contact electrification, although descriptions are still at a rather generic level. In principle, it can handle both metallic and dielectric contacts, and so may contribute to the debate how tribo-charging works between dielectrics.

We also demonstrated that a model battery robustly reproduces generic features seen in macroscopic battery discharge such as dependence on temperature or external load, and relaxation and recharge behaviour. Once parametrised in a material-specific way, redoxSQE can be particularly useful in studying the processes near the interface of electrolyte and electrode. It is the first empiric approach to model an entire battery atomistically. For more details, we refer the reader to the original literature.

Acknowledgements

The authors gratefully acknowledge computing time on JUROPA.

References


We investigate the thermal transition from nuclear matter to a quark gluon plasma by simulations of lattice QCD with two quark species using the Wilson twisted mass formulation. A new code based on OpenCL is presented, which can be used on either CPU or GPU of any vendor. To optimise efficiency, we thermalise our Monte Carlo simulations using highly parallelised code on CPU machines, whereas production runs are done using multiple Monte Carlo chains on a GPU cluster, fitting an entire lattice on each GPU with zero overhead for communication. We discuss preliminary results for the thermal transition obtained in this way.

1 Introduction

The fundamental theory of the strong interactions governing nuclear and subnuclear forces is Quantum Chromodynamics (QCD). Its fundamental degrees of freedom are quarks and gluons, which combine into numerous tightly bound states, the hadrons, among them the familiar nucleons. A key feature of the theory is asymptotic freedom, according to which the coupling strength depends on the energy scale of a scattering process. For energies below 1 GeV, the coupling is too large to allow for weak coupling approximations and analytic predictions. On the other hand, at large temperatures the average energy per particle is large and the theory enters a weak coupling regime, where the quarks and gluons form a plasma rather than bound states. The transition from the hadronic to the plasma regime takes place at temperatures of about 170 MeV and still belongs to strong coupling physics. However, the theory can be reformulated on a space-time lattice, whereupon it is amenable to Monte Carlo simulations.

In this contribution we present a study of the thermal QCD transition with two mass-degenerate quark species. The main interest is in the critical temperature and the nature of the transition. In the limit of massless quarks, the theory has a chiral symmetry under mixing of the quark species. This symmetry is broken spontaneously, so that the lightest particles, the pions, are exactly massless. At finite temperature a non-analytic transition takes place in which this symmetry gets restored. Since massless QCD cannot be simulated, the order of this phase transition is not known to date. On the other hand, for finite quark mass the chiral symmetry is explicitly broken and the pions are massive. The phase transition then gets weakened to an analytic crossover. A first order chiral transition disappears gradually in a $Z(2)$ critical point, whereas a second order transition disappears immediately for non-zero mass, Fig. 1.
Figure 1. Possible scenarios for the chiral phase transition as a function of pion mass. In the chiral (massless) limit there must be a true transition of first or second order, whereas for finite masses the transition is merely a smooth crossover. The boundary between the two corresponds to a critical point.

2 Twisted Mass Lattice QCD

Consider euclidean spacetime discretised on a hypercube with lattice spacing $a$. We denote the spatial and temporal extent of the system with $N_{\sigma}$ and $N_{\tau}$, respectively. The QCD action $S_{\text{QCD}}$ is then replaced by a lattice version afflicted by discretisation errors,

$$S_{\text{LQCD}} = S_{\text{QCD}} + aS_1 + a^2S_2 + \ldots ,$$

and continuum physics can be obtained in the limit $a \to 0$.

The central object in statistical physics is the partition function $Z$ of the system, and on the lattice, an expectation value of some observable $A$ reads:

$$\langle A \rangle = Z^{-1} \int DUD\psi D\psi A \exp \{-S_{\text{QCD}}[U]\}$$

$$= Z^{-1} \int DUA \det D[U] \exp \{-S_{\text{gauge}}[U]\} .$$

(2)

Here, $\psi$ and $U$ denote the fermion and gluon fields, respectively. The latter are represented by so-called links on the lattice. If one identifies $S_{\text{LQCD}} = \beta H$, the exponential in the first line is the Boltzmann factor. The fermion fields $\psi$ can be integrated out exactly due to their Grassmann nature, and the resulting determinant of the fermion matrix $D$ is expressed in terms of (pseudo fermions) $\phi$,

$$\det D[U] \sim \int D\phi^\dagger D\phi \exp \{-\phi^\dagger D^{-1}[U]\phi\} ,$$

(3)

yielding the effective action $S_{\text{eff}}[U,\phi] = S_{\text{gauge}}[U] + \phi^\dagger D^{-1}[U]\phi$. Importance sampling methods are used to evaluate this high-dimensional integral. Using the Boltzmann-weight $p[U,\phi] = \exp \{-S_{\text{eff}}[U,\phi]\}$ as probability measure, an ensemble of $N$ gauge configurations $\{U_m\}$ is generated. Then, $\langle A \rangle$ may be approximated by

$$\langle A \rangle \approx \frac{1}{N} \sum_{m} A[U_m] .$$

(4)

The standard simulation algorithm to generate QCD gauge configurations is the Hybrid Monte-Carlo (HMC) algorithm, where the effective action is embedded in a fictitious classical system evolved over a time $\tau$ according to the hamiltonian equations of motion. Since the numerical integration is not exact, a Metropolis step is carried out in the end, thus ensuring detailed balance. For details see Ref. 6. In order to invert the high-dimensional,
sparse fermion matrix $D$, iterative Krylov space methods are used, i.e. $D^{-1}$ is determined by equations like

$$D\phi = \psi \Rightarrow \phi = D^{-1}\psi .$$

This is the most cost-intensive part of a simulation and it is crucial to have a well tuned implementation of the derivative term $D$.

In our studies$^4,5$ we employ the so-called two flavour twisted mass Wilson fermions$^6$. Their fermion matrix reads:

$$D_{\text{tm}}^\pm = (1 \pm 2i\kappa\gamma_5)\delta_{xy}\delta_{\alpha\beta}\delta_{ab} - \frac{\kappa}{2} \sum_\mu (1 - \gamma_{\pm\mu})_{\alpha\beta} U_{\pm\mu}(x)_{ab}\delta_{n+\mu,y}$$

$$= M_{\text{diag}}^\pm + \bar{D} ,$$

with $\gamma_{-\mu} = -\gamma_\mu$ and $U_{-\mu}(x) = U_\mu(x - \vec{\mu})$. $\gamma_\mu$ denotes Dirac matrices and $a, b, \alpha, \beta$ are colour and spinor indices, respectively. The sign in $M_{\text{diag}}^\pm$ corresponds to “up” and “down” quarks. In the gauge sector, the tree-level Symanzik improved Wilson action is used,

$$S_{\text{sym}} = \frac{\beta}{N_c} \sum_x \left(c_0 \sum_{\mu,\nu > \mu} \{1 - \text{Re Tr}(P_{\mu\nu}(x))\} + c_1 \sum_{\mu,\nu} \{1 - \text{Re Tr}(R_{\mu\nu}(x))\} \right) .$$

Here, $P_{\mu\nu}(x)$ and $R_{\mu\nu}(x)$ denote path-ordered plaquette and rectangle products of link variables. The parameters are the lattice coupling $\beta = 6/g^2$, $c_0 = 1 - 8c_1$ and $c_1 = 1/12$. For particular values $\kappa = \kappa_c(\beta)$, corresponding to “maximal twist”, the $O(a)$ discretisation effects vanish$^6$ and the quark mass is solely determined by $\mu$.

A finite temperature $T$ can be introduced by identifying

$$T = (a(\beta)N_c)^{-1} .$$

Thus, a scan in temperature equals a scan in $\beta$, and $a(\beta)$ is needed to set the scale. For this, we interpolate $T = 0$ data by the ETM collaboration$^8$ as in Fig. 2. ETMC also provides a formula from chiral perturbation theory to estimate $m_\pi(\mu)$ at maximal twist. The general phase structure of Twisted Mass fermions has been investigated in Ref. 5. The critical temperature and order of the chiral transition was studied in Ref. 4. Simulations
were performed at three different pion masses ($300 \lesssim m_\pi \lesssim 500$ MeV) on lattices of size $32^3 \times 12$ and $32^3 \times 10$. The resulting critical temperatures are shown in Fig. 3. Extrapolations to the chiral limit for different orders of the transition with their associated scaling behaviour and critical exponents,

$$T_c(m_\pi) = T_c(0) + Am_\pi^{2/(\beta\delta)},$$

are also shown. However, the combinations of exponents that go into the fit are numerically very similar, and much lower pion masses are needed to clarify the situation. This is exceedingly difficult, as it also implies larger lattices to fit the large correlation length of light pions on the lattice, while the numerical costs for the HMC scales like $V^{5/4}$ and $m_\pi^{-6}$.

### 3 LQCD Using OpenCL

In recent years, *Graphics Processing Units* (GPUs) have become an integral part of many modern computing clusters and are used in many LQCD applications. Tab. 1 shows an overview of available GPUs and CPUs. GPUs surpass CPUs in peak performance as well
as in memory bandwidth, however, one also notes the drop in performance when going from single to double precision on the GPU. Current LQCD applications utilising GPUs are predominantly written using NVIDIA CUDA\textsuperscript{9}, and many routines are publicly available\textsuperscript{10}. However, CUDA is only applicable to NVIDIA hardware. Currently we have access to two compute clusters with GPUs. One is the LOEWE-CSC (University of Frankfurt)\textsuperscript{7}, consisting of nodes with two 12-core AMD Magny-Cours CPUs and one AMD Radeon HD 5870 GPU. The other is the SANAM supercomputer (GSI Darmstadt), which has two AMD S10000 and two Intel Xeon E5-2650 CPUs per node.

For these architectures we developed a new HMC for twisted mass Wilson fermions\textsuperscript{2,1,3}: CL\textsuperscript{2}QCD. It is based on OpenCL\textsuperscript{11}, an open standard for parallel computing that provides an alternative to the vendor-bound CUDA. All operations are carried out in double precision. Fig. 4 demonstrates excellent performance of the $\mathcal{D}$ for the GPUs used in LOEWE-CSC and SANAM. 70 and 100 GFLOPS are achieved, respectively, over a wide range of lattice sizes. Also shown are results on NVIDIA GTX 680. Due to lack of optimisation the performance is poor, but it demonstrates the platform independence of OpenCL, which can also be run on CPUs. The limitation of applicability is mostly given by the GPU memory. For example, the AMD Radeon HD 5870 has only 1 GByte of memory which limits its performance when the entire lattice is put on one card.

The good performance of the $\mathcal{D}$ carries over to the full HMC, Fig. 5. It shows three different setups corresponding to different pion masses\textsuperscript{1} executed with CL\textsuperscript{2}QCD compared to a reference code\textsuperscript{6} run on the CPUs of one node in LOEWE-CSC. The performance shows a speedup of two for the AMD Radeon HD 5870, and a speedup of four for the newer AMD S10000 with respect to the reference code. This means that one GPU is able to perform the HMC algorithm much more efficiently than two CPU nodes. In addition, the acquisition costs of a GPU is typically lower than those of a server CPU.
Figure 5. HMC performance compared to reference code tmlqcd\textsuperscript{6} for different setups on a $24^3 \times 8$ lattice\textsuperscript{3}.

4 First Results for $m_\pi \approx 270$ MeV

To further improve the results obtained in Ref. 4, we started simulations at a smaller pion mass of around 270 MeV on $32^3 \times 12$ lattices. The extrapolations of Fig. 3 predicts a critical $\beta$ near 3.85.

We thermalised HMC chains using the tmlqcd CPU code\textsuperscript{6} in highly parallelised fashion on LOEWE-CSC and JUGE\textsuperscript{E}N\textsuperscript{E} in Jülich\textsuperscript{14}. On the other hand, the computer SANAM can fit the entire lattice in the memory of each GPU. We thus used this machine for production in “pedestrian parallelism”: From each thermalised chain we started new chains at the same parameters, separated by sufficiently many trajectories to rule out autocorrelations. In order to reduce wall-time, we have also started to implement Multi-GPU usage in CL\textsuperscript{2}QCD, too\textsuperscript{3} and are using it in current runs. However, this is only expected to be efficient once lattices are too large to fit into the memory of one unit.

Figure 6. The susceptibility of the chiral condensate, $\langle \bar{\psi} \psi \rangle$. The left curve corresponds to the preliminary $m_\pi \approx 270$ MeV results, the right one to the previous $m_\pi \approx 316$ MeV.
In this way we could effectively gather statistics of $O(60k)$ trajectories for the 10 $\beta$ values around the estimated $\beta_c$. The order parameter for the chiral transition is the chiral condensate $\langle \bar{\psi} \psi \rangle$. A peak in its susceptibility,

$$\sigma_{\bar{\psi}\psi}^2 = \frac{V}{T} \left( \langle (\bar{\psi}\psi)^2 \rangle - \langle \bar{\psi}\psi \rangle^2 \right),$$

(10)

signals maximal fluctuations and thus the location of the transition. This is shown in Fig. 6, corresponding to the lowest mass point in Fig. 3. Unfortunately, the new data do not show a pronounced peak, only a plateau around $\beta \approx 3.85$. On the other hand, we note that simulations for $\beta < 3.83$ become increasingly unreliable in the current setup, since the lattice spacing is only known for $\beta \geq 3.9$ and thus the pion mass cannot be held fixed reliably for smaller values. Clearly, additional simulations are necessary in order to clarify this issue.

Fig. 7 shows the renormalised chiral condensate

$$\langle \bar{\psi}\psi \rangle_{\text{ren}} = \langle \bar{\psi}\psi \rangle_T(\mu) - \langle \bar{\psi}\psi \rangle_0(\mu) + \langle \bar{\psi}\psi \rangle_0(0,0) \langle \bar{\psi}\psi \rangle_0(0,0),$$

(11)

which may serve as an indicator for the transition. Again, the signal for an inflection point around $\beta \approx 3.85$ is very weak, but consistent with that of the susceptibility.

5 Conclusions

We have developed a lattice QCD code for twisted mass Wilson fermions based on OpenCL\(^1\), CE\(^2\)QCD, which is able to utilise GPUs and CPUs in a vendor-independent way. It shows excellent performance on various generations of AMD GPUs. An HMC can be shown to be up to four times as fast as a reference code running on the CPUs of a whole LOEWE-CSC node. Currently, we are refining Multi-GPU usage\(^3\).

We have used CE\(^2\)QCD on SANAM together with highly parallel CPU code on JUQUEEN to investigate the chiral transition in two-flavour QCD at finite temperature. Unfortunately, the current setup does not allow to investigate the full temperature range

$$m_\pi = 270 \text{ MeV}$$

Figure 7. Preliminary results for the renormalised chiral condensate, $\langle \bar{\psi}\psi \rangle_{\text{ren}}$, for the lightest pion.
needed for the transition at the lightest pion mass. Thus, enlarging the temporal extent $N_t$ is the next step. In addition, an analysis of the generated data of the Polyakovloop regarding the deconfinement transition is under way.

Acknowledgements

O. P. and C. P. are supported by the German BMBF grant *FAIR theory: the QCD phase diagram at vanishing and finite baryon density*, 06MS9150. M. B., O. P, and C. P. are supported by the Helmholtz International Center for FAIR within the LOEWE program of the State of Hesse. M. B. and C. P. are supported by the GSI Helmholtzzentrum für Schwerionenforschung. Some of the calculations have been performed on LOEWE-CSC\textsuperscript{7}, SANAM and JUGENE\textsuperscript{13}. The authors thank the administrating teams for all the support.

References

   http://www.lattice2013.uni-mainz.de/presentations/4G/Bach.pdf
   http://dx.doi.org/10.1007/s00450-011-0161-5
9. NVIDIA, NVIDIA CUDA C Programming Guide.
   http://developer.nvidia.com/
   nvidia-gpu-computing-documentation
1 Introduction

The standard model (SM) of particle physics comprises the electroweak and the strong interactions. The electroweak theory, which unifies the electromagnetic and the weak interactions within a chiral gauge theory, exploits the concept of spontaneous symmetry breaking, known as the Higgs mechanism to provide masses for all quarks, leptons and the weak gauge bosons (W, Z), postulating the existence of the Higgs boson.

In July 2012, the ATLAS and CMS experiments at CERN announced the discovery of a Higgs-like particle, with the mass being about 125 GeV. The observation of a Higgs-like state has been corroborated by many more experimental data and a rather precise value for the mass of this scalar boson (spin zero) has been given. This discovery appears to confirm the prediction of the SM and completes it.

The basic idea of the Higgs mechanism is illustrated in Fig. 1. The potential of the Higgs field develops a minimum at a non-zero value of the radial degree of freedom of the Higgs field providing thus a vacuum expectation value $v$ for the Higgs field. It is through this vacuum expectation value that the particles acquire their masses, because a non-vanishing $v$ induces mass-like terms in the Lagrangian, e.g. a term proportional to $v\bar{\Psi}\Psi$ for the quarks.

However, old questions about the special rôle of spin-zero fields remain. Scalar fields have much more severe divergences in a quantum field theory than the fermion fields and the gauge fields. Also new questions emerge from the observed Higgs mass, for instance concerning the stability of the vacuum. Another question is whether this field is in fact an elementary one or whether the Higgs boson rather is a bound state.

Attempting to answer these questions, a key issue is to have a precise theoretical control of the SM theory. This precision physics requires the accurate determination of its fundamental parameters as well as the theoretical study of the Higgs sector of the SM. For the determination of SM parameters the strong interactions, described by QCD, are the major challenge because they are intrinsically non-perturbative. But also in the Higgs sector, one needs to question what can be computed by the expansion in the couplings, i.e. by “perturbation theory”, and how well this can be done.

These non-perturbative questions can be answered with techniques developed in lattice field theory where space and time are made discrete and a 4-dimensional finite grid.
with lattice spacing $a$ is introduced. This setup allows for numerical simulations which in turn provide \textit{ab initio} calculations that are not restricted to approximation methods such as perturbation theory. Of course, in the end the continuum limit has to be carried out numerically and one has to assure that symmetries broken by the discretisation are recovered in that limit.

2 Theoretical Studies of the Higgs Boson

The discovery of the Higgs boson and the subsequent measurements of the decay modes of the Higgs particle have led to the urgent need for understanding a list of theoretical issues which are relevant for the consistency of the SM and the possible extensions of it. Some of these theoretical problems are of non-perturbative nature, and these are the main targets of the lattice field theory approach to Higgs physics.

2.1 Non-Perturbative Studies of the Higgs-Yukawa Model

The upper and lower bounds of the Higgs boson mass are one main question of non-perturbative nature. For the lower bound, non-perturbative computations are highly desirable since the perturbative calculations rely on the instability of the effective potential, while it can be shown that the effective potential is, in fact, convex. Using a chirally-invariant lattice Higgs-Yukawa model the NIC group has performed non-perturbative computations to address these Higgs boson mass bounds. We refer to Refs. 1, 3 for reviews of lattice Higgs Yukawa models.

These calculations have been carried out for a standard model top quark mass of 175 GeV and also for heavier fermion masses up to 700 GeV in order to test the possibility of a fourth generation of quarks. As one important result of these previous computations\textsuperscript{2,4} it was found that the suggested extension of the standard model by a heavy fourth fermion
generation is not consistent with the lower Higgs boson mass bound values, given that the mass of such a fourth generation quark is at least 300 GeV as discussed in Ref. 7.

It is important to note that in the calculations a detailed study of finite size effects was performed. This becomes necessary since the model contains massless particles, the Goldstone bosons, and hence the finite size effects are algebraic and not exponentially suppressed. This led to simulations of lattice sizes of up to $40^4$ which then finally allowed to carry out an infinite volume extrapolation of the Higgs boson and fermion masses, and of the vacuum expectation value $v$.

An another important ingredient has been that the numerical simulations have been complemented by analytical calculations, i.e. lattice perturbative calculations of the effective potential. This allowed to guide the simulations and helped e.g. in determining the cut-off dependence of the Higgs boson mass. The next step of this research is to explore the influence of a higher dimensional $\phi^6$ operator on the Higgs boson mass bounds. This is of very significant phenomenological impact, linking the discovery of the Higgs-like particle to physics beyond the SM. These above discussed Higgs boson mass bounds depend on the cut-off of the theory which represents the scale of some new, so far unknown, physics beyond the standard model. For a given value of the Higgs boson mass there will be a crossing point with one of these mass bounds.

In the SM and for a Higgs boson mass close to 125 GeV such a crossing will happen with the lower Higgs boson mass bound. As illustrated by Fig. 2 the crossing seems to happen at a very large value of the cut-off, a scenario which is much discussed in the literature. The interesting question is, whether the addition of a $\phi^6$ term in the action can substantially shift the lower Higgs boson mass bound. First results by the NIC group in this direction indicate that a $\phi^6$ term induces a rich phase structure of the model with first order phase transitions for certain combinations of the coupling parameters. The NIC group will carry out a detailed study of this situation with a combined effort of lattice perturbative and non-perturbative numerical computations. This effort is carried out in close collaboration with the group of Dr. D. Lin from Taiwan.

2.2 Vacuum Stability in Perturbation Theory

A quantitative shortcoming of the non-perturbative investigations in the Higgs-Yukawa model is that – at least for the time being – important interactions in the theory have to be dropped in the simulations. Due to various studies it is expected that in particular the strong interactions (QCD) can be important. Through the coupled renormalisation group equations, QCD influences the relationship between the renormalised Higgs mass, Higgs self-coupling and top-quark Yukawa coupling (at large renormalisation scales) to the observed Higgs, W-boson and top-quark mass. The recent experimental masses together with the inclusion of higher loop corrections into the perturbative renormalisation group equations have lead to a strong modification of the Higgs potential in Fig. 1. The so-computed effective potential is just stable, but close to being unstable. One speaks of the stability/instability of the vacuum i.e. the ground state of the world. That the situation is really as sketched above cannot be considered as an established fact as yet since several assumptions enter into the logics. (i) the new boson observed at the LHC is interpreted to be the Standard model Higgs. (ii) the effective potential is assumed to be described accurately by perturbation theory in all couplings. (iii) the input parameters for the perturbation theory
are sufficiently well known. Assumption (ii) is put to a test in our study of the Higgs-Yukawa model. Concerning (iii) the most relevant uncertainties in the input parameters are the top-quark mass and the value of the strong coupling \( \alpha_s \) at a high scale, say, of around 100 GeV.

## 3 The Strong Coupling

Apart from the above question of the stability of the electroweak vacuum, the value of the strong coupling is important in many high energy processes at the LHC. In particular, together with the so-called parton distribution functions, which describe how quarks and gluons are “distributed” inside the protons, \( \alpha_s \) is the most important input parameter in the theoretical description of the production of the Higgs boson and other states from proton-proton collisions at the LHC.

The coupling is difficult to determine from experiments due to confinement. This property of QCD means that quarks and gluons exist only bound inside hadrons, such as the proton, as their constituents. One can therefore not directly determine the strength of a quark–gluon interaction in an experiment. Moreover, the question arises how one even defines the strength of the strong interactions.

### 3.1 Defining the Strength \( \alpha \) of the Strong Interactions (QCD).

In quantum electro dynamics (QED), the coupling is usually defined by the low energy limit of the scattering of photons on (free) electrons, the Thomson cross section. For the
aforementioned reasons, an analogous definition is not possible in QCD, but the potential energy between two static colour charges at large distances is very similar to Thomson scattering. This potential is nothing but the Coulomb potential governing all of atomic physics with quantum corrections included, such as multiple photon exchanges and virtual $e^+ e^-$ pair creation.

In QCD, the static potential is defined as the energy of a system with a static quark source at $x$ and an anti-quark at $y$. Static quarks have a divergent, unobservable, self energy in the quantum theory. One therefore discards in the potential the piece which is independent of the distance $r = |x - y|$ and considers the force,

$$F(r) \equiv \frac{d}{dr} V(r) = \frac{g_0^2}{4\pi} \frac{4}{3} \frac{1}{r^2} + O(g_0^4). \quad (1)$$

Here, we have also given the expansion of the force in terms of the bare coupling, $g_0$. This coupling appears in the Lagrangian of the theory – in exact analogy to the coupling between photons and electrons in QED. The bare coupling is again not observable; it suffers from the infamous divergences of the quantum field theory. In contrast,

$$\alpha_{qq}(r) \equiv \frac{\bar{g}^2_{qq}(r)}{4\pi} = \frac{3}{4} r^2 F(r), \quad (2)$$

is finite since it is defined in terms of the force and it is related to the bare coupling by a perturbative series $\bar{g}^2_{qq}(r) = g_0^2 + O(g_0^4)$. In the weak coupling region, where $\bar{g}_{qq}$ is small, it is connected to all kind of other definitions of the coupling in such a perturbative fashion; but it is also properly defined when it is large! By a proper definition we mean in particular that it is finite and invariant under gauge transformations which form the basic, defining, symmetry of QCD. The physical picture and Feynman graph behind Eq. 1 is that the leading term corresponds to the exchange of a single gluon between a (static) quark and an anti-quark. Apart from the factor $4/3$ originating from the SU(3) group of QCD, the leading term is equal to the one-photon exchange of the Coulomb potential.

### 3.2 Computing the Interaction Strength $\alpha$

The $r$-dependence of $F(r)$ – and therefore of $\bar{g}^2_{qq}(r)$ – can be computed self-consistently in perturbation theory when $\bar{g}_{qq}(r)$ is small\(^9,10\):

$$r \frac{d}{dr} \bar{g}_{qq}(r) \overset{r \to 0}{\sim} b_0 \bar{g}^3(r) + b_1 \bar{g}^5(r) + \ldots , \quad (3)$$

where $b_0 = (4\pi)^{-2} (11 - \frac{2}{3} N_f)$ and $N_f$ denotes the number of quark flavours “active” at the scale $r$. Asymptotically all quarks count, i.e. $N_f = 6$. Small coupling corresponds to small $r$, or equivalently to high momenta, via the relation

$$\bar{g}^2_{qq}(r) \sim \frac{1}{-2 b_0 \log(\Lambda_{qq} r)} . \quad (4)$$

The integration constant of the differential equation Eq. 3 is parameterised by $\Lambda_{qq}$ and is not accessible to perturbation theory.

This property, called asymptotic freedom, allows to analytically compute the evolution of $g_{qq}(r)$ to arbitrarily small distances once a starting value is known which is small
enough. We “only” have to enter the beginning of this asymptotic regime with a non-perturbative lattice simulation and connect the asymptotic region non-perturbatively to the low-energy properties of QCD, such as the proton mass. More precisely, one wants to express $r$ in units of $m^{-1}_{\text{proton}}$ for values of $r$ where $\alpha_{qq}$ is small, say around $\alpha_{qq}(r) = 0.2$ or smaller. Here, $m_{\text{proton}}$ has been chosen as a typical QCD scale. Others are possible and in fact a scale extracted from the weak decays of the lightest hadrons is technically advantageous \textsuperscript{11}. In a straight-forward attempt to perform this computation, one would simulate an $L^4$-world with $L > 6$ fm (where finite size effects are known to be small) and a certain lattice spacing $a$. One then determines the proton mass in lattice units, $m_{\text{proton}}a$, as well as $\alpha_{qq} = f(r/a)$. Setting $r \equiv \rho/m_{\text{proton}}$, one then knows $\alpha_{qq}(r) = f(\rho/(am_{\text{proton}}))$ up to effects due to the discretisation of the theory. Such a straightforward strategy is, however, not reliable since short distances of around $r \approx 0.1$ fm, where $\alpha_{qq}(r)$ is small, correspond to $r \approx L/60$. On the other hand, lattices with $L/a = 64$ are about at the limit of what can be simulated at present. In such a naive setup one would then determine $\alpha_{qq}(r)$ at $r \approx a$, where the discretisation effects can be very large. Obviously a better strategy is needed, which allows for $r \gg a$ and a continuum limit.

In the strategy developed and continuously improved by the ALPHA-collaboration and in particular the NIC group, the scale problem is solved in the following way \textsuperscript{12–14}. One introduces a definition\textsuperscript{a} of the coupling $\bar{g}_{\text{SF}}(L)$ which explicitly depends on the linear size $L$ of the system, i.e. $\bar{g}_{\text{SF}}(L)$ “runs” with $L$. For $L \approx 0.5$ fm to 1 fm, one connects $L$ to $m_{\text{proton}}^{-1}$. Then one performs so-called step scaling, namely one determines $\bar{g}_{\text{SF}}^2(L)/\bar{g}_{\text{SF}}^2(L/2)$ from two simulations and iterates according to the following scheme:

All of this is possible on lattices with moderate $L/a$ and moderate cost\textsuperscript{15,16}, apart from the simulations to determine the physical scale in lattice units, $am_{\text{proton}}$. For the latter, the “only” requirement is now $am_{\text{proton}} \ll 1$ to get discretisation effects under control. In each step a continuum limit can be taken by a controlled extrapolation of the numerical results.

### 3.3 Improving the Precision

The strategy involves the simulation of many lattices and high precision on each one of them. In a simplified version of the theory with only two quarks, the full strategy was carried out, yielding a $O(1\%)$ precision for the coupling at 100 GeV. For a realistic prediction relevant for LHC physics, the theory with at least three quarks needs to be considered and

---

\textsuperscript{a}Due to space limitations, we cannot describe here how this is done in detail. We just mention that the exact boundary conditions play an important rôle and that there are analogies to the electromagnetic Casimir effect.
one would also like to further improve the precision. The latter can now be achieved with a modified definition of the coupling. It is based on a flow equation\textsuperscript{17,18}, which evolves the fields in the path integral into fields which behave as classical fields in many ways. At leading order in the coupling, the flow equation is a heat equation and the quantum fields, $A_\mu(x)$, are related to the fields at flow time $t$ by

$$ B_{\mu,1}(x,t) = \int d^4y \ K_t(x-y) \ A_\mu(y), $$

where $K_t(z) = (4\pi t)^{-2} \exp \left\{ -z^2/(4t) \right\}$ is a heat kernel. Thus, the fields are smoothed over distances $\sqrt{8t}$. As a consequence the typical short distance singularities of the quantum field theory are removed\textsuperscript{18} and completely new observables become available. Furthermore, also the fluctuations of these new observables in our simulations of the path integral, are tamed, leading to very high statistical accuracy.

The group has recently studied the application of these general ideas to a new definition of a QCD coupling in a finite volume\textsuperscript{19}. Very good precision can indeed be reached\textsuperscript{19,20}. The challenge of the overall calculation of $\alpha(100 \text{ GeV})$ is then entirely reduced to the large volume computation (at small quark masses) of a scale such as $a m_{\text{proton}}$. This part has just been started, within a collaboration of the NIC group with other partners in the so-called Coordinated Lattice Simulations initiative. The initial simulations are supported by two PRACE projects, one focusing on reaching the physical quark masses and the other one on the extrapolation to the continuum limit.

From the combination of these simulations with the step scaling for the new coupling we expect to obtain the QCD coupling with unprecedented precision, both systematic and statistical. It will help to further probe the Standard Model of Particle Physics.

## 4 Summary

In this contribution we have presented two important aspects of testing the SM by non-perturbative lattice simulations: through studies of the mass bounds and vacuum stability in the Higgs sector, and through the precise determination of the strong coupling constant in QCD.

An other research topic of the NIC research group “Elementary Particles” at DESY is heavy flavour physics, where the lattice computations provide the basis for many precision tests of the SM through the determination of the relevant hadronic matrix elements.

The progress of lattice computations for all these research directions strongly depends on developments and improvements on various frontiers. These include innovative theoretical approaches, efficient simulation algorithms, sophisticated data analysis methods, and of course the efficient use of the growing performance of the computer platforms.

## Acknowledgements

We thank the John von Neumann Institute for Computing for its support of the research group Elementary Particle Physics through several projects on the supercomputers of the Gauss Centre, the JSC as well as the PAX cluster in Zeuthen. We thank the research group (http://nic.desy.de/members) for the excellent work and the pleasant working
atmosphere. In particular, we thank A. Nagy who has provided Fig. 1. Unfortunately, due to space limitations we here could only illuminate two aspects of the broad research of the BIC group. The collaboration with the group from Taiwan is financially supported by a DAAD travel grant.

References

Astrophysics
Astrophysics:
The Very Small and Very Large Universe

Peter L. Biermann

MPI for Radioastronomy, Bonn, Germany
E-mail: plbiermann@mpifr-bonn.mpg.de
Inst. Exp. Nucl. Phys., Karlsruher Institut für Technologie KIT, Germany
Dept. of Phys. & Astron., University of Alabama, Tuscaloosa, AL, USA
Dept. of Phys. & Astron., University of Bonn, Germany

What remains to be discovered out there is one of the most tantalising questions in science. Here we briefly describe recent attempts to understand the gravitational wave signal from neutron star binary mergers, the plasma environment around a comet relevant for the past and future of our Earth and its space weather, a large volume simulation of our cosmos to comprehend better where normal matter is, what dark matter may be, and what the elusive dark energy might be, and finally the complex environment of massive stars, which provide the baseline for the acceleration of the most energetic Galactic cosmic ray particles, at energies far beyond the LHC at CERN.

From Neutron Stars to the Highest Energies in our Galaxy

Gravitational waves remain yet to be directly discovered, and one of the best candidates to detect them could be the merger of a tight binary neutron star system, forming a new black hole as a result. Together with the gravitational wave emission there should be other emissions, in electromagnetic waves, and quite possibly also in neutrinos. Neutrinos have now been detected as a presumably extragalactic background, and so the hunt is on. Simulations on super computers can be used to understand the kind of in-spiral of the two neutron star ahead of the merger, and the concomitant other emissions. Since it is plausible that quantum gravity - an as yet undeveloped theory - also plays a role in the formation of black holes, these simulations are extremely important so as to get an observational handle on the key process of black hole formation. Short Gamma Ray Bursts might have to be explained as neutron star mergers, and in that case we have some observations already, albeit at very large distances.

Comets are famous as the first known witnesses for the existence of the Solar wind; comets may have delivered much of the water to our young Earth. Nowadays we are able to send space craft to comets to explore them directly. The magnetic ionised gas flow around comets allows us to test plasma physics in the limit, when the collisional mean free path is larger than the system, and when injection of new initially neutral particles takes place. Furthermore, since the flow from the Sun is so unsteady, sometimes referred to as space weather, we additionally learn about the effect of the extreme outbursts from the Sun, as they affect the flow around a comet. In the case of the comet Comet 67P/Churyumov-Gerasimenko we will have in situ observation by the ROSETTA space craft, and we also
have plasma simulations in advance so as to maximise our understanding. Since space weather can seriously impact our civilisation, all attempts must be made to understand the unsteady plasma flow from the Sun.

The universe expands, and is composed of normal baryonic matter, dark matter, and dark energy; the sum of these three contributions is unity, implying that the geometry is “flat”, which says, that the sum of the angles around an arbitrary triangle in the universe is exactly 180 degrees - as long as one does not get too close to a black hole along the way. We do not know, what dark energy is, what dark matter is, and where most of the normal matter resides. We know about less than one percent. And yet, using simulations we can describe the evolution of structure in the universe remarkably well, as long as the length scales are large. Therefore it is of outmost importance to develop such very large scale simulations to find the limits, so as to discover boundary conditions for our attempts to describe dark matter and dark energy. The Nobel prize of 2011 was given for the discovery of dark energy, a force that pushes the universe into accelerated expansion. Ultimately we need to embed dark matter and dark energy into our world of physical understanding, from the smallest to the largest scales.

Very massive star in their life and in their explosions have provided all the heavy chemical elements starting with the atom essential for life, Carbon. Already during their life these stars drive very powerful winds, with their accumulated energy output coming close to their explosion energy at the end; these winds expose the deeper layers of the star throwing out soot in some cases. These stars occur in binary systems, and in groups, and as a system, move throughout the interstellar medium. Such stars probably were the very first in the universe to also eject magnetic fields. Given magnetic fields their explosions also accelerate energetic particles - nuclei - to energies far beyond anything doable at the LHC at CERN. In short such stars, their life and their explosions provide the key drivers of evolution in our Galaxy. All these effects can be simulated, tested, further developed, and need to in order to comprehend our Galaxy.

All this work requires the use of super computers.

References

Neutron star binaries are ubiquitous in the universe, and are expected to be prominent sources of both electromagnetic and gravitational radiation in the final stages of their evolution, where the binary is driven to merger by the emission of gravitational radiation. In order to measure the properties of the binaries from the detected gravitational wave signals and elucidate what electromagnetic signals one can expect, one needs accurate numerical modelling of the final stages of neutron star binary evolution. Here we present simulations that expand the parameter space of neutron star binary simulations to include eccentric orbits and spinning neutron stars. We also show that a formulation of the Einstein equations we have recently developed leads to more accurate simulations than the current standard formulation.

1 Introduction

Neutron stars are the end result of supernova explosions of stars with masses around 10–20 solar masses ($M_\odot$), and are supported against gravitational collapse by neutron degeneracy pressure. Neutron stars are extreme objects in many ways, with up to twice the mass of the Sun squeezed into a ball with a radius of about 10 km, resulting in maximum densities of a few times the density of atomic nuclei on Earth. Indeed, neutron stars are the only places in the universe where one can probe theories of cold, dense matter at such densities. There is considerable theoretical uncertainty in the description of neutron star interiors, and thus an attendant uncertainty in the predictions for almost all observables\(^1\). It is hoped that comparing these theoretical predictions against future observations will reveal the properties of cold matter at these densities. Neutron stars also have the largest magnetic fields observed in the universe (up to $\sim 10^{15}$ times the Earth’s magnetic field) and impressively short rotation periods for such large objects (the shortest observed to date\(^2\) is $\sim 1.4$ ms, so the surface of this star is moving at $\sim 20\%$ of the speed of light).

Just as with all other types of stellar objects, one frequently finds neutron stars in binaries with all different types of stars and compact objects. In particular, there is a collection of double neutron star binaries, including the famous Hulse-Taylor pulsar and the double pulsar, which have provided some of the most exquisitely precise tests of general relativity, notably providing indirect evidence for the existence of gravitational radiation (the gravitational equivalent of electromagnetic radiation, predicted by Einstein’s theory of general relativity). The gravitational radiation carries away energy and angular momentum from the binary, and one observes the effects of these losses on the binary’s orbit. Eventually, these losses drive the binary to coalescence, by which time the orbit will have circularised in all but a few exceptional cases. The last phases of this evolution are highly relativistic and the theoretical predictions that are the focus of our project can only be made using
numerical techniques that solve the full field equations of general relativity (Einstein’s equations). Binary neutron star simulations in general are reviewed in Ref. 3.

One of the primary reasons we model these systems is to study their gravitational wave signals: Neutron star binaries are a prominent source for ground-based gravitational wave detectors such as GEO600\(^4\), the Laser Interferometer Gravitational-wave Observatory (LIGO)\(^5\), and the Virgo Observatory\(^6\). However, realistic sources are distant (often many millions of light years away), so even though these systems radiate strongly in gravitational waves, the gravitational wave signals one expects to receive on Earth are very weak. One therefore needs numerical modelling to develop accurate templates for the radiation one expects from such systems in order to be able to detect them in the detector’s noise and then infer their properties from the detected waveform. In particular, the properties of cold, dense matter are imprinted on the gravitational wave signal through their influence on stellar structure, most notably through the effects of tidal deformations\(^7\).

Additionally, the mergers of binary neutron stars are thought to produce significant electromagnetic and neutrino emission (e.g., short gamma-ray bursts), as well as enriching the surrounding interstellar medium with heavy r-process elements through their ejecta. Modelling these signatures accurately requires the addition of considerably more physics besides general relativistic hydrodynamics, so we do not consider them in detail, though we do look at the bulk properties of the ejecta in our simulations.

The very strong gravity present in neutron stars (Newtonian surface gravity of \(\sim 10^{11}\) times the Earth’s) requires Einstein’s theory of general relativity for an accurate description. For the modelling of highly dynamical scenarios involving neutron stars, such as binary merger, one needs the methods of numerical relativity, which involve writing the Einstein equations in a form amenable to numerical treatment, and then performing simulations on a large supercomputer: The Einstein equations are highly nonlinear, and even simply evaluating the “right-hand side” of the evolution equations requires thousands of floating point operations for each grid point. Additionally, there are only a very small number of reformulations of the Einstein equations which allow for successful simulations; we discuss our recent addition to this collection in Sec. 3.

Our binary neutron star simulations have focused on three different situations: First, performing highly accurate simulations of quasicircular binary neutron stars to measure the tidal effects in the waveform\(^8\), which we use to calibrate the analytic waveform models that will be employed to detect these binaries and measure their properties\(^9\). We will not discuss these simulations any further here. Second, simulating eccentric neutron star binaries: Our group presented the first such simulations\(^10\), and we are now working to make our simulations of these systems more accurate and realistic, as described in Sec. 4. Third, simulating spinning neutron star binaries. This is a project we have just started recently, using the initial data that one of our collaborators constructed\(^11,12\) in the past few years; these simulations are described in Sec. 5.

2 Computational Setup and Code Performance

For our fully general relativistic simulations we use the BAM code\(^13,14\), which combines state-of-art methods to deal with black hole spacetimes and general relativistic hydrodynamics. The code is based on the method of lines, and uses high-order finite difference stencils for the spatial discretisation of the geometric variables (up to 10th order for pure
vacuum simulations and 6th order for matter simulations), while high resolution shock-capturing methods are used for the hydrodynamic variables. The time integration is done with an explicit Runge-Kutta method. The BAM infrastructure also supplies adaptive mesh refinement via a combination of fixed and moving boxes.

We have recently upgraded our MPI parallelisation scheme to hybrid OpenMP/MPI (OMP/MPI) parallelisation, which helps to increase memory efficiency and reduce communication overhead by decreasing the number of MPI jobs that have to be started. We demonstrate the scaling of our code for up to \( \sim 3000 \) cores on JUROPA in the left panel of Fig. 1. The benchmark simulation consists of a single black hole covered by seven refinement levels, with \( 260^3 \) points in each level. The scaling up to such a high number of processors is close to ideal, but seems to decrease slightly for the highest numbers of cores. However, \( \sim 1000 \) cores are typically used only when we need very high accuracy, i.e., a large number of grid points. Thus in the right panel of Fig. 1 we show a weak scaling test. This shows that the code’s performance increases with an increase in the number of cores up to 1536 cores.

3 The Z4c Formulation of General Relativity

The Einstein equations are initially cast as an equation for the entire spacetime, so the first step in any numerical treatment is necessarily rewriting them in the form of an evolution equation. This was first done in the 1950s, though the resulting equations are not well suited for numerical evolutions. However, later work has developed two better-behaved variants that have been used for all numerical relativity production simulations to date: The BSSNOK formulation, which is used for the other simulations presented here, and is the choice of the majority of numerical relativity groups, and the generalised harmonic formulation, which is used by a few groups. See the review by Sarbach and Tiglio\textsuperscript{15} for details about these formulations.

Each of these formulations has certain advantages and shortcomings: BSSNOK allows for a free choice of gauge (letting one select a gauge that allows for simple evolutions of
black holes), but does not provide an optimal treatment of constraint violations or have successfully implemented constraint-preserving boundary conditions. (General relativity has constraint equations that must be satisfied at each evolution step; these are analogous to the divergence equations in electrodynamics.) Conversely, the generalised harmonic formulation has a scheme to exponentially damp away constraint violations, and radiation controlling constraint-preserving boundary conditions, but comes tied to a particular gauge, requiring a more cumbersome method of black hole evolution.

In Ref. 16 we proposed a formulation, known as Z4c, with the advantages of both the BSSNOK and generalised harmonic formulations. We have now completed the development of this formulation with the derivation of constraint-preserving boundary conditions, and have tested it on compact binaries, finding that it improves the accuracy of both binary black hole and binary neutron star simulations\textsuperscript{17}. In particular, one sees smaller constraint violations (see Fig. 2), cleaner convergence of the waveforms with resolution, and better conservation of energy with Z4c than with BSSNOK. We thus anticipate that Z4c will become our standard for production simulations in the near future, and several other groups have now also started to implement and use it\textsuperscript{18}.

### 4 Eccentric Neutron Star Binaries

Most neutron star binaries are expected to have orbits that are very close to circular near merger, due to the circularising effects of gravitational wave emission, as discussed in Sec. 1. However, there are certain situations in which a binary could form with a sufficiently tight and eccentric orbit that it merges without first shedding all its eccentricity. Such situations take place in dense stellar environments, and recent work has shown that there likely exists a small population of such eccentric neutron star binary mergers\textsuperscript{19}.

In Ref. 10, we presented the first numerical relativity simulations of highly eccentric neutron star binaries (work led by R. Gold, who was at that time working on this project),
and found several interesting features. Most notably, we confirmed the Newtonian prediction by Turner\textsuperscript{20} of tidally induced oscillations in the neutron stars after their close encounters. We also found that the accretion disks could be quite massive, up to $\sim 0.3M_\odot$, making these systems potential short gamma-ray burst progenitors.

![Figure 3](image.png)

Figure 3. Tracks (left) and gravitational waveforms (right) for a $(1.3, 1.4)M_\odot$ eccentric neutron star binary with the SLy EOS, showing the tidally induced oscillations of the neutron stars in the gravitational wave signal, and the difference between the cases with and without a thermal component in the EOS, in a preliminary low-resolution run. (The tracks are computed using the coordinate position of each star’s maximum density.)

However, these simulations were carried out using constraint-violating initial data (the constraint violations were small, but not at the level of truncation error), and a very simple description of the neutron star matter, viz., a polytropic equation of state (EOS). In addition, we only considered the simplest case of an equal-mass binary and did not make careful convergence tests or error estimates. We have since been working to obtain constraint-solved initial data, and to quantify and improve the accuracy of our simulations, including obtaining relatively clean convergence in a simple test case. We have also been experimenting with adding more physics using our original method of initial data construction. In particular, we have now considered the unequal-mass case with a realistic equation of state (the Douchin and Haensel\textsuperscript{21} SLy model), including phenomenological thermal effects. As illustrated in Fig. 3, in a preliminary run, we find that including thermal effects has a significant impact on the orbital dynamics, which we are still investigating.

5 Spinning Neutron Star Binaries

While there is a population of highly spinning neutron stars, the known neutron star binaries all have much more modest spins, so previous numerical studies of neutron star binaries have focused primarily on the irrotational case, since most of the observed spins would have a negligible effect on the orbit. There have also been a small number of simulations in the corotational case (i.e., with a vanishing fluid velocity in the frame corotating with the binary), though it is known that the viscosity in neutron stars cannot be high enough to lead to corotation\textsuperscript{22}. However, one member of the double pulsar has a rotation period of 22.7 ms\textsuperscript{23}, which will only have increased to $\sim 27$ ms at merger. Such an orbital frequency
is still nonnegligible compared to the orbital frequency of a few milliseconds during the last orbits, and strongly motivates numerical simulations of such systems.

Numerical relativists have only started to simulate spinning binary neutron star mergers very recently\(^\text{24, 25}\), and these simulations used constraint-violating initial data as well as an improperly normalised matter velocity field. However, consistent and constraint solved initial data for spinning binary neutron stars have been constructed\(^\text{11, 12}\), and we have evolved such data for the first time.

![Figure 4](image.png)

**Figure 4.** Comparison of the matter’s rest-mass and momentum densities for binary neutron stars with antialigned (left column), irrotational (middle column), and aligned spins (right column). In all cases the initial proper distance of the stars is \(\sim 45\) km.

Fig. 4 shows the matter’s rest-mass and momentum densities during the different phases of the evolution: inspiral, hypermassive neutron star (HMNS)-phase, and black hole with accretion disk. The left and right columns show equal-mass, equal-spin systems with antialigned and aligned spins, respectively, with a spin period of around \(13.4\) ms. The middle column illustrates an irrotational model. In all three cases, each binary component has a mass of \(\sim 1.525M_\odot\) in isolation. Note that an orbital period of 13.4 ms corresponds to a dimensionless spin (i.e., the body’s angular momentum scaled by the maximum angular momentum of a black hole of that mass) of \(\sim 0.045\), which is small compared to black hole simulations where the dimensionless spin is approximately one order of magnitude higher. However, imprints of the spin are visible during the evolution. Comparing the figures in the first row of Fig. 4, we see the orbital hang-up effect, i.e., the neutron stars are closer for the antialigned than for the aligned configuration (note that the binary is orbiting
in a counterclockwise direction). This effect was first observed in binary black hole evolutions\textsuperscript{26}, but has recently also been seen for binary neutron stars\textsuperscript{25}. The next two rows show the HMNS and (in the antialigned and irrotational cases) the final black hole.

The most important difference between the three configurations is the time needed to form a black hole. The black hole forms earlier in the antialigned case than in the irrotational case, and has not formed by the end of the simulation in the aligned case. The different lifetimes are due to differing amounts of centrifugal support of the HMNS: During the HMNS-phase the neutron star exceeds the maximal mass a non-rotating cold neutron star could have, so the HMNS will eventually collapse to a black hole as it loses rotational and thermal support. Thus, increasing the total angular momentum by adding spins to the individual neutron stars aligned with the orbital angular momentum increases the time needed to form a black hole. However, when the HMNS survives as long as in the aligned spin-case, additional physics we have not simulated (e.g., neutrino emission and magnetic fields) plays an important role and has to be included in further studies.

6 Conclusions

Binary neutron stars are important astrophysical systems, and potential sources for electromagnetic, gravitational wave, and neutrino detectors. Numerical simulations are the only way to theoretically probe the final stages of these systems’ evolution, and we have a well-developed program of performing such simulations. In particular, we have shown that a new formulation of the Einstein equations we have developed improves the accuracy of these simulations, compared with the current standard formulation. We have also given the results from some of our explorations into hitherto unconsidered corners of the binary neutron star parameter space. In particular, we have presented the first simulations of eccentric neutron star binaries and the first simulations of spinning neutron star binaries with consistent initial data. We continue to explore both of these portions of parameter space.

Acknowledgements

This work was supported by the DFG SFB/TR7 and Graduiertenkolleg 1523 as well as the Graduierten-Akademie Jena. Computations were performed on JUROPA, SuperMUC, and Louhi.

References

5. http://www.ligo.caltech.edu/
The Plasma Environment of Comet 67P/Churyumov-Gerasimenko

Christoph Koenders, Karl-Heinz Glassmeier, and Ingo Richter

Institut für Geophysik und extraterrestrische Physik, Technische Universität Braunschweig, 38106 Braunschweig, Germany
E-mail: {c.koenders, kh.glassmeier, i.richter}@tu-braunschweig.de

In 2014 the Rosetta spacecraft will arrive at comet 67P/Churyumov-Gerasimenko. Among others, 5 plasma instruments are on board and will study the evolution of the cometary plasma environment. In order to prepare the measurements we use the A.I.K.E.F. code to study the interaction. The A.I.K.E.F. code is based on the hybrid model and is able to describe important kinetic effects. In this article we will describe the used code and the important structures and boundaries in the plasma environment which will occur during the mission.

1 Introduction

ESA’s spacecraft Rosetta has started its unique mission to comet 67P/Churyumov-Gerasimenko (CG) in March 2004. The goal of the mission is to study the origin of the solar system and the evolution of the comet during the journey around the sun. In summer next year, at a heliocentric distance of about 3.5 AU, Rosetta will arrive at the comet and will start its prime mission. A major objective is the smooth landing of the lander Philea on the surface of the nucleus in November 2014, which will be the first landing on a comet. Afterwards, Rosetta will stay close to the comet for about one and a half year and the experiments on board will perform their measurements.

Among these experiments, the Rosetta Plasma Consortium (RPC), a package of 5 instruments, will study the interaction between the solar wind and the comet and their evolution during the mission. This interaction is based on the emanating neutral gas from the nucleus due to heating by insolation. This neutral gas forms an extended exosphere which is mixed with the impinging solar wind and the interplanetary magnetic field. Solar UV radiation leads to the ionisation of the neutral molecules, which then interact with the solar wind. The type, position and other properties of the resulting structures depend on many parameters which will vary during the mission. In order to prepare the measurements, the experiments need to know the location of the structures in space and in time.

Furthermore, most of these structures will be observed for the first time, although other spacecraft already visited comets, e.g. the Giotto mission to comet 1P/Halley. But all the former missions differ from the Rosetta mission in two main facts: 1) the spacecraft have only performed a single flyby, thus the experiments could only take snapshots of the interaction and 2) the gas production rate of the comets, a major factor of the interaction, were higher as in the case of CG. This is why models have to be used to study the interaction region of CG prior to the main mission phase.

Due to the fact that comet CG has a small gas production rate, the triggered structures will also be small and their scales will be comparable to the scales of the ion motion. Since the single ion motion differs from the fluid motion, this interaction cannot properly be...
described by fluid models. This is why one has to ensure that the kinetic effects of the ions are taken into consideration.

2 The Hybrid Model and the A.I.K.E.F. Code

An ion in a plasma gyrates around the magnetic field. At standard solar wind conditions at about 3.5 AU, a cometary ion has a gyroradius of about 50000 km. In contrast to that, the main interaction region of the comet has only a radius of about 300 km. This is why the kinetic aspect has to be taken into account in order to model the comet, which makes the modelling of weak outgassing comets much more complicated.

This project uses the A.I.K.E.F. (Adaptive Ion Kinetic Electron Fluid) code which was developed by Müller et al. (2011) and is based on the hybrid model, in which the ions are described as particles. Due to that kinetic effects can be described properly. In order to save computational resources, the ions are gathered to macroparticles, which can move to all positions inside the simulation box. In total, up to 7 billion macroparticles are in a simulations box with a very high resolution. The Lorentz force acts on the macroparticles and the equations of motion

\[
\frac{d}{dt} \vec{x}_i = \vec{v}_i \tag{1}
\]

\[
\frac{d}{dt} \vec{v}_i = \frac{q_i}{m_i} \left( \vec{E} + \vec{v}_i \times \vec{B} \right) \tag{2}
\]

are solved for each macroparticle at each simulation step. The charge, mass, position and the velocity of a macroparticles are labelled by \( q_i, m_i, \vec{x}_i \) and \( \vec{v}_i \). The entities \( \vec{E} \) and \( \vec{B} \) are the electric and magnetic field at the position of the ion. Based on the locations and the velocities of the ions, the A.I.K.E.F. generates the densities \( n_{SW} \) of the solar wind ions and \( n_{CI} \) of the cometary ions and mean velocities of each ion species \( \vec{u}_{SW} \) and \( \vec{u}_{CI} \) at the nodes of a numerical mesh by the Cloud-In-Cell method.

Since the scales of the electron motion are orders of magnitudes smaller than the scales of the ions, the hybrid model described them as a massless fluid. In addition the model assumes quasi-neutrality, which means that the density of the electrons is equal to the sum of the ion density at the location \( n_i \). The assumption of massless electrons allows us to use the momentum equation of the electron fluid to gain an expression for the electric field

\[
\vec{E} = -\vec{u}_i \times \vec{B} + \frac{1}{\mu_0 m_i} \left( \nabla \times \vec{B} \right) \times \vec{B} - \nabla \frac{p_e}{n_i}, \tag{3}
\]

where \( \vec{u}_i \) is the mean velocity of all ions and \( p_e \) the pressure of the electrons. This equation is solved by using the CAM method by Matthews.

The magnetic field can be obtained by using Faraday’s law

\[
\frac{\partial \vec{B}}{\partial t} = -\nabla \times \vec{E}. \tag{4}
\]

In order to close the system of equations, an adiabatic law is assumed for the pressure of the electrons. However, observations at comet 1P/Halley showed hot electrons in the
solar wind and cold electrons in the inner cometary environment. Based on that the model assumes an electron fluid with different temperatures

\[ \nabla p_e \approx \frac{p_e,SW,0 \nabla n_{SW} + p_e,CI,0 \nabla n_{CI}}{(n_{SW} + n_{CI})}, \]

where \( p_e,SW,0 \) and \( p_e,CI,0 \) denote the initial pressure of the solar wind and the cometary electrons.

As mentioned above the macroscopic quantities are calculated and stored on the nodes of a numerical mesh. This mesh can be a uniform Cartesian, but also a hierarchical or adaptive one. To use the hierarchical or adaptive mesh the A.I.K.E.F. code uses the Hybrid-Block-AMR method, details see Müller et al. (2011). However, in order to simulate the cometary environment, a hierarchical or adaptive mesh are highly needed because the triggered boundaries have different spatial scales. As an example, the comet at a heliocentric distance of 1.3 AU has an extended exosphere with a radius of about \( 1.7 \times 10^6 \) km, but the bow shock, the largest shock structure in the environment, has only a radius of about 2000 km. Closer to the nucleus the scales of important boundaries become even smaller, for example, the cometary ionopause has a radius of 40 km.

However, although we use the hierarchical mesh the total interaction region does not fit into our simulation box. Thus, we have to use an extended upstream boundary model, which describes the interaction of the flow in front of the simulation box. This can be done by a semi-kinetic analytical model because no main structures occur in these upstream regions. Details on the extended upstream boundary model are described in Koenders et al. (2013).

The main interaction process of the comet is the ionisation of cometary neutrals, which will be picked-up by the impinging solar wind. In A.I.K.E.F. this process is modelled by injecting new cometary ions into the simulation box at each time step. So far, the positions of the new ions are set to fulfil a isotropic distribution

\[ N_s = \frac{\nu Q}{4\pi u_{ng} r^2}, \]

which describes the isotropic radial emanating of the neutral gas and the photoionisation \( \nu \). \( Q \) denotes the gas production rate, \( u_{ng} \) the speed of the neutral gas, and \( r \) the radial distance. Other important processes are the collision, charge exchange and the recombination of ions, which are modelled by a Monte Carlo processes, since the probabilities of a reaction differ strongly between the ions and the processes.

### 3 The Evolution of the Interaction

At the time of the arrival of the Rosetta spacecraft at the comet in Summer 2014, the expected gas production rate is about \( Q \approx 10^{26} \) s\(^{-1}\). This small rate will only lead to a small perturbation of the solar wind and the interplanetary magnetic field. The result of a hybrid simulation of this gas production rate is shown in Fig. 1a and d. Other parameters of the interaction are listed in Tab. 1.

Since the solar wind and the interplanetary magnetic field are only slightly disturbed, the electric field, Eq. 3, is dominated by the convective part

\[ \vec{E}_{con} = -\vec{u}_i \times \vec{B}. \]
Here, $\bar{u}_i$ is the mean velocity of all ions at this position, which is nearly the solar wind speed in most regions of the simulations box. Since all ions (index $s$) are acted by the weakly disturbed magnetic field and the convective electric field
\begin{equation}
\vec{F}_L = q_s \left( \vec{u}_s - \bar{u}_i \right) \times \vec{B},
\end{equation}
the newborn cometary ions perform a cycloidal motion because their initial velocity can be neglected in comparison to the solar wind velocity. However, due to the properties of the solar wind and the magnetic field at these large heliocentric distances the arcs of this motion have a length of about 100 000 km. The beginning of this cycloidal pick-up ion tail can be observed in the southern hemisphere of Fig. 1c. Here also a small increase of the magnetic field strength can be observed.
Table 1. Parameters used for the simulations presented in this study.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gas production rate $Q$</td>
<td>$10^{26} - 5 \cdot 10^{27} \text{ s}^{-1}$</td>
</tr>
<tr>
<td>cometary ion mass $m_i$</td>
<td>19 amu</td>
</tr>
<tr>
<td>ionisation rate $\nu$</td>
<td>5.88 $\text{s}^{-1}$</td>
</tr>
<tr>
<td>neutral gas velocity $u_{\text{ng}}$</td>
<td>1 km/s</td>
</tr>
<tr>
<td>solar wind number density $n_{\text{sw}}$</td>
<td>6 cm$^{-3}$</td>
</tr>
<tr>
<td>solar wind velocity $u_{\text{sw}}$</td>
<td>400 km/s</td>
</tr>
<tr>
<td>strength of interplanetary magnetic field $B_{IMF}$</td>
<td>4.9 nT</td>
</tr>
<tr>
<td>Parker angle $\theta$</td>
<td>$90^\circ$</td>
</tr>
</tbody>
</table>

An additional increase in the field can be seen behind the nucleus on the northern hemisphere. Below, the field strength is decreased. This structure is called a Mach cone since the opening angle is correlated with the Mach number of the solar wind flow. It is triggered by deflected solar wind ions which pass the region close to the nucleus. There the mean ion velocity $\bar{u}_i$ is reduced due to the new and resting cometary ions. Thus, the Lorentz force pushes the solar wind particle to the northern hemisphere. Since the magnetic field is frozen into the solar wind, the enhanced density in this hemisphere leads to an enhanced magnetic field. In contrast, the upward deflected solar wind ions produce a void in the southern hemisphere, where also the magnetic field is reduced.

On the way to the sun, the gas production rate will increase and more cometary ions will be ionised, which will reduce the mean ion velocity $\bar{u}_i$. Therefore, the deflection of the impinging solar wind ions becomes stronger. Based on that, the Mach Cone becomes asymmetric, which can be seen in Fig. 1b. Furthermore, the perturbation in the inner region leads to a modified cometary ion tail. First, the ions move into the void of the Mach cone, but there the ions are successively decelerated to the southern hemisphere, see Fig. 1e.

Close to the perihelion the comet will produce so many ions that the Mach cone will transfer into a bow shock. In contrast to the smooth transition of the plasma quantities at a Mach cone, the bow shock is a discontinuity at which the field strength jumps on several quantities. One of them is the tangential velocity of the flow, which jumps from supersonic to subsonic speeds and allows the flow to stream around the obstacle in the inner coma. The RPC experiments intend to perform measurements of this important structure and the mission needs information about its location. Since the bow shock originates from a Mach cone and since the kinetic effect can not be excluded, Koenders et al. study the location of the bow shock during the Rosetta mission with several hybrid simulations. One of the major results is that the position of the shock changes vastly when small changes in the solar wind conditions occur. For example, the position of the bow shock changes from about 1500 km at an IMF strength of 2 nT to about 3300 km at an IMF strength of 9 nT. This result based on hybrid simulations contrast with those of magnetohydrodynamics models, but it can be explained with the inertia of the cometary ions. Based mainly on these results the Rosetta spacecraft will make an excursion to the bow shock in autumn 2015.
However, the spacecraft will mostly stay close to the comet during the mission. This is why this project sets a major focus on the inner coma, too. In order to resolve the innermost region of the cometary interaction and the boundaries therein, the hybrid simulations have to use a hierarchical mesh. The simulation presented in Fig. 2 has a global box size of about 11000 km edge length to avoid boundary effects. In regions far away from the nucleus, the simulation uses a coarse mesh resolution of 75 km to resolve the bow shock, but the closer to the nucleus the higher the resolution. The maximum resolution of 2 km is reached at about 75 km distance to the nucleus. Based on the small cell volume in the inner coma and the particle character of the hybrid simulations, up to 7 billion macroparticles in one simulation are needed to describe the processes in the inner coma.

One of the features resolved by a high resolution hybrid simulation is the magnetic pile up region, see Fig. 2d. Within this region two plasma flows collide; from the sun-
ward direction the solar wind, massloaded by cometary ions, and from the comet a purely cometary plasma flow approach this region, and due to that several transitions occur. At a distance of about 120 km the solar wind density drops and only the massloaded cometary ions enter the innermost part. This boundary is called ion composition boundary. The impinging cometary ions are further decelerated and at about 75 km subsolar of the nucleus the plasma stagnates. Since the magnetic field is mostly frozen-in to the flow, the magnetic field is piled up in this region. The enhanced and draped magnetic field can be seen in Fig. 2d.

Close to the maximum in the magnetic field strength, the magnetic field drops within 40 km from 70 nT to zero. This transition is called cometary ionopause and this boundary separates the magnetic pile up region and the diamagnetic cavity. A force balance between the magnetic forces in the magnetic pile up region and a drag force are balanced at this boundary. The drag force is based on collisions between ions and the neutral gas, which moves radially outward. So far, our simulations assume an isotropic outgassing, but we expect that local inhomogeneities, e.g. local jets, disturb the boundaries and that waves will be triggered in this region. However, our simulations are the first global hybrid simulations of the cometary interaction region which resolve the magnetic pile up region and the diamagnetic cavity.

4 Summary and Outlook

This project studies the plasma environment of comet CG with the A.I.K.E.F. code prior to the main phase of the Rosetta mission. The code is based on the hybrid simulations and the ions are described as particles. This feature requires lots of numerical resources, but based on that particle description, the model can describe important kinetic effects of the ions. Since our model is the only available model, which allows a global simulation of the cometary environment and considers the kinetic effect of the ions, it is used to prepare the measurements of the RPC experiments.

One of the major scientific objectives of these experiments is the study of the evolution of the cometary environment in the inner solar system. As presented above, the comet will trigger a Mach cone at the beginning of the escort phase. While the comet approaches the sun the gas production rate will increase and the Mach cone will change into a bow shock. In addition to the large scale structures, our simulations detects several boundaries in the inner coma of CG at 1.3 AU. These are the first hybrid simulations which could resolve these structures properly. Based on our success, we intend to study the formation of these boundaries prior to the arrival of Rosetta at the comet.

Acknowledgements

The work was financially supported by the German Bundesministerium für Wirtschaft und Technologie and the Deutsches Zentrum für Luft- und Raumfahrt under contract 50 QP 1001 for Rosetta. The simulations were performed with a grant of computer time on JUROPA provided by the NIC and with a grant by the the system of the North-German Supercomputing Alliance.
References


A Coherent Hubble Volume Simulation
for All-Sky ISW Predictions and Large Scale Surveys

Stefan Gottlöber\textsuperscript{1}, Jose Maria Diego\textsuperscript{2}, William Watson\textsuperscript{3},
Ilian Iliev\textsuperscript{3}, and Gustavo Yepes\textsuperscript{4}

\textsuperscript{1} Leibniz-Institut für Astrophysik (AIP), An der Sternwarte 16, 14482 Potsdam, Germany
E-mail: sgottloeber@aip.de
\textsuperscript{2} IFCA, Instituto de Física de Cantabria (UC-CSIC),
Avda, Los Castros s/n, 39005 Santander, Spain
E-mail: jdiego@ifca.unican.es
\textsuperscript{3} Astronomy Centre, Department of Physics & Astronomy, University of Sussex,
Falmer, Brighton, BN1 9QH, United Kingdom
E-mail: \{w.watson, i.t.iliev\}@sussex.ac.uk
\textsuperscript{4} Departamento de Física Teórica, Modulo C-XI, Facultad de Ciencias,
Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain
E-mail: gustavo.yepes@uam.es

We present first results from the Jubilee ISW project, a large $N$-body cosmological simulation
with a volume of $V = (6\text{Gpc}/h)^3$ and 216 billion particles.

1 Introduction

In 1965 Arno Penzias and Robert Wilson detected the cosmic microwave background radiation. More than 13 billion years ago this radiation was imprinted on the sky, only a few 100,000 years after the Big Bang. In 1992 the COBE satellite detected anisotropies in the temperature of the CMB radiation. Meanwhile these temperature fluctuations are measured with very high precision by satellites (WMAP\textsuperscript{7}, Planck\textsuperscript{8}) as well as many ground based observations. The measured temperature fluctuations tell us that shortly after the Big Bang the Universe was almost homogeneous with tiny density fluctuations of the order of $10^{-5}$. Comparing the power spectrum of measured temperature fluctuations with theoretical models, cosmologists conclude that the Universe is spatially flat and consists at present of about 68\% of some unknown Dark Energy, 27\% of also unknown Dark Matter and 5\% of baryons. In the evolved universe one can directly observe the distribution of baryons and indirectly deduce (gravitational lensing, velocity measurements) the distribution of Dark Matter. We see huge clusters of galaxies with masses up to a few $10^{15}$ solar masses in the knots of the cosmic web which comprise galaxies in a wide range of masses from dwarfs ($10^9$ solar masses) to massive elliptical galaxies ($10^{13}$ solar masses). All these structures have evolved from tiny fluctuations generated during the early inflationary phase and measured in the CMB background. The formation of structure on large scales is well understood within the concordance model of cosmology. The initial small perturbations grow by gravitational instability and form bound objects called halos which decouple from the expansion of the universe. These bound objects grow hierarchically by accretion of matter and merging with smaller halos. The gravitational clustering becomes
increasingly non-linear. Dark Matter is more abundant and hence becomes the most important for formation of large scale structures where gravity dominates. On smaller (galactic) scales baryons play an important role. They interact not only gravitationally but form a gas with a certain pressure and temperature. In the gravitational potential wells of the Dark Matter halos the originally hot gas cools via radiative cooling and finally stars are formed of the cooled gas. Almost all chemical elements above helium are formed in stars and then redistributed into the cosmic medium. Star formation and the feedback of the stars on the gas are the most important processes in galaxy formation. Nevertheless due to the domination of gravity at early times galaxies are closely associated with the Dark Matter halos.

The vastness of scales and the non-linearity of gravitational clustering are the reasons why numerical simulations and the intensive use of the largest supercomputers are the only methods suited to study the gravitationally driven growth of structures down to the high over-densities in halos and the formation of galaxies therein. Cosmological simulations follow the clustering of matter by numerically solving the gravitational interaction based on an $N$-body approach. Additionally one needs to model hydrodynamical processes, radiative cooling, star formation and the feedback of stars, in order to simulate in detail the formation of galaxies in their different environments. However, in large volumes and for a large number of galaxies it is practically impossible to follow the hydrodynamical processes. Therefore, Dark Matter only simulations are performed with very high resolution so that the evolution of halos and subhalos can be followed numerically. These halos and sub-halos host the galaxies. In a second (postprocessing) step the properties of the galaxies can be determined by semianalytical or abundance matching algorithms. Semianalytical methods are based on the formation history of the dark matter halos and on recipes which predict the properties of galaxies formed in these halos. Abundance matching algorithms compare observed luminosity functions with halo mass or velocity functions deduced from dark matter only $N$-body simulations.

In order to simulate the formation of structure in the universe very high mass- and spatial-resolutions are necessary which imposes a strong challenge for present day computational algorithms. Moreover, such simulations need a large amount of computational time and have huge requirements to the total available memory as well as to the possibilities to store the obtained results for further analysis. Therefore, they can be performed only at the largest supercomputer centres.

2 Dark Matter and Dark Energy

One of the main unsolved puzzles in modern cosmology concerns the nature of dark matter and dark energy. Already in 1932 the Dutch astronomer Jaan Oort studied the dynamics of the brightest stars in the disk of the Milky Way and deduced that the total density exceeds the density of the visible stellar populations by a factor of up to 2. He concluded that some “dark matter” must be present in the Milky Way. In the following years it became clear that there must be an undetected (and non-baryonic) component in the universe that is responsible for a very significant (in fact, the most significant) fraction of the matter. This elusive matter remains undetected today and is called Dark Matter. It has been detected indirectly through its gravitational effects. Like ordinary matter (baryonic matter), dark matter creates gravitational potential wells and through detailed observations of the space...
distortions around massive gravitational wells it has been possible to determine some of the characteristics of dark matter. In addition, the puzzle of small scale structure formation is related to the properties of the dark matter particles.

Over the last 20 years, compelling evidence has been growing that suggests the presence of another mysterious component of matter that acts in a direction that opposes gravity: its main effect is to accelerate the expansion of the universe while gravity always decreases the speed of expansion. This mysterious component is called dark energy. Among the few things that we know about the dark energy is that it behaves in a fashion very similar to what Albert Einstein called the cosmological constant. This was introduced by him in 1917 in order to achieve a static universe within general relativity. The cosmological constant represents the minimum amount of energy that remains in a volume of the space, should this space be deprived of all kinds of matter and radiation. Since its energy density is constant with time it becomes increasingly important in an expanding universe (whereas other constituents – such as matter – are diluted by the expansion). The first strong evidence of dark energy came from observations of distant supernovae. Supernovae behave like standard candles, so that their observed brightness can be linked with their real distance. Combining observations of nearby and distant supernovae, it is possible to trace the expansion rate of the universe. For the first time these observations showed that the universe has accelerated its expansion rate over the last 5 billion years. The supernovae results have been confirmed by other independent cosmological tests, in particular by the high-precision measurements of the CMB fluctuations.

The cosmic microwave background radiation originated more than 13 billion years ago. The photons of the cosmic microwave background serve as a superb background light since they cross the entire universe before reaching us. In their stunning voyage from the infancy of our universe towards us, the photons of the CMB interact with the structures they cross. These interactions leave imprints on the cosmic microwave background that can be used to infer information about the structures they cross. Two of these effects are quickly becoming the source of interesting research. One is the integrated Sachs-Wolfe effect and the other one is the gravitational lensing effect. The integrated Sachs-Wolfe effect contains information about the dark energy; the gravitational lensing effect is used for both dark energy and dark matter studies.

Gravitational lensing is today one of the best techniques to study dark matter. On large scales, the gravitational redshift affects the cosmic microwave background by shifting the paths of the photons by a few arcminutes. This effect is small but has been recently measured in cosmic microwave background observations and opens the door to new ways of seeing the dark matter directly. A key ingredient to advance our understanding in this area is the undertaking of simulations of lensing not only in the microwave background itself but also in the associated catalog of galaxies that can be used to correlate with the cosmic microwave background. Our simulation is starting to produce such results and will be a valuable tool for future work where present and future codes can be validated with our simulations. Our lensing effect maps are useful not only for the cosmic microwave background but can also be used in future space missions like Euclid that will survey a large fraction of the sky searching (among other things) for the small lensing distortions produced in background distant galaxies.
3 The Simulation

The Jubilee project is based on a large-scale structure $N$-body simulation with $6000^3$ (216 billion) particles in a volume of $(6 \, h^{-1}\text{Gpc})^3$. We assumed a WMAP5 cosmology with the following cosmological parameters: $\Omega_m = 0.27$, $\Omega_\Lambda = 0.73$, $h = 0.70$, $\Omega_b = 0.044$, $\sigma_8 = 0.80$, $n_s = 0.96$. Thus the particle mass is $7.49 \times 10^{10} \, h^{-1} \text{M}_\odot$, yielding a minimum resolved halo mass (with 20 particles) of $1.49 \times 10^{12} \, h^{-1} \text{M}_\odot$, corresponding to galaxies slightly more massive than the Milky Way.

The simulation and most analyses were performed on the JUROPA supercomputer at Jülich Supercomputing Centre and required approximately 1.5 million core-hours to complete. The initial conditions were generated by Zel’ dovich approximation to place particles in their positions at redshift $z = 100$. The simulation was run on 8,000 computing cores.

Figure 1. Slice through the simulation at redshift $z = 0$ with a zoom in on the most massive cluster, an object with a mass of $7 \times 10^{15}$ solar masses, moving with a velocity of 707 km s$^{-1}$.
(1,000 MPI processes, each with 8 OpenMP threads) using the CUBEP\textsuperscript{3}M $N$-body code, a P\textsuperscript{3}M (particle-particle-particle-mesh) code\textsuperscript{2}. CUBEP\textsuperscript{3}M calculates the long-range gravity forces on a 2-level mesh and short-range forces exactly, by direct summation over local particles. The code is massively-parallel, using hybrid (combining MPI and OpenMP) parallelisation and has been shown to scale well up to tens of thousands of computing cores.

A substantial part of the data analysis of the simulation is halo finding. We used a Friends-of-Friends (FOF) algorithm as well as spherical overdensity halo-finders (SO). For some outputs also sub-halos have been identified. The total number of halos identified in the simulation at redshift $z = 0$ was 412 Million. The Jubilee simulation has been used along with a suite of other simulations to investigate the high mass end of the halo mass function\textsuperscript{4}. A redshift-parameterised fit for the SO mass function and the FOF mass function has been provided. For very large halo masses – which can be studied only in large volumes as provided by the Jubilee simulation – a lower collapsed mass fraction has been found than predicted by the Tinker fit, which was based on smaller volumes.

The data handling requirements for analysing the Jubilee simulation were particularly challenging. For each output slice the simulation’s particle data totalled around 4 TB. These outputs were then analysed and converted into density and then potential fields. The mesh used for the potential fields was $6000^3$ in size ($(1\ h^{-1}\text{Mpc})^3$ per cell) so each output slice in redshift for the potentials was 800 GB in size. Overall, the data for the potential fields used in this analysis totalled over 15 TB and was reduced from particle data that was 100 TB in size. For the weak lensing outputs five derivatives of the potential were calculated, resulting in another 75 TB of data.

In Fig. 1 we show a slice through the simulation centred at the most massive cluster found in the simulation box. This cluster has a mass of $7 \times 10^{15} M_\odot$ and is represented by more than 66,000 particles which allows to study the substructures of the cluster. In the four insets we zoom into this cluster. Structures on small scales can be well seen.

4 The ISW Effect

On the largest scales that are probed by the cosmic microwave background, the effect of dark energy on the photons of the cosmic microwave background is to increase or reduce their energy when they cross a large volume of the universe. In general, as the universe expands faster and faster due to the presence of dark energy, gravitational potentials are diluted and decay. This leaves an imprint in the photons of the cosmic microwave background that depends on whether they are travelling through an overdense or underdense region. When a photon sees a variation in the gravitational potential it gains or loses energy depending on whether the photon falls into a gravitational well (gain of energy) or the photon has to climb out of that well (loss of energy). If the gravitational potential is static, the gain and loss compensate each other resulting in no net gain or loss of energy. However, with the presence of dark energy the potential is evolving with time, and there are gains or losses of energy where infall potentials are weaker or stronger than outfall ones. Over the path the photons travel from the CMB last scattering surface to the Earth there may exist net gains or losses of temperature from the successive addition and subtraction of energy from over- and underdense regions, which we observe as the integrated Sachs-Wolfe (ISW) effect.

The best experiment today to study the cosmic microwave background on large scales...
is the European-led ESA’s Planck mission. Recent results by Planck demonstrate the power of constraining dark energy through the cosmic microwave photons. However, some discrepancies have been observed and remain unsolved. One of these discrepancies is the unusually strong signal detected around some prominent structures seen in optical sky surveys. In order to test this controversial result and understand the possible systematic effects, simulations are needed that mimic the effect of dark energy on the cosmic microwave background which are also capable of producing synthetic catalogs that mimic the optical observations. This is the main motivation of our work where our main goal was to produce both the all-sky ISW effect together with its associated optical catalog of galaxies. Our simulations provide the means to test the significance of the results obtained with real data and opens new doors of analysis since it provides the synthetic data against which to test more powerful data analysis codes.

5 Results and Outlook

In Fig. 2 we summarise our first results. The map shown on top of the figures represents an all-sky spherical projection (mollweide) of the density an observer would measure in a shell at that distance. In the middle panel we show the Integrated Sachs-Wolfe effect map from the same slice. The red regions (corresponding to energy gains in the cosmic microwave background) correlate well with peaks in the matter density (see top) while the blue regions (energy losses) correlate with underdense regions in the same plot. Finally, we want to compare the simulated ISW with the gravitational lensing effect corresponding to structures in the same slice. Whilst the large-scale ISW effect is governed by the dark energy-driven time variability of the gravitational potential – and is therefore observed in the radial direction – variations in the tangential direction of the potential results in achromatic path distortions of the photons (i.e with no gain or loss of energy). These tangential distortions are the gravitational lensing effect. Lensing distortions concentrate on the small scales (of the order of a few arcminutes) and hence complement the large-scale ISW effect. The lensing effect does not depend (at least not to first order) on dark energy but is very sensitive to the distribution of the total mass. Gravitational lensing can provide independent and robust estimates of the cosmological model, in particular constraints can be set on the spatial curvature, dark energy or neutrino masses that are normally degenerate when only the CMB power spectrum is available. The amplitude of the deflection angles due to lensing (bottom panel) shows very clearly the large overdensity affecting light rays in its region of the sky. Therefore, features in this map can be seen to correspond to ones in the ISW map (middle). The largest effect occurs around peaks in the matter density (top).

In future the ISW signal will be used to help to discriminate between cosmological models. At present the ISW effect does not constrain the $\Lambda$CDM model to anything like the precision of the standard datasets (CMB and BAOs). However, for a universe containing an amount of warm dark matter or one with a temporally varying dark energy component, the ISW effect will be an aid in constraining the theoretical models and predictions. For alternative cosmological models a variety of expectations of the ISW signal arise. For example the effect of massive neutrinos on the ISW-LSS correlation signal, along with the expectations of different coupled dark energy models, shows that model discrimination typically involves a difference in the expected height of the peak in the cross-correlations. The Jubilee ISW project will help determine the best strategies to discriminate among
Figure 2. Density map in a shell (top), ISW effect (middle) and lensing predictions (deflection angle) from the same shell (bottom).
models since the necessary tools (ISW maps and associated catalogs) will be available in the Jubilee database.

We plan to extract from the simulations predictions of the Rees-Sciama effect. Although a second order effect at large scales compared with the linear ISW effect, this non-linear part of the ISW signal becomes dominant at small scales, especially near galaxy clusters. Like the ISW effect, the Rees-Sciama effect leaves an imprint on the photons of the CMB that gain or lose energy after crossing the region of varying potential. The Rees-Sciama effect arises from small scale time variations of the gravitational potential that are triggered for instance by the collapse of the haloes, the merging of two or more haloes and the transverse velocity (with respect to the line of sight) of the halo. Moreover, we plan to search Bullet cluster candidates in our simulation. The Bullet Cluster, consisting of a large host dark matter halo and a smaller, but still very massive, sub halo that has shot through the host in a high velocity collision, presents an interesting challenge for cosmology. Current estimates place the impact velocity of the bullet at close to or perhaps above the acceptable limit for a collision of this type in a LCDM universe. The best way to assess the likelihood of such a collision is via very large N-body simulations that cover a large cosmological volume, thereby containing a number of collision events with extreme relative velocities.

Acknowledgements

Simulations described here would be impossible without the excellent user support and help by JSC Jülich.

References

6. The Jubilee home page and database, see http://jubilee.ft.uam.es/
8. http://www.esa.int/Our_Activities/Space_Science/Planck
The Circumstellar Medium of Massive Stars in Motion

Jonathan Mackey, Norbert Langer, Dominique M.-A. Meyer, Vasilii V. Gvaramadze, Shazrene Mohamed, Hilding R. Neilson, and Andrea Mignone

1 Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
E-mail: {jmackey, nlanger, dmeyer}@astro.uni-bonn.de

2 Sternberg Astronomical Institute, Lomonosov Moscow State University, Universitetskij Pr. 13, Moscow 119992, Russia
E-mail: vgvaram@mx.ikey.rssi.ru

3 South African Astronomical Observatory, P.O. Box 9, Observatory, Cape Town, 7935, South Africa
E-mail: shazrene@saao.ac.za

4 East Tennessee State University, Box 70652, Johnson City, TN, 37614, USA
E-mail: neilsonh@etsu.edu

5 Dipartimento di Fisica, Università degli Studi di Torino, Via Pietro Giuria, 1, 10125 Torino, Italy
E-mail: mignone@ph.unito.it

The circumstellar medium around massive stars is strongly impacted by stellar winds, radiation, and explosions. We use numerical simulations of these interactions to constrain the current properties and evolutionary history of various stars by comparison with observed circumstellar structures. Two- and three-dimensional simulations of bow shocks around red supergiant stars have shown that Betelgeuse has probably only recently evolved from a blue supergiant to a red supergiant, and hence its bow shock is very young and has not yet reached a steady state. We have also for the first time investigated the magnetohydrodynamics of the photoionised H II region around the nearby runaway O star ζ Oph. Finally, we have calculated a grid of models of bow shocks around main sequence and evolved massive stars that has general application to many observed bow shocks, and which forms the basis of future work to model the explosions of these stars into their pre-shaped circumstellar medium.

1 Introduction

Massive stars are the main drivers of the evolution of the gaseous component of galaxies. They emit huge quantities of far-ultraviolet photons that ionise and heat surrounding gas; their strong winds drive shocks and gas flows; and their final explosions as supernovae or gamma-ray bursts generate powerful blastwaves that eject chemically enriched matter into the interstellar medium (ISM). They evolve from a hot main sequence star to a more extended red supergiant (RSG) or blue supergiant (BSG), and sometimes to more exotic Wolf–Rayet or Luminous Blue Variable phases. These changes induce strong variations in stellar wind properties, luminosity, and temperature, profoundly affecting the star’s circumstellar medium (CSM) interaction and leading to a highly structured CSM by the time the star ends its life.

Massive stars are born in groups and star clusters, but a large fraction of them are ejected from their birth sites either by dynamical interactions with other stars inside the
Figure 1. Far-infrared \textit{IRAS} (60 \,\mu m) and \textit{AKARI} (65 \,\mu m) images of the bow shock and linear bar in the circumstellar medium of Betelgeuse, taken from Fig. (1) of Mohamed et al. Betelgeuse itself is the overexposed central object; the elliptical feature surrounding it in the \textit{IRAS} image is an artefact. North is up, East is to the left, and the star is moving in a North-Easterly direction.

cluster, or by the disruption of a binary system when one of its components explodes. These stars, known as runaways or exiles, move through the ISM supersonically and this modifies the otherwise spherical CSM to a roughly axisymmetric solution with an upstream bow shock and a downstream wake. These structures have been observed for a number of runaway stars in both our Galaxy and the Magellanic Clouds. Examples are shown for the RSG Betelgeuse in Fig. 1, for the nearby main sequence star \zeta Oph in Fig. 4, and for the interaction of Supernova 1987A with its pre-shaped CSM in Fig. 7.

Our research group has worked since the 1990s to develop a comprehensive understanding of the interaction between massive stars and the various environments in which they form. Our current project has focused on how relative motion between the star and its surroundings affects the star–CSM–ISM evolution, investigating the differences between these models and previously calculated stationary star models. We are also beginning to address the interaction of the star’s supernova explosion with this structured CSM, to see what observational consequences this will have. In this report we highlight the main results obtained in the past two years.

2 The Bow Shock Around the Nearby Red Supergiant Betelgeuse

Betelgeuse is one of the two nearest RSGs to Earth, and is the brightest star in the constellation Orion, “The Hunter”. It is also a well-known runaway star, and emits a powerful stellar wind that (because of the star’s supersonic motion) generates a cometary shaped bow shock at its interface with the surrounding ISM (see Fig. 1). Its CSM has a number of unexplained features that we aimed to investigate: (1) The bow shock appears very smooth compared to numerical predictions; (2) it is close to circular, in contrast to
predictions\textsuperscript{3}; and (3) there is a mysterious bar-shaped structure perpendicular to the star’s direction of motion, at a larger radius than the bow shock\textsuperscript{4}.

In Mohamed, Mackey, & Langer\textsuperscript{9} we performed the first three-dimensional (3D) hydrodynamic simulations of the formation and evolution of Betelgeuse’s bow shock, using a modified version of the hydrodynamics code GADGET-2\textsuperscript{11}. We simulated RSGs moving at velocities ranging from 28 to 73 km s\textsuperscript{−1} through a uniform ISM. Results from two simulations are shown in Fig. 2, showing bow shocks at two different velocities from three viewing angles.

These ground-breaking simulations highlighted areas of significant agreement and disagreement with observations of Betelgeuse. The bow shock appeared more smooth and layered for higher velocity stars, in closer agreement with observations. We calculated the bow shock mass from AKARI observations\textsuperscript{8} and found a surprisingly low value ($\approx 3 \times 10^{-3} M_\odot$) compared to the simulated bow shocks that are about 30 times more massive in steady state ($\approx 0.1 M_\odot$). The shape of the simulated bow shock deforms from approximately circular at early times to the predicted cometary shape\textsuperscript{3} when it reaches steady state. Furthermore, the bow shock is smoother when it is young because instabilities have a finite growth time. All of these findings support the hypothesis that Betelgeuse has entered the RSG phase only “recently” (within the last 25 thousand years), and that the young bow shock is still forming and expanding into its surroundings. The low mass of the bow shock was subsequently confirmed by more detailed observations\textsuperscript{10}.

3 \textbf{Bow Shocks Around Moving Stars as they Evolve}

If Betelgeuse has only recently evolved from a BSG to a RSG, some remnants of the bow shock from this previous stage of evolution may still be imprinted in its CSM. The linear bar outside the bow shock in Fig. 1 is an obvious candidate, being oriented almost exactly.
perpendicular to Betelgeuse’s direction of motion. To test this possibility, we incorpo-
rated an evolving stellar wind module into the radiation-hydrodynamics code PION\textsuperscript{13}, and
performed two-dimensional (2D) simulations of bow shocks around a moving star as it
evolves from a blue to a RSG\textsuperscript{12}. The evolving wind properties were obtained directly from
a stellar evolution model of a similar-mass star to Betelgeuse.

The results of one of the simulations are plotted in Fig. 3. Four snapshots are plotted
corresponding to: (A) the BSG phase, in which the hot star produces a fast and strong wind
that supports a large bow shock; (B) the beginning of the RSG phase where the star begins
to emit a slow, dense and powerful wind; (C) the interaction phase where the collapsing
BSG bow shock generates a second, inner bow shock around the expanding RSG wind;
and (D) the end of the star’s life, by which time the two bow shocks have merged.

We identify panel (C) of Fig. 3 with the current evolutionary phase of Betelgeuse and
its CSM. In this interpretation, the circular bow shock around Betelgeuse is the interaction
of the newly-expanding dense RSG wind with the collapsing bubble of hot gas left by the
(now extinguished) BSG wind. The linear bar in Fig. 1 corresponds to the remnant BSG
bow shock that is still upstream from the inner bow shock (but see also Ref. 10). The inner
bow shock’s mass, shape, thickness and size are consistent with the observed properties of
Betelgeuse’s bow shock, in contrast to steady-state models.

4 \textbf{H II Regions Around Moving Main Sequence Stars}

The nearest example of an H II region around an exiled star is Sh 2-27, powered by the
main sequence O star ζ Oph. The left-hand panel of Fig. 4 shows the Hα image of this H II
region from the Southern Hα Sky Survey Atlas (SHASSA)\textsuperscript{15}. The right-hand panel shows
a spectacular infrared image of the bow shock around ζ Oph, on a much smaller scale. In
Figure 4. Left: SHASSA H\alpha image of the H\textsc{ii} region around \(\zeta\) Oph, from Fig. 11 in Ref. 14. The star is at the centre of the circle, and the arrow indicates its direction of motion relative to the ISM. The image is oriented with Galactic longitude (in units of degrees) increasing to the left and Galactic latitude increasing upwards. Right: Spitzer Space Telescope image of the bow shock on a much smaller scale (the image spans about \(1.5^\circ\)) at infrared wavelengths of \(3.6 - 24\,\mu\text{m}\) (Credit: NASA, JPL-Caltech, Spitzer Space Telescope).

Figure 5. Projections through the 3D simulations of H\textsc{ii} regions around a runaway O star for a non-magnetised ISM (left) and a magnetised ISM (right) where the magnetic field is along the line of sight (from Figs. 4 and 6 of Ref. 14). The star is marked with a cross, and is moving from left to right in the plane of the image. Projected H\alpha emitted intensity is plotted on the linear colour scale and neutral H column density \((N_{\text{HI}})\) as contours. The mean column density of the undisturbed grid is subtracted off (to remove grid edge effects), so underdense regions with negative column density are shown with grey contours and overdense regions with green contours.

Gvaramadze, Langer & Mackey\textsuperscript{16} we showed that the relative sizes of the H\textsc{ii} region and bow shock could be used to constrain the mass-loss rate of \(\zeta\) Oph.

In Mackey, Langer & Gvaramadze\textsuperscript{14} we used the \textsc{pion} code to make the first 3D radiation-magnetohydrodynamics simulations of a star similar to \(\zeta\) Oph moving through the ISM with a velocity of \(26.5\,\text{km}\,\text{s}^{-1}\). Synthetic observations through the simulations are shown in Fig. 5, where we plot the H\alpha emission from ionised gas (units: \(10^{-16}\,\text{erg}\,\text{cm}^{-2}\,\text{s}^{-1}\,\text{arcsec}^{-2}\)) and neutral hydrogen column density contours (with spac-
We found that the H\textsc{ii} region generates a dense expanding shell expanding from its lateral surfaces, leaving a conical shell (with a large momentum) enclosing an underdense wake trailing behind it. This shell should be observable in neutral hydrogen emission at 21 cm, although its properties are quite sensitive to the interstellar magnetic field. The ionisation front in the upstream direction is also unstable, and may be more stable for the magnetised case than for the hydrodynamic case.

5 Evolution and Explosion of Massive Stars in their Environment

We are using the magnetohydrodynamics code PLUTO\textsuperscript{17} to simulate bow shocks around massive stars moving through the ISM for their full evolution. Nine systems have been simulated, representing the complete stellar evolution of three stars (with masses 10, 20, and 40 \(M_\odot\)) moving with three different velocities through the ISM. We included thermal conduction by hot electrons as well as heating and cooling physics appropriate for the different stellar evolutionary phases. A paper based on our results is now at an advanced stage of preparation (Meyer \textit{et al.}, 2013). Fig. 6 shows six of these simulations for the RSG stage of evolution. The gas density field is plotted for the 10 and 20 \(M_\odot\) stars at three different velocities, increasing from top to bottom. The stability of the bow shocks is

\[ \Delta N_{\text{HI}} = 0.5 \times 10^{20} \text{ cm}^{-2} \]
broadly consistent with other works, and we are investigating the radiative emission from these bow shocks to compare to observations such as those of Betelgeuse.

We also calculate the CSM evolution up to the pre-supernova stage when the star is just about to explode. This structured CSM then provides the initial conditions for our next project, to explode supernovae from the centre of the simulation and study the emission properties of the blast wave as it ploughs through the RSG wind and bow shock and on into the undisturbed ISM. Following an established method, we first model the one-dimensional spherically-symmetric explosion of a supernova into a dense stellar wind. This is then mapped onto the centre of the 2D CSM grid and the blast wave expansion is followed to much larger radii. An example of the pre-supernova bow shock and the subsequent supernova-CSM interaction is shown in Fig. 7, right-hand panel. The blast wave is already distorted by the asymmetric CSM, and this should have observable consequences.

6 Concluding Remarks

The interaction of massive stars with their surroundings is an incredibly rich subject, where a combination of the different stellar masses (with associated variations in evolution), different interstellar environments, and different stellar space velocities, produces a huge diversity of circumstellar structures. Many of the closest such structures, such as the CSM of Betelgeuse and ζ Oph, have been observed in great detail with modern observatories, and yet we still do not have a definitive understanding of these objects. The use of large supercomputers such as JUROPA has enabled us to make models with sufficient detail that we are now beginning to make quantitative comparisons to circumstellar structures, for both nearby stars and those in other galaxies. This is significantly increasing our understanding of the physical processes involved and of the properties of the massive stars themselves, including some of the brightest and best known stars in the sky.
Acknowledgements

We acknowledge the John von Neumann Institute for Computing for a grant of computing time on the EUROPA supercomputer at Jülich Supercomputing Centre. JM and HRN acknowledge funding by fellowships from the Alexander von Humboldt Foundation. Figs. 1 and 2 reproduced with permission from Astronomy & Astrophysics, © ESO.

References


Computational Biology and Biophysics
Over the past decades, molecular dynamics (MD) simulations have developed not only into a powerful method to study macromolecular dynamics in condensed phase, with a particular focus at biological macromolecules, but also into a powerful tool for an astonishingly diverse set of other challenges. The use of MD simulations for structure refinement, e.g., – both from crystallographic as well as from nuclear magnetic resonance (NMR) data – has a long and impressive track record. More recently, MD simulations have been successfully employed to aid the structural interpretation of an increasingly broad variety of experimental approaches, including atomic force spectroscopy, fluorescence spectroscopy, small angle x-ray scattering, electrophysiology, various NMR-derived data beyond the traditional NOE interproton distance data, and many others. Further, whereas the de novo structure prediction by MD simulation is still in its infancy – notwithstanding recent impressive break-throughs for small proteins – the prediction of relatively small structural changes such as mutations or posttranslational changes in proteins is currently becoming an increasingly wide-spread and successful approach. Wherever needed, the traditional force fields are supplemented by on-the-fly quantum-mechanical calculations (QM/MM), to include chemical reactions or charge transfer into the simulation.

The four examples in this chapter have been selected to chart this rapidly expanding territory.

The first example aims at the development of a biosensor for inorganic phosphate – an essential tool for large-scale bioreactors. The idea is to report structural changes induced by phosphate-binding to an E. coli phosphate binding protein via fluorescence changes of an attached pair of dye molecules. Whereas the structure of the unlabelled and unphosphorylated protein has been solved by x-ray crystallography, neither the structure of the attached dyes, nor the structural changes due to phosphorylation are known. Jens Dreyer et al. demonstrate how these complex and collective structural transitions can be predicted by MD simulations. Specifically, the simulations predict a transition of initially separated dyes into a stacked configuration, thus explaining the spectroscopic data in molecular terms.

The second example turns our attention to riboswitches – regulatory mRNA stretches with aptamer domains which bind small molecules with amazing specificity. Their structural dynamics and flexibility, as well as the effect of mutations, have so far been studied to a much lesser extent than those of their protein counterparts, both due to lack of structural data as well as due to typically longer relaxation times. In their contribution, Hanke and Gohlke have addressed the dynamics of the guanine sensing riboswitch aptamer domain, both its wild type an a mutant, with microseconds simulations. They find an unexpected dynamic coupling between the binding site and quite distant regions of the molecule, which might point to a complex network of entropic control. It will be exciting to see the structural determinants and thermodynamics of this network.
The third example by Lange illustrates state-of-the-art automated ensemble refinement of NMR NOESY data, simultaneously exploiting and combining multiple sources of information. With a sophisticated and highly optimised tool-chain exploiting advanced GRID-computing, Lange demonstrates that, by coupling automatic assignment techniques with ROSETTA-based iterative conformational sampling, highly accurate (RMSD<2Å) structures can be obtained, in a fully automated manner, for proteins in the 15-20 kDa range.

The fourth example illustrates the potential medical relevance of MD simulations. Köhler et al. focus at fibrinogen, a key multi-protein compound in blood coagulation. Its adsorption properties to inorganic surfaces, e.g., of implants, bear obvious medical relevance. Despite the considerable size of this complex, the authors managed to carry out large-scale MD simulations of its adsorption to both graphite and to mica surfaces, which also has been studied via atomic force microscopy. The simulations underscored the crucial importance of electrostatic interactions for this complex process, which was suggested, but not understood from previously observed strong salt concentration effects. The simulations also provide an impressive example on how simple intuition, here on the mechanical properties of fibrinogen, may be highly misleading, and can only be corrected by a combined experimental / simulation approach.

I trust the reader will find these four examples helpful, in navigating the tools and methods that Computational Biology and Biophysics has to offer. At the interface between physics, chemistry, biology, computer science and mathematics, the clever combination of complementary information and methods using high performance computing is the key to success.
Molecular Simulation-Based Structural Prediction of a Rhodamine-Labelled Biosensor

Jens Dreyer\textsuperscript{1,2}, Marcos Brown Gonçalves\textsuperscript{1}, Emiliano Ippoliti\textsuperscript{1,2}, and Paolo Carloni\textsuperscript{1,2}

\textsuperscript{1} Computational Biophysics, German Research School for Simulation Sciences
\textsuperscript{2} Institute for Advanced Simulation, Forschungszentrum Jülich, D-52425 Jülich, Germany

\textsuperscript{‡} E-mail: p.carloni@grs-sim.de

Fluorescent sensors consisting of biological components may be used to recognise certain materials. Here, we predicted the structure of a protein-based biosensor for inorganic phosphate (Pi). This is the phosphate binding protein from \textit{Escherichia coli} labelled with two rhodamine fluorophores. Classical molecular dynamics and hybrid Car-Parrinello/molecular mechanics simulations provide molecular models of the biosensor both in the presence and in the absence of Pi. In the latter case, the rhodamine fluorophores maintain a stacked conformation. On binding Pi, a protein conformation change prevents significant stacking of the two rhodamines. In both states, the rhodamine fluorophores form hydrophobic contact with LEU291, without establishing significant hydrogen bonds to the protein. The accuracy of the models is established by a comparison between calculated and experimental absorption spectra.

1 Introduction

Fluorescence probes are routinely used to investigate the structural basis of biochemical processes, yet details of the structural interactions that affect their optical properties are still poorly understood. A typical system is the rhodamine-based fluorescent biosensor suitable for real-time measurements of inorganic phosphate (Pi), a product of many enzymatic reactions and an important assay target for study of cellular activities\textsuperscript{1}. The sensor uses the phosphate binding protein (PBP) from \textit{Escherichia coli} as a scaffold\textsuperscript{1}. PBP binds Pi very specifically and tightly with a well-defined conformation change. Two tetramethylrhodamines (RHO) were covalently attached to mutation-generated cysteines (A17C, A197C) on the surface of the protein, using 6-iodoacetamidotetramethylrhodamine. A17 and A197 were selected to obtain a large change in the relative positions of the two fluorophores upon the conformation change due to Pi binding\textsuperscript{1}. In the resulting adduct, RHO\textsubscript{2}−PBP·Pi, the fluorescence emission of the rhodamines increases 18-fold when Pi binds to the sensor\textsuperscript{1,2}. These changes suggested that Pi binding causes the disruption of a stacked dimeric conformation of the RHOs (Fig. 1, left), which is known to quench their emission\textsuperscript{3}. Similar features have been observed for a rhodamine dimer in water\textsuperscript{4}. The enhanced fluorescence upon Pi binding to the sensor protein is accompanied by a change in the visible absorption spectrum of the rhodamines\textsuperscript{1}. The visible absorption spectrum shows only a partial change and remains much perturbed from the monomer spectrum. Thus, the two rhodamines are likely to keep a certain degree of electronic interaction even in the Pi-bound form of the complex presumably because the two rhodamines remain held in proximity by the protein scaffold.

\textsuperscript{1}Joint venture of RWTH Aachen University and Forschungszentrum Jülich, Germany.
Figure 1. The phosphate binding protein from *Escherichia coli* (blue) binds inorganic phosphate (Pi) in the cleft separating two domains, which leads to a conformational change of the protein. The stacked arrangement of the two rhodamine fluorophores (green) covalently attached to the protein largely quenches fluorescence (left). The conformational change induced by phosphate binding disrupts the stacked rhodamine dimer (right). The fluorescence emission of the monomeric rhodamines increases about 18-fold\(^1\).

Experimental difficulties have prevented elucidation of an atomic resolution structure of RHO\(_2\)–PBP by X-ray crystallography or NMR spectroscopy, so structural features that underlie the absorbance and fluorescence changes could not be investigated. Here, we use molecular simulations to predict structural determinants of RHO\(_2\)–PBP and RHO\(_2\)–PBP·Pi\(^5\). We use classical molecular dynamics (MD) simulations to construct the molecular models. Then, a hybrid density functional theory (DFT) approach employing Car-Parrinello\(^6\)/Molecular Mechanics (CP/MM) simulations\(^7\) at room temperature refines the models. Finally, optical spectra are calculated so that comparison with experimental data allows us to validate the models. For the simulation of optical spectra, time-dependent DFT (TDDFT)\(^8\) is employed on snapshots from the CP/MM trajectory. The approach used here takes fully into account environmental effects of the protein and the aqueous solution as well as inhomogeneous spectral broadening due to room temperature fluctuations of the entire system.

2 Computational Methods

Details of the computational methods are provided in Ref. 5. In brief, the structures of RHO\(_2\)–PBP and RHO\(_2\)–PBP·Pi were built based on the X-ray structure of T141D PBP\(^9\) and of PBP in complex with Pi\(^10\), respectively. The complexes were inserted in a box containing \(~14,700\) water molecules, in total \(~48,700\) atoms. The AMBER parm99 force field\(^11\) was used to describe the protein. The TIP3P model was adopted to describe water molecules\(^12\). In the hybrid CP/MM MD simulations the positively charged 3,6-bis(dimethylamino)xanthylium cation moiety was considered as the QM part (38 atoms). In RHO\(_2\)–PBP, both xanthylium cations were included in the QM part (76 atoms). For RHO\(_2\)–PBP·Pi the two rhodamines were separated during the classical MD, maintaining a configuration with largely reduced interaction. Therefore, the chromophores were considered independently as the QM part in two different CP/MM MD simulations. The QM part was described at the DFT-BLYP level, while the rest was treated at the force field level as in the initial classical MD simulations. The full Hamiltonian approach\(^7\) was used to interface the QM and MM regions. Absorption spectra were computed by averaging results.
from TDDFT calculations on 40 CP/MM snapshots taken from all calculated CP/MM trajectories. Line spectra obtained for single snapshots were convoluted with a Gaussian function.

3 Results and Discussion

3.1 Structural Model for RHO₂–PBP·Pi

The structure of the protein is well maintained during the MD simulations, as is shown by a superposition of the initial structure for RHO₂–PBP·Pi taken from X-ray experiments for PBP⁹,¹⁰ and the final structure of the classical MD simulation (Fig. 2a).

In the X-ray structure¹⁰ Pi forms hydrogen bonds (H-bonds) to THR10, THR141, GLY140, SER38, and PHE11 NH groups, to SER139, THR141, and SER38 OH side chains, as well as to the guanidinium group of ARG135 and the side-chain carboxylate of ASP56¹⁰. The Pi coordination is fairly well preserved during the dynamics, although the Pi H-bonds are longer in the MD structure than in the X-ray structure (Fig. 3). In addition, ASP56 forms 2 H-bonds in the MD structure whilst only one is observed in the initial experimental structure. However, THR10 and PHE11 form 3 H-bonds in the initial model, but not in the MD structures.

The two RHOs in RHO₂–PBP·Pi are positioned in a monomer-like configuration (Figs. 2a, 4a) during the entire MD. They experience large fluctuations as is evidenced by the RMSD variations of the fluorophores that are particularly large for RHO17. Likewise, the dihedral angle between the xanthylium and the aryl group fluctuates between 65° and 130° in RHO17, whereas the it remains stable at about 65° in RHO197. This supports the conclusions from fluorescence experiments¹, which attribute the large increase of fluorescence upon Pi binding to disruption of the stacked RHO dimer conformation assumed for RHO₂–PBP, thus inhibiting emission quenching. The two fluorophores do not form stable H-bond interactions. However, RHO197 forms hydrophobic interactions with LEU291,
Figure 3. Representative of the most populated cluster from the classical MD simulation for Pi coordination in RHO$_2$—PBP·Pi, which is in close contact with the tail of the RHO197 benzoate ring (Fig. 4a) over the entire trajectory, as well as with GLN201, ASN202 and ASN203. In contrast, RHO17 does not show any persistent interactions with PBP. The calculated absorption spectrum (Tab. 1) is shifted to lower wavelength with respect to the experimental one by about 70 nm ($\approx$0.4 eV). We mostly attribute this shift to the BLYP functional within the TDDFT approach, for which similar absolute errors have been found previously$^{13}$. It should be noted that the choice of functionals within our computational approach based on plane wave expansions of the wave function is restricted to gradient-corrected functionals of similar quality as BLYP. The spectrum shows a broad band with an implied shoulder on the red side, but no individual peaks as in the experimental spectrum are resolved (Fig. 5).

Figure 4. Representatives of the most populated clusters from the classical MD simulations showing the connection of two rhodamines to the respective cysteine residues as well as interactions between the rhodamines and PBP. a) RHO$_2$—PBP·Pi. b) RHO$_2$—PBP.
Table 1. Experimental and calculated excitation energies.

<table>
<thead>
<tr>
<th></th>
<th>Experiment(^{1})</th>
<th>Simulation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[nm]</td>
<td>[eV]</td>
<td>[nm]</td>
</tr>
<tr>
<td>RHO(_2)−PBP·Pi</td>
<td>556 (w)</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>517 (s)</td>
<td>2.4</td>
<td>450</td>
</tr>
<tr>
<td>RHO(_2)−PBP</td>
<td>554 (vw)</td>
<td>2.2</td>
<td>490</td>
</tr>
<tr>
<td></td>
<td>515 (s)</td>
<td>2.4</td>
<td>435</td>
</tr>
</tbody>
</table>

An analysis of the electronic spectra for RHO\(_2\)−PBP·Pi shows that the broad band arises from contributions of intense and narrow \(\pi \rightarrow \pi^*\) HOMO→LUMO excitations overlapped with weak, but widely spread, combinations of two transitions. The first was \(\pi \rightarrow \pi^*\) HOMO-1→LUMO and HOMO→LUMO centred in the red part of the band and the second was \(\pi \rightarrow \pi^*\) HOMO-1→LUMO, centred in the blue part of the band. An energetic separation in terms of two different centres of these contributions, as evident in the experimental spectrum, cannot be deduced from the simulation data. However, the simulated band shape does reflect the experimentally observed intensity redistribution between the two peaks in going from a free rhodamine monomer in water to protein-bound RHOs in a monomer-like conformation as in RHO\(_2\)−PBP·Pi. TDDFT studies were combined with a polarisable continuum model (PCM) on a related rhodamine chromophore in water using vertical excitations at a ground state optimised geometry\(^{14}\). These studies gave a single peak in the relevant spectral region with a similar description of the character of the electronic transition. The method was extended to a multilevel approach combining classical MD simulations with subsequent single-point TDDFT calculations on a selection of MD snapshots\(^{15}\). In contrast, this revealed a broad band with two shoulders as observed experimentally though with a different intensity distribution.

### 3.2 Structural Model for RHO\(_2\)−PBP

The structure of RHO\(_2\)−PBP is similar to the experimental X-ray structure\(^{16}\) and remains stable along the MD (Fig. 2b). The two RHO molecules adopt a stacked dimer arrangement.
with the carboxylate groups pointing in the same direction (Figs. 2b and 4b). In RHO197, the rhodamine of the dimer makes contact with the protein, the carboxylate group points away from the protein surface towards the solvent. This preference of rhodamines, presumably driven by hydrophobic effects, has been noted previously. The degree of $\pi - \pi$ interaction between the two xanthyliums is evaluated by measuring the distance between the geometric centres of the three rings in the xanthylium moieties as well as the angle between the normal vectors of the respective ring planes. In the first 5 ns of the classical MD trajectory distances vary between 4 and 9 Å. Thereafter the inter-ring separation becomes stabilised, with a skewed relationship between the two xanthylium moieties, characterised by a gradual increase in the separation between the individual paired rings of the two groups. The angle between the normal vectors, which is 0° for ideal $\pi - \pi$ stacking with parallel ring planes, fluctuates between 0 and 70° with an average value of 23±11°. This analysis thus reveals a stable and slightly angled stacked conformation of the two xanthylum systems.

The fluorophore in contact with the protein (RHO197) forms direct hydrophobic interactions with TYR198 and LEU291 (Fig. 4b). The two aromatic rings of RHO197 and TYR198 maintain $\pi - \pi$ interactions and the TYR198 side chain. LEU291 always stays close to RHO197 which can be attributed to hydrophobic interactions. The RHO197 fluorophore interacts with TYR198 and LEU291, the latter being the only residue in contact with the RHOs in both RHO$_2$−PBP and RHO$_2$−PBP·Pi. This suggests that mutations of these two residues could have considerable influence on the structural properties of the fluorophore-protein complex, which is expected to be reflected in a change of optical properties.

The simulated absorption spectrum for RHO$_2$−PBP plotted in Fig. 5 is blueshifted by about 60-80 nm from the experimentally observed peak, and a similar shift was seen in the spectrum calculated for RHO$_2$−PBP·Pi (Tab. 1). The spectrum shows two peaks with an intensity ratio in agreement with experiment and a splitting of about 0.3 eV, which is somewhat larger than observed experimentally (≈0.2 eV). The shift between the two intense peaks in the RHO$_2$−PBP·Pi and RHO$_2$−PBP spectra was ≈20 nm and it is possibly due to the different model adopted, that considers the two xanthylum chromophores separately in different calculations with a charge of +1e each versus a xanthylum dimer with a total charge of +2e.

Analysis of the most intense transitions for RHO$_2$−PBP among all snapshots reveals that the strong peak at about 435 nm can be assigned to the combination of HOMO−1→LUMO and HOMO→LUMO+1 transitions. Both transitions exhibit the same $\pi - \pi^*$ character, being HOMO-1→LUMO in RHO17 and HOMO→LUMO+1 in RHO197. The small peak around 490 nm can be assigned to the same transitions though with different contributions. In all of the transitions contributing to the spectra the $\pi - \pi^*$ excitations are localised within one of the two xanthylum cation chromophores. Charge transfer transitions within the dimer are not found to provide a contribution to the spectra.

4 Conclusions

Our calculations allow us to establish that, in the absence of Pi, the two fluorophores rearrange so as to be stacked (Fig. 2b). Indeed, in RHO$_2$−PBP the two RHO fluorophores maintain a stacked dimer conformation. In contrast, in the system with Pi the
two fluorophores are largely separated from each other with appreciable conformational freedom concerning their mutual orientation as well as their orientation towards the protein (Fig. 2a). Whereas RHO17 maintains persistent interactions with PBP in neither RHO2−PBP-P nor in RHO2−PBP, RHO197 interacts with LEU291 in both configurations and strongly with TYR198 in case of RHO2−PBP. This suggests that mutations of these residues would lead to considerable structural changes that are expected to be reflected in altered optical spectra. Optical absorption spectra have been simulated with TDDFT methods for the xanthylium chromophores taking the protein environment as well as the conformational flexibility of the systems in aqueous solution at room temperature into account. The absolute values for calculated excitation energies are blueshifted by 0.3-0.4 eV (80-60 nm), which we attribute to well-known deficiencies\textsuperscript{13} in the accuracy of the DFT/BLYP method. The offset between the spectra for RHO2−PBP-P and RHO2−PBP as well as the small spacing between the spectral contributions in case of RHO2−PBP-P is attributed to the different chromophore models adopted for the two systems. The overall shape of the simulated spectra, however, are in fair agreement with experimental spectra.

The approach presented here, also has the possibility of providing understanding of the detailed mechanism of biosensors signal production, such as the rhodamine-PBP. This can potentially aid design of novel reagentless biosensors, based on the same principle, by allowing detailed predictions of the interaction between fluorophore and protein, or as in this case, two fluorophores on the surface of a protein.

5 Acknowledgements

We thank Prof. John E. T. Corrie and Prof. Martin R. Webb for experimental support. Computer time provided by the Jülich Supercomputing Centre on JUROPA is gratefully acknowledged.

References


How Tertiary Interactions Between the L2 and L3 Loops Affect the Dynamics of the Distant Ligand Binding Site in the Guanine Sensing Riboswitch

Christian A. Hanke and Holger Gohlke

Institute for Pharmaceutical and Medicinal Chemistry, Heinrich-Heine-University, 40225 Düsseldorf, Germany
E-mail: {christian.hanke, gohlke}@hhu.de

In order to investigate how tertiary interactions in the L2/L3 loop region of the guanine sensing riboswitch aptamer domain (Gsw) affect the domain’s ability to bind ligands, molecular dynamics simulations of wildtype Gsw and a G37A/C61U mutant of in total 9 µs length are performed. The simulations reveal a dynamic coupling between the loop region and the distant ligand binding site suggesting that there exists a complex pathway for the transmission of stability information through the aptamer domain. This finding may have important implications for understanding how Gsw functions at a molecular level.

1 Introduction

Riboswitches are \textit{cis}-acting mRNA regulatory elements that modulate gene expression through their ability of binding small molecules with high specificity. They are mostly located in the 5'-untranslated region of bacterial mRNA\textsuperscript{1-3} and usually consist of two domains: the aptamer domain, which binds the ligand molecule, and the expression platform, which undergoes conformational changes upon binding of the ligand and, thereby, determines the expression of the genes under its control. Riboswitches in bacteria can act either on the translational or transcriptional level. In the latter case, the expression platform is involved in the formation of an intrinsic terminator or antiterminator, leading to the termination or activation of transcription, respectively.

In order for such a riboswitch to be an effective regulator of gene expression, a decision must be made at a branchpoint \textit{during} transcription in favour of one of the two folding pathways\textsuperscript{4}. This requires the unbound state of the riboswitch to maintain ligand-binding competence but at the same time to be able to follow the default pathway in the absence of the ligand. Another level of complexity arises from the observation that riboswitches involved in transcription regulation function through primarily a kinetically controlled mechanism\textsuperscript{5}, that is, the aptamer does not reach equilibrium between the unbound and ligand-bound state before the genetic decision. For understanding how the regulatory decision made at the branchpoint leads to one of the two structural states and the role of kinetic discrimination in this, the nature of the unbound riboswitch state and the folding pathways must be known in atomic detail.

The guanine sensing riboswitch is one of the smallest riboswitches known, involved in transcription regulation, and experimentally well studied\textsuperscript{6}. Crystal structures of the guanine sensing riboswitch aptamer domain (Gsw) from the xpt-pbuX operon of \textit{B. subtilis} in the ligand-bound state\textsuperscript{7} revealed that the Gsw is built from three paired regions (P1, P2 and P3), two loops (L2 and L3) capping the P2 and P3 region and forming tertiary interactions,
and three joining regions (J1/2, J2/3 and J3/1) connecting the paired regions and forming the ligand binding site (Fig. 1A), in which the ligand is deeply buried\(^7\). The stable tertiary interactions between the L2 and the L3 loop are important for the structural stability of the Gsw as well as for its ligand binding ability: Mutations that replaced L2 and L3 with stable UUCG tetraloops, that way eliminating the tertiary interactions, abolished the ligand binding ability of the Gsw\(^7\). Introducing a destabilising G37A/C61U double mutation in the loop region resulted in a structure displaying a Mg\(^{2+}\) concentration dependence of the formation of the loop-loop interactions as well as of the ligand binding ability\(^8\text{-}10\). Information in atomic detail how formation and stability of the loop-loop interactions affect the ligand binding ability in the Gsw has remained elusive, however.

To this end, we performed explicit solvent molecular dynamics (MD) simulations of in total 9 \(\mu\)s length of the wild type (Gsw\(^{\text{apt}}\)) and the G37A/C61U mutant (Gsw\(^{\text{loop}}\)) of the guanine sensing riboswitch aptamer domain in the ligand-unbound state at different Mg\(^{2+}\) concentrations. These simulations reveal a dynamic coupling between the loop region and the distant ligand binding site, suggesting that there exists a pathway transferring stability information through the Gsw.

2 Methods

The starting structure for the MD simulations of Gsw\(^{\text{apt}}\) was taken from the X-ray structure of the \emph{B. subtilis} guanine sensing riboswitch aptamer domain bound to hypoxanthine (PDB code 4FE5\(^7\)). The starting structure for the MD simulations of Gsw\(^{\text{loop}}\) was taken from the X-ray structure of this mutant bound to thioguanine (PDB code 3RKF\(^10\)). In both cases, the ligands were removed as were all ions and water molecules found in the structures.

For both structures, three simulation systems with different Mg\(^{2+}\) concentrations (0, 12, and 20 Mg\(^{2+}\) ions per Gsw) were set up according to experimental findings on the Mg\(^{2+}\) dependence of Gsw\(^{\text{loop}}\) properties\(^8\text{-}10\). In order to allow for a sufficient equilibration of the Mg\(^{2+}\) ions, which may be hampered by the slow exchange times of first shell ligands of Mg\(^{2+}\)\(^11\), the Mg\(^{2+}\) ions were initially placed as hexahydrated complexes. These systems were then solvated using TIP3P water molecules, resulting in system sizes of \(\sim50,000\) atoms. The parm99 force field\(^12\) was used for the Gsw.

For each of the six systems, three independent MD simulations at 300 K of 500 ns length each were performed with the GPU version of pmemd\(^13\) of the Amber package\(^14\) summing up to a total simulation time of 9 \(\mu\)s.

3 Results and Discussion

Mg\(^{2+}\) ions initially show a high mobility during MD simulations and occupy sites in very good agreement with those found in X-ray structures

In order to investigate the Mg\(^{2+}\) dependence of the formation of the loop-loop interactions and the ligand binding ability in Gsw\(^{\text{apt}}\) and Gsw\(^{\text{loop}}\), we performed MD simulations with three different Mg\(^{2+}\)/Gsw ratios (0, 12, 20 Mg\(^{2+}\) ions per Gsw). Our careful initial placement of the Mg\(^{2+}\) ions (see Methods section) resulted in all ions showing a pronounced mobility in the first 100 ns of the simulations (Fig. 1B), with several ions remaining mobile even after 500 ns (data not shown). This indicates that the ions can sufficiently equilibrate.
Figure 1. A: Structure of the Gsw bound to hypoxanthine (magenta spheres); the Gsw structure is coloured according to secondary structure elements, which were assigned according to Ref. 7: grey: P1; green: P2; orange: P3; red: L2; blue: L3; yellow: J1/2; cyan: J2/3; brown: J3/1. M marks the area of the G37A/C61U mutation; L marks the ligand binding site. B: Positions of 12 Mg$^{2+}$ ions from an MD simulation over a simulation time of 100 ns; colours correspond to different Mg$^{2+}$ ions; RNA in grey. C: Comparison of preferred sites of occupancy of Mg$^{2+}$ ions during 100 ns of MD simulation (red) to experimentally determined Mg$^{2+}$ binding sites (green/magenta: binding sites of cobalt hexammine ions in X-ray structures with PDB ID: 4FE57 /3RKF10); grey: RNA.

prior to making direct contacts with the Gsw, which is important considering the slow exchange kinetics of first shell ligands of Mg$^{2+}$ on the order of $\mu$s. In addition, we occasionally observed Mg$^{2+}$ ions swapping their positions (data not shown). As a result, preferred occupation sites of the ions identified in the simulations are in very good agreement with those found in X-ray structures (Fig. 1C). These findings show that our setup of the Mg$^{2+}$ ions results in simulation systems that should be well suited for investigating the influence of the concentration of Mg$^{2+}$ ions on the structure and dynamics of the Gsw.

The G37A/C61U mutation destabilises the native hydrogen bond network in the loop region

In MD simulations of Gsw$^{apo}$, we observe that the hydrogen bond network connecting the L2 and L3 loops is stable (data not shown). In contrast, in MD simulations of Gsw$^{loop}$, hydrogen bonds involved in the loop-loop interactions repeatedly break and then partially reform during the simulation time (data not shown). Thus, our MD simulations confirm the local destabilising effect of the G37A/C61U mutation on the hydrogen bond network connecting the L2 and L3 loops, which allows us to use Gsw$^{loop}$ as a model system to investigate the influence of the stability of the loop-loop interactions on the overall structure and dynamics of the Gsw.
The G37A/C61U mutation leads to overall structural destabilisation and increases the dynamics of Gsw\textsubscript{loop}, which is counteracted by the presence of Mg\textsuperscript{2+} ions.

In order to investigate how the destabilisation of the native hydrogen bond network in the loop region by the G37A/C61U mutation affects the overall structure and dynamics of the Gsw in a Mg\textsuperscript{2+}-dependent manner\textsuperscript{9}, we simulated Gsw\textsuperscript{apt} and Gsw\textsuperscript{loop} in the presence of 0, 12 and 20 Mg\textsuperscript{2+} ions per Gsw molecule. For reasons of space limitations, we only report results for 0 and 20 Mg\textsuperscript{2+} per Gsw here.

In the presence of Mg\textsuperscript{2+} ions, the mean radius of gyration of both Gsw\textsuperscript{apt} and Gsw\textsuperscript{loop} is smaller by \(~1\) Å than in the absence of Mg\textsuperscript{2+} ions (Fig. 2A), demonstrating a higher compactness of the Gsw structures in the former case. This arises from the Mg\textsuperscript{2+} ions preferentially occupying the space between the RNA backbones, which decreases the electrostatic repulsion between the phosphate groups. Notably, the mean radii of gyration obtained from the MD simulations in the presence of Mg\textsuperscript{2+} ions differ by <0.1 Å with respect to the radii of gyration computed for the respective X-ray structures, which also contained Mg\textsuperscript{2+} or [Co(NH\textsubscript{3})\textsubscript{6}]\textsuperscript{3+} ions (data not shown). Regarding the dynamics of the systems, Gsw\textsuperscript{loop} shows a higher mobility than Gsw\textsuperscript{apt} as demonstrated by mean root mean

![Figure 2. Comparison of the radius of gyration and RMSF for simulations of Gsw\textsuperscript{apt} and Gsw\textsuperscript{loop} in the absence (red) and presence of 20 Mg\textsuperscript{2+} ions per Gsw (blue)](image-url)
square fluctuations (RMSF) that are larger by 0.5 Å in the former case (Fig. 2B). In the presence of Mg$^{2+}$ ions, mean RMSF are observed that are lower by 0.5 Å (0.7 Å) for Gsw$^{apt}$ (Gsw$^{loop}$) than in the absence of Mg$^{2+}$ (Fig. 2B). This results in the mean RMSF of Gsw$^{loop}$ only being larger by $\sim$0.3 Å than that of Gsw$^{apt}$ in the presence of Mg$^{2+}$ ions then.

In summary, these findings indicate that the mutation in Gsw$^{loop}$ does not influence the gross structure of the RNA but rather destabilises the RNA compared to Gsw$^{apt}$, which results in an increased dynamics in the case of Gsw$^{loop}$. This destabilising effect can be counteracted by the presence of Mg$^{2+}$ ions, with a stronger influence of Mg$^{2+}$ observed in the case of Gsw$^{loop}$. This result is in agreement with experimental findings according to which the ability to bind a ligand is restored for Gsw$^{loop}$ if Mg$^{2+}$ ions are present$^9$.

The destabilising effect of the G37A/C61U mutation is most pronounced in the distant ligand binding site

In order to gain insights which parts of the Gsw structures are most influenced in terms of the dynamics by the G37A/C61U mutation and the presence of Mg$^{2+}$ ions, we compared differences in RMSF values on a per-residue level (Fig. 3). Regarding Gsw$^{loop}$ and Gsw$^{apt}$ (Fig. 3A), the most pronounced difference is found for residue 74, which is part of the joining region J3/1, more than 25 Å away from the loop region, and has been found crucial for RNA-ligand interactions$^{15}$: this residue’s RMSF values are $\sim$1.5 Å higher in simulations for Gsw$^{loop}$ than for Gsw$^{apt}$ in the absence of Mg$^{2+}$ ions. The second largest differences are found for the joining region J2/3 (Fig. 3A), which opposes residue 74 and has been implicated to act as an entrance gate to the ligand binding pocket$^{15}$. This region is also the most flexible one on an absolute scale in all simulations (except for the P1 re-

![Figure 3. Difference in RMSF values determined for each residue projected on the RNA. A: Gsw$^{loop}$ - Gsw$^{apt}$ in the absence of Mg$^{2+}$ ions; residue numbers are denoted; B: Gsw$^{apt}$ with 0 Mg$^{2+}$ - Gsw$^{apt}$ with 20 Mg$^{2+}$ ions; C: Gsw$^{loop}$ with 0 Mg$^{2+}$ - Gsw$^{loop}$ with 20 Mg$^{2+}$ ions.](image-url)
region, which contains the termini of the RNA strands; Fig. 2B), which is in agreement with experiments\textsuperscript{16,17}. The third largest difference is found for the joining region J1/2, which is also part of the ligand binding site\textsuperscript{7}. Except for residue 66, which is part of L3 in the loop region, all other differences in the RMSF values between Gsw\text{loop} and Gsw\text{apt} are marginal (< 1 Å).

Regarding the influence of the presence of 20 Mg\textsuperscript{2+} ions per Gsw versus the absence of Mg\textsuperscript{2+}, only small (< 1 Å) differences in the per-residue RMSF values are found for Gsw\text{apt}, particularly in regions J1/2, J2/3, and the loop region (Fig. 3B). In contrast, pronounced differences occur for Gsw\text{loop}, with the largest influence of Mg\textsuperscript{2+} exerted on the regions J3/1 (residue 74 shows the overall largest difference of 2.3 Å), J2/3, J1/2, and part of L3 (Fig. 3C), i.e. those regions that become most destabilised due to the G37A/C61U mutation.

The increased mobility of residues involved in ligand binding in Gsw\text{loop} versus Gsw\text{apt} provides an explanation why Gsw\text{loop} is not able to bind the ligand productively in the absence of Mg\textsuperscript{2+} ions; at the same time, these residues show the largest decrease in the mobility upon addition of Mg\textsuperscript{2+}, which explains why the binding ability is restored in the presence of these ions\textsuperscript{9}.

4 Conclusion

We performed explicit solvent molecular dynamics simulations of the wildtype Gsw and its G37A/C61U mutant in the ligand-unbound state at three different Mg\textsuperscript{2+} ion concentrations in order to investigate how tertiary interactions in the L2/L3 loop region affect Gsw’s ability to bind ligands. Initially, we validated our simulation setup by monitoring Mg\textsuperscript{2+} ion mobility, sites in the Gsw occupied by Mg\textsuperscript{2+} ions, and the local hydrogen bond network in the loop region. On a global scale, we observed that the G37A/C61U mutation leads to overall structural destabilisation and increases the dynamics of Gsw\text{loop}, which is counteracted by the presence of Mg\textsuperscript{2+} ions. In contrast, structural differences between Gsw\text{apt} and Gsw\text{loop} are small. On a local scale, the destabilising effect of the G37A/C61U mutation is most pronounced in the distant ligand binding site, and the presence of the Mg\textsuperscript{2+} ions restores the stability of this site almost to the level of Gsw\text{apt}. These findings yield a possible explanation on an atomic level as to why Gsw\text{loop} is not able to bind the ligand productively in the absence of Mg\textsuperscript{2+} ions but can do so in the presence of Mg\textsuperscript{2+}, as observed experimentally\textsuperscript{9}.

Our findings furthermore reveal a long-range transmission of stability information through the Gsw from the loop region to the ligand binding site, which is ∼25 Å away. This suggests that both sites are dynamically coupled. This may have important implications for understanding how Gsw functions at a molecular level. Thus, we will next set out to characterise the pathway of information flow through the Gsw applying Constraint Network Analysis\textsuperscript{18}, which was successfully used for predicting signal transmission pathways in the ribosomal exit tunnel already\textsuperscript{19}.

Acknowledgements

We gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUROPA at Jülich Su-
percomputing Centre (JSC) (NIC project 4722). Additional computational support was provided by the “Center for Information and Media Technology” (ZIM) at the Heinrich-Heine-University of Düsseldorf (Germany).

References

Automatic NOESY Assignment Using an Iterative CS-Rosetta Scheme

Oliver F. Lange

Biomolecular NMR and Munich Center for Integrated Protein Science, Department Chemie, Technische Universität München, 85747 Garching, Germany
E-mail: oliver.lange@tum.de

We have developed an approach for simultaneous structure calculation and automatic NOE assignment to solve NMR structures from unassigned NOESY data. The approach, autoNOE-RASREC, integrates RASREC Rosetta NMR calculations with algorithms for automatic NOE assignment. The method was applied to two proteins in the 15-20kDa size range for which both, NMR and X-ray data is available. The autoNOE-RASREC calculations converge for both proteins and yield accurate structures with RMSD of 1.9 Å to the X-ray reference structures.

The method greatly expands the radius of convergence for automatic NOE assignment, and should be broadly useful for NMR structure determination.

1 Introduction

Structure determination by nuclear magnetic resonance (NMR) spectroscopy is largely driven by distance information gathered through Nuclear Overhauser Effect (NOE) spectroscopy. To use NOE data as distance restraints the NOE crosspeaks in multidimensional spectra have to be assigned to individual atoms of the biomolecular system (Fig. 1). The NOE crosspeak assignment and structure generation steps are usually performed in an integrated manner over several iterations to maximise the number of conformational restraints while guaranteeing self-consistency of all distance restraints.

Chemical shift assignments of individual spins and the positions of cross-peaks in NOE spectra (peak-picking) can often be obtained accurately without explicit 3D structural modelling, whereas resolving the high ambiguity in NOE cross-peak assignments requires almost always structural models. The main challenge is thus to obtain converged and sufficiently accurate initial 3D structures, despite the high ambiguity and low-fidelity of initial automatic NOE cross-peak assignments (Fig. 1D). If accurate enough, these initial 3D models can be used to refine the assignments, and the protein structure can be determined accurately in high-resolution.

Resolution Adapted Structural RECombination (RASREC) is an iterative sampling strategy for restraint guided structure determination in ROSETTA. As shown previously, RASREC requires less data than standard algorithms to converge and has been shown to allow structure determination for proteins up to 20kDa from RDC and expert-assigned backbone NOE data. Using additional ILV methyl-methyl NOE data RASREC can determine structure of proteins up to 40kDa. Most importantly, RASREC requires less NOE data and is more robust against inaccurate restraints. These properties make RASREC an ideal partner for automatic NOE assignment methods, as one expects a considerable number of miss-assigned or ambiguous assigned NOEs in the initial attempt to assign cross-peaks. Moreover, the lower restraint-density required by RASREC compared to conventional structural modelling algorithms allows to filter assignments more aggressively to
Figure 1. Illustration of the NOESY Crosspeak assignment problem. A) Shown are several CH strips from a 3D NOESY experiment, whose cross-peaks align well across strips. B) The cross-peaks assigned in panel A give rise to the shown network of NOE distance restraints which yield valuable information on the packing of the protein core. C) a typical network of NOE contacts as they occur in the beta-strand regions of proteins. D) The shown strip is centred on the directly correlated C and H dimensions of a measured crosspeak (green circle) of a 3D NOESY experiment. Circles show all proton resonance frequencies that fall within the tolerances (box) around the measured peak. The correctly assigned proton has the frequency shown as red circle.

further reduce the number of erroneous or doubtful assignments.

Thus we sought to combine RASREC-Rosetta structure determination with automatic NOE assignment methods within the ROSETTA3 software suite. As starting point for the NOE assignment algorithm we took established algorithms such as, ARIA\textsuperscript{6,7}, AutoStruct\textsuperscript{8} and CANDID\textsuperscript{9}. CANDID is implemented in the popular programs CYANA and UNIO, the other two algorithms are implemented in program packages of the same name. The general approach of these algorithms is very similar. First, atoms are initially assigned to a 2D, 3D or 4D NOESY cross-peak based on the known chemical shift resonances. This yields on the order of 10-20 initial assignments for a typical 3D NOESY peak (Fig. 1D). Subsequently, all initial assignments of a given peak are ranked according to different heuristics, including the chemical shift compatibility, network anchoring\textsuperscript{9}, symmetry considerations, and compatibility with preliminary structural models.

In subsequent rounds of iterative automatic NOE assignment one usually assumes that the correct fold is obtained with the preliminary structural models, and thus all restraints that are significantly violated by these models, can be assumed to be wrong and are removed for subsequent calculations\textsuperscript{9,8,6,7}. The inherent danger with this approach is to remove correct information too early, when sampling has not yet been sufficient to reach conformations consistent with most of the correct restraints. To avoid premature filtering
we dedicate an extended initial phase of RASREC to fragment assembly structure calculations with unfiltered NOE based distance restraints. This first phase (Phase I) consists of RASREC stages I-IV and usually requires 20 iterations of fragment assembly to complete. Sampling barriers are usually overcome eventually and structural accuracy suffices to remove incorrect NOEs safely. In a transition phase (Phase II) we start to filter-out consistently violated NOE assignments, but do this only for regions of the structure where heterogeneity in the low-energy models is low already, i.e., in regions where the structure calculation has converged. For sampling in Phase II, we repeat RASREC stages III-IV. Finally, - in Phase III - we switch to RASREC stages V-VI and remove any restraint that violates a certain fraction (usually 50%) of the low-energy conformers. The coupling between the RASREC and automatic NOE assignment modules is illustrated in detail in Fig. 2.

2 Methods

2.1 Iterative Structure Calculation in Rosetta

Structure calculations coupled to automatic NOE assignment require an iterative approach. The high computational demand of ROSETTA fragment assembly calculations requires us
Figure 3. Illustration of the MPI-framework for job-control and iterative structural sampling. Each blue box is a distinct process and arrows (1-4) indicate Message Passing Interface (MPI) based communication. Arrows (5-6) are file-IO operations. The individual communications are (1) (up) queueing of new batches (down) notification of finished results and underflow of the queue (2) (right) start jobs (left) report success/failure (3) write to file (4) lock/release files before reading with (6) to avoid competition with writing (5). The architecture is modular and the RASREC protocol is implemented by specialising the base-class Archive. The MPI-based JobDistributor is now also used in non-iterative contexts within ROSETTA to efficiently execute a number of tasks within a single job.

The software design is illustrated in Fig. 3. Three processes are dedicated to organisational tasks; these are single-point File-IO, job distribution and structural analysis, respectively. The computationally expensive structure calculations are carried out in parallel. An internal job-queue within the dedicated JobDistributor process is used to keep the worker nodes always busy.

The single-point File-IO (MPIFileBuffer) exports virtual files to individual worker processes within the ozstream class (a derivative of the C++ ostream) which is already used for all File-IO within existing code in Rosetta, such that without any further changes to the high-level structure calculation code the File-IO of individual worker processes is rerouted via MPI to the dedicated output process. This has several advantages. First, access of the file system for writing by individual processes has caused heavy-IO, often bringing down our in-house cluster, which has a similar but smaller architecture to JUROPA. Second, to avoid scrambling of multiple blocks of output data the individual output frames have to be written to individual files, or a locking scheme has to be used such that only one process writes at a time. The latter scheme was implemented in ROSETTA previously, but results
in long wait times for IO if more than a few processes compete. Writing to individual files, on the other hand, had the disadvantage, that too much time was lost in the analysis stage of iterations to collect the frames from individual files.

The ArchiveManager process maintains a pool of structures, which reflect the best observed results in all previous iterations. This process is notified by the JobDistributor via MPI if more than 50 new decoys are available for reading from one of the current batches. These structures are subsequently read and evaluated. If the score of an incoming structure is better than the worst score in the pool the structure replaces the worst scoring structure otherwise the new structure is discarded. If the number of queued jobs gets critically low, the JobDistributor will prompt the ArchiveManager process to generate a new batch of jobs.

Extensive profiling has been carried out of the MPI based iterative structure calculation framework to identify bottlenecks. For instance, the evaluation of incoming structures carried out by the ArchiveManager to decide whether a conformation is better than previously seen conformations did require application of the Rosetta energy function to all incoming decoys. Even though this step only required a couple seconds per structure, it quickly became the bottleneck in simulations with more than 50 processes. A large part of the time was spent in building the internal representation of the protein conformation from the serialised stream data. We were able to speed up the re-construction of protein conformations from file-data 2-3 fold, but not sufficiently to remove this bottleneck entirely. Instead, the bottleneck was finally overcome by carrying out the necessary evaluation steps at the end of a structure calculation on the individual worker processes. The final values are stored together with the output frame that also contains the protein conformation. In this manner the final energy evaluation is much faster, as the conformation is already represented in the correct fashion, and since the evaluation is distributed across all processes it has no impact on the runtime within the central ArchiveManager process.

2.2 Automatic NOE Assignment of SsR10 and DrR1470

Fragments were picked by the Rosetta3 fragment picker\(^\text{13}\) using the chemical shift data from the BMRB. Homologous proteins using an e-value cutoff of 0.05 (sequence identity > 20\%) were excluded from fragment picking. After removal of a C-terminal His-Tag, the target SsR10 (PDB accession 3h9x) is 118 residues long. For target DrR1470 the first 8 residues, as well as the C-terminal His-Tag are removed due to flexibility (according to TALOS+ computed RCI-S2 < 0.7). The remaining sequence from residue 9-155 of the NMR construct is 147 residues long. AutoNOE-RASREC has been run on both targets with cst_strength of 25 using 4 HPC compute nodes equipped with four 2.6GHz AMD Opteron processors (12 core) each. For DrR1470 the structure calculation was completed after 5h, for SsR10 after 2.5h.

2.3 AutoNOE-RASREC

RASREC structure calculations\(^\text{2}\) were run with a reduced pool-size of 100 conformers (command-line flag -iterative:pool_size 100) compared to the standard protocol\(^\text{2}\). This speeds up convergence considerably and reflects the reduced need for structural exploration when NOE data is present. Recombination-Stages were terminated when the acceptance ratio into the pool dropped below 10 (-iterative:accept_ratio 0.1) and the cycle
factor was set to 2.0 (-increase_cycles 2). The protocol was modified to add chemical shift pseudo-energies with a weight of 5.0 to the ROSETTA energy to bias the RASREC pool of low-energy structures towards conformations in agreement with the experimental chemical shifts. Chemical shifts were computed from conformations using SPARTA+ and compared to the experimental chemical shifts to yield a pseudo-energy as described previously. The original RASREC algorithm comprised of stages I-VI. The first four stages are run with Phase I parameters for NOE assignment. Subsequently, we re-run stages III and IV of the original RASREC algorithm with Phase II parameters, followed by stages V and VI with Phase III parameters. Renumbering the stages we thus get stages I-VIII as shown in Fig. 2.

Both, the chemical shift calculations as well as the structure-based automatic NOE crosspeak assignment requires all-atom models. Thus, a shortened refinement procedure that uses only 1 of the usual 5 relax cycles was applied in the centroid stages (stage I-VI) of RASREC. SPARTA+ was implemented as a module of ROSETTA to allow computation of chemical shift pseudo-energies during RASREC iterations.

3 Results and Discussion

3.1 AutoNOE-RASREC Calculations of DrR1470 and SsR10

In the following, we demonstrate autoNOE-RASREC on two protein targets from the North East Structural Genomics Consortium. Both proteins have an X-ray structure as reference and their NMR data was taken from a previously published benchmark study made available through an NESG website (http://psvs-l_4-dev.nesg.org/MR/dataset.html). After trimming flexible termini (Methods) the lengths are 118 and 142 residues, for SsR10 and DrR1470, respectively.

For SsR10 3 peak lists are available, 3D aliphatic 13C, a 3D aromatic 13C, and a 3D 1H-15N HSQC-NOESY. In total 5131 3D peaks. For target DrR1470, a single 3D peak-list is available with both 13C and 5N NOE cross-peaks comprising 8579 entries, of which 3737 have non-zero intensity. The zero-intensity peaks are ignored by the program. Chemical shift assignments are downloaded from the BMRB and CYANA 3.0 specifies their completeness to 98.8% and 84.8% for SsR10 and DrR1470, respectively.

The autoNOE-RASREC calculations converge for both targets (Fig. 4). The Cα-RMSD of the 10 final models to the respective Xray structure is 1.5±0.2 Å and 2.4±0.4 Å, for SsR10 and DrR1470, respectively. As shown in Fig. 4, multiple RASREC stages are required for both targets to converge to the correct fold. Only after stage IV a convergence to the correct fold is observed. These models are further improved in the following stages, which is driven by all-atom refinement in ROSETTA. However, the removal of violated NOE restraints in phase II and phase III is crucial here to allow ROSETTA a refinement of the structures without being held back by wrong restraints.

As pointed out above, the CANDID algorithm is quite similar to the automatic NOE assignment employed here. CANDID has been the basis of popular programs for automatic NOE assignment, such as CYANA and UNIO, albeit the methods for automatic assignment and peak calibration have evolved since its original publication. The main difference between autoNOE-RASREC and CANDID is thus the tight coupling with RASREC-ROSETTA structural sampling. To illustrate the benefit from RASREC-ROSETTA we
Figure 4. Structural ensembles obtained with automatic NOE assignment. The reference structure (grey) is superimposed with NOE-based models depicted with a colour gradient reflecting the sequence position from N-terminus (blue) to C-terminus (red). (a-e, g-k) Ensembles during autoNOE-RASREC calculation of targets SsR10 (a-e) and DrR1470 (g-k), respectively. Shown are (top to bottom) the 30 lowest energy conformations after RASREC stage III, stage IV, stage VI, stage VII and stage VIII, respectively. Final models for SsR10 and DrR1470 reach a $C_{\alpha}$-RMSD of 1.5 Å and 2.3 Å to the respective Xray structures. (f+l) Final CYANA models of SsR10 (f) and DrR1470 (l) with a $C_{\alpha}$-RMSD to the Xray structure of 4.5 and 14.5, respectively.
have also run CYANA (Version 3.0) to calculate structures for SsR10 and DrR1470 from the same input data. As shown in Fig. 4f, CYANA converges for SsR10 to a tight structural bundle (average backbone RMSD to mean: 0.8 Å) but the resulting structure is not accurate and displays a wrong relative orientation of the helices. For DrR1470 (Fig. 4l), CYANA converges to a structural bundle with some remaining heterogeneity (average backbone RMSD to mean: 3.4 Å). However, the fold reflected in the structural bundle is inaccurate with an RMSD of 14.5 ± 0.3 Å to the reference structure. These results illustrate the benefit of the computationally expensive but enhanced structural modelling of RASREC-ROSETTA. A more comprehensive benchmark of 50 data sets has been carried out subsequently, and its results confirm these observations on a broad basis.\(^{17}\)

4 Concluding Remarks

We have implemented an algorithm for automatic NOE assignment in ROSETTA and coupled it with the iterative conformational sampling method RASREC. RASREC has been shown previously to yield highly accurate structures even from sparse NMR data, but required assigned NOEs or that an initial fold can be determined with CYANA.\(^{4}\) By tightly coupling automatic NOE assignment (as in CYANA) with RASREC we were able to unlock synergies between both approaches and move a significant step further in automatic NOE assignment. Here the method has been shown to yield accurate structures (1.9 Å) for two data sets of proteins in the 15-20kDa size range, whereas CYANA does not even yield an accurate initial fold from the same data set. An independent and more comprehensive benchmark on 50 NOE data sets also shows significantly improved performance for the new method (ZZ and OL, manuscript in preparation). The method is available within the ROSETTA3 software suite (www.rosettacommons.org) and will be released in versions 3.6 or higher. The supporting tool-chain is highly recommended and requires CS-Rosetta toolbox version 2.0 or higher. This toolbox and additional documentation and user support can be found on the CS-Rosetta portal (www.csrosetta.org).

Acknowledgements

We gratefully acknowledge support with computing time from the Jülich Supercomputing Centre.

References


The Flexibility of Fibrinogen and its Initial Adsorption Stages at the Graphite and Mica Surface

Stephan Köhler¹, Friederike Schmid¹, and Giovanni Settanni¹,³

¹ Institut für Physik, Johannes Gutenberg–Universität Mainz,
Staudingerweg 7–9, D–55128 Mainz
E-mail: {koehlst, friederike.schmid, settanni}@uni-mainz.de

² Graduate School Materials Science in Mainz, Staudingerweg 9, D–55128 Mainz

³ Max Planck Graduate Center mit der Johannes Gutenberg-Universität Mainz,
Staudingerweg 9, D–55128 Mainz

Fibrinogen is a multi-protein complex which, when activated, aggregates to form fibrin, a net-shaped molecular formation which is fundamental for the coagulation of blood following, e.g., a wound or when an extraneous body comes into contact with blood (e.g., medical implants). Thus, the adsorption of fibrinogen on material surfaces plays an important role in the viability of those materials for medical applications. It is known that fibrinogen is very flexible in solution, a fact that made it hard to crystallise. The flexibility of fibrinogen can also be observed in adsorption studies. Here we examine the flexibility of fibrinogen using atomistic molecular dynamics simulations. Simulations of fibrinogen in solution and adsorbing on inorganic surfaces are used to evaluate the behaviour of fibrinogen on material surfaces and study the initial adsorption stages. This, in turn, may have implications for medical applications such as material design for implants.

1 Introduction

Fibrinogen (Fg) is a 340kD glyco-protein complex which can polymerise into fibrin, the main component of blood clots. When an external device (graft, catheter) is introduced in the blood stream, it is quickly covered by a layer of adsorbed blood proteins including Fg. Fg plays a particularly important role in the determination of the biocompatibility of device materials, because adsorption-induced conformational changes of Fg are thought to trigger immune response and potentially fatal blood clotting¹. Thus, the understanding of the mechanisms of Fg adsorption on material surfaces represents a significant research challenge, which may have important implications in the design of medical apparatus.

The elongated structure of human Fg, as shown by the crystallographic data², is formed by two symmetric units which dimerise through a central globular E region. Each symmetric unit (protomer) is constituted by 3 peptide chains Aα, Bβ and γ which depart from their N-terminal region (E region), form an elongated coiled coil region, and end into two globular domains forming the D region (Fig. 1). Here, we report the results of extensive molecular dynamics (MD) simulations performed on Fg in solution and on adsorbing surfaces, otherwise not possible without the use of JUROPA and other computational facilities.
Figure 1. (a) Schematic representation of the fibrinogen molecule. The three chains of Fg, Aα, Bβ and γ are shown in red, green and blue, respectively. (b) Van der Waals representation of the crystallographic structure (pdb 3GHG) of Fg, colour coded as in (a). Carbohydrates are in orange. The αC region and the FpA and FpB peptides were not resolved in the crystal structure.

2 Methods

Atomistic Molecular Dynamics simulations

All our simulations are based on the crystal structure of human Fg (PDB ID: 3GHG). Simulations with and without carbohydrate groups have been performed to investigate the effect the carbohydrate chains have on the dynamics of fibrinogen. Protomer-protomer interactions were investigated using simulations of both the full fibrinogen dimer (6 protein chains) and the protomer system (3 chains). Rectangular simulation boxes with explicit TIP3P water and physiological ion concentration (15mMol NaCl) were prepared using VMD (Tab. 1 for box sizes).

Isobaric-isothermal simulations were set up at a temperature of 310K and pressure of 1atm using NAMD with a Langevin thermostat and a Langevin piston barostat. The covalent bonds involving hydrogen atoms were fixed in length and a 2fs timestep was used. The CHARMM force field was used with its recent extension to mica surfaces and sugars in combination with ParamChem (http://www.paramchem.org) and the CHARMM generalised force field (CGenFF). The van der Waals forces were cut off at 12Å while PME was used for long range electrostatic interactions. After minimisation of hydrogen atoms and water molecules, the system was heated and equilibrated for 10ns. Production runs statistics are given in Tab. 1. We employed collective variable constraints to keep the main axis of the molecule aligned to the simulation box and verified no influence on the overall dynamics by comparison with unconstrained simulations.

To identify the collective motions of the protein we performed a Principal Component Analysis (PCA) using wordom. DynDom was used to identify rigid domains and hinges of motion. The overlap between spaces spanned by the dominant PCA modes of different simulations was used to quantify the similarity of the observed dynamics. The
The electrostatic potential of the globular domains of fibrinogen was calculated by solving the Poisson-Boltzmann equation with the APBS software over a series of similar aligned structures and the electric potential was then averaged along the structures. The time averaged potential for each domain was calculated by averaging the potential at each grid point over all snapshots.

For the adsorption simulations, the graphite surface was prepared as a 6-graphene-layer sheet using the VMD Carbon Nanostructure Builder. The model for a mica unit cell that contains a mica double layer with realistic surface defect distribution was kindly provided by H. Heinz. This model was multiplied to produce a surface of the desired dimension. After addition of TIP3P water the system was minimised and equilibrated for 0.6ns before the equilibrated protomer, including the complete E-domain, was placed in the box in different orientations, rotated around its main axis by multiples of 120°. The protein structure had a minimal distance of 8Å from the surface for all orientations. Ionic strength was set to 0.15mMol NaCl. The dimensions of the surface systems are given in Tab. 1. The system was minimised, heated and equilibrated as above before production. No collective variable constraints were used.

### 3 Results and Discussion

#### Hinge in the coiled-coil region

All our simulations show large bending motions of Fg. PCA is used to quantify these motions. It reveals that the dominant modes of the Fg protomer are the same in all the sampled trajectories (large PCA overlap). The first mode is associated with a bending motion around an hinge identified using DynDom (Fig. 2). The hinge region coincides with a break in the α-helical structure of the C chain, which is facilitated by two proline residues (C70 and C76).

The hinge bending can be described more intuitively by a bending angle γ and a dihedral angle ϕ defined using selected groups of atoms on the Fg structure. The distribution of the angles observed in the simulations is given in Fig. 2. The γ and ϕ angles strongly

<table>
<thead>
<tr>
<th>System</th>
<th>Initial box size [nm]</th>
<th>N</th>
<th>Simulation time [ns]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimer, glycosilated</td>
<td>13.27 × 48.59 × 12.70</td>
<td>788173</td>
<td>77+88</td>
</tr>
<tr>
<td>Dimer, unglycosilated</td>
<td></td>
<td>786811</td>
<td>25+20</td>
</tr>
<tr>
<td>Protomer, glycosilated</td>
<td>12.28 × 27.89 × 11.57</td>
<td>381397</td>
<td>199+188</td>
</tr>
<tr>
<td>Protomer, unglycosilated</td>
<td></td>
<td>380169</td>
<td>135+109+100+82+51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphite 0°</td>
<td>12.18 × 27.36 × 13.03</td>
<td>440893</td>
<td>45+20</td>
</tr>
<tr>
<td>Graphite 120°</td>
<td></td>
<td>440875</td>
<td>31+17</td>
</tr>
<tr>
<td>Graphite 240°</td>
<td></td>
<td>440698</td>
<td>49+47</td>
</tr>
<tr>
<td>Mica 0°</td>
<td>17.87 × 28.78 × 14.21</td>
<td>462950</td>
<td>47+47</td>
</tr>
<tr>
<td>Mica 120°</td>
<td></td>
<td>462887</td>
<td>34+35</td>
</tr>
</tbody>
</table>

Table 1. Initial simulation box sizes, particle numbers and trajectory length for the different simulated Fg systems.
correlate with the projections of the dominant PCA modes. As mentioned above, a significant overlap between the dominant PCA modes in glycosilated and unglycosilated Fg trajectories is observed. Thus, carbohydrate clusters do not affect the large scale dynamics of Fg in solution. Similarly, the dimerisation state does not induce any noticeable change in the large scale motion of Fg. Dimer and monomer simulations show consistent hinge bending and PCA overlap between isolated protomer and dimerised protomer is large.

Coarse grained model

A simplified model of Fg has been developed to investigate the effects of the flexibility of Fg as emerging from the atomistic simulations on its adsorption behaviour. The model is constituted by a central rod, representing the stiff coiled-coil regions, which connects the two hinges. The E domain is placed at the centre of the rod. The D-domains are connected to the hinges with rods, that can pivot around the hinge. The dimensions of the model components can be extracted from the crystal structure (Fig. 3). The globular domain are represented as spheres. In the model, adsorption occurs if and only if the distance of the E and the two D domains of a Fg molecule from the adsorbing surface is lower than a threshold \( h_{\text{max}} \) and both hinges lie above the surface. We use a simple Monte Carlo (MC) algorithm to generate a large set of coarse grained adsorbing Fg conformations and for each of them we measure the \( \alpha \) angle (Fig. 3) between the globular domains, as it would be observed in an AFM experiment. We then discretise the set of generated conformations. We identify a limited number of regions corresponding to the areas of \( \gamma \varphi \)-plane sampled by MD (Fig. 2). Each adsorbing Fg conformation is characterised by the two discrete indexes \( i \) and \( j \) of the \( \gamma \varphi \) region where its two hinges occur. We then determine the bending distribution of \( P_{ij}(\alpha) \) observed in the MC sampled conformations associated with Fg conformations from each region of the \( (\gamma, \varphi) \) plane. We then assume
that the observed experimental distribution $P(\alpha)$ is a superposition of the distributions from the possible Fg hinge conformations $P(\alpha) = \sum_{ij} a_{ij} P_{ij}(\alpha)$. The $a_{ij}$ can then be fitted to the experimental distribution. If the two hinges behave independently, then, given the symmetry, we can use a simpler expression $P(\alpha) = \sum_{i} a_{i} P_{i}(\alpha)$.

With this model, based on MD simulation results, we fitted the data coming from two sets of experiments. The first set (Fig. 4 right) from an electron microscopy experiment where Fg was deposited on a hydrophilic carbon film, the model with independent hinges produces a good fit. On the other hand, a correlation between the hinges is necessary to reproduce the distribution of Fg on mica observed in AFM experiments.

This finding prompted us to look for a possible reason for the correlation between the Fg hinges observed on mica. If the two protomers of Fg had the same conformation and the dimer would be placed on a surface, the protomers would face the surface with the opposite sides. Then, the correlation observed between the hinges implies that the two sides of the protomer have substantially different interaction with the surface. Since the correlation is observed on mica surface, which is hydrophilic and negatively charged, it is reasonable to think that the differences between the sides of Fg has an electrostatic origin.
Fibrinogen electrostatics

To verify our hypothesis on the different electrostatic properties of the two sides of the protomer, we calculated the electrostatic potential of fibrinogen. In Fig. 5 we can identify 2 large negatively charged patches per protomer located on one side of the D-domain, but absent on the other side. It has previously been noted that such patches should contribute to the Fg-Fg association during fibrin fibril formation19. Here we can identify the D-domain patch as the γ-hole binding site for the fibrinogen α knob described in Ref. 20.

Figure 5. Left: Typical bend conformation of the fibrinogen dimer with the electrostatic potential at D- and E-domains. Drawn are the isosurfaces at ±26.7mV (blue/red). Right: Number of contacts the glycosilated Fg protomer with full E-domain forms with the mica surface in 4 runs with 2 different initial orientations (red, blue).

Fibrinogen on graphite and mica surfaces

The adsorption simulations revealed several Fg adsorption events (Fig. 6). We analysed the type of contacts made during the early adsorption stages on the two different surfaces (see Tab. 2).

Figure 6. Snapshots of conformations of the Fg protomer where hinge bending lead to the initial adsorption event on graphite (left) and where interaction between mica and the D-patch (grey) reoriented the protomer for adsorption on mica (middle, right).
The charged patches identified with the electrostatic analysis above, contribute significantly more to the interaction of Fg with the hydrophilic mica surface than they do with graphite, especially considering the fact that the surface of the patch covers only 4.7% of the total protein surface. On the other hand, as expected, hydrophobic residues play a less important role on mica, than on graphite. Thus, the simulations of early adsorption support our hypothesis that different electrostatic interactions at the two opposite sides of Fg are responsible for the conformational distribution experimentally observed for adsorbed Fg on mica.

### 4 Conclusion

The classical atomistic molecular dynamics simulations of the Fibrinogen in solution reveal the extraordinary flexibility of the molecule and prompt a revision of earlier models. The picture coming from the simulations indicates that Fg flexibility is due to two hinges located on the two symmetrical coiled-coil regions of the complex, where large bending motions take place. Furthermore, simple considerations on the revised model of Fg lead to hypothesise correlations in the behaviour of the molecule at the two hinges upon adsorption on mica. Reason for the correlations is reasonably found in the asymmetric distribution of charged patches on the surface of the molecule and in particular on the D globular domains. Results from the early adsorption stages of Fg on mica and graphite surfaces, investigated by atomistic MD simulations, support this view.

### Acknowledgements

SK gratefully acknowledges financial support from the Graduate School Materials Science in Mainz. GS gratefully acknowledges financial support from the Max-Planck Graduate Center with the University of Mainz. We gratefully acknowledge support with computing time from the Jülich Supercomputing Centre.

### References


124


Chemistry
The Noble Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt, and Arieh Warshel “for the development of multiscale models for complex chemical systems”. The work awarded focuses on the development (and application) of computational-chemistry methods that combine Schrödinger’s quantum physics with Newton’s classical mechanics. Today, these methods are known as QM/MM methods, where QM stands for “quantum mechanics” and MM for “molecular mechanics”. The QM part is taken care of by ab initio quantum chemistry while empirical force fields are used for the MM part. Furthermore, the work on multiscale modelling was also awarded, in which information about groups of atoms and molecules is included instead of information of the individual atoms (as in atomistic modelling).

The three “chemistry” contributions in the present proceedings of the NIC Symposium 2014 give examples of purely atomistic simulations – from first principles – of complex chemical systems, with emphasis on molecular dynamics (MD) and reactivity. Also the contribution by W. B. Dapp and M. H. Müser may be viewed as “chemistry” contribution. These authors have developed a new atomistic method to model (non-equilibrium) redox reactions using empirical force fields for use in MD simulations. All four contributions together provide beautiful examples of the size and complexity of the systems that today can be modelled theoretically by computer simulations, using either ab initio QM or empirical MM methods. Progress in these fields is still rapid and in particular multiscale modelling is a timely and active field of research.

In their contribution Force-activated Reactivity of Disulfides or the Strange Case of Dr. Jekyll and Mr. Hyde in the field of “covalent mechanochemistry”, P. Dopieralski, J. Ribas-Arino and D. Marx report on computer simulations performed in order to understand the data of reactive single-molecule force-clamp atomic force microscopy (AFM). Intriguing experimental results come from the study of a biomolecular nucleophilic substitution reaction at a disulfide bond at constant external tensile stress. The reactivity changes abruptly at a certain value of the external force when the reaction rate is measured as a function of that force. As a model system, the diethyl disulfide molecule was simulated in water being attacked by an OH− anion. These simulations were conducted using the CPMD code.

G. Bekçioğlu, Ch. Allolio and D. Sebastiani report on Molecular Dynamics Simulations of Heterocyclic Chromophores in Aqueous Solution, aiming at the understanding of the ultrafast (solvation) dynamics in complex hydrogen-bonded systems, with the chromophore in its ground- as well as excited state. For this purpose, as model system of 7-hydroxyquinoline (7-HQ) surrounded by 340 water molecules was simulated using the
CP2K software package. The study of these authors focused on the cis- and trans-rotamers of 7-HQ.

In their contribution *Ab initio Molecular Dynamics Simulation of the Interaction of Liquid Water with Sub-Picosecond High-Intensity THz Pulses*, P. Kr. Mishra, O. Vendrell and R. Santra report on CP2K simulations as well. These authors have found that sub-picosecond intense THz pulses, as they are achievable at modern light-source facilities, are able to transfer a large amount of energy to liquid water, which implies a very high temperature. This study and the two above mentioned “chemistry” contributions provide quite different examples of the variety of modern, atomistic *ab initio* molecular dynamics (AIMD) simulations.

The contribution *Atomistic Modelling of Redox-Reactions in Non-Equilibrium* by W. B. Dapp and M. H. Müser is different in this respect, as it employs empirical force fields in place of *ab initio* quantum-chemical potentials. These authors have developed the redox split-charge equilibration (redoxSQE) method that assigns a discrete ionisation state to each atom. In this method, atoms can swap charges and partial charges can be exchanged across bonds. With redoxSQE, the discharge behaviour of a nano-battery was simulated, and it was demonstrated that this simulation qualitatively reproduced the generic properties of a macroscopic battery.

In conclusion, the three “chemistry” contributions in the present proceedings of the NIC Symposium 2014 provide wonderful examples of the variety of applications of first-principles atomistic simulations that are feasible today. Furthermore, the contribution by W. B. Dapp and M. H. Müser shows how an entire battery can be modelled atomistically using empirical force fields.
Force-Activated Reactivity of Disulfides
or the
“Strange Case of Dr. Jekyll and Mr. Hyde”

Przemyslaw Dopieralski¹, ², Jordi Ribas-Arino³, and Dominik Marx¹

¹ Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, 44780 Bochum, Germany
E-mail: {przemyslaw.dopieralski, domink.marx}@theochem.rub.de

² Faculty of Chemistry, University of Wroclaw, Joliot-Curie 14, 50-383 Wroclaw, Poland
E-mail: mclar@elrond.chem.uni.wroc.pl

³ Universitat de Barcelona, Av. Diagonal 647, 08028, Barcelona, Spain
E-mail: jribasjr@yahoo.es

As the influence of external mechanical forces on free energy surfaces of stretched molecules in solution is largely unknown it is extremely hard to understand the data of reactive single-molecule force-clamp AFM experiments without the aid of computer simulations. Most recent enigmatic experimental results come from the study of a bimolecular substitution reaction at constant external tensile stress, where a disulfide bond in a stretched-out protein is reduced by nucleophiles. In particular, an unexpected abrupt reactivity “switch” at about 0.5 nN has been observed when measuring the reaction rate as it increases as a function of force. Although many possible explanations have been proposed in various publications, all of them remained inconclusive. This issue is addressed here with advanced ab initio simulations of a model system, namely a diethyl disulfide molecule (CH₃–CH₂–S–S–CH₂–CH₃) in water being attacked by hydroxide (OH⁻). In particular, thanks to force-transformed free energy landscape calculations via metadynamics and capability computing, we unravel unambiguously that the reactivity switch is due to a force-induced conformational transition. At low forces, the S–S reaction centre is “open” to nucleophilic attack and thus increasing force greatly accelerates the reaction, whereas at about 0.5 nN the preferred C–C–S–S torsional angle is changed to a “closed” state, thus counteracting the reaction as such. Given the ubiquitous nature of torsional degrees of freedom and thus different conformational states in chemistry and biochemistry, the change from Dr. Jekyll’s to Mr. Hyde’s regime at sub–nN forces is expected to be significant beyond the specific case.

1 Introduction: Covalent Mechanochemistry of Disulfide Bonds

Only during the last decade became it possible to exert nano-Newton forces on molecules in rather controlled ways as required for triggering and influencing site-specific chemical reactions¹⁻³. Thanks to a flurry of recent experiments in the domain of “covalent mechanochemistry” (CMC), it is now well appreciated that mechanical forces can not only promote but also steer truly complex chemical processes. Examples include pericyclic, redox and enzymatic reactions, even if such reactions might be regarded as impossible in the context of “standard” thermal chemistry⁴⁻⁶, e.g. “thermally forbidden” in the Woodward-Hoffmann sense. Recent milestone experiments based on single-molecule atomic force microscopy (AFM) techniques to exert mechanical forces on covalent bonds⁷ thus open a novel avenue to initiate, accelerate, and control site-specific chemical reactions, which defines the emerging field of CMC and “molecular nanomechanics”¹⁻³.
Just recently, AFM experiments carried out at constant force, so-called “force-clamp spectroscopy”, on a mutated 27th domain of cardiac titin unambiguously showed that the rate of the reduction reaction of embedded disulfide bonds exhibits an abrupt change at a critical mechanical force of about 500 pN when using hydroxide (OH⁻) or other nucleophiles to break the S–S bond. But even the thermal chemistry of redox reactions involving disulfide bonds is amazingly challenging, as demonstrated by a century of experiments and decades of theoretical work.

In order to explain the reactivity switch at about 0.5 nN, various putative scenarios to explain the two vastly different reaction regimes have been proposed in the literature or are at least conceivable:

• the reduction of the disulfide bond is described by a (free) energy profile composed of two consecutive energy barriers;
• mechanical forces above 500 pN cause an abrupt conformational change in the substrate disulfide bond involving the C–S–S–C dihedral;
• mechanical forces above 500 pN cause a drastic change of the conformation of the protein containing the probed S–S bond;
• other reaction mechanisms are involved, e.g. C–S bond rupture,
• trimolecular reaction mechanisms might be possible where both ends of the breaking bond could react simultaneously.

As we have recently communicated nucleophilic substitution at sulfur, here abbreviate by “SN₂ at S”, was found to feature the experimentally observed two distinct regimes of disulfide reactivity below 500 pN and above. First of all, no change of mechanism has been found at the critical force value. Next, the “switch” – observed by us for the first time computationally – has been traced back to be a manifestation of the subtle interplay of two antagonistic effects. In the low force regime the pulling force greatly accelerates the rate of the reaction thus showing Dr. Jekyll’s face, whereas stretching out the molecule even more causes an unfavourable conformational change for the SN₂ reaction to occur at about 500 pN and thus exposes Mr. Hyde behaviour.

2 Isotensional Stretching of Molecules

Before presenting our representative results on disulfide reactivity we explain the basic idea of our particular ab initio simulation approach in a few words.

Our earlier work on CMC was exclusively focused on static descriptions of the response of isolated molecules to constant external force, thus providing “isotensional” stretching conditions in vacuo. The central concept of this formalism is the force-transformed potential energy surface (FT-PES), which, given an external constant force $F_0$, is rigorously defined as

$$V_{EFEI}(x, F_0) = V_{BO}(x) - F_0q(x), \quad (1)$$

where $V_{BO}(x)$ is the usual Born-Oppenheimer PES as a function of all nuclear cartesian coordinates $x$ and $q$ is the mechanical coordinate, i.e. a structural parameter being a generalised coordinate in terms of $x$, on which the force acts. For more details we refer to our original work or to our review.
As thermal activation effects are not negligible at room temperature, and as they must eventually override mechanical activation effects at some point, we have recently implemented this so-called EFEI approach, where the “External Force is Explicitly Included”, into an \textit{ab initio} molecular dynamics\textsuperscript{16} framework using the CPMD code\textsuperscript{20}. Importantly, when furthermore combined with \textit{ab initio} metadynamics\textsuperscript{21,22}, this methodology will allow us for the first time to properly study force-transformed free energy surfaces (FT-FES), as initially applied to mechanical ring-opening in the gas phase\textsuperscript{23}. This was the basis for exploring possible reaction pathways as a function of applied external constant force at finite temperatures while also taking the solution environment explicitly into account. The novel technique in conjunction with using the power of capability platforms, such as the Blue Gene/Q machine JUQUEEN at Forschungszentrum Jülich, allowed us to go a major step beyond just computing the FT-PES \textit{in vacuo}.

\section{Disulfide Reactivity: Confirming the Reactivity Switch}

Understanding the reactivity of disulfide compounds is surprisingly challenging. As even the thermal reactivity of disulfides is very complex due to several competing reaction mechanisms, adding a new dimension of complexity, which is mechanical stretching of disulfides, might easily generate a change of mechanism and thus explain the switching in its turn. Prompted by these considerations, we started to study disulfide bond breaking reaction mechanism without applied external force in bulk aqueous environments before starting to pull the molecule.

We have thus examined the hydrolysis reaction of the central disulfide bond of a carefully preselected model system in aqueous solution, which is diethyl disulfide in water attacked by a hydroxide anion. Let us start by justifying this computational model. All previous attempts to theoretically describe the nucleophilic substitution reaction at sulfur, or in general a $S_N2$ substitution at an atom such as S, C, or N using an isolated molecule in vacuum predicted unreasonable transition states and energetic barriers. Microsolvation models, i.e. using a few water molecules in selected static arrangements involving the reaction centre and the nucleophile, is also unable to capture the scene. Implicit solvation schemes like polarisable continuum models (PCM) seem to fail to reproduce the experimentally observed trends as well. Thus it was clear that solvent plays a critical role in such substitution reactions and might even dictate which mechanism will be favoured.

Thus, the obvious conclusion is that one must definitively go beyond the isolated molecule approach and include the solvation effects, where the reactant state including the properly solvated $OH^-$ (aq) species, the transition state and the products are solvated. These considerations together with the reactive nature of the process considered calls for treating not only the reactive species, but also the solvent by electronic structure (here DFT) methods. In the present case, one diethyl disulfide substrate, one $OH^-$ nucleophile, and 70 solvating water molecules were hosted in a 14 Å cubic box subject to periodic boundary conditions. One must notice that already this “small” molecular model system can only be studied in the sense of computing free energy landscapes as a function of force when using capability platforms such as JUQUEEN.

Our \textit{ab initio} molecular dynamics simulations, together with thermodynamic integration (leading to the data shown in Fig. 1) and metadynamics sampling (see Fig. 1) at room temperature, reveal for the first time that the free energy barrier for the nucleophilic
Figure 1. Panel a: Free energy profiles as a function of external force for nucleophilic cleavage of diethyl disulfide in bulk water by OH$^-$ from isotensional \textit{ab initio} thermodynamic integration at $T = 300$ K. The reaction coordinate is defined as $d_{S-S} - d_{S-OH}$, thus reactant and product states are reached in the limit of negative and positive values, respectively. Panel b: Approximate reaction rates obtained from the data in panel a via $k(F) \sim \exp[\Delta A^\ddagger(F)/k_B T]$ (red points) plotted on a logarithmic scale using arbitrary units compared to the experimental data\textsuperscript{8} in the bottom panel (blue); the lines connect the data linearly to guide the eye.

Substitution at sulfur is about 26 to 28 kcal/mol in the thermal limit, which is in good agreement with the experimental value of 21 kcal/mol, and that the reaction occurs in a single concerted step, which is in accordance with the accepted view. Next, when a constant stretching force is applied to the terminal carbon atoms, the free energy barrier is found to drop down (see Fig. 1a) and a significant acceleration of the reaction is observed when converting the activation free energies from thermodynamics integration to rates via $k(F) \sim \exp[\Delta A^\ddagger(F)/k_B T]$, see top panel of Fig. 1b. Most importantly, the computed reaction rates as a function of force do show the same sort of switching behaviour in the sub-nN regime as the force-clamp AFM measurements\textsuperscript{8}, see bottom panel of Fig. 1b.

Having full access to the atomistic details allows us in a next step to analyse the cause of the hitherto enigmatic switching behaviour. The authors of the AFM experiments\textsuperscript{8} suggested that the reactivity switch could be connected to a change of the reaction mechanism in the low-force regime when accessing the $F > 0.5$ nN regime. Our exhaustive study of the force-dependent reactivity, however, shows upon detailed analysis that the mechanism, which is nucleophilic substitution at sulfur, remains the same up to the highest force studied. This calls for further analyses in order to pinpoint the reason behind the switch.
4 Disulfide Reactivity: Explaining the Reactivity Switch

In order to understand the nonlinear behaviour of the computed activation free energies as a function of force we have performed additional simulations using metadynamics. The key idea is that the most favourable approach for an “$S_N 2$ at S” reaction is a back-side attack of the nucleophile with an attack angle close to $180^\circ$. The first collective variable (CV1) for metadynamics is the difference between the $S$–$S$ distance and the distance between the preselected sulfur atom and the attacking hydroxide oxygen, whereas CV2 is the dihedral angle $S$–$S$–$C$–$C$ at the side of the reaction centre. Two force-transformed free energy surfaces as well as the thermal $F = 0$ nN reference landscape as obtained from this setup are depicted in Fig. 2.

![Figure 2](image)

Figure 2. Force-transformed free energy surfaces, FT-FES, for nucleophilic attack at sulfur by hydroxide ion in the absence of force (top) and at two selected external forces (middle, bottom) from isotensional *ab initio* metadynamics sampling at $T = 300$ K. CV1 is the difference between the $S$–$S$ distance and the distance of the attacked sulfur, $S^\ast$, with respect to the attacking hydroxide oxygen, thus negative (positive) CV1-values correspond to reactant (product) states. CV2 is the $S$–$S^\ast$–$C$–$C$ dihedral angle on the side of the reaction centre where $180^\circ$ corresponds to the “closed” conformation (see text) of diethyl disulfide.
We can immediately see that while at zero force the FES features the three minima along the dihedral coordinate CV2 (see dark blue regions for CV1 ≈ −2 Å, which is the reactants’ conformations), already at force of \( F = 0.45 \text{ nN} \) only a single minimum remains at CV2 ≈ 170° in the reactant basin. In the force regime from 0.2 to 0.3 nN the transition from the thermal behaviour to the single-well free energy landscape is observed. This change of dihedral and thus of the population of different conformers is the reason for the experimentally observed reactivity switch at a force of about 500 pN. But why so? Up to the transition state region, i.e. for CV1 up to about 1 Å, the preferred conformational structures of the reactive complex are different at high force compared to zero force, cf. CV2 in Fig. 2 in this regime.

Upon analysing further the ensemble of transition state structures at zero and higher forces the picture sketched in Fig. 3 is uncovered as follows: At the transition state structure of the nucleophilic substitution reaction at sulfur, the angle created by the S–S bond and the attacking hydroxide oxygen, i.e. the S–S–OH angle, must be close to linear for easy attack. Thus the dihedral angle S–S–C–C (and not C–S–S–C as analysed in ample detail in the literature) turns out to be crucial not only in the sense of being an important “slow coordinate” that must be accelerated by metadynamics, but it is also the very mechanistic reason that leads to the observed biphasic behaviour of the reaction rate15.

Based on this analysis it is obvious that during the hydrolysis reaction there is a subtle interplay between two effects taking place. Initially, applying external mechanical forces accelerate the reaction as the system is (slightly) stretched while being in the “open” state for attack, but eventually the preferred conformation is changed from “open” to “closed” states due to stretching the molecule, thus partially “closing” the reaction cone for the “S\( _{\text{N}} \)\( _{2} \) at S" reaction. Note that already at force of about 500 pN the population of the “closed” conformer (with a S–S–C–C dihedral CV2 ≈ 170°) is about 95 %. Thus, Dr. Jekyll transforms into Mr. Hyde by closing the reaction cone, thereby hindering nucleophilic attack and thus damping the increase of reactivity due to mechanical activation. The observed partial closing of the reaction cone up to about 500 pN, as depicted in Fig. 3, together with having no indication for a change of the underlying “S\( _{\text{N}} \)\( _{2} \) at S” reaction mechanism and the qualitatively agreement of computed and measures reaction rates conclusively explains the switching behaviour recently discovered in experiment8.

---

**Figure 3.** Sketch of the reaction cone for the observed “S\( _{\text{N}} \)\( _{2} \) at S” reaction for the “open” and “closed” conformations (see text) of the S–S–C–C dihedral angle due to stretching the fully solvated diethyl disulfide molecule.
5 Conclusions

In summary, the experimentally observed two distinct regimes of disulfide reactivity, i.e. below about 500 pN and above, as observed for a mutated 27th domain of cardiac titin using force-clamp AFM, is attributed here to one and the same mechanism up to the highest force investigated, which is S_N2 disulfide bond reduction. The observed switch is the manifestation of an antagonism involving two effects: stretching the molecule accelerates the reaction but induces at the same time a conformation change of a decisive dihedral angle (which is S–S–C–C and not C–S–S–C as widely discussed by others). This force-induced conformational change partially blocks the reaction cone for the S_N2 reaction and thus hinders substitution at sulfur. The phenomenon emerges as a change of reactivity around 0.5 nN, where the population of the disfavoured “closed” state approaches 100 %. The basic finding that applying force not only straightforwardly lowers activation barriers, but also distorts the stretched molecule with severe impact on the structure of reactive complexes is of broad importance in covalent mechanochemistry.

Acknowledgements

We are grateful to Deutsche Forschungsgemeinschaft (Reinhart Koselleck Grant to D. M.), Alexander von Humboldt Stiftung (Humboldt Fellowship to J. R. A.) and the Catalan Government (Beatriu de Pinós Fellowship to J. R. A.) for partial financial support. The authors gratefully acknowledge the Gauss Centre for Supercomputing (GCS) for providing computing time for a GCS Large Scale Project on the GCS share of the supercomputer JUQUEEN at Jülich Supercomputing Centre (JSC). GCS is the alliance of the three national supercomputing centres HLRS (Universität Stuttgart), JSC (Forschungszentrum Jülich), and LRZ (Bayerische Akademie der Wissenschaften), funded by the German Federal Ministry of Education and Research (BMBF) and the German State Ministries for Research of Baden-Württemberg (MWK), Bayern (StMWFK) and Nordrhein-Westfalen (MIWF).

References

Ab initio Molecular Dynamics Simulations of Heterocyclic Chromophores in Aqueous Solutions

Gül Bekçioğlu¹ ², Christoph Allolio², and Daniel Sebastiani²

¹ Dahlem Center for Complex Quantum Systems, Fachbereich Physik, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany
² Institut für Chemie, Martin-Luther-Universität Halle-Wittenberg, Von-Danckelmann-Platz 4, 06120 Halle (Saale), Germany
E-mail: daniel.sebastiani@chemie.uni-halle.de

This project aims at understanding of the ultrafast dynamics of water molecules especially hydrogen bonded networks. Here, we have performed a first series of first principles molecular dynamics simulations and quantum chemical calculations to elucidate the relationships between local hydrogen bonding structure, and spectroscopic observables. Our approach is based on density functional theory and molecular dynamics simulations, which allows an accurate simulation of the solvation dynamics in complex hydrogen bonded systems.

1 Introduction

1.1 Importance of Functionalised Chromophores

Chromophores in aqueous solutions play a decisive role in determining large variety of chemical and biological processes such as the proton conductivity, acid-base neutralisation rearrangement, and proton pumping through membrane protein channels¹ ⁴.

A molecular probe, typically a fluorescent dye can be employed to obtain site specific data, e.g. from molecular binding pockets or channels. N-methyl-6-quinolone (MQ), for instance, has been used as such a probe for the local terahertz spectrum of water and other polar solvents. It is especially attractive for this purpose, because of its small size, which allows insertion into DNA or proteins (replacing nucleic acids or tryptophan) and even is a photoacid when protonated, so that a flexible atomistic treatment of the excited state also gives access to proton transfer dynamics (see Fig. 1)⁵, ⁶.

The study of photoinduced chemical reactions at a molecular level comprises a key step towards obtaining a deeper understanding of the dynamical effects of the solvent environment on charge transfer reactions in solution, and in broader sense of the effects of solvent dynamics on chemical reactions. The dynamics of photoinduced chemical reactions in aqueous medium depend firmly on the geometrical structure of the solute molecule (distance between donor and acceptor terminals), and the nature of the solvent, especially with respect to the ability of hydrogen bond formation and on hydrogen bond strength.

Despite considerable experimental and theoretical efforts, these chemical reactions mediated by water still remain elusive on the atomistic level.

1.2 Hydroxyquinolines in Water

The photoinduced dynamics and mechanisms of chromophores have been mostly investigated with photoacids which exhibit increased acidity when they promoted to their excited state⁷ ⁹.
Hydroxyquinolines (HQs) are known to undergo proton transfer reactions which are initiated by optical excitation and also by intermolecular interactions with the solvent molecules (see Fig. 2).\textsuperscript{10–15} HQs possess two hydrogen bond attachment points. In the excited state, the hydroxy group acts as a proton donor and the nitrogen acts as a proton acceptor. In our study, we specifically research on 7-HQ.

![N-Methyl-6-Quinolone (MQ) in aqueous solution](image)

**Figure 1.** N-Methyl-6-Quinolone (MQ) in aqueous solution (snapshot from an \textit{ab initio} based molecular dynamics simulation). The two main hydrogen bonding sites are illustrated with coloured spheres.

![7-HQ together with ammonia or water bridges](image)

**Figure 2.** 7-HQ together with ammonia or water bridges.
2 Motivation

The study of solvation dynamics at a molecular level comprises a key step towards obtaining a deeper understanding of the dynamical effects of the solvent environment on charge transfer reactions in solution, and in broader sense of the effects of solvent dynamics on chemical reactions.

In principle, it should be possible to extract comprehensive information about topology, reorganisation, and dynamics of the solvent and solute molecules. Ab initio molecular dynamics simulations are a powerful tool to look at microscopic structures in a detail. Being complementary to the experiment, the ab initio molecular dynamics simulations enable us to study the correlation between microscopic configurations and spectroscopic parameters. Thus, the first principles calculations give access to structure-property relationship and spectroscopic parameters especially in the ground state. Here we study the ground and solvation dynamics of 7-HQ, focusing on the identification of hydrogen bond network topologies.

3 Computational Details

We have computed the atomistic modelling of local solvation dynamics via highly accurate first principles, ab initio density functional theory calculations. In particular, we used Born-Oppenheimer molecular dynamics simulations to investigate 7-HQ as a chromophore to elucidate the complex hydrogen bond networks of liquid water.

3.1 System Setup of Solvated 7-HQ

The molecular dynamics were staged in a cubic, periodic box with a side length of about 21.45 with 7-HQ and 340 H2O at a density of d=1.00 g/cm$^3$. The solvation box is thermostated firstly for 12 ps, using a Nosé-Hoover thermostat with a time constant of 50 fs then continued 18 ps of a run with time constant of 600 fs for ground state. With the increase in temperature we hope to counter overstructuring effects, found in water simulations at lower temperature$^{16}$. For the very same reason, we used the DFT-D2 dispersion corrections$^{17}$. The simulation temperature of 350 K corresponds to a physical temperature of 300-320 K.

For all calculations we used GPW$^{19}$ scheme as implemented in the CP2K$^{20}$ software package. The BLYP functional with a TZVP valence basis set, Goedecker$^{21}$ potentials and a 350 Rydberg plane-wave cutoff.

4 Results and Discussions

4.1 Benchmarking: Electronic Structural Evaluation of 7-HQ

When 7-HQ is triggered to its excited state (one $\pi$ electron is promoted to $\pi^*$ molecular orbital) the first thing that happens is intramolecular charge transfer from the oxygen atom to the aromatic ring system. This charge delocalisation of the oxygen lone pair into the ring system weakens the OH bond, causing the proton dissociation from hydroxy group and/or proton acceptance from solvent molecule to nitrogen feasible.
Table 1. Mulliken charges of Oxygen and Nitrogen atoms of isolated 7-HQ for ground and excited states in elementary charges.

<table>
<thead>
<tr>
<th>Atom</th>
<th>O</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS</td>
<td>-0.3</td>
<td>-0.09</td>
</tr>
<tr>
<td>ES</td>
<td>-0.2</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

Additional support can be found from the Mulliken atomic charges for 7-HQ on the atoms O and N are displayed in Tab. 1 from the first principle calculations in the ground and excited state. We find the negative charge on Oxygen decreases upon excitation whereas increases on Nitrogen which confirms the intramolecular charge transfer in the ring system.

4.2 Rotational Isomerisation of 7-HQ

Two rotational isomers are possible around C-O bond axis of HQs so called cis- and trans-HQs (see Fig. 3). Here we investigated the rotational isomerisation of 7-HQ and its influence on hydrogen bonding. We started our molecular dynamics simulations with cis-7-HQ and observed the rotation of the OH group which formed the trans-7-HQ after 10 ps in the ground state. From the first-principle quantum chemical calculations we also confirmed that trans-7-HQ is more stable than cis-7-HQ by 1.6 kJ mol$^{-1}$.

The rotational fluctuations of the OH group can be seen in the Fig. 4 which shows the dihedral distribution of carbon-carbon-oxygen-hydrogen dihedral angle during ground state.

As a second step we computed coordination numbers of the $N_{HQ-H_{aq}}$ and $H_{HQ-O_{aq}}$ to elucidate the effect of rotation on hydrogen bonding. We found that the coordination number of $H_{HQ-O_{aq}}$ remains at approximately 1 in the case of both isomers. On the other hand, for the case of $N_{HQ-H_{aq}}$, we have the increased coordination number from 1.3 to 1.5 for trans- and cis-7-HQ, respectively.

5 Conclusion

We have presented the results on 7-Hydroxyquinoline in aqueous solutions in the field of *ab initio* molecular dynamics simulations. Concerning 7-HQ in water, we analysed the
hydrogen bond network around 7-HQ. We specifically investigated the hydrogen bonding between the oxygen of the OH group and nitrogen in the distal ring. It has been found that hydrogen bonding becomes stronger in the excited state irrespective of rotation of hydroxyl group. Further researching is required to reveal the more detailed understanding of the role of the solvent in proton transfer reactions.

For dynamics of aqueous hydrogen bond networks and topologies, the main problem is to establish a computational description of the 7-HQ molecule not only in its electronic ground state, but also in its first excited state. These ab initio simulations are strongly computing time demanding and they require the usage of massively parallel codes. For the continuation of this project, we need more computation time to allow a decent sampling over the available phase space.

### Acknowledgements

The molecular dynamics simulations presented here were performed with a partially approved grant of CPU time by the NIC supercomputers of the Research Centre Julich, under project code HBU20. This work as such was supported by Leibniz Graduate School of Molecular Biophysics Berlin.
References


The advent of new accelerator-based light sources producing ultrashort and intense bursts of light in a wide frequency range opens the door to detailed studies of complex molecules and biological structures. Liquid water is the most common solvent of most of the chemical and biological processes. The sub-picosecond response of liquid water to an intense and ultrashort THz pulse has been investigated to understand the energy flow and structural changes by \textit{ab initio} molecular dynamics simulations with time. We find that bulk water gains a large amount of energy from the electromagnetic pulse in about 500 fs, reaching a kinetic temperature of $\sim 900$ K after the pulse. Each water monomer absorbs $\sim 25$ THz photons during the pulse. The transformation of time resolved X-ray diffraction (TR-XRD) pattern from double to single peak with time confirms the depletion of strong hydrogen bond network and also corresponds to a state at a very high temperature. Radial distribution functions ($g_{O-O}(r)$ and $g_{O-H}(r)$) also indicate hydrogen bond depletion and a huge structural change in the liquid state which becomes a gas like state. This complete process is isochoric. THz pumping of liquid water might create a convenient environment for pump-probe experiment on thermal chemical reactions.

1 Introduction

Atomistic molecular dynamics simulations of complex systems possible at supercomputer facilities offer the possibility to “experiment” and learn from the microscopic world \textit{in silico}. \textit{Ab initio} molecular dynamics (MD) is a branch of MD methods in which the forces between atoms are calculated from quantum mechanics first principle theories. These methods are highly accurate, but also very time consuming for moderate size system, meaning that supercomputers play a decisive role in this type of research.

Water, the “matrix of life”\textsuperscript{1}, has forced researchers to explore it computationally and experimentally more and more from last several decades because of its numerous anomalous behaviour and interesting properties\textsuperscript{2}. In recent years, it has been discovered that liquid water does not behave only as an inert and passive environment around the solute\textsuperscript{1}. This leads to the dynamics of water, as a solvent around the solute during chemical and biological process\textsuperscript{3} which play a key role in solvation and stabilisation of reaction intermediates. The strong hydrogen bond network is one of the main fundamental reasons for all these fantastic properties of water.

In recent years, the ultrafast pump-probe spectroscopic studies on liquid water in a broad frequency range are exemplifying the growing interest to understand the interaction of water with electromagnetic radiation. The intermolecular vibrational modes lie in the frequency domain from 1 THz to 10 THz ($\approx 30$-300 cm$^{-1}$)\textsuperscript{4,5}. This feature makes the
THz radiation a special frequency range to investigate the hydration dynamics around the solute\textsuperscript{6}. There have been several experimental and theoretical studies of THz spectroscopy on water dynamics at low intensity for solutes like carbohydrates and proteins\textsuperscript{7,8}. THz absorption has revealed cooperative effects extending over several solvation shells (∼7-20 Å) which have influence in its hydrogen bond dynamics in the ps time scale.

2 Motivation

The advent of high intensity, high frequency and ultrashort light sources has turned the dream into reality for pump-probe experiments with femtosecond resolution and high brilliance\textsuperscript{9}. Free electron lasers (FELs) operating in the THz domain have been available for some time now. Recently, the generation of THz radiation from X-ray FELs (XFELs) in full synchronisation with the X-ray pulse has been demonstrated\textsuperscript{10,11} at many modern light sources such as Linac Coherent Light Sources (LCLS), USA and FLASH, Germany. This provides an interesting platform for pump-probe experiments where THz can excite the intra-molecular modes and interrogates the system at well defined delay time by an ultrashort X-ray pulse. The combination of X-ray probe with THz pump scheme has already been used in pump-probe experiments in material science specifically in light induced superconductivity applications\textsuperscript{12}. Liquid water has also been studied in ultrafast infrared spectroscopy and nonlinear vibrational spectroscopy\textsuperscript{13,14} where infrared has been used to excite the intramolecular librational modes, OH stretching and bending and follow the dynamics with a fs X-ray probe pulse. In the IR domain, the intense pulse excites the intramolecular vibrations and follow the energy dissipation processes using fs IR lasers. It has also been studied that THz light couples to low energy collective modes of the liquid at low intensity\textsuperscript{7,15}. But all these studies have been performed at low intensity where mostly one photon absorption processes take place. However, there is not much known about the sub-picosecond response of liquid water to intense and ultrashort (one cycle) THz pulse which is mostly due to the lack of high intensity THz sources with pulse duration in the fs regime. Availability of highly intense and ultrashort THz pulse in full synchronisation with X-ray FEL has motivate us to perform time-resolved pump-probe study on liquid water.

3 Method

Molecular Dynamics (MD) simulations have been vastly performed to investigate the energy relaxation, redistribution and time-resolved structural changes in the liquid phase. Ab Initio Molecular Dynamics (AIMD) simulations were performed by using CP2K package in which nuclear motion is described classically and the electronic structure is described quantum mechanically by density functional theory (DFT). The AIMD Born-Oppenheimer trajectories have been propagated via Quickstep electronic structure module as implemented in CP2K. Gaussian and plane waves (GPW) method of quickstep module was used to study the electronic structure during geometry optimisation (GO) and MD. The Perdew-Burke-Ernzerhof (PBE) functional was used together with Geodecker-Teter-Hutter (GTH) pseudopotential to define the electronic density. We used plane waves up to 400 Ry and the TZV2P basis set to expand the density and orbitals respectively. All the simulations were carried out under periodic boundary conditions in all three spatial directions. 1 fs long timestep was used in all simulations.
Figure 1. Average CPU time for each step MD simulation (excluding first step) for 32, 64, 128 and 256 water molecules at JUROPA. Each node contain 8 processors.

The computation cost of AIMD simulations on CP2K depends many factors like basis set, pseudopotential and size of the system etc. The first step of the AIMD simulation is \(\sim 2.5\) times more costly than the rest. Here we report the cost of AIMD simulation for each step (excluding the first step) with above mentioned tools for 32, 64, 128 and 256 water molecules (Fig. 1). The simulations are performed on the JUROPA supercomputer, at Jülich Supercomputing Centre, Germany where each node contain 8 processors.

A cubic box containing 128 water molecule at a density of 1 g/cm\(^3\) was used to optimise the geometry and then the optimised configuration was equilibrated thermally at room temperature of 300 K with canonical (NVT) MD with nose-hoover thermostat. The thermally equilibrated configuration was propagated for 5ps long time to collect 15 initial configurations (positions and velocities) at certain time intervals which are statistically uncorrelated to each other. Each initial configuration was propagated microcanonically (NVE) for 1.5 ps long time in the presence of THz pulse. During NVE-MD, the AIMD Born-Oppenheimer trajectories were propagated with explicit inclusion of electric field in Quickstep. Such THz pulse is given by

\[
E(t) = \epsilon(t)u_z \cos(\omega_c t + \phi),
\]

where \(\epsilon(t)\) is a Gaussian envelope with a full width at half maximum (fwhm) of 250 fs and intensity \(10^{10}\) W/cm\(^2\). This intensity is achievable at state-of-the-art THz sources. Assuming a focus of the THz light to a spot of 1 mm\(^2\) represents a total pulse energy of 25 \(\mu\)J\(\times 10^6\). We assume a photon energy \(\omega_c = 100\) cm\(^{-1}\) (\(\approx 3\) THz). At this photon energy, one cycle of the electromagnetic field oscillation takes about 330 fs, meaning that our pulse is between a half and a full cycle long. \(u_z\) is the polarisation direction of the electric field and \(\phi\) is the carrier to envelope phase (CEP), which we set to \(\pi/2\). We also tried other CEPs and obtained very similar results. Pulses with these characteristics match the design specifications of THz sources available at FELs. The THz pulse envelop is centred at \(t=0\) and trajectories start at \(t=250\) fs. During NVE MD, thermostating was switched off.
The same set of simulations has also been performed with box sizes of 32 and 64 water molecules at density of 1 g/cm$^3$ to gauge the possible box-size effects. Time resolved X-ray diffraction (TR-XRD) patterns of the system were obtained by Fourier transform of the electron density of snapshots during the dynamics. We averaged over the 15 available trajectories at every time delay between THz and X-ray probe pulse and convolved the set of patterns with a 20 fs Gaussian envelope to represent the finite length of the probing X-ray pulse. The Gaussian and augmented plane waves (GAPW) method of Quickstep was used instead of GPW for the calculation of the TR-XRD patterns. GAPW is an all electron method, which is crucial to describe the electron density near the nuclei.

4 Results

We found that the THz pulse transfers a large amount of energy to liquid in a sub-picosecond time-scale. At time $t=-250$ fs, water is in thermal equilibrium at the temperature $T=300$K in the absence of an electric field. We partitioned the total kinetic energy (KE) of the water monomer into 3 different distributions:

1. Translational (TE) : Translation motion of the centre of mass of each molecule
2. Rotational (RE) : Rotation of molecule around its axis
3. Vibrational (VE) : Internal vibrations of a molecule

VE includes intra-molecular bending and stretching motions which have not been separated here. TE and RE contributions are from hindered rotations and librational motions in the water monomer. Fig. 2(a) shows the mean value of the total KE per water monomer as a function of time, as well as the TE, RE and VE contributions. At time $t=-250$ fs, KE is equipartitioned among all the degrees of freedom. All three distributions correspond to equilibrium Maxwell-Boltzmann distribution at $T=300$K. The probability distribution of TE, RE and VE has been presented together with Maxwell-Boltzmann distribution at 300 and 330 K in Fig. 2(b).

![Figure 2](image-url)
The total KE per monomer with time shows that the monomer gains a large amount of energy quickly during the pulse and remains constant after the pulse for a long time. The total length of the pulse is \( \sim 500 \text{ fs} \). The KE increases more than \( 2000 \text{ cm}^{-1} \) which translates into a final kinetic temperature \( T_k \) of \( \sim 900K \). The Total KE per water monomer for the box of 32, 64 and 128 water molecules has also been presented in Fig. 3(a) to address the box-size effect which shows the nice convergence with box size. Each monomer absorbs an energy from the pulse equivalent to \( \sim 25 \text{ THz photons} \). Energy partitioning shows that as the pulse starts propagating, the translational mode starts gaining energy. This is due to sudden modification of intermolecular potential and specially due to modification of the strength of the hydrogen bonds by THz pulse. During the pulse, most of the KE is gained by translational mode. Just before the pulse is over, the TE starts to redistribute among the other modes. The large amount of energy gain by translational modes increases the mobility of the water molecules, which causes the increment in molecular collisions. Both the RE and VE contributions starts to increase due to molecular collisions. Rotational and vibrational modes keep on gaining energy after the pulse is over which indicates that THz does not excite the high frequency intramolecular mode of water (or only to very small extent).

We have also performed simulations on isolated water molecules. It has been found that THz pulse has no effect on the vibrational mode of molecule (Fig. 3(b)) and there is only a very small energy transfer to rotational mode as compared to amount of energy gained by rotational mode of water monomer in the bulk. Small clusters of water such as dimer and trimer have no vibrational mode in the THz spectral range. For such clusters, we observed that THz pulse induces the break up of the hydrogen bond among the molecules. This is because the pulse has disturbed the inter-molecular potential through polarisation along the hydrogen bond. Such effect has already been reported for water clusters forming the hydrogen bond network exposed to static field electric field\(^{18}\).

THz pulse transfer large amount of energy to liquid water which also causes large structural changes. A convenient probe of the structural transformations of a material is the coherent X-ray diffraction (XRD) pattern which is given by the modulus squared of

![Graphs showing energy distribution](image)

Figure 3. (a) Total kinetic energy per water monomer for the box of 32 water (green curve), 64 water (blue curve) and 128 water (red curve) (b)Vibrational Energy Gain (\( \Delta E_{\text{vib}} \)) per molecule as a function of time for bulk water (magenta) with electric field (EF) in Z-direction and single water molecule with EF in X (green), Y (blue) and Z-direction (red).
the Fourier transform of the electron density. XRD pattern has been studied theoretically and experimentally by many researchers for bulk water for a broad range of temperatures and pressures\textsuperscript{19,20}. XRD pattern of liquid water contains double peak at T=300K\textsuperscript{17}. The peak at $\sim 2$ Å is related to a typical oxygen-oxygen distance in tetrahedrally coordinate water, whereas lower peak at $\sim 3$ Å is related to features in the density connected to the hydration-bonding structure. XRD pattern at $t=-250$ fs is clearly showing the double peak behaviour at T=300K. The TR-XRD (Fig. 4) shows transitions into a pattern with a single peak at about 2.2 Å with in time between -50 and 150fs.

This is a strong indication of dramatic modification of the tetrahedral structure of the liquid. This is also a signature of disruption of the hydrogen bond network of water. The TR-XRD pattern obtained after the pulse is similar to the XRD pattern of water at temperature of 1500 K and pressure of 12 GPa\textsuperscript{19} where a complete disruption of the shell structure of liquid has been reported. This strongly indicates that a kind of extreme condition which

![Figure 4. Coherent X-Ray diffraction intensity at different time delays between the THz and X-ray pulse. Time is given in fs.](image-url)

![Figure 5. (a) Radial distribution functions (a) $g_{O-O}(r)$ and (b) $g_{O-H}(r)$ at different time delays under the effect of THz pulse. Times are in fs.](image-url)
belongs to a very high temperature and pressure has been achieved within a sub-ps time scale. Fig. 5(a) shows the $g_{O-O}(r)$ radial distribution function (RDF) calculated as a function of time. As in the TR-XRD, $g_{O-O}(r)$ transitions from two peaks, which is indicative of the first two solvation shells around each molecule, into a flat RDF typical of a gas. This change occurs between -50 and 150 fs after start of the THz pulse. Similarly, the $g_{O-H}(r)$ RDF and Fig. 5(b) shows a very fast disruption of the hydrogen bond network, clearly seen by the disappearance of the peak at ~2 Å which is related to the oxygen-hydrogen distance in a hydrogen bond. We remind that the structural changes occur isochorically since there is no time for the bulk to change its volume in the ps time scale of the excitation.

5 Concluding Remarks

In conclusion, a very hot and structureless environment of liquid water is created by a sub-ps intense THz pulse, achievable at modern light source facilities. Such sub-ps pulses are able to transfer a large amount of energy to the liquid water which translates to a very high temperature. The water molecules mostly acquire energy as they start to move relative to each other in response to changes in the hydrogen bond strengths and through impacts the translational energy redistributes in a few hundred fs into monomer librations and vibrations. These changes can be monitored experimentally by time-resolved measurements of X-ray diffraction patterns of the liquid. Even theoretically, radial distribution functions are also indicating the same observations. The transient and hot gas like liquid environment achieved by the THz pulse can have interesting properties as a matrix to study activated chemical processes and and further large scale ab initio molecular dynamics simulations will be key in guiding future experimental attempts in this direction.

Acknowledgements

We are thankful to the “Virtual Institute of the Helmholtz Association: Dynamic Pathways in Multidimensional Landscapes” for financial support. All these computations were performed with a grant of computing time on JUROPA cluster provided by Jülich Supercomputing Centre, Jülich.

References

Elementary Particle Physics
Elementary Particle Physics: Towards Precision Physics

Gernot Münster
Institut für Theoretische Physik, Universität Münster, 48149 Münster, Germany
E-mail: munsteg@uni-muenster.de

The constituents of atomic nuclei are themselves composed of smaller particles, the quarks. In nature, quarks do not appear isolated but bind together to form hadrons, consisting of two or three quarks. The force between quarks is mediated by so-called gluons. The physics of quarks and of the strong interactions between them is theoretically described by Quantum Chromodynamics (QCD). This theory has been applied very successfully to high energy collisions in particle accelerators. However, many physically important questions in QCD are still open. This is due to a fundamental feature of QCD, namely the fact that the interaction strength depends on energy and gets rather large at low energies. Consequently, fundamental properties of hadrons like their masses and other basic observables cannot be calculated by analytical methods. The numerical approach to solve non-perturbative problems of QCD is based on lattice QCD. Space-time is discretised on a four-dimensional lattice and quantum theoretical expectation values are calculated by evaluating the Feynman functional integral numerically with the Monte Carlo method. In the end, of course, the results have to be extrapolated to the continuum limit. This represents an approach to quantum field theory from first principles.

In previous decades the primary aim of lattice QCD was to establish basic characteristics of QCD and to test the theory against experiment. It has for example been possible to calculate the masses of several elementary particles with quite small numerical errors and to check that the results are in agreement with the experimental values.

In recent years lattice QCD has entered the era of precision physics. There are many quantities in hadron physics which are very important for the analysis and interpretation of experimental results, but have been theoretically unaccessible so far. Simulations of lattice QCD on supercomputers offer the possibility to obtain precise results for such quantities. An example is the mass splitting between the proton and the neutron, which is due to the small difference between the masses of up- and down-quarks and to electromagnetic effects. More generally, a number of properties of elementary particles within the same multiplet vary due to these effects. This issue is discussed in the articles of Fodor et al. and of Schierholz, where results of investigations concerning the role of quark mass splittings and of electromagnetic effects are presented. In order to obtain the desired precision, different sources of systematic errors have to be controlled, as explained in Fodor’s contribution. In the article of Wittig et al. other physical quantities, the knowledge of whose precise values is highly desirable, are discussed. They include the anomalous magnetic moment of the muon and hadronic form factors. Using various technical improvements, results relevant for precision physics could be obtained.
The subject of the contribution of Katz et al. is the physics of hadronic matter at high temperatures, as is investigated in experiments at e. g. the LHC. QCD predicts a phase transition from hadronic matter to a quark-gluon plasma. The nature of this transition and the behaviour of characteristic observables as a function of temperature is studied in the framework of lattice QCD.

Although much work in lattice field theory is devoted to QCD, there are other field theoretical models offering non-perturbative problems that can be studied by means of numerical simulations. One of them is supersymmetric Yang-Mills theory, which is the subject of the contribution of Bergner and collaborators. Supersymmetric theories play a central role in attempts to formulate models for the physics beyond the Standard Model of elementary particle physics. Supersymmetry puts certain restrictions on the structure and properties of particle multiplets. The article reports on progress towards confirming the non-perturbative existence of a supersymmetric continuum limit from the lattice and investigating the properties of the low lying supermultiplets.

It should be noted that the reports about the work of the NIC Research Groups by Bleicher et al., Philipsen et al. and Jansen et al., which can be found in this volume, also address topics from the theory of elementary particles.
Isospin Splittings in the Light Baryon Octet from Lattice QCD+QED at the Physical Mass Point

Zoltan Fodor\textsuperscript{1,2,3}, Stefan Krieg\textsuperscript{1,2}, and Thomas Lippert\textsuperscript{1,2}

\textsuperscript{1} Bergische Universität Wuppertal, Fachbereich C - Physik, D-42119 Wuppertal, Germany
E-mail: fodor@physik.uni-wuppertal.de

\textsuperscript{2} IAS, Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany
E-mail: {s.krieg, th.lippert}@fz-juelich.de

\textsuperscript{3} Institute for Theoretical Physics, Eötvös University, H-1117 Budapest, Hungary

Electromagnetic effects and the up-down quark mass difference have small but highly important effects on octet baryon masses. A prominent example is the stability of the hydrogen atom against beta decay. Here we report on a calculation\textsuperscript{1} that includes these effects by adding them to valence quarks in an $N_f=2+1$ lattice Quantum Chromodynamics calculation based on ensembles with 5 lattice spacings down to 0.054 fm, lattice sizes up to 6 fm, and average up-down quark masses all the way down to their physical value. This large parameter space allows us to gain control over all systematic errors, with the exception of the one associated with neglecting electromagnetism in the sea. We compute the octet baryon isomultiplet mass splittings, as well as the individual contributions from electromagnetism and the up-down quark mass difference. Our results for the total splittings are in good agreement with experiment.

1 Introduction

All observed particle physics phenomena are accurately described by an $SU(2)_L \times U(1)_Y \times SU(3)_c$ relativistic quantum gauge theory known as the Standard Model. In this theory, the $SU(2)_L \times U(1)_Y$ component explains the weak and electromagnetic interactions, while the $SU(3)_c$ component, known as Quantum Chromodynamics (QCD), describes the strong interaction between quarks and gluons. For light or heavy-light hadron processes, the Standard Model reduces to an $SU(3)_c \times U(1)_{em}$ gauge theory, where the $U(1)_{em}$ stands for Quantum Electrodynamics (QED).

Simulations of QCD use the lattice regulated theory, called lattice QCD. Present-day, state-of-the-art lattice QCD computations are performed in the isospin limit, in which it is assumed that the $u$ and $d$ quarks are mass degenerate (i.e. $m_u = m_d = m_{ud} \equiv (m_u + m_d)/2$) and in which electromagnetism is neglected. This framework is known as $N_f=2$ or $N_f=2+1$ (if strange sea quarks are included) lattice QCD. In this framework, we have recently obtained important results amongst which are those for the light hadron spectrum\textsuperscript{2}, for the average up-down and strange quark masses\textsuperscript{3,4}, for the $SU(3)$-flavour breaking effects in the ratio of leptonic decay constants $f_K/f_\pi$\textsuperscript{5}. These results were obtained with fully controlled, combined statistical and systematic errors on the few percent level.

Given this important progress and the fact that we are now reaching percent level accuracies in our QCD computations, it is becoming critical to include QED and quark-mass isospin breaking effects, the last ingredients required to claim to have a full Standard Model description of quark processes at low energies.
While QED and isospin breaking effects are small for most hadronic quantities, their consequences far surpass their numerical size. For instance, they are strongly believed to be responsible for the fact that neutrons are heavier than protons, thereby ensuring the existence of stable atoms and, more generally, of the large majority of visible matter in the universe. Moreover, they are required to determine the individual up and down quark masses \(m_u\) and \(m_d\) \textit{ab initio}. This is important because a vanishing up quark mass, \(m_u = 0\), would provide a very elegant solution to the strong CP problem. Though this possibility is very unlikely given present knowledge\(^6\), it will only be ruled out for certain when isospin breaking corrections are fully calculated. The determination of the individual \(u\) and \(d\) quark masses is intimately related to the corrections to Dashen’s theorem\(^7\), which have been the object of heated debates ever since its formulation. In addition, thanks to the progress made recently in lattice simulations, a number of very important theoretical predictions for particle physics have errors in the range of a few percent. With isospin breaking corrections parametrically on the order of about 1\%, it is clear that for further progress to be made, they will have to be included.

Here, we report on a calculation\(^1\) aiming to compute the aforementioned neutron-proton mass difference as well as the remaining octet baryon isomultiplet mass splittings. The isospin quark mass splittings are included in the partially quenched approximation and electromagnetic effects in quenched QED. Our results for the individual quark masses and the implications on Dashen’s theorem will be discussed in an upcoming publication\(^8\).

2 Simulation

Our simulation setup is the following\(^1,9-11\): in order to study QED and isospin-breaking \((m_u - m_d = \delta m < 0)\) effects on hadron properties, we simulate three flavours of quarks (at physical mass parameter values) and QED in the valence sector, using our 47, 2010 ensembles\(^3,4\) with 5 lattice spacings down to 0.054 fm, lattice sizes up to 6 fm and average up-down quark masses all the way down to their physical value. In this way, we take the dominant effects induced by QED and isospin breaking into account: since mass and e.m. isospin symmetry breaking corrections are small and of comparable size, it is legitimate to expand the standard model in powers of \(\delta m\) and \(\alpha\), assuming \(O(\delta m) \sim O(\alpha)\) (and \(O(\delta m^n), \delta m\) to be normalised by a typical QCD mass scale). Given the magnitude of the expansion parameters, this expansion is expected to converge very rapidly, with each subsequent order contributing \(\sim 1\%\) of the previous one. Considering the typical size of other uncertainties in our calculation, we can safely work at LO in this expansion, i.e. at \(O(\delta m, \alpha)\).

2.1 Simulation Parameters

We have to fix the four parameters of three flavour QCD, the quark masses \(m_u, m_d, m_s\) and the lattice spacing \(a\), setting the bare \(\alpha\) to its renormalised value, which is justified in a quenched QED calculation. In order to set the quark masses, we use the observables \(M_\pi^+\) to set the average up- and down-quark mass \(m_{ud}\), \(M_K^2 \equiv (M_{K^+}^2 + M_{K^0}^2 - M_{\pi^+}^2)/2\) to set the strange quark mass \(m_s\), \(\Delta M_K^2 \equiv M_{K^+}^2 - M_{K^0}^2\) for the isospin breaking of the light quarks \(\delta m\), and either \(M_{\Omega^-}\) or the isospin averaged \(\Xi\) mass to set the scale. To match valence and sea calculations, we tune \(m_{ud}\) and \(m_s\) so that \(M_{\pi^+}\) matches the sea pion mass
and $M_{K^+}$ reproduces its sea value. To that end, we generated three datasets (see Fig. 1). For the first valence dataset we tuned the individual bare up- and down-quark masses such that they are approximatively both equal to the sea light mass. To perform this tuning we had to determine the critical mass shifts in the up- and down-quark mass coming from
the e.m. self-energy\textsuperscript{12}. In the second valence dataset, \( m_d \) is set to be heavier than in the first one simulating the physical splitting \( \delta m \); in the third set we vary \( \alpha \) in order to be able to separate chiral dependencies with good precision. The lattice spacings are determined simultaneously from a combined fit of the data with \( \Delta M \simeq 0 \), using techniques described earlier\textsuperscript{1–4}. The isospin mass splitting \( \Delta M_X \) of a hadron \( X \) is naturally described by the LO isospin expansion:

\[
\Delta M_X = A_X \alpha + B_X \Delta M^2 ,
\]

where \( \Delta M^2 \) substitutes for \( \delta m \). The coefficients \( A_X \) and \( B_X \) still depend on the isospin symmetric parameters of the theory, e.g. \( m_{ud} \) or \( m_s \). We find that their dependence on these parameters is well described by a linear expansion in \( M_{SU}^2 \) and in \( M_{\tilde{K}}^2 \) for the range of masses considered here.

The separate e.m. and \( \delta m \) contributions to the baryon mass splittings are interesting in their own right. In order to compute their individual magnitudes, we use the masses of the quark-connected pseudo-scalar mesons \( \bar{u}u \) and \( \bar{d}d \), with \( \delta m = 0 \) now implying that their mass difference vanishes: \( \Delta M^2 = M_{\bar{u}u}^2 - M_{\bar{d}d}^2 = 0 \) (with the remaining parameters tuned to their physical values). Using \( \chi PT \) results\textsuperscript{13}, it is straightforward to show that the difference of these squared masses is \( \Delta M^2 = 2B_2 \delta m + O(\alpha m_{ud}, \delta m m_{ud}, \alpha \delta m, \alpha^2) \), where \( B_2 \) is the \( N_f=2 \) quark condensate parameter. Close to the physical point, \( O(m_{ud}) \) can be counted like \( O(\delta m) \). This definition of the e.m. contribution, therefore, differs from any other valid one only by higher order terms.

The \( \delta m \) contribution can be obtained by working with \( \alpha = 0 \), with the other parameters again tuned to their physical values. In particular, the physical value of \( \Delta M^2 \) is obtained from the analysis of \( \Delta M_{\tilde{K}}^2 \) and by computing the value of \( \Delta M^2 \) corresponding to the physical \( \Delta M_{\tilde{K}}^2 \) value\textsuperscript{1,8}.

In Fig. 1 we show the range of parameters used in our study. Our extensive dataset allows us to gain full control on the \( \delta m \) and \( \alpha \) dependence of the baryon mass splittings. This is shown in Fig. 2, where we display the \( \Delta M^2 \) dependence of \( \Delta M_{\Xi} \equiv \Xi^0 - \Xi^- \) and the fully controlled interpolation to the physical value of \( \delta m \), determined from the experimental measurement of \( \Delta M_{\tilde{K}}^2 \) as described above.

We then perform similar interpolations for \( \Delta M_{\tilde{K}}^2 \), \( \Delta M_N \equiv M_p - M_n \) and \( \Delta M_{\Sigma} \equiv M_{\Sigma^+} - M_{\Sigma^-} \) as well as all of the other interpolations and extrapolations required to tune to the physical \( m_{ud} \) and \( m_s \) masses, and to the continuum and infinite volume limits.

2.2 QED

We generate an e.m. field \( A_\mu(x) \) for each QCD configuration, using the non-compact e.m. action (in a Coulomb gauge). The action is quadratic and, therefore, the generation of the field straightforward in Fourier space\textsuperscript{a}. The gauge potential is then fast Fourier transformed back to position space and exponentiated as \( U_\mu^{\text{QED}}(x) = \exp (i q A_\mu(x)) \). Unlike the QCD links, our QED links are not smeared before being coupled to quarks. Also, we have not added a clover improvement term for the \( U(1) \) field. The \( U(1) \) fields are subsequently multiplied with the \( SU(3) \) gauge variable on each link and inserted into the Wilson Dirac operator associated with the quark of charge \( q \) whose propagator we wish to compute.

\textsuperscript{a}Here, periodic boundary conditions require subtraction of the zero Fourier mode, \( \tilde{A}_\mu(p = 0) \).
Figure 2. Interpolation of $\Delta M_{\Xi}$ to the physical value of $\delta m$ holding $\alpha$ fixed to its physical value, $(\Delta M_{\text{phys}})^2$ denotes the physical value.

2.3 Finite-Volume Effects

Finite-volume (FV) effects are particularly important in a QED calculation, because of the presence of the massless photon. Using published techniques\textsuperscript{14}, and performing appropriate asymptotic expansions, it is straightforward to show that the leading finite-volume term in scalar and spinor QED is proportional to $1/L$. These corrections are typically large, as can be expected. In boxes with $L = 1.6 \div 2.6$ fm, which is the largest range of sizes considered in all but one\textsuperscript{15} previous studies, the correction to $\Delta M_{\Xi}$, the QED contribution to $\Delta M_{\Xi} \equiv M_{\Xi^+} - M_{\Xi^-}$, ranges from 123 to 76%. This is illustrated in Fig. 3, where we plot our results for $\Delta M_{\Xi}$ as a function of $1/L$. Similar FV corrections are found for QED contributions to other splittings. It is clear that with such corrections one cannot claim to control the extrapolation of QED contributions to infinite volume.

In our calculation $L$ extends up to 6 fm, where we find 36% FV corrections. While still large, these corrections are sufficiently small that they may be described with a low-order polynomial in $1/L$. This is confirmed by the data in Fig. 3, which show no sensitivity to terms beyond linear order in $1/L$. The same features are observed in our results for $\Delta M_N \equiv M_p - M_n$, but with larger statistical errors. Thus we find it sufficient to extrapolate these quantities linearly to the infinite volume limit. The situation is different for $\Delta M_{\Sigma} \equiv \Delta_{[\Delta I_3=2]} M_{\Sigma} = M_{\Sigma^+} - M_{\Sigma^-}$, where the $1/L$ dependence is very small, as expected.

2.4 Discretisation Effects

Concerning discretisation effects, the improvement of the QCD action implies $O(\alpha_s \alpha, a^2)$ corrections to $A_X$ and $B_X$. However, due to the lack of improvement in the coupling of
the photon to quarks, discretisation effects on $A_X$ are $O(a)$. In our analysis, we include $O(a)$ QED discretisation effects to $A_X$ as well as $O(\alpha_s a, a^2)$ QCD ones to $B_X$.

Combining all of this information yields a 9 parameter description of each of the mass splittings. In the notation of Eq. 1, this corresponds to:

$$A_X = a_0^X + a_1^X [M_\pi^2 - (M_{\text{phys}}^\pi)^2] + a_2^X [M_{K^\pm}^2 - (M_{\text{phys}}^{K^\pm})^2] + a_3^X a + a_4^X \frac{1}{L},$$

$$B_X = b_0^X + b_1^X [M_\pi^2 - (M_{\text{phys}}^\pi)^2] + b_2^X [M_{K^\pm}^2 - (M_{\text{phys}}^{K^\pm})^2] + b_3^X f(a)$$

where the $a_i^X$ and $b_i^X$ are the parameters and $f(a) = \alpha_s a$ or $a^2$, alternatively. These functional forms characterise the dependence of the mass splittings on the parameters required to reach the physical point and to separate them into $\delta m$ and e.m. contributions. However, the many competing dependencies make this study particularly challenging.

In our fits we keep only parameters whose fitted values are more than one standard deviation away from zero. For $\Delta M_{K^\pm}^2$, all parameters are relevant. We also allow for different parameter combinations if they satisfy the previous requirement and not eliminated through their poor fit quality.
2.5 Error Estimation

We follow our histogram based analysis strategy\(^2\)\(^-\)\(^5\) to control the systematic uncertainties. Here, we consider the following variations in our analysis procedure. We use two different initial times in our correlator fits, one for which we expect negligible excited state contributions and a second more aggressive one, allowing us to control excited state effects. With the \(\Omega^-\) and the isospin averaged \(\Xi\) we have to ways to set the scale. The uncertainty associated with the truncation of the Taylor expansion used to interpolate these two masses to physical \(M_{\pi^+}\), is estimated by varying the fit ranges excluding all data with a pion mass above 400 or 450 MeV. To estimate part of the same uncertainty for the isospin splittings, we consider cuts at \(M_{\pi^+} = 450\) and 500 MeV, since their \(M_{\pi^+}^2\) dependence is very mild. Furthermore, we include either \(\alpha_s\) or \(\alpha^2\) discretisation errors in order to control the systematic uncertainty of our continuum extrapolation. Finally, to estimate any additional uncertainty arising from the truncation of these expansions, we consider the result of replacing either \(A_X\) or \(B_X\) by Padé expressions. These are obtained by considering that the expansions of \(A_X\) and \(B_X\) in Eqs. 2-3 are the first two terms of a geometric series which we resume. This resummation is not applied to the FV corrections. Instead we try adding a \(1/L^2\) term to either the Taylor or Padé forms. In all cases, we find the coefficient of this term to be consistent with zero.

In total, these different procedures lead to \(2^7 = 128\) different fits for each of the isospin splittings and parameter combinations. Correlating these with the 128 fits used to determine \((\Delta M_{\text{phys}})^2\), and allowing various parameter combinations but discarding fits with irrelevant parameters, we obtain between 64 and 256 results for each observable. The central value of a splitting is then the mean of these results, weighted by the \(p\)-value; the systematic error is the standard deviation. As usual, the procedure is repeated for 2000 bootstrap samples and the statistical error is the standard deviation of the weighted mean over these samples. The unweighted results differ from the weighted ones by fractions of the respective calculated errors.

Isospin breaking effects not included in the sea are NLO and can, therefore, be safely neglected. The quenching errors related to the neglected \(O(\alpha)\) sea-quark contributions are of order \(O(1/N_c)\), when large-\(N_c\) counting is used. Combining the two suppression factors yields an estimate \((M_Z - M_N)/(N_c M_N) \approx 0.09\). Smaller estimate can be obtained\(^9\)\(^,\)\(^16\), however, in the absence of direct quantitative evidence, it is safer to assume that the e.m. contributions to the splittings carry an \(O(10\%)\) QED quenching uncertainty.

3 Results and Conclusion

Our main results for the total light hadron octet isospin splittings and their decomposition into QCD \((\delta m)\) and QED\((\alpha)\) contributions are shown in Tab. 1 and Fig. 4, where we also plot the experimental values. We find good agreement of our results for the total splittings with the ones from experiment.

Through our careful analysis of the different sources of systematic uncertainties, we were able to control all systematic errors with the exception of those due to QED effects on sea quarks. We consider our results to be an encouraging step toward a precise determination of octet baryon splittings, which would constitute an \textit{ab initio} confirmation that the proton cannot decay weakly.
Table 1. Isospin breaking mass differences in MeV for members of the baryon octet. The first error is statistical and the second is systematic. QED quenching uncertainties on the e.m. contributions are estimated to be $O(10\%)$. Propagating the uncertainty in $\Delta_{\text{QED}} M^2_K$ yields an $O(4\%)$ error on the $\delta m$ contributions. The quenching uncertainties on the total splittings can then be obtained by adding those of the e.m. and $\delta m$ contributions in quadrature (not included in the results).

<table>
<thead>
<tr>
<th>$X$</th>
<th>$\Delta M_X$</th>
<th>$\Delta_{\text{QED}} M_X$</th>
<th>$\Delta_{\text{QCD}} M_X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>$-0.68(39)(36)$</td>
<td>$1.39(30)(35)$</td>
<td>$-2.28(25)(7)$</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>$-7.84(87)(72)$</td>
<td>$0.08(12)(34)$</td>
<td>$-7.76(79)(105)$</td>
</tr>
<tr>
<td>$\Xi$</td>
<td>$-7.16(76)(47)$</td>
<td>$-1.29(15)(8)$</td>
<td>$-5.87(76)(43)$</td>
</tr>
</tbody>
</table>

Figure 4. Summary of our results for isospin mass splittings. The total, physical splittings are shown in blue, the QCD ($\delta m$) contributions in red and the QED ($\alpha$) contributions in green. On the points, the error bars are the statistical and total uncertainties (statistical and systematic combine in quadrature). The experimental results are shown as black points.

Acknowledgements

The authors gratefully acknowledge computing time provided by the John von Neumann Institute for Computing, the Gauss Centre for Supercomputing, and PRACE on the Blue Gene/Q (JUQUEEN) located at JSC, Forschungszentrum Jülich. Further HPC resources were provided by GENCI-[IDRIS/CCRT] (grant 52275), as well as on clusters at Wuppertal and CPT CNRS Marseille, France. This work was supported in part by the OCEVU Excellence Laboratory, by CNRS grants GDR $n^0$2921 and PICS $n^0$4707, by EU grants FP7/2007-2013/ERC $n^0$208740, MRTN-CT-2006-035482 (FLAVIAnet) and by DFG grants FO 502/2, SFB-TRR 55.
References

Numerical Simulation of Supersymmetric Yang-Mills Theory

G. Bergner¹, I. Montvay², G. Münster³, U. D. Özgurel³, and D. Sandbrink³

¹ Institut für Theoretische Physik, Universität Frankfurt, Max-von-Laue-Str. 1, D-60438 Frankfurt
E-mail: bergner@th.physik.uni-frankfurt.de

² Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, D-22603 Hamburg
E-mail: montvay@mail.desy.de

³ Institut für Theoretische Physik, Universität Münster, Wilhelm-Klemm-Str. 9, D-48149 Münster
E-mail: {munsteg, oezugurel, dirk.sandbrink}@uni-muenster.de

We report on large-scale numerical simulations of supersymmetric Yang-Mills (SYM) theory by the DESY-Münster Collaboration. The spectrum of light composite particles is investigated and confronted with theoretical expectations based on unbroken supersymmetry for large volumes and small gaugino masses.

1 Introduction

The recently found Higgs scalar particle at the Large Hadron Collider (LHC)¹,² completes the Standard Model (SM) of elementary particle interactions. All known matter is composed of a small number of fundamental constituents, the quarks and leptons. Subatomic interactions of these particles, namely strong, weak and electromagnetic interactions, are described in the Standard Model within the framework of Quantum Field Theory. The scalar field corresponding to the Higgs particle is the source of masses of quarks and leptons and of the vector boson fields mediating the interactions. All known experimental data in the presently available energy range can be described by the Standard Model.

In spite of the completeness and beauty of the Standard Model the question what happens at still higher energies cannot be answered with certainty. The simplest possibility would be that the Standard Model in the present form is valid at all energies till infinity. Otherwise there are physical laws Beyond the Standard Model (BSM). One argument in favour of the existence of BSM physics is the large number of free parameters in the Standard Model. By just a minimal change of some free parameters the entire world would become completely different from the one we know. For instance, the existence and stability of atoms and ordinary matter relies heavily on the fact that neutrons are slightly more massive than protons. The reason for this small difference, besides the different electromagnetic self-interactions, is the small mass difference of up- and down-quarks as constituents. It is an intriguing question, why this small mass difference happens to be just the one realised in our world.

A theoretical hint towards the existence of BSM physics comes from the investigation of the change of effective (“running”) couplings as a function of the energy. It is a long known fact that the three running couplings (electromagnetic, weak and strong) become almost equal at high energies above, say, $10^{16}$ GeV. At this Grand Unification Theory (GUT) scale the basic symmetry group underlying the Standard Model changes
from SU(3)⊗SU(2)⊗U(1) to a larger embedding group like SU(4)⊗SU(4), SU(5)⊗SU(5), or SO(10). The small discrepancy of the three running couplings near the GUT scale disappears if the Standard Model is extended by supersymmetry (SUSY) to a Supersymmetric Standard Model (SSM). Supersymmetry is an extension of the Poincaré symmetry of space-time corresponding to an extension of the Poincaré algebra by one or several supersymmetry charges to a super-Poincaré algebra that relates bosons to fermions (for a review see Ref. 6).

The supercharges change the spin by \( \frac{1}{2} \), hence the supersymmetry multiplets contain particles with different spins, in particular, bosons and fermions at the same time. Since at present energies no such supermultiplets with degenerate masses are observed, supersymmetry – if it is realised in Nature – has to be a broken symmetry.

Another important reason why BSM physics has to exist, is the triviality of the infinite cut-off limit: some renormalised couplings such as the electromagnetic coupling, quartic scalar coupling and (generalised) Yukawa couplings have to vanish, i.e. such interactions cannot be realised in the framework of a relativistic Quantum Field Theory. In case of the electromagnetic coupling this problem is traditionally described by the emergence of a Landau pole at high energies, where the running electromagnetic coupling diverges. As a consequence of triviality in the Standard Model, for a given cut-off there is a region in parameter space of renormalised couplings that is allowed. In the limit of an infinite cut-off this region shrinks to the origin at zero renormalised couplings. Since the renormalised couplings in the Standard Model are known from the experimentally known particle masses, there is a cut-off value at which the couplings are at the border of the allowed region. Higher cut-offs are impossible and therefore BSM physics has to appear not later than at this energy scale. The allowed region of renormalised quartic and Yukawa couplings can be investigated in numerical simulations. (See for instance Ref. 8.) If the Higgs boson mass is about 125–126 GeV, it is in particular the lower Higgs boson mass bound that can imply the values of the energy scale where new BSM physics has to appear.

Supersymmetry may also relieve the constraints arising from triviality. For instance, in \( \mathcal{N} = 2 \) supersymmetric Yang-Mills (SYM) theory every renormalisable quartic and Yukawa coupling is proportional to the non-abelian gauge coupling, which is asymptotically free and can, therefore, describe non-trivial interactions (see for instance Ref. 10,11). In case of the Standard Model, an advantage of GUTs is that the U(1) gauge symmetry corresponding to the electromagnetic interaction becomes part of a non-abelian gauge symmetry and therefore the Landau pole problem disappears. Also, from the point of view of the triviality problem of quartic and Yukawa couplings, a SUSY GUT is better (less restrictive) than a GUT without supersymmetry. More generally, in SUSY theories the radiative corrections are less important and the infinities in renormalisation are fewer.
Every supersymmetric extension of the Standard Model contains as a building block the $\mathcal{N} = 1$ SYM theory, which is the object of investigation in our project. It is the supersymmetric extension of Yang-Mills theory describing the carriers of gauge interactions, the gauge particles, together with their supersymmetric partners, the gauginos. Gauginos are massless Majorana fermions, which are in the adjoint representation of the gauge group. In the continuum the (on-shell) Lagrangian of the theory is

$$\mathcal{L} = \text{tr} \left[ -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{i}{2} \bar{\lambda} \gamma^\mu D_\mu \lambda - \frac{m_g}{2} \bar{\lambda} \lambda \right] ,$$  

where $F_{\mu\nu}$ is the non-Abelian field strength formed out of the gauge fields $A_\mu(x)$, $\lambda(x)$ is the gaugino field, and $D_\mu$ denotes the gauge covariant derivative in the adjoint representation. The supersymmetry of the theory is broken softly by the gaugino mass term. We are presently investigating the simplest non-abelian SYM theory with gauge group SU(2). Our nonperturbative studies are concentrating on the properties of the light particle spectrum. In particular, we determine the masses of the lightest composite particles by performing numerical simulations in the lattice-regularised theory. For our recent publications see Refs. 12–15.

## 2 Supersymmetry on the Lattice

Poincaré (Lorentz) symmetry is broken by lattice regularisation. Since SUSY generators define an extension of the Poincaré algebra, it is not surprising that SUSY is, in general, also broken by the lattice. There are some recently exploited exceptions to this rule, especially in case of extended SUSY with several SUSY generators ($\mathcal{N} > 1$) and in most cases in lower dimensions (for a review see for instance Ref. 16). An interesting example in four dimensions is $\mathcal{N} = 4$ SYM theory, which may be discretised in such a way as to preserve one scalar supersymmetry at nonzero lattice spacing\(^\text{17}\). (The other 15 supersymmetries are still broken by lattice artefacts of $O(a)$, where $a$ denotes the lattice spacing.) In case of $\mathcal{N} = 1$ SYM there is no such possibility, hence SUSY is broken in any known lattice formulation.

Our numerical calculations are based on the Curci-Veneziano lattice action\(^\text{18}\), which is built in analogy to the Wilson action of QCD\(^\text{19}\) for the gauge field ("gluon") and Wilson fermion action for the gaugino ("gluino"). Both supersymmetry and chiral symmetry are broken by lattice artefacts but are expected to be restored in the continuum limit if the gaugino hopping parameter (i.e. bare mass) is tuned to a critical value. The breaking of the chiral symmetry for nonzero lattice spacings could be avoided by using domain-wall\(^\text{20, 21}\) or overlap\(^\text{22}\) fermions and then there is no need of parameter tuning, but the SUSY breaking remains and the required numerical effort for simulations would substantially increase.

### 2.1 SYM Theory on the Lattice

The Curci-Veneziano lattice action for SYM theory is given by

$$S = S_\text{g} + S_\text{f} .$$  

171
Here $S_g$ is the gauge field action

$$S_g = \beta \sum_{pl} \left( 1 - \frac{1}{N_c} \text{Re} \text{Tr} U_{pl} \right),$$

with the gauge coupling $\beta \equiv 2N_c/g^2$ for an SU($N_c$) gauge field. $U_{pl}$ is the product of the gauge link fields along a plaquette. The fermionic part of the action (Eq. 2) is

$$S_f \equiv \frac{1}{2} \lambda Q \lambda \equiv \frac{1}{2} \sum_x \left\{ \bar{\lambda}_x^a \lambda_x^a - K \sum_{\mu=1}^4 \left[ \bar{\lambda}_x^{a+\mu} V_{ab,x\mu} (1 + \gamma_\mu) \lambda_x^b + \bar{\lambda}_x^{a-\mu} V_{ab,x\mu} (1 - \gamma_\mu) \lambda_x^b \right] \right\} \cdot$$

Here $K$ is the hopping parameter which determines the gaugino mass, $\gamma_\mu$ denotes a Dirac matrix and $V_{ab,x\mu}$ is the gauge field variable in the adjoint representation of the gauge group, which is obtained from the gauge field links in the fundamental representation $U_{x\mu}$ by

$$V_{ab,x\mu}^{ab} \equiv 2 \text{Tr} \left( U_{x\mu}^T T_a U_{x\mu} T_b \right)$$

($T^a$ are the generators of SU($N_c$)). The gaugino field $\lambda_x$ satisfies the Majorana condition

$$\bar{\lambda}_x = \lambda_x^T C,$$

with the charge conjugation Dirac matrix $C$.

Performing the path integral over the fermion field $\lambda$ results in a Pfaffian:

$$\int [d\lambda] e^{-\frac{1}{2} \lambda Q \lambda} = \int [d\lambda] e^{-\frac{1}{4} \lambda M \lambda} = \text{Pf}(M),$$

where $M$ is the antisymmetric matrix defined as

$$M \equiv C Q = -M^T.$$

The square of the Pfaffian $\text{Pf}(M)$ is equal to the determinant of the fermion matrix $Q$:

$$\det(Q) = \det(M) = |\text{Pf}(M)|^2.$$

The Monte Carlo simulations are performed by importance sampling with respect to a positive measure. Since for finite lattice spacing $a$ the Pfaffian is not always positive, its sign has to be taken into account separately. Taking the non-negative square root of the determinant, the effective gauge field action is

$$S_{CV} = \beta \sum_{pl} \left( 1 - \frac{1}{N_c} \text{Re} \text{Tr} U_{pl} \right) - \frac{1}{2} \log \det Q[U].$$

The factor $\frac{1}{2}$ in front of $\log \det Q$ can be interpreted as corresponding to a flavour number $N_f = \frac{1}{2}$ of Dirac fermions. The gauge configuration for this fractional flavour number can be created, for instance, by the two-step polynomial Hybrid Monte Carlo (TSPHMC) algorithm, which is our choice for Monte Carlo updating.

The omitted sign of the Pfaffian can be taken into account by reweighting:

$$\langle A \rangle = \frac{\langle A \text{ signPf}(M) \rangle_{CV}}{\langle \text{signPf}(M) \rangle_{CV}},$$

172
where \( \langle \ldots \rangle_{CV} \) denotes expectation values with respect to the effective gauge action \( S_{CV} \).

The reweighting with the Pfaffian sign in Eq. 11 may lead to a sign problem if a strong cancellation occurs among contributions with opposite sign. This could lead to very large statistical errors. However, as we have shown in previous papers by monitoring the sign of the Pfaffian \(^{24, 25}\), for positive gaugino masses practically no sign problem occurs because the positive contributions dominate.

### 3 Light Particle Spectrum in SYM Theory

The gauge coupling in SYM theory is asymptotically free at high energies and becomes very strong in the infrared limit. The low energy particle spectrum is expected to consist of hadron-like colourless states due to confinement, similar to QCD. The difference to QCD is that in SYM the quarks are replaced by Majorana fermions in the adjoint representation. Because of supersymmetry, the particles should belong to mass degenerate SUSY multiplets. The verification of this expectation is a central task for nonperturbative studies in the lattice regularisation.

#### 3.1 Pseudoscalar and Scalar Mesons

The colourless states can be created from the vacuum state by gauge invariant operators which are built from the gluon and gluino field operators. (In this section we shall call, in analogy to QCD, the gauge field “gluon field” and the gaugino as the “gluino”.) A simple example of colourless composite states are the adjoint mesons. (The name “adjoint” refers to the fact that the composing fermions are in the adjoint representation.) The adjoint mesons are composite states of two gluinos with spin-parity \( 0^+ \) and \( 0^- \). We denote the former by \( a^-\eta' \) and the latter by \( a^-f_0 \), where the prefix \( a \) refers to the adjoint representation. For projecting to these states we use the gluino bilinear operators \( O = \bar{\lambda} \Gamma \lambda \) where \( \Gamma = \gamma_5, 1 \) respectively. The resulting meson propagator consists of connected and disconnected contributions:

\[
C_\Gamma(t) = \frac{1}{V_s} \sum_{\vec{x}, \vec{y}} \left\langle \text{disconnected} \begin{aligned}
&\text{Tr}_{sc}[\Gamma Q_{xx}^{-1}] \text{Tr}_{sc}[\Gamma Q_{yy}^{-1}] - 2 \text{Tr}_{sc}[\Gamma Q_{xy}^{-1} \Gamma Q_{yx}^{-1}] \\
&\end{aligned}
\right\rangle
\]

\[
\left( 1 - \frac{1}{V_s} \left\langle \frac{1}{T} \sum_t \sum_{\vec{x}} \text{Tr}_{sc}[\Gamma Q_{xx}^{-1}] \right\rangle^2 \right)
\]

where \( \text{Tr}_{sc} \) denotes a trace over spin and colour indices. The connected term can be used to extract the mass \( m_{a^-\pi} \) of the adjoint pion, which is an unphysical state in SYM (the last term in Eq. 12 is zero for \( \Gamma = \gamma_5 \)). The vanishing pion mass is used to signal the chiral limit.

The numerical evaluation of the disconnected propagators is rather demanding. In order to reduce the large variance, the disconnected part has been calculated using the stochastic estimator method\(^{26}\). As it is the case in QCD, the disconnected diagrams are intrinsically noisier than the connected ones and dominate the level of noise in the total correlator.

An additional difficulty for calculating the mass of the scalar meson (\( \Gamma = 1 \) in Eq. 12) is that in this case, in contrast to the pseudoscalar meson (\( \Gamma = \gamma_5 \)), the vacuum expectation
value appearing in the last term of Eq. 12 is non-zero. This has to be subtracted or has to be fitted as an additional contribution in the correlator. This can be done, for instance, by the method of fitting and error determination described in Sec. 3.2 of Ref. 27, but the effect of the additional fit parameter is an increase of the statistical error.

Scalar particles with the same quantum numbers as $a\cdot\eta'$ and $a\cdot f_0$ can also be created by purely gluonic operators built from the gauge links as in pure gauge theory or QCD. It is possible that the gluon operators create the same states as the above gluino operators. For this case the low energy effective theory was defined in Ref. 28. The other possibility is that there are two low-mass SUSY multiplets created by the two sorts of operators, which can eventually also be mixed with each other\textsuperscript{29}.

### 3.2 Gluino-Glue

For completing a (chiral) SUSY multiplet, besides a scalar and pseudoscalar particle, a Majorana fermion particle is also required. Such particles are provided by the gluino-glueballs.

An operator for the gluino-glue particle is in the continuum

$$\hat{O}_{g\bar{g}} = \sum_{\mu,\nu} \sigma_{\mu,\nu} \text{tr} \left[ F_{\mu,\nu} \lambda \right] ,$$

(13)

where $\sigma_{\mu,\nu} = \frac{1}{2} \left[ \gamma_\mu, \gamma_\nu \right]$ and $F_{\mu,\nu}$ is the field strength tensor. A lattice version of this, which can be used in numerical simulations, is

$$O_{g\bar{g}}^\alpha = \sum_{i<j,\beta} \sigma_{i,j}^{\alpha,\beta} \text{tr} \left[ P_{ij} \lambda^\beta \right] ,$$

(14)

where the indices $i$ and $j$ stand for the spatial directions. A choice for $F_{ij}$ with the proper parity and time reversal transformation properties is the antihermitian part of the clover plaquette $U^{(c)}$

$$P_{ij} = \frac{1}{8g_0} \left( U^{(c)}_{\mu,\nu} - \left( U^{(c)}_{\mu,\nu} \right)^\dagger \right) .$$

(15)

For its definition and more details see Ref. 13.

The gluino-glue correlator has been obtained using different smearing techniques. The link fields are smeared using APE smearing, the fermionic fields using Jacobi smearing. In order to decrease lattice artefacts and statistical fluctuations in the Wilson-Dirac fermion matrix $Q$ of the lattice action (Eqs. 4-5), the gauge link variables $U_{x,\mu}$ have been replaced by stout smeared links\textsuperscript{30}.

### 4 Conclusions and Outlook

The results for light particle masses are summarised in Fig. 1. The masses are shown as a function of the squared mass of the adjoint pion, which for small values is proportional to the gluino mass. Also shown are the extrapolations to the limit of vanishing gluino mass. All figures include the gluino-glue mass for comparison.

The detailed investigation of finite volume effects showed\textsuperscript{13} the lattice volumes are sufficiently large, such that the finite volume effects are smaller than the statistical errors. The
Figure 1. Light particle masses in SYM as a function of the squared mass of the adjoint pion. $r_0$ is the Sommer scale parameter.

Simulations are performed at non-zero gluino masses where the supersymmetry is softly broken. The extrapolation to vanishing gluino mass, where supersymmetry is expected, is consistent with the emergence of a mass-degenerate chiral supermultiplet. Of course, besides the soft breaking by non-vanishing gluino masses, there are also additional SUSY breaking lattice artefacts. This should and could be diminished in future simulations closer to the continuum limit, that is at larger values of $\beta$.

References

Flavour Physics of Up, Down and Strange Quarks from Dynamical QCD $\times$ QED

Gerrit Schierholz

Deutsches Elektronen-Synchrotron DESY, 22603 Hamburg, Germany
E-mail: gerrit.schierholz@desy.de

Lattice simulations of QCD are now reaching a precision, where isospin breaking effects can be investigated. These effects are caused by (i) mass differences between the up, down and strange quarks and (ii) electromagnetic effects due to the different charges of the quarks. So far most lattice QCD simulations are performed neglecting electromagnetic effects. In order to compute physical observables to high precision, it is important to include and control contributions from QED. In previous work we have outlined a program to systematically investigate the pattern of flavour symmetry breaking. The program has been successfully applied to meson and baryon masses involving up, down and strange quarks. In this project we extend the investigations to include matrix elements, charm quarks and electromagnetic effects.

1 Introduction

One of the most profound open questions in particle physics is to understand the pattern of isospin and flavour symmetry breaking and mixing. Lattice simulations are now reaching a precision, where these effects can be investigated. They are due to two causes:

- The mass differences between the up, down and strange quarks
- Electromagnetic effects due to the different charges of the up, down and strange quarks

Bienholz et al.\textsuperscript{1,2} have outlined a program to systematically investigate the pattern of flavour symmetry breaking. The program has been successfully applied to meson and baryon masses involving up, down and strange quarks. In this project we extend the investigations to include matrix elements, charm quarks, the $u$, $d$ mass difference and electromagnetic effects.

A distinctive feature of our simulations is the way we tune the light and strange quark masses. We have our best theoretical understanding when all three quark flavours have the same mass, because we can use the full power of SU(3) flavour symmetry. Starting from the SU(3) symmetric point, our strategy is to keep the singlet quark mass

$$\bar{m} = (m_u + m_d + m_s)/3$$

fixed at its physical value, while

$$\delta m_q \equiv m_q - \bar{m}$$

is varied, with $\delta m_u + \delta m_d + \delta m_s = 0$. As we move from the symmetric point $m_u = m_d = m_s$ (where the pion mass is $\approx 411$ MeV) to the physical point along the path $\bar{m} = \text{const}$, the $s$ quark becomes heavier, while the $u$ and $d$ quarks become lighter. These two effects tend to cancel in any flavour singlet quantity. The cancellation is perfect at the symmetric point, and we have found that it remains good down to the lightest points we have simulated so far. (In contrast, the procedure followed by most other collaborations is to first tune the strange quark mass $m_s$ to its physical value and then vary the up and down quark masses.) This procedure leads to highly constrained extrapolations and reduces the number of free parameters drastically.

The dependence of hadronic matrix elements on up, down and strange quark masses has been worked out group-theoretically to LO, see for example Cooke et al., similarly to the case of hadron masses. That leads again to highly constrained extrapolations for nonsinglet quantities. When confronted with numerical calculations, this provides us with invaluable information on the pattern of flavour symmetry breaking. Flavour symmetry breaking effects in electroweak matrix elements are a key issue in precision tests of the Standard Model. A strong feature of our approach is that along the entire trajectory $\bar{m} = \text{const}$ both kaon and hyperon $V - A$ transition form factors are expected to vary at most quadratically in $\delta m_q$.

We choose the electromagnetic coupling large enough so as to achieve a significant effect on hadron masses and matrix elements. The result will then be interpolated to the physical fine structure constant.

2 QCD

Let us first consider the case of pure QCD and highlight some of the salient features of our approach, along with a selection of results. We have space for hadron masses and the pattern of flavour symmetry breaking only.
With clover fermions the quark masses are defined by the distance from $\kappa_c$, the critical value of the hopping parameter $\kappa$. The bare quark masses then read

$$am_q = \frac{1}{2\kappa_q} - \frac{1}{2\kappa_c},$$

where vanishing of the quark mass along the SU(3) flavour symmetric line determines $\kappa_c$.

![Graph](image-url)

Figure 1. Top: the pseudoscalar meson octet “fan” plot of $M_{\pi O}^2/X_\pi^2$ for $\pi_O = \pi, K, \eta_s$ versus $\delta m_l$. Bottom: the baryon octet “fan” plot of $M_{N O}^2/X_N^2$ for $N_O = N, \Sigma, \Xi, N_s$ versus $\delta m_l$. Both sets of data are normalised to the singlet quantities $X_\pi^2 = (2M_K^2 + M_\pi^2)/3$ and $X_N^2 = (M_N^2 + M_\Sigma^2 + M_\Xi^2)/3$, respectively.
The initial value on this line, $\kappa_0$, is found by looking, for example, where $2m_K^2 + m_\pi^2$ is equal to its physical value.

Using symmetry arguments, we get the mass formula for the outer pseudoscalar mesons

$$M^2(ab) = M_0^2 + \alpha (\delta m_a + \delta m_b)
+ \beta_0 (\delta m_a^2 + \delta m_b^2 + \delta m_c) + \beta_1 (\delta m_a^2 + \delta m_c^2) + \beta_2 (\delta m_a - \delta m_b)^2. \tag{6}$$

For the outer octet baryons we obtain

$$M(aab) = M_0 + A (2\delta m_a + \delta m_b)
+ B_0 (\delta m_a^2 + \delta m_b^2 + \delta m_c^2) + B_1 (2\delta m_a^2 + \delta m_b^2) + B_2 (\delta m_a - \delta m_b)^2 \tag{7}$$

and for the decuplet baryons we find

$$M(abc) = M_0 + A (\delta m_a + \delta m_b + \delta m_c)
+ B_0 (\delta m_a^2 + \delta m_b^2 + \delta m_c^2)
+ B_1 (\delta m_a^2 + \delta m_b^2 + \delta m_c^2)
+ B_2 (\delta m_a^2 + \delta m_b^2 + \delta m_c^2 - \delta m_a \delta m_b - \delta m_a \delta m_c - \delta m_b \delta m_c). \tag{8}$$

In Fig. 1 we show the pseudoscalar octet and nucleon octet masses together with a combined fit, where the up and down quarks have been assumed mass degenerate,

$$m_a = m_d \equiv m_\ell. \tag{9}$$

In this case only one variable is needed to parameterise the symmetry breaking, as $\delta m_s = -2\delta m_\ell$. Typical ‘fan’ plots are seen with results radiating from the common SU(3) symmetric point. We also see an absence of any curvature in the data and the fits, predicting $\beta_0, \beta_1, \beta_2 \approx 0$ as well as $B_0, B_1, B_2, B_3 \approx 0$. This shows that the Gell-Mann–Okubo relations work all the way from the SU(3) symmetric to the physical point. In Fig. 2 we show these results together with other recent results in a plot taken from Kronfeld, to which we refer to for more details.

**Figure 2.** Results for the hadron masses at the physical point by various collaborations. Our points are the orange squares. The figure is taken from Kronfeld.
In investigating the effects of SU(3) breaking due to quark mass differences, the group-theoretical analysis of the mass dependence greatly helped us to organise our results. We now do the same with charge effects. We find in fact that the group theory for the two cases is very similar, and we can often simply read off the form of the electromagnetic effects from our quark mass results.

The symmetry of the electromagnetic current is similar to the symmetry of the quark mass matrix. The simplifications that we get in the mass case by imposing the constraint \( m_u + m_d + m_s = \text{const} \) are similar to the simplifications that come from the identity \( e_u + e_d + e_s = 0 \), which reduces the number of allowed terms in the three-flavour case, when compared with two or four flavours. One difference between quark mass expansions and electromagnetic expansions is that in the mass expansion we can have both odd and even powers of \( \delta m_q \), but in the expansion of hadron masses we are only allowed even powers of the quark charges. We can therefore read off the leading QED polynomials from Tab. 3 of Bietenholz et al. This should be all we need for simulations at the symmetric point. Away from the symmetric point we might want to consider mixed polynomials of the order \( e^2 \delta m_q \), so that we can describe (for example) differences in the electromagnetic mass between the proton (\( uud \)) and the \( \Sigma^+ (uus) \), or between neutron (\( udd \)) and \( \Xi^0 (uss) \).

We can use symmetry arguments, just like those of Sec. 4 of Bietenholz et al., to write down the leading order electromagnetic contributions \( M_{EM} \) to the masses of the outer octet mesons and nucleons and the decuplet baryons. We just drop the linear terms, and keep the quadratic terms, and change masses to charges. For the outer mesons we have

\[
M_{EM}^{2}(ab) = \beta_{EM}^0 (e_u^2 + e_d^2 + e_s^2) + \beta_{EM}^1 (e_a^2 + e_b^2) + \beta_{EM}^2 (e_a - e_b)^2 + (\beta_{EM}^1 + \beta_{EM}^2)(e_a^2 + e_b^2) - 2 \beta_{EM}^2 e_a e_b .
\]

The bottom form of the mass equation can be directly matched up with different classes of Feynman diagrams shown in Fig. 3. The first set of diagrams, with both ends of the photon line attached to the same valence quark (Fig. 3a), contributes to \( \beta_{EM}^0 + \beta_{EM}^2 \). The second set of diagrams, with the photon crossing between the valence lines (Fig. 3b), only contributes to \( \beta_{EM}^2 \). The final set of diagrams, with the photon attached to a sea quark bubble (Fig. 3c), only contributes to \( \beta_{EM}^0 \). This last set of diagrams would be missed out if the electromagnetic field was quenched instead of dynamical.

![Figure 3. Examples of Feynman diagrams contributing to the meson electromagnetic mass.](image-url)
Our first dynamical $\text{QCD} \times \text{QED}$ run was made with $\kappa_u = \kappa_d = \kappa_s$, the same $\kappa$ values we used for the symmetric point in pure QCD, and $e^2 = 1.25$. Our strategy is to simulate at an artificially large coupling, $\alpha_{\text{EM}} \approx 1/10$, and then interpolate between this point and pure QCD to the physical value. The first point to mention is that diagrams like the one shown in Fig. 4 have a big effect on the $\text{QCD} \times \text{QED}$ vacuum. In Fig. 5 we show the effect on the average plaquette.

When QED is added the meson masses become much heavier, especially the $u\bar{u}$. We attribute this to a shift in $\kappa_c$ for the quarks due to their electromagnetic self-interaction. ($\kappa_c$ works rather like an additive renormalisation of the quark masses.) The up quark turns out to be considerably heavier than the two other, which is to be expected, because it has a larger charge. So, to keep the quark masses the same, we will need to simulate at a different set of $\kappa$ values than those used in pure QCD.

![Figure 4. Example of a Feynman diagram contributing to the vacuum.](image)

![Figure 5. The average plaquette for pure QCD (top) and QCD × QED (bottom).](image)
However, even with the present large quark masses we can already see some physical effects of QED, beyond this shift in $\kappa_c$. Extrapolated to the physical value of the fine structure constant, $\alpha_{EM} = 1/137$, our present data give

\begin{align}
\beta_0^{EM} &= 0.034, \\
\beta_1^{EM} &= 0.076, \\
\beta_2^{EM} &= 0.031.
\end{align}

$\beta_0^{EM}$ receives contributions from quark-line disconnected diagrams, like that shown in Fig. 3c. It turns out that this contribution alone accounts for $\approx 2\%$ of the mass of the pseudoscalar mesons. From PCAC and/or the leading flavour expansion we expect that $M_{ab}^2 = (M_{aa}^2 + M_{bb}^2)/2$. Violations of this relation are 27-plet and cannot be present at leading order in the quark mass. Using Eq. 10, we see that the $\beta_0^{EM}$ and $\beta_1^{EM}$ terms cancel and the only term which contributes is $\beta_2^{EM}$, $M_{ud}^2 - (M_{uu}^2 + M_{dd}^2)/2 = \beta_2^{EM}$.

The sign of $\beta_2^{EM}$ is sensible. Opposite charges attract, like charges repel. So we would expect electromagnetic effects to raise the energy of the $u\bar{d} (\pi^0)$ meson (with a repulsive electromagnetic force between the valence quarks) relative to the $u\bar{u}$ and $d\bar{d}$ mesons (with attractive electromagnetic force between the valence quarks), and that is exactly what we find.

There is another difference between the up quark and the other two. The $Z_m$ renormalisation factor will now depend on both the QCD coupling and the QED coupling, and the up quark will have a different $Z_m$, and a different anomalous dimension $\gamma_m$, from the other two quarks. This means that the ratio $m_u/m_d$ now depends on renormalisation scheme and scale (even in the continuum). Likewise, isospin violating mass splittings, for example $M_u - M_d$, are scheme independent, but the question of how much of the splitting is due to the quark mass difference $m_d - m_u$, and how much is due to electromagnetic effects, becomes dependent on scheme and scale. This effect might be minor with $\alpha_{EM} \approx 1/137$, but might be more relevant with $\alpha_{EM} \approx 1/10$.

In pure QCD we can impose perfect SU(3) symmetry simply by making all three $\kappa$ values equal. With QED present, there is no way to have perfect SU(3) symmetry, and so no completely unique way to define a line, where all three quark masses are equal. In particular, we cannot tune the $\kappa$ values to make all members of an SU(3) multiplet degenerate. However, a physically reasonable definition is to look for the line, where the following neutral pseudoscalar meson masses are equal: $s\bar{d}$, $d\bar{s}$ (real particles) and $u\bar{u}$, $s\bar{s}$, $u\bar{d}$ (partially quenched mesons, with annihilation diagrams switched off, so that they do not mix with each other). This line will have $\kappa_s = \kappa_d \neq \kappa_u$.

This symmetric line will end at a point, where all the neutral pseudoscalar mesons are massless. We define this to be the chiral point, the point where all our quark masses are zero. In the case of the down and strange quark masses it is clear that this is the correct definition. Even with QED present, there is a chiral SU(2) symmetry connecting strange and down quarks. So, if both quarks are massless, there will be a massless Goldstone boson from the spontaneous symmetry breaking. Although the neutral mesons will be massless at the chiral point, the charged mesons can have a mass from electromagnetic effects, even when all the quark masses are zero. The charged axial currents are no longer conserved after QED is added to the action, so there is not a Goldstone theorem for the charged pseudoscalars.
Among the quantities we are currently looking at are the splittings of nucleon and kaon masses, as well as the kaon decay constant and the form factors of the semileptonic $K_{\ell 3}$ decay. Typically, the strong isospin violation and electromagnetic corrections are of the same order of magnitude. Of particular interest is the muon anomalous magnetic moment. It is one of the most precisely measured quantities in particle physics. Recent high precision measurements at Brookhaven Lab reveal a deviation of $\approx 3\sigma$ from the Standard Model, which could be a hint for new physics.

4 Outlook

Flavour symmetry and isospin breaking effects in hadron masses and matrix elements are among the most fundamental phenomena in particle physics. Within the Standard Model, they are described by essentially five parameters, the masses of up, down and strange quarks and the strength of the strong and electromagnetic interactions. It appears that these parameters need to be finely tuned to allow life. For example, a slight increase of the ratio $m_d/m_u$, and/or decrease of $\alpha_{EM}$, would make the deuteron unstable and render nuclear fusion impossible. It is conceivable that some of the low-energy parameters of the Standard Model are uniquely determined by an underlying dynamical principle, similar to the prediction of the top quark mass\(^9\). Ultimately, that might be driven by an infrared fixed point of QCD $\times$ QED. (Indeed, a very recent calculation of the SU(3) beta function\(^{10}\) suggests that QCD with two flavours of massless quarks has an infrared fixed point at $\alpha_{QCD} \approx 0.5$.) To shed light on this problem, and perhaps resolve the difference in mass between the up and down quark eventually, a first principles lattice calculation of QCD $\times$ QED is needed. Because hadrons are formed from bound states of quarks, there is no systematic way to treat electromagnetic effects in weak coupling perturbation theory.

Unfortunately, space limitations did not allow me to go into greater detail of this project and give full account of the present status of the calculations. But I am sure we will hear more about it at this Symposium.

References

10. R. Horsley et al., DESY 13-127 (September 2013), to be published.
More Results on Finite Temperature QCD with Wilson Fermions

Szabolcs Borsányi¹, Stephan Dürr¹,², Zoltán Fodor¹,²,³, Christian Hoelbling¹, Sándor D. Katz³, Stefan Krieg¹,², Dániel Nógrádi³, Kálmán K. Szabó¹, Bálint C. Tóth¹, and Norbert Trombitás³

¹ University of Wuppertal, Department of Physics, Wuppertal D-42097, Germany
E-mail: szabo@uni-wuppertal.de

² Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich D-52425, Germany

³ Eötvös University, Institute for Theoretical Physics, Budapest 1117, Hungary
E-mail: szaboka@general.elte.hu

We investigate 2 + 1 flavour QCD thermodynamics using dynamical Wilson fermions in the fixed scale approach. Our previous study at a pion mass of 545 MeV is extended with two additional pion masses, approximately 440 MeV and 280 MeV. We perform simulations using 3 or 4 lattice spacings at each fixed pion mass and measure the renormalised chiral condensate, strange quark number susceptibility and Polyakov loop as a function of the temperature. We observe a decrease in the light chiral pseudo-critical temperature as the pion mass is lowered while the pseudo-critical temperature associated with the strange quark number susceptibility or the Polyakov loop is only mildly sensitive to the pion mass. These findings are in agreement with previous results obtained in the staggered formulation.

1 Introduction

In Refs. 1, 2 we have started a study of 2 + 1 flavour QCD thermodynamics using the Wilson fermion formulation. Even though continuum extrapolated results with physical pion masses are already available within the staggered formulation⁴–⁸ the theoretical uncertainty related to the so-called rooting trick necessitates a comprehensive study with a theoretically sound fermion formulation. One example is the Wilson fermion formulation. In Ref. 2 the pion mass was rather large, around 545 MeV. A careful continuum extrapolation was performed at this fixed pion mass and a comparison was made with similarly continuum extrapolated staggered results and nice agreement was found between the two approaches. In the current work we lower the pion mass and add 440 MeV and 280 MeV to our data set.

The motivation for lowering the pion mass is clearly that we would like to approach the physical pion mass point. But intermediate pion masses, between the physical point and the rather heavy 545 MeV are also important on their own. This is mainly because it was observed in simulations with staggered fermions that the pseudo-critical temperature associated with the light chiral condensate is decreasing as the pion mass is lowered. In order to confirm this phenomenon we have simulated at the intermediate pion masses of 440 MeV and 280 MeV.

The results for these intermediate masses confirm the picture that emerged from the staggered simulations and indeed the light chiral pseudo-critical temperature is decreasing as the pion mass is lowered. On the other hand the pseudo-critical temperature associated
with the confinement-deconfinement transition of the strange quark (given by the strange quark number susceptibility) and the Polyakov loop is only mildly sensitive to the pion masses, again in accordance with the staggered results.

Apart from being theoretically sound another attractive feature of the Wilson formulation using the fixed scale approach is that it complements the staggered formulation using the fixed \( N_t \) approach in terms of cut-off effects. More precisely, as a function of temperature the fixed \( N_t \) approach has small cut-off effects for high temperature (because of a small bare coupling) and has larger cut-off effects at low temperature (because of a larger bare coupling) while in the fixed scale approach the situation is exactly the opposite: low temperatures correspond to large \( N_t \) and hence lead to small cut-off effects while high temperatures correspond to low \( N_t \) and hence larger cut-off effects. For further studies of finite temperature QCD using the Wilson formulation see Refs. 9–15.

The organisation of the paper is as follows. In Sec. 2 we summarise the simulation setup, parameters and algorithms that were used. In Sec. 3 the measured observables are given and their renormalisation properties are discussed and in Sec. 4 we present the results of our investigations.

### 2 Simulation Points and Techniques

The Symanzik tree level improved action\(^{16,17}\) is used in the gauge sector while in the fermionic sector the clover action further improved by six steps of stout smearing is adopted\(^{19}\). The clover coefficient is set to its tree level value \( c_{SW} = 1 \) and the stout smearing parameter is chosen at \( \varrho = 0.11 \). For more details see Ref. 2.

The light quarks \( u \) and \( d \) are assumed to be degenerate and a 2 + 1 flavour algorithm is used. The HMC algorithm\(^{20}\) is adopted for the light quarks and the RHMC algorithm\(^{21}\) for the strange quark. Various algorithmic improvements are applied for speeding up the simulation: the Sexton-Weingarten multiple time scale integration\(^{22}\), the Omelyan integration scheme\(^{23}\) and even-odd preconditioning\(^{24}\).

Finite temperature simulations of QCD can be carried out in two main approaches (or a mixture of the two). First, in the fixed-\( N_t \) approach the bare coupling is used to change the temperature at given temporal lattice extent \( N_t \). Then increasing \( N_t \) corresponds to the continuum limit. Second, the fixed scale approach where the temperature is changed by changing \( N_t \) at fixed bare coupling \( \beta \). The continuum limit in this case corresponds to increasing \( \beta \). While the former is better suited for staggered fermions the latter is more convenient for Wilson fermions and we hence use the latter.

The calculations were performed at the same four gauge couplings as in Ref. 2 \( \beta = 3.30, 3.57, 3.70 \) and 3.85. The scale was set by \( m_\Omega = 1672 \) MeV. The temperature at each fixed bare coupling \( \beta \) is varied in discrete steps by varying \( N_t \).

In our past work\(^2\) the pion mass was relatively heavy, around 545 MeV. We will call the simulations at this mass the “heavy” pion mass point.

Two sets of simulations were performed in the current work each corresponding to a fixed \( m_\pi/m_\Omega \) and \( m_K/m_\Omega \) mass ratio. In the first set, which we call “medium” pion mass, the quark masses were tuned to \( m_\pi/m_\Omega = 0.264(3) \) and \( m_K/m_\Omega = 0.341(2) \). These correspond to about \( m_\pi = 440 \) MeV and \( m_K = 570 \) MeV. At this pion mass the simulations were performed at all 4 lattice spacings. Finite volume effects are expected to be small since \( m_\pi L > 7 \) at each lattice spacing.
Table 1. Bare parameters for the “medium” pion mass (top) and “light” pion mass (bottom) simulations. The $N_t$ values used for the finite temperature runs and the values used for the zero temperature runs are separated by a comma.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$am_{ud}$</th>
<th>$am_s$</th>
<th>$N_s$</th>
<th>$N_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.30</td>
<td>-0.1122</td>
<td>-0.0710</td>
<td>32</td>
<td>6 - 16, 32</td>
</tr>
<tr>
<td>3.57</td>
<td>-0.0347</td>
<td>-0.0115</td>
<td>48</td>
<td>6 - 16, 64</td>
</tr>
<tr>
<td>3.70</td>
<td>-0.0181</td>
<td>0.0</td>
<td>48</td>
<td>8 - 24, 48</td>
</tr>
<tr>
<td>3.85</td>
<td>-0.0100</td>
<td>0.0050</td>
<td>64</td>
<td>8 - 36, 64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$am_{ud}$</th>
<th>$am_s$</th>
<th>$N_s$</th>
<th>$N_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.30</td>
<td>-0.1245</td>
<td>-0.0710</td>
<td>32</td>
<td>6 - 16, 32</td>
</tr>
<tr>
<td>3.57</td>
<td>-0.0443</td>
<td>-0.0115</td>
<td>48</td>
<td>8 - 24, 64</td>
</tr>
<tr>
<td>3.70</td>
<td>-0.0258</td>
<td>0.0</td>
<td>64</td>
<td>8 - 24, 96</td>
</tr>
</tbody>
</table>

Figure 1. The various pion and kaon masses used in our past and current work. The heaviest pion mass is from our past work\(^2\), the 4 red data points correspond to 4 lattice spacings. For the medium pion mass (this work) also 4 lattice spacings are used, while for the lightest pion mass (this work) we have simulated at 3 lattice spacings. The physical point is also shown for comparison. The scale is set by $m_\Omega = 1672$ MeV.

In the second set, which we call “light” pion mass, the meson masses were tuned to $m_\pi/m_\Omega = 0.171(1)$ and $m_K/m_\Omega = 0.315(3)$, corresponding to about $m_\pi = 280$ MeV and $m_K = 520$ MeV. At these pion masses the simulations were performed at 3 lattice spacings and for the finite volume of the system $m_\pi L > 5.4$ holds.

At each lattice spacing, i.e. fixed $\beta$, the mass of the strange quark $m_s$ is fixed at its physical value across all three pion masses “heavy”, “medium” and “light” and the physical point is approached by changing $m_{ud}$ only.

A summary of the various pion and kaon masses used in our past and current work is shown on Fig. 1. The bare quark masses, spatial and temporal lattice extents are shown in Tab. 1 while the measured meson, baryon and PCAC masses are shown in Tab. 2. As can be seen $m_\Omega$ and hence the lattice spacing depends rather mildly on the light quark masses.
At each finite temperature point around 1000-1500 equilibrated trajectories were generated while around 1000 at zero temperature points.

3 Renormalisation

The temperature dependence of three quantities is determined in the current work, the renormalised light chiral condensate, the strange quark number susceptibility and the renormalised Polyakov loop.

3.1 Chiral Condensate

The bare light chiral condensate requires both additive and multiplicative renormalisation. The details of the full renormalisation procedure is given in Ref. 2 following Refs. 25, 26 and will be summarised below.

Additive renormalisation at $T > 0$ is implemented by the subtraction of $T = 0$ quantities as this difference is free from polynomial divergences. Multiplicative renormalisation is then achieved by the multiplication of the PCAC mass $m_{PCAC}$ and the finite renormalisation constant $Z_A$. The latter were determined in the chiral limit from 3-flavour simulations in Ref. 2 and can be taken from there directly for each $\beta$. Finally the Ward identity establishes a relationship between the chiral condensate and the integrated pion correlator leading to the final expression for the fully renormalised condensate at finite temperature,

$$m_R \langle \bar{\psi} \psi \rangle_R(T) = 2 N_f m_{PCAC}^2 Z_A^2 \Delta_{PP}(T) ,$$

where,

$$\Delta_{PP}(T) = \int d^4 x \langle P_0(x) P_0(0) \rangle(T) - \int d^4 x \langle P_0(x) P_0(0) \rangle(T = 0) ,$$

where $P_0(x)$ is the bare pseudo-scalar condensate; for more details see Ref. 2. The final result in Ref. 2 was shown for $m_R \langle \bar{\psi} \psi \rangle_R(T)/m_\pi^4$ since this combination is dimensionless. However when comparing different pion masses as in the current work this normalisation is not convenient because it introduces an artificial pion mass dependence through the $4^{th}$ power. It turns out that the normalisation $m_R \langle \bar{\psi} \psi \rangle_R(T)/m_\pi^2/m_\Omega^2$ is more suitable as can
be inferred from the GMOR relation as well. All results related to the chiral condensate will be presented with the latter normalisation and also the final result in Ref. 2 will be converted into it for comparison.

3.2 Strange Quark Number Susceptibility

The strange quark number susceptibility \( \chi_s \) can be made dimensionless by considering \( \chi_s/T^2 \) and can be improved at tree level by the division of its infinite volume and massless Stefan-Boltzmann limit at each finite \( N_t \). The Stefan-Boltzmann values for each \( N_t \) were listed in Ref. 2. The strange quark number susceptibility is sensitive to the confinement-deconfinement temperature of the strange quark and as we will see is only mildly dependent on the pion mass.

3.3 Polyakov Loop

Our renormalisation procedure for the Polyakov loop also follows Ref. 2. The additive divergence of the free energy can be removed by the following renormalisation prescription: a fixed value \( L_\star \) can be fixed for the renormalised Polyakov loop at a fixed but arbitrary

Figure 2. The pion mass dependence of the renormalised chiral condensate at four different lattice spacings, \( \beta = 3.30 \) (top left), \( \beta = 3.57 \) (top right), \( \beta = 3.70 \) (bottom left), \( \beta = 3.85 \) (bottom right). The data for the “heavy” pion mass (545 MeV) is from Ref. 2. Clearly as the pion mass is decreased the pseudo-critical temperature is also decreasing.
temperature \(T_c > T_s\). This prescription leads to the following renormalised Polyakov loop \(L_R\) in terms of the bare quantity \(L_0\),

\[
L_R(T) = \left( \frac{L_s}{L_0(T_s)} \right)^{T_s} L_0(T).
\]

We choose \(T_s = 0.143\ m\Omega\) and \(L_s = 1.2\) similarly to Ref. 2 while other choices would simply correspond to other renormalisation schemes.

4 Results

At each fixed pion mass the simulations were performed at several, 3 or 4, bare couplings. These in principle allow for a controlled continuum extrapolation similarly to Ref. 2 but this will be left to a future publication. We however do show one example of the continuum extrapolation which is for the chiral condensate, see below.

The renormalised chiral condensate is shown on Fig. 2 for the four lattice spacings corresponding to the four bare couplings \(\beta\). On each plot the three pion masses are shown and clearly as the pion mass decreases the pseudo-critical temperature corresponding to the chiral crossover of QCD is seen to decrease as well. This feature is visible at all four lattice spacings.

![Figure 2. The pion mass dependence of the strange quark number susceptibility at four different lattice spacings. \(\beta = 3.30\) (top left), \(\beta = 3.57\) (top right), \(\beta = 3.70\) (bottom left), \(\beta = 3.85\) (bottom right). The data for the "heavy" pion mass (545 MeV) is from Ref. 2. As can be seen the pseudo-critical temperature is only mildly sensitive to the pion mass.](image-url)
Figure 4. The pion mass dependence of the renormalised Polyakov loop at four different lattice spacings. $\beta = 3.30$ (top left), $\beta = 3.57$ (top right), $\beta = 3.70$ (bottom left), $\beta = 3.85$ (bottom right). The data for the “heavy” pion mass (545 MeV) is from Ref. 2. As can be seen the pseudo-critical temperature is not sensitive to the pion mass at all.

Figure 5. Comparison of the continuum renormalised chiral condensate results for the three pion masses 545 MeV (heavy), 420 MeV (medium) and 280 MeV (light). A downward shift in the pseudo-critical temperature with decreasing pion masses is clearly visible.
spacings we expect the same to hold in the continuum as well. Indeed, as shown on Fig. 5 the continuum results corresponding to the three pion masses also show this behaviour.

The strange quark number susceptibility is shown on Fig. 3 again for all four lattice spacings separately. At each lattice spacing the data for the three pion masses show only mild dependence on the pion mass itself. We have not yet performed a continuum extrapolation of the data but certainly expect that also the continuum result at each pion mass will be only mildly sensitive to the pion mass.

On Fig. 4 our similarly presented data for the Polyakov loop is given. At each of the four lattice spacings the data corresponding to the three pion masses is shown. Just as with the strange quark number susceptibility very little sensitivity to the pion mass is seen. Again, this feature is expected also for the continuum results.

Acknowledgements

Computations were carried out on both GPU clusters at the Universities of Wuppertal and Budapest and also on the Blue Gene/Q computer in Forschungszentrum Jülich. This work was supported by the EU Framework Programme 7 grant (FP7/2007-2013)/ERC No 208740, by the Deutsche Forschungsgemeinschaft grants FO 502/2 and SFB/TRR55 and by the grant OTKA-NF-104034 by OTKA.

References

7. S. Borsanyi et al., Is there still any $T_c$ mystery in lattice QCD? Results with physical masses in the continuum limit III, JHEP, 1009, 073, 2010.
Low-Energy Precision Physics and Lattice QCD

B. B. Brandt\textsuperscript{1}, A. Francis\textsuperscript{2}, V. Gülpers\textsuperscript{2,3}, G. Herdoiza\textsuperscript{3}, G. von Hippel\textsuperscript{1}, H. Horch\textsuperscript{3}, B. Jäger\textsuperscript{2,3}, H. B. Meyer\textsuperscript{2,3}, T. Rae\textsuperscript{3}, and H. Wittig\textsuperscript{2,3}

\textsuperscript{1} Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany
\textsuperscript{2} Helmholtz Institute Mainz, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany
\textsuperscript{3} PRISMA Cluster of Excellence and Institut für Kernphysik, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany

E-mail: wittig@kph.uni-mainz.de

We present results for pion and nucleon form factors, as well as the hadronic vacuum polarisation contribution to the muon\((g - 2)\) obtained from lattice QCD. By using \(O(a)\) improved Wilson quarks for near-physical pion masses and three values of the lattice spacing we achieve good control over systematic errors associated with lattice artefacts and extrapolations to the physical pion mass. Several technical improvements are discussed, including the efficient calculation of quark-disconnected diagrams, the reduction of unwanted excited-state contributions in baryonic correlation functions, and the impact of using partially twisted boundary conditions.

1 Introduction

Lattice QCD has emerged as a versatile tool for tackling a wide range of topics in strong interaction physics. Lattice calculations of the light hadron spectrum have contributed significantly to validating QCD as the theory of the strong interaction. Furthermore, lattice QCD makes precise predictions for Standard Model parameters, such as quark masses and the strong coupling constant, as well as for mesonic decay constants and form factors, which are relevant for obtaining accurate estimates of the elements of the Cabibbo-Kobayashi-Maskawa matrix. These successes have established lattice QCD as a mature field, whose status is further underlined by the fact that an international collaboration is now preparing global averages of lattice results, very much in the spirit of the Particle Data Group\textsuperscript{1}.

Lattice calculations are also increasingly important for the interpretation of experimental results on hadron structure and tests of the Standard Model\textsuperscript{2}. In this contribution we report on our results for form factors of the pion and the nucleon, as well as on the hadronic vacuum polarisation contribution to the muon’s anomalous magnetic moment. As will become clear, there are several technical issues which must be addressed before the overall accuracy of these quantities can be claimed to be comparable or even better than what can be achieved in other phenomenological approaches.

2 Hadron Structure on the Lattice

The internal structure of hadrons has been the subject of a major experimental programme at accelerator facilities worldwide, which must be matched by equally precise theoretical analyses. In particular, one wants to gain a quantitative understanding of structural properties of the nucleon in terms of QCD. To this end, one confronts precise measurements of form factors and structure functions to the corresponding predictions of the theory. Lattice
calculations of baryonic observables are technically much more difficult compared to their mesonic counterparts. Therefore, quantities such as pion form factors offer an ideal testing ground for state-of-the-art lattice methods before they are applied in the baryonic sector.

Chiral Perturbation Theory (χPT) is another theoretical tool for studying the strong interaction at low energies. While lattice QCD seeks to describe hadronic properties in terms of the fundamental constituents, i.e. quarks and gluons, χPT is an effective theory based on hadronic degrees of freedom. Lattice QCD and χPT complement each other: on the one hand, lattice simulations are typically performed at unphysically large light-quark masses and thus χPT is used to extrapolate lattice data to the physical point; on the other hand, lattice simulations allow one to compute matrix elements that can also be calculated in χPT, and thus to determine the low-energy parameters of χPT from first principles.

2.1 The Electromagnetic and Scalar Form Factors of the Pion

The pion – the lightest bound-state in the spectrum of QCD – is best suited to perform a matching between lattice QCD and χPT. The non-perturbative phenomena governing the structure of hadrons is encoded in form factors depending on the momentum transfer $Q^2$. While the electromagnetic form factor of the pion provides information on the distribution of its charged constituents, namely valence and sea light quarks, the scalar form factor of the pion describes the coupling of the pion to the Higgs boson.

A comprehensive account of our study of the pion electromagnetic form factor has recently been published\(^3\). Here, we briefly report on the most salient aspects of this work. Ensembles generated by the CLS initiative with two dynamical flavours of non-perturbatively O($a$)-improved Wilson fermions were used. The computation was performed at three different values of the lattice spacing in the range $0.05 - 0.08$ fm and pion masses between $280$ and $630$ MeV at $m_\pi L \geq 4$. The use of partially twisted boundary conditions\(^4, 5\) has allowed for a determination of the form factor with a very fine resolution of the momentum dependence. Fig. 1 shows a comparison of our results from two ensembles, corresponding to $m_\pi = 325$ MeV and $280$ MeV, to determinations from other lattice collaborations and

![Figure 1](image-url)

**Figure 1.** Compilation of results for the pion electromagnetic form factor in dynamical lattice QCD and as determined from experiment. Our measurements are labelled by the name on the ensembles, F6 and F7, corresponding to $m_\pi = 325$ MeV and $280$ MeV, to determinations from other lattice collaborations and

\[^{3}\text{Nguyen et al.}\]
\[^{4}\text{RBC/UKQCD, m}_\pi = 330\text{ MeV}\]
\[^{5}\text{F6 F7}\]
from experiment. The dense set of data points near vanishing momentum transfer allows for a precise and model-independent determination of the pion’s charge radius \( \langle r^2 \rangle \) from the slope of \( f_{\pi \pi}(Q^2) \) at \( Q^2 = 0 \).

In order to better constrain the mass and \( Q^2 \)-dependence of the pion form factor, it is useful to perform simultaneous fits to the form factor, the pion decay constant and the pion mass, based on the expressions of \( \chi PT \). We then observe that \( \chi PT \) at next-to-leading-order (NLO) fails to produce a consistent description of our lattice data for the entire set of pion observables in the studied mass range \( m_\pi \geq 280 \text{MeV} \). While individual fits to the pion mass and the pion decay constant lead to a coherent picture, inconsistencies arise when data for the form factor are included as well. By contrast, the NNLO expressions allow for a fully consistent description of all three observables, at the current statistical precision. The resulting estimate for the charge radius at the physical pion mass reads

\[
\langle r^2 \rangle = 0.481(33)(13) \text{fm}^2,
\]

the first error is statistical, while the second is an estimate of the total systematic uncertainty. This estimate is in very good agreement with the result \( \langle r^2 \rangle = 0.452(11) \text{fm}^2 \) quoted by the PDG.

The correlation function relevant for the determination of the scalar form factor of the pion receives contributions from quark-disconnected diagrams, which are notoriously difficult to evaluate with good statistical accuracy. Several techniques have been developed to address this challenging computation. In our recent analysis\textsuperscript{7} we have employed stochastic all-to-all propagators in combination with a hopping parameter expansion\textsuperscript{8}, in order to evaluate the disconnected contribution. Our findings indicate that the contribution from disconnected diagrams is far from being negligible, particularly so, near the physical pion mass. This is illustrated in Fig. 2 for the case of the pion scalar radius, which is related to

![Figure 2](image-url)

**Figure 2.** The scalar radius of the pion plotted against the pion mass squared. Dark red points denote the results obtained from the full (i.e. connected and disconnected) contributions, while yellow points represent the connected contribution only. Labels denote the ensembles computed on JUQUEEN. We observe that the contribution from disconnected diagrams is far from being negligible, particularly so, near the physical pion mass.
the derivative of the form factor at $Q^2 = 0$. After performing a chiral extrapolation to the physical pion mass, based on $\chi$PT at NLO, we find

$$\langle r^2 \rangle_s = 0.637(23) \text{ fm}^2,$$

where the error is statistical. We stress that this estimate is consistent with a phenomenological determination based on $\pi\pi$ scattering\cite{9}, but only after including the disconnected contribution.

### 2.2 Nucleon Structure

Lattice QCD has produced an impressive collection of phenomenologically relevant results for masses and decay properties of hadrons. However, the picture for many hadron structure observables of the nucleon – such as the nucleon axial charge $g_A$ or the momentum fraction $\langle x \rangle_{u-d}$ – is somewhat less satisfactory. Moreover, the $Q^2$-dependence of isovector electromagnetic form factors of the nucleon obtained in lattice calculations mostly disagrees with the experimental findings\cite{10,11}. Furthermore, calculations of the nucleon axial charge, $g_A$, tend to underestimate this quantity by typically $10 - 15\%$. There is a broad consensus that systematic effects are non-negligible for these quantities.

Among the common sources of systematic error are lattice artefacts and the influence of finite-volume effects. An obvious question is whether the chiral behaviour is sufficiently controlled in the calculations performed so far, or whether much smaller pion masses are required in order to make contact with the experimental value. One major issue addressed by our group\cite{12,13} is the possible contamination of the ground state of correlation functions by contributions from higher excited states. This is particularly problematic for baryon correlation functions, since their bad signal-to-noise ratio does not allow for long Euclidean time separations between the interpolating operators and local currents and densities. To address this issue, our group has advocated the use of “summed operator insertions”\cite{14}, in which excited state contributions are parametrically more strongly suppressed compared to the conventional ratios of correlation functions.

The construction of interpolating operators which maximise the overlap with the ground state in correlation functions is crucial for any effort to address the issue of excited states contamination. Source smearing is widely used in order to create operators with improved projection properties. The intuition behind this is that a hadron should be best described by a state created by a spatially extended operator rather than a point-like one, guided by the principle that the spatial profile of the extended operator resembles the shape of the hadron in question. We have recently proposed new types of smearing which allow to achieve a reduction in the noise-to-signal ratio in correlation functions at non-zero momentum\cite{15} or to preserve the shape of the smearing function when performing the continuum-limit extrapolation\cite{16}.

Overall, significant progress in addressing the systematic effects present in the quantities related to baryon structure has been achieved over the last few years. In Fig. 3 the pion mass dependence of $g_A$ obtained from summed insertions is compared to the conventional method. The summation method clearly produces estimates in much better agreement with experiment. A chiral extrapolation of our results to the physical pion mass yields\cite{13}

$$g_A = 1.223(63)(^{+35}_{-50})$$

(3)
the first error is statistical. This compares very well with the experimental determination of $g_\Lambda = 1.2701(25)$. Our study has revealed that the agreement between experiment and lattice data for $g_\Lambda$ can be substantially improved when excited state contributions are properly taken into account.

3 Hadronic Contribution to the Anomalous Magnetic Moment of the Muon

The magnetic moment of a charged lepton is extracted from the vertex function describing the interaction between the lepton and a photon in the limit of vanishing photon momentum. The corresponding anomalous magnetic moment $a_l$ is then defined as half the difference between the gyromagnetic factor $g$ and its classical value of 2, i.e. $a_l = (g_l - 2)/2$. In the case of the electron, the quantity is dominated by QED contributions. The anomalous magnetic moment mediates helicity flip transitions, which implies that quantum corrections due to heavier particles of mass $M$, in the Standard Model or beyond, are proportional to $m_l^2/M^2$. For this reason the muon anomalous magnetic moment $a_\mu$ is regarded as a sensitive probe for effects of nearby New Physics. However, by the same argument, given that $m_\mu \leq m_\pi$, the hadronic contributions to $a_\mu$ are larger and notoriously difficult to quantify.

While the experimental and theoretical estimates have both reached similar levels of precision of 0.5 ppm, a tension by 2 or 3 standard deviations between theory and experiment persists. Before invoking “new physics” as the reason for this tension the theoretical result and, in particular, all contributions due to hadronic effects, must be corroborated.
The uncertainty is dominated mainly by the leading order hadronic vacuum polarisation contribution $a_{\mu}^{HVP}$ and secondly by the hadronic light-by-light contributions. Currently $a_{\mu}^{HVP}$ is estimated via a phenomenological approach based on the evaluation of a dispersion integral. In the low-energy regime, the spectral function in the integrand must be determined experimentally, either from the cross section $e^+e^\to \to hadrons$ or from the rate of hadronic $\tau$-decays. Despite the different systematics, both methods produce results in broad agreement, provided that isospin breaking effects are properly accounted for\cite{17}. None of them, however, reduces the discrepancy between theory and experiment on $a_{\mu}$. A purely theoretical estimate of $a_{\mu}^{HVP}$ from a first-principles approach is clearly desirable.

Our group has pursued a research programme\cite{18} aiming at an accurate determination of $a_{\mu}^{HVP}$ using lattice QCD. In the lattice approach, the hadronic vacuum polarisation contribution to $(g-2)\mu$ is determined by a convolution integral

$$a_{\mu}^{HVP} = 4 \alpha^2 \int_0^\infty F(Q^2) \left( \Pi(0) - \Pi(Q^2) \right) dQ^2 , \quad (4)$$

where $\Pi(Q^2)$ is the vacuum polarisation function (VPF) computed on the lattice from the Fourier transform of the current-current correlator. The kernel $F(Q^2)$ is a known analytic function, and $\alpha$ is the fine-structure constant. An important ingredient in our approach is the use of partially twisted boundary condition, in order to enhance the $Q^2$-resolution near the origin. This is crucial for the determination of the additive contribution, $\Pi(0)$, which enters the convolution integral. Fig. 4 shows the VPF measured on an ensemble with our lowest pion mass, $m_{\pi} \approx 190$ MeV, satisfying $L \approx 4$ fm. By comparing the $Q^2$-position and the number of red and blue data points, corresponding to periodic and twisted boundary conditions, respectively, one clearly observes the advantages provided by the lat-

![Figure 4. Momentum dependence of the vacuum polarisation with periodic and twisted boundary conditions shown for an ensemble with $m_{\pi} \approx 190$ MeV, $L \approx 4$ fm and a lattice spacing $a = 0.063$ fm. The different momentum ranges have been separated by different colours.](image)
The subtracted vacuum polarisation $\hat{\Pi}(Q^2) = \Pi(Q^2) - \Pi(0)$ and the Adler function $d\hat{\Pi}(Q^2)/dQ^2$ using the recently proposed time-momentum method (coloured bands). The data shown in black were obtained using the standard momentum-space method.

This figure also indicates the importance of reaching the small $Q^2$ regime and, in particular, the left-most yellow region corresponding to $Q^2 \leq m_{\mu}^2$, where the integrand in Eq. 4 is peaked. The right panel of Fig. 4 shows the relative contribution from data in different $Q^2$ intervals.

In order to obtain an accurate lattice estimate of $a_{\mu}^{\text{HVP}}$, further attempts are necessary to achieve good control of the low-$Q^2$ region. We have recently tested an alternative method based on the time-momentum representation of the vector-vector correlator, in which $Q^2$ is a tunable parameter. Preliminary results are shown in Fig. 5, where they are compared to the standard method on the same ensembles. We expect that by combining the two methods by which we have computed the VPF, a significant improvement in the determination of $a_{\mu}^{\text{HVP}}$ can eventually be achieved.

4 Conclusion

Lattice QCD calculations are increasingly important for the interpretation of experimental results on hadron structure and for constraining the limits of the Standard Model. In this contribution we have presented a status report of our ongoing projects aimed at determining hadronic form factors and the leading-order hadronic vacuum polarisation contribution to the muon $(g-2)$ with high precision. Currently, efforts are still focused on controlling various sources of systematic error, such as excited state contamination, contributions from disconnected diagrams, or the uncertainty associated with the small-momentum region. It is expected that the technical improvements discussed above will soon allow for determinations of these quantities with the accuracy required for having a big impact on phenomenological studies.
Acknowledgements

Our simulations were performed on JUQUEEN and JUROPA as part of NIC project HMZ21, and on the dedicated QCD cluster “Wilson” at the Institute for Nuclear Physics, University of Mainz. We thank Dalibor Djukanovic and Christian Seiwerth for technical support.

References

Materials Science
Materials science is a field that addresses problems of potential technological relevance. The families of “materials” involved can be vast in number, as is the range of properties, and yet there are features that are common to all. Francis Crick noted in his autobiography: “If you want to study function, study structure,” and changes in the atomic arrangement in families of compounds (their structure) have certainly helped in understanding the changes in individual properties (their function). Structures can, in principle, be determined from the energy surface $E(R_I)$, where $E$ is the total energy of the system in question and $R_I$ represents the coordinates of all atoms. The energy minima correspond to the stable (and metastable) structures, and the barriers between them determine the paths and heats of reactions.

The dramatic improvement in computing power in the last 20 years has made possible materials simulations that were simply unimaginable when our first supercomputers became available. This is particularly evident in the case of density functional (DF) calculations. Twenty years ago, we were proud to be able to calculate the energy surfaces of different states of the ozone molecule $O_3^2$. Today it is possible to perform (in much less time) DF simulations of hundreds of atoms over a nanosecond time scale. Apart from an approximation for one term in the energy expression, the method contains neither adjustable parameters nor experimental information and can be used for almost all elements in the periodic table. This is a crucial advantage in studies of families of related compounds, including the effect of exchanging one element with another of the same group. The availability of leadership computers, such as JUQUEEN, has expanded greatly the range of materials and the length and time scales accessible to such calculations.

The role of DF calculations in materials science is underscored by the fact that the four projects presented here all use DF calculations to attack interesting (and technologically relevant) problems. It is also notable that these projects use program packages (Quantum Espresso, VASP, GPAW) that have been developed during the last 20 years by teams of users. While these packages cannot (and should not) be viewed as “black boxes” that simply need feeding with appropriate input, the development of high-quality programs and numerical methods has reduced significantly the time needed to start projects using the DF approach. It is perhaps unnecessary to note that many applications still lie outside the present simulation and time scales for DF calculations mentioned above, and classical simulations with parametrized force fields will be needed for many years to come.

The discovery of graphene, single layers of hexagonally bonded carbon atoms, led to the award of the 2010 Nobel Prize for Physics to Andre Geim and Konstantin Novoselov and much excitement among materials scientists. In the field of microelectronics, there is
optimism that the growth of graphene on silicon surfaces might open the way to advances that can take advantage of the familiar, low cost Si technology. This is the focus of the work of Dąbrowski and coworkers, who studied the growth of graphene on several substrates [hexagonal BN, mica, Ge(001)]. They show that defects play an essential role in these processes.

Magnetic recording applications are possible outcomes of the self-assembled arrays of FePt and CoPt nanoparticles studied by Gruner and Entel. The nanoparticles contained up to 1415 atoms (12792 valence electrons), and the authors focused on the structural phase transitions at the interface between the high-temperature, high symmetry structure and a low-temperature, low symmetry structure. Large-scale calculations of the evolution of structural phases of Si electrodes in Li-ion batteries are discussed by Rohrer and Albe. Although Si possesses a large Li storage capacity, it cannot yet be implemented commercially for this purpose, and the phases studied here provide insight into the degradation process. Finally, Brügner and Walter study materials (metal-phthalocyanines) that are promising candidates for organic solar cells. Crucial to this study is the use of time-dependent DF theory to study the excitation process caused by incident photons. This aspect of density functional calculations has been applied widely to “chemical” problems in recent years.

These articles provide just a sample of the wide range of materials properties being studied using the facilities of NIC. Much of this work seeks answers to problems originating in the world of technology, and coming generations of supercomputers and the development of new algorithms and programming techniques will play crucial roles in finding the solutions.

References

**Ab initio** Modelling of Growth of Graphene for Silicon-Compatible Microelectronics

Jarek Dąbrowski¹, Andrzej Fleszar², Gunther Lippert¹, and Grzegorz Łupina¹

¹ IHP, Im Technologiepark 25, 15236 Frankfurt(Oder), Germany
E-mail: {dabrowski, lippert, lupina}@ihp-microelectronics.com

² Universität Würzburg, Institut für Theoretische Physik und Astrophysik, Am Hubland, 97074 Würzburg
E-mail: fleszar@physik.uni-wuerzburg.de

Graphene, a two-dimensional crystal consisting of at most a few atomic layers of sp²-bonded carbon, is supposed to be the material that can open the way to microelectronics working at terahertz speeds and produced by the traditional, low-cost Si technology. In practice, this hope is likely to be realised only if a technologically acceptable method to grow graphene on Si wafers is developed. This project uses ab initio density functional theory simulations of the interaction between C atoms, small graphene crystallites, and various substrates to build up understanding of the physical mechanisms that are critical for such a growth process. The calculations are combined with experiments: deposition from atomic source in a molecular beam epitaxy chamber and characterisation of the films. Here we describe the computed similarities and differences between the growth process on various substrates, from hexagonal BN inert to C, through the SiO₂-like surface of mica that can be reduced by C, to the Ge(001) surface that strongly adsorbs C. In particular, our results mean that growth of graphene on wide band-gap surfaces requires nearly defect-free substrates (which may be difficult to realise for oxides) and that the growth of graphene on Si-compatible Ge(001) surfaces is governed by the factors that can be to certain extent controlled in the deposition environment: by Ge atoms ejected from the substrate and by bonds formed between the substrate and the edge of graphene crystallites.

1 Introduction

Increasing the speed of the circuitry drives the semiconductor industry. But for the well-established, low-cost silicon technology there is no obvious way to cross the terahertz speed boundary. A solution would be to use add-on materials, such as graphene¹.

Graphene may be used to build the channel of a field effect transistor (FET)² or the base electrode of a heterojunction transistors (HBT)³. In a FET, the load current flows in graphene (i.e., parallel to the graphene sheet) and the key property of graphene is the exceptionally high mobility of carriers that guarantees very short transition times of carriers through the graphene channel separating the source and drain electrodes. In a HBT, the load current flows across the graphene base (i.e., in the direction perpendicular to the graphene sheet) and the key property is the exceptionally small film thickness that suffices for the sheet resistivity of the base electrode not to slow down the control of the carrier flow between the emitter and collector electrodes. In both cases, the idea is that modules containing graphene transistors are designed in such a way that they can be integrated in the same silicon chip that contains other, more traditional elements comprising most of the complete device and accounting for most of its full production costs.

Several methods to fabricate good graphene are well established. In particular, graphene can be grown on a metal foil and subsequently transferred onto the target Si
wafer\textsuperscript{4}, or it can be formed on silicon carbide (SiC) islands previously grown for this purpose on the silicon wafer\textsuperscript{5}. However, the transfer step is difficult to integrate with the chip production process, as it damages graphene at numerous places, while the use of SiC/Si requires non-standard orientation of Si wafers, so that a module with SiC/Si devices cannot be included as a part of a chip manufactured by the usual technology. Graphene can also be grown on Si or on SiO\textsubscript{2} by plasma-enhanced chemical vapour deposition (CVD)\textsuperscript{6}. Yet these films are too thick for FET-like applications, while the interaction of C with Si results in the formation of carbide, excluding the usage of such graphene in HBT-like devices.

In general, it seems that the difficulties to grow graphene in a technology-friendly environment are related to the way C atoms interact with the substrate. Production of graphene is thus one of the subjects where a combination of fundamental studies by theoretical and experimental methods can give support to technology development. The first results provided by us in this context were of such fundamental character: we demonstrated that few-layer graphene may be grown from atomic beam by molecular beam epitaxy (MBE) and we interpreted this process on the basis of \textit{ab initio} calculations\textsuperscript{7}. We also formulated and theoretically evaluated the concept of a Graphene Base Transistor (GBT)\textsuperscript{3}, a device akin in many aspects to an HBT. Our current experimental and theoretical studies show that graphene can be grown by MBE on Ge(001)\textsuperscript{18}, which is a Si-compatible substrate. Very recently, this has been complemented by the proof that good graphene can be produced on Ge by CVD\textsuperscript{10}. Our future work will thus be focused on guidance for and on interpretation of experiments aimed to integrate graphene grown on Ge(001) into a production process of Si-compatible modules containing GBTs. In this report we analyse the interaction of carbon with insulating surfaces and with Ge substrates. The former brings us to the conclusion that Si-compatible growth of graphene on oxide substrates is difficult to achieve, and the latter provides the theoretical substance for the search of the ways to solve the major problems related with the production of graphene for a GBT module: lowering of the growth temperature and improvement of electrical contacts.

\section{Computational Setup}

Electronic structures, total energies, forces, and atomic positions were calculated in the framework of the Density Functional theory (DFT) as implemented in the \textit{Quantum Espresso} package designed for massively parallel computations\textsuperscript{11}. The exchange and correlation energy of electrons was approximated by the gradient corrected functional of Perdew, Burke and Ernzerhof\textsuperscript{12}. Spin polarisation was accounted for. The atoms were described by the pseudopotential method using custom ultrasoft potentials and a plane wave basis set with the energy cut-off of 30 Ry for the wave functions and 240 eV for the electron density. The calculations for Ge(001) substrates (rectangular surface cell) were done with two special \(k\)-points from the Brillouin zone (or with a sampling derived from this \(k\)-point set), while the calculations for substrates with hexagonal surface cells (hexagonal boron nitride, model mica, graphene) were done with the \(\Gamma\) sampling (or with a \(\Gamma\)-derived sampling).\textsuperscript{a} Structures were assumed to be converged when the maximum force dropped down below 0.03 eV/Å.

\textsuperscript{a}Derived means here "using the same set but taken from Brillouin zone of a larger surface cell". This is done with caution, as it costs computing time. For example, calculation using the \(\Gamma\) point from the 24×24 cell of graphene (this is a fully converged \(k\)-point set) requires 10 times more wall time (or 10 times more cores if the same wall
below 0.01 eV/Å and the expected change of the total energy by further relaxation was insignificant for the final conclusion. Energy barriers on diffusion and reaction paths and the corresponding atomic configurations have been obtained using the Nudged Elastic Band algorithm\textsuperscript{15,16}, which searches for the saddle point on the energy surface. Where applicable, the energies and barriers are compatible with those obtained in previous DFT studies\textsuperscript{13,14}.

These tasks are well suited for a supercomputer. Many structures are optimised simultaneously, which requires many CPU units but little data exchange. Under such conditions, CPU time scales nearly linearly with the number of cores and parallelisation is economical.

3 Interaction between Carbon and Hexagonal Boron Nitride

Hexagonal boron nitride (h-BN) consists of sheets built of $sp^2$-bonded B and N atoms. Its lattice is akin to that of graphene and has lattice constant close to that of graphene (the mismatch is 1.7\%, cf. Fig. 1), but in contrast to graphene this material has a wide, indirect band gap (4.5-5.8 eV). It is an attractive substrate for graphene growth, because graphene exfoliated onto h-BN substrates retains its high mobility\textsuperscript{17} and because the chemical interaction between graphene and h-BN appears to be weak enough to (a) prevent substantial damage of the substrate by the impinging atoms or molecules and to (b) eliminate the hazard that platelets are formed that stand vertically on the substrate instead of lying on it.

Atomic C deposited at 800°C-900°C and with amounts corresponding to several graphene layers produced films covering only a fraction of the surface\textsuperscript{8}. This can be due either to very efficient surface diffusion of C atoms on h-BN (C ad-atom finds a graphene island before it finds another C ad-atom), or to small sticking coefficient of carbon to h-BN and much larger sticking coefficient of carbon to graphene surfaces (the growth starts from defects in the substrate and proceeds by expansion of islands nucleated in this way).

Our simulations\textsuperscript{8} support the latter hypothesis. The adsorption energy of C on h-BN and the barrier for desorption of this atom are low (0.87 eV) and only slightly higher than the diffusion barrier (0.68 eV). So if the substrate temperature is 850°C, a typical C atom time is requested) than the standard calculation done with the $\Gamma$ point from the 6×6 cell (this is a poor set for adsorption energies, but a reasonable one for diffusion barriers). Therefore, the structures are computed using the fundamental k-point set and only in the most interesting cases refined using the fully converged set.
remains adsorbed (Fig. 2a) for less than a nanosecond before it desorbs. During its short life time $\tau_{\text{ads}}$, it can make only a few jumps from one adsorption site to another. In order to contribute to graphene, the atom must thus fall (from the beam) within atomic-scale distance from the graphene island. Withal, the life time $\tau_{\text{ads}}$ for atomic C on h-BN is nine orders of magnitude shorter than the average time it must stay on the surface before another C atom drops close to it. This means that if the h-BN surface is free of defects, graphene can nucleate only if the C flux is nine decades higher than that achievable in ultra-high vacuum MBE equipment. Such flux would correspond to nearly atmospheric pressure.

When a C atom falls on a graphene island instead, its adsorption and desorption energies increase to 1.2 eV and its barrier for diffusion decreases to 0.43 eV, resulting in notably longer distances $r_{\text{walk}}$ by which the atom can depart from the impact site before it desorbs: below a nanometre on h-BN, many tens nanometres on graphene. And when the graphene is doped (e.g., by the defects produced during growth), the barrier for desorption increases greatly (to about 2.5 eV independently of the doping type), which translates into huge increase of $r_{\text{walk}}$: into the micrometer or – at lower temperatures – even into the millimetre regime. Hence, graphene may extend from seeds on which the C atoms land.

Some of the C carbon atoms on h-BN get trapped into an interstitial site (Fig. 2b-c). But since this is associated with a kinetic barrier (2.0 eV) much higher than the desorption barrier (0.87 eV), only a small fraction of the adsorbed atoms can be trapped (e.g., one in a hundred thousand at 850°C). As the interstitial site is on the path for doping BN with C, it follows that doping from the atomic beam does not produce on perfect h-BN any significant density of new seeds. Efficient doping at defect sites is however not excluded.
4 Interaction between Carbon and Mica or Silica-Like Surfaces

We will now consider deposition of atomic C on the surface of another layered insulator: mica. The surface of mica is basically a hexagonal SiO$_2$ surface with some admixture of aluminium oxide. Its remarkable feature is that it may develop regions of macroscopic size that are virtually free of defects. We observed}\(^7\) that high quality graphene crystallites can be produced by MBE on mica substrates, but the growth is strongly non-homogeneous: wide areas of the substrate are not affected by the carbon flux, other wide areas are in turn covered by relatively thick graphitic film, and the graphene is found only between these two areas. This disparity is much more pronounced on mica than on h-BN.

As discussed in the previous Chapter, one might try to attribute this result to inefficient sticking of atomic carbon to perfect areas of the mica surface and to much lower density of defects on mica than on h-BN. The apparent difficulty is here however that carbon is well known to react strongly with oxygen: it is not easy for an adsorbed C atom to surpass the stability of a CO molecule, as the latter is isoelectronic to the extremely stable N$_2$.

We investigated numerous hypotheses offering an explanation to the observed character of carbon adsorption on mica. These hypotheses can be divided into three groups:

a there exists an adsorption site easily accessible to the impinging atom and separated from other similar adsorption sites by barriers that are low enough for fast migration of C and much lower than the barriers for CO desorption from the diffusion path;

b albeit a C atom falling onto a perfect mica surface reduces it, producing an O vacancy and CO, later on another C atom is adsorbed at the O vacancy site and can diffuse on the surface in a manner that obeys the conditions specified in the hypothesis a;

c something prevents the C atom from coming into chemical contact with the surface.

Fig. 3 illustrates one of the calculated diffusion paths of interstitial carbon, pertinent to the hypothesis (a). On this path, the carbon atom is protected by the SiO$_2$ skin that forms the uppermost atomic layer of mica.\(^b\) The associated diffusion barrier is close to

\(^b\)The skin contains also admixture of Al oxide, but this does not influence the effects discussed here.
Figure 4. (a) Main adsorption site (interstitial) of C on mica. The C atom is capped by a H atom taken from the hydroxide subsurface layer. The barrier for CO emission from here is 2.6 eV. (b) Secondary adsorption site (substitutional). The C atom substitutes a surface O atom, which then saturates it. The barrier for CO emission is 1.3 eV. (c) Evolution of carbon $2p^2$ orbitals (the states around 0 eV) during a molecular dynamics simulation of C adsorption from molecular beam, run with fixed orbital occupation. The excitation of the $2p^2$ shell begins around $t=0$: the kinetic energy is adsorbed by the split-off $p_z$ orbital up to $t \approx 40$ fs, then the atom turns back.

3 eV. This is higher than the barrier for spontaneous decay of the carbon occupying the main adsorption site (in the middle of Fig. 3, cf. also Fig. 4a) into CO and a surface oxygen vacancy. The C atom may also acquire another adsorption site (Fig. 4b), in which it has nearly the same energy as in the main site but is more mobile. Yet the CO emission barrier from this site drops then to 1.3 eV, which is below the diffusion barrier. No paths consistent with hypothesis (a) were found. Hypothesis (b) failed as well: the barrier for vacancy-assisted diffusion of C is lower than for CO emission from states on the diffusion path.

The remaining hypothesis (c) requires that a mechanism that prevents the adsorption is named. We find that this prevention can be attributed to an excitation of $p$ electrons that lives at least about 60 ps from the moment the C atom begins to feel the presence of the surface (Fig. 4c). As the atom approaches a surface oxygen from vacuum, its degenerate $2p^2$ orbitals split. The (partially occupied) $2p_z$ orbital and of the $2s^2$ orbital are repelled from oxygen orbitals. If the de-excitation of $2p_z$ through electron-hole and/or plasmon channels takes longer than about 60 fs needed for the atom to be reflected, the $p$ orbital adsorbs and then returns the energy and momentum of the impact, making the atom jump back instead of sticking to the surface (this resembles, e.g., the role of the air pressure in a football). Rough estimates suggest that the excitation may live long enough for that; more exact calculations are difficult and expensive, but may be attempted in near future.

The excitation is expected to become weaker and to live shorter at surfaces with a narrower band gap and at defect sites with many surface states and/or with unsaturated orbitals. Within the approximations used here (DFT and fixed or smeared occupation), this is confirmed by the simulations done for the adsorption on h-BN, graphene, and Ge(001).

The conclusion is thus that in spite of high reactivity of C against oxygen, it is expected that graphene may be grown on oxides even from C atomic beam. The necessary condition is, however, that most of the virgin surface is free of dangling bonds and of similar defects. This condition may be difficult to fulfil for dielectrics compatible with the Si technology.
5 Interaction between Carbon and Ge(001) Surface

We have recently observed\textsuperscript{18} that Ge(001) can be uniformly covered with graphene at temperatures between 800°C and the melting temperature of Ge. The films are closed but nanocrystalline. Their resistivity decreases strongly with growth temperature, weakly with C coverage, and reaches the range necessary for terahertz GBT. The activation energy of the surface roughness is low (0.7 eV) and independent on the growth temperature.

The calculations\textsuperscript{18} indicate that the major processes responsible for the growth mode are: substitution of Ge in surface dimers by C atoms, interaction between C clusters and Ge ad-atoms, and formation of bonds between graphene edge atoms and surface Ge atoms. Adsorption of atomic C on top of a surface dimer proceeds without barrier, the energy gain is high (4.6 eV). The adsorbed C atom subsequently kicks out a Ge atom from the dimer. The barrier for this is only 0.65 eV. The resulting Ge ad-atom is highly mobile. Molecular dynamics, analysis of the energies of C occupying various interstitial and substitutional sites under the surface and in the bulk of Ge, and comparison of energy barriers for diffusion and for structural transformation of C residing on and under the surface indicate that the amount of ejected Ge is of the same order of magnitude as the amount of C atoms that hit the surface of Ge. This analysis also shows that the C atoms stay close to the surface and retain sufficient mobility to form graphene. Diffusion of C atom takes place under the surface; any C atom that finds itself on top of the Ge(001) promptly kicks a Ge atom out.

The C atoms on the edge of graphene tend to form bonds with the substrate dimers (Fig. 5a), which can make them stand vertically. Albeit graphene has the lowest energy when bonded on all edges (as also happens to nanoribbons on Si)\textsuperscript{19}, it is difficult for a randomly nucleated piece to find a low-energy path to this state. Since new C atoms arrive to the graphene from under the surface, they are preferentially attached at the edge already bonded to the substrate, so that the graphene/Ge bonding line – if it is long enough – gradually shifts from one dimer row to the neighbouring one, but it does not necessarily round up so that it may fasten the whole edge to the substrate. Many pieces would thus stand up as in Fig. 5a. Such behaviour would hinder the growth of good graphene. This tendency is however reduced by the kicked-out Ge ad-atoms that diffuse on top of the surface. These relatively large atoms stick one by one to the edge preferentially at the corners (Fig. 5b), gradually bringing the graphene molecule to a more horizontal position.
6 Concluding Remarks

From a broad theoretical and experimental study of the interaction of C with various substrates we learned that growth of graphene on wide band-gap surfaces is governed by defects and that the growth on Si-compatible Ge(001) is governed by the Ge atoms ejected from the substrate and by the bonds formed between the substrate and the graphene edge.

To obtain a single diffusion barrier of C on Ge(001), 10 to 20 structures, each with about 200 atoms, must be simultaneously optimised under common constraints. Depending on the quality of the starting guess, such a task may be completed within about half a day to a day if running on about 300 cores. Thus, supercomputing power is indispensable.

Acknowledgements

Financial support from the European Commission through the GRADE project (No. 317839) and computing time support from the Jülich Supercomputing Centre of the John von Neumann Institute for Computing (project hfo06) are gratefully acknowledged.

References

2. Y. Wu et al., Nano Letters 12, 3062 (2012).
7. G. Lippert et al., physica status solidi b 248, 2619 (2011); G. Lippert et al., physica status solidi b 249, 2305 (2012); G. Lippert et al., Carbon 52, 40 (2013).
Nanometre-Scale \textit{ab initio} Investigations of Functional Magnetic Materials

Markus Ernst Gruner and Peter Entel

Faculty of Physics and Center for Nanointegration, University of Duisburg-Essen, D-47048 Duisburg, Germany
E-mail: Markus.Gruner@uni-due.de

The present survey summarises recent advances in the understanding of the functional properties of application-grade magnetic materials achieved with large-scale density functional theory calculations on the IBM Blue Gene supercomputers at Forschungszentrum Jülich. The discussion focuses on two areas of the project: Structural properties and orbital magnetism of hard-magnetic nanoparticles for data recording applications and the evolution of a nanoscale microstructure in technologically relevant magnetic shape memory systems.

1 Introduction

Complementing experimental methodology, the exponential increase in high-performance computing power meanwhile opens another route to characterise, understand and improve nanostructured functional materials. Since most of the calculations presented here, which were until recently only possible on the largest computing systems present in the world, could become generally available soon (possibly within less than a decade), it is not far-fetched to predict that the computational design and characterisation of functional materials will become one essential standard step in materials research and development – on par with experimental techniques like transmission electron microscopy or state-of-the-art x-ray analysis methods. In their combination, theory and experiment can yield a complete picture ranging from the electronic structure of an individual atom to the macroscopic functional properties, which will provide us with what we call understanding and – derived from this – heuristic guidelines for the development of better materials. The present survey demonstrates in two examples, how modern electronic structure theory in the framework of density functional theory (DFT) can take part in this quest and lead to new insights of immediate technological relevance. Since large systems had to be involved, including up to more than one thousand spin-polarised transition metal atoms, the calculations were carried out on the IBM Glue Gene/P at Forschungszentrum Jülich making use of the VASP\textsuperscript{1,2} package; for technical details, see, e. g., Refs. 3–5.

2 FePt Nanoparticles for Recording Applications

Over the years, large expectations were put into self-assembled arrays of FePt and CoPt nanoparticles as a future, commercially relevant ultra-high density medium for magnetic recording applications. This is owed to the very large magnetocrystalline anisotropy of the respective bulk materials, which would allow producing extremely fine grains on the nanoscale still capable of storing information for decades (for an introduction including
further references, see, e.g., Ref. 6). This is not possible with conventional recording materials, since comparatively fine structures of these would become superparamagnetic at ambient conditions, thus losing their information in short time simply due to thermal fluctuations. So far, the hopes to obtain FePt nanoparticles with diameters in the range of 3 to 5 nm with stable ferromagnetism at room temperature have not come true. One central difficulty lies in the realisation of a single crystalline, well ordered L1_0 phase which should by present in the largest fraction of the particle, without disturbing defects as grain or antiphase boundaries. In literature, two reasons are discussed why this aim is difficult to achieve in experiment: First, the presence of surfaces, which implies a lower coordination of the respective atoms, disturbs the thermodynamic ordering process, since the driving force for ordering comes from the interatomic interaction. Since the stable easy axis of magnetisation is bound to the crystal lattice, the presence of twins or grains with different crystallographic orientations within one particle effectively compromises the desired hard magnetic behaviour as well. Twins and grains might occur “by accident”, due to the formation kinetics, or “on purpose”, since they belong to the formation principle of a structure which is energetically preferred. This is the case for the so-called “multiply twinned” particles, which are characterised by a regular arrangement of twin defects, typically with a five-fold symmetry. Large parts of the project work was thus dedicated to the competition of single crystalline particles with morphologies containing internal structural defects like twin or grain boundaries. This investigation was based on massively parallel first-principles calculations of nanoscale transition metal clusters containing up to 1415 atoms or 12792 valence electrons, respectively. The calculations include the full relaxation of the geometric structures which is important to provide a realistic estimate of the relative stability of different paradigmatic morphologies. Such calculations are very demanding with respect to computing power and memory requirements and could only be completed in a reasonable time frame on world-leading computer hardware.

We find for ordered near-stoichiometric FePt particles a crossover between hard magnetic single crystalline bulk-like structures and multiply twinned morphologies taking place at particle diameters in the range of about 4 nm (cf. Fig. 1). This yields already one important explanation for the experimental difficulties to form ferromagnetic particles in this size range exhibiting a sufficiently large magnetocrystalline anisotropy energy for ultra high density data recording purposes. One possible strategy to prevent undesired twinning is alloying of the material with an element which suppresses the formation of twin boundaries. From the electronic structure, one can derive as a heuristic guideline, that a reduced electron count in the minority spin 3d states will act in this way and, indeed, substitution of Fe by Mn turns out as a suitable measure in this respect. However, due to its antiferromagnetic interaction with its neighbours, Mn acts adversely on the magnetic properties in a direct way and can only be used in small amounts.

For building a functional device, the particles need to be assembled on a surface including a protective layer to prevent chemical degradation. Due to the high surface-to-volume ratio such interfaces can affect the magnetic properties of small particles significantly – either directly through hybridisation of the electronic states at the new interface or indirectly through the interface stress. This was paradigmatically investigated for the case of monolayer capping with Al, Cu and Au, explicitly including orbital magnetism being responsible for the magnetocrystalline anisotropy. The hybridisation of Al-sp and Fe-d states leads to a degradation of spin- and orbital magnetic properties, in close correspon-
Figure 1. Size dependent evolution of the energetic order of FePt L1₀ cuboctahedra and multiply twinned icosahedra with different kinds of order up to 1415 atoms. The lines connecting the data points are polynomial fits to a phenomenological power law⁷. The largest isomers corresponding to the respective motifs are depicted on the right (dark brown spheres refer to Pt, bright blue to the 3d metal). The onion-ring icosahedron is shown in cross section. The broken yellow lines indicate the internal interfaces between the 20 twins of the icosahedra. Figure adapted from Ref. 8.

dence to simultaneous synchrotron experiments using X-ray magnetic circular dichroism to determine the spin-to-orbital moment ratio. Protective layers of late transition metals appear to be a much better alternative as the hard magnetic properties of the free particle could be largely preserved. This result depends on the type of interface termination, i.e., a Cu-cap should preferably be brought in contact with an Fe-layer, while for a Au-cap, we prefer a Pt-terminated FePt core. From the inspection of the site resolved magnetic properties, which are uniquely accessible through a large scale DFT approach, it becomes clear that in particular the surface moments of the heavy component makes the largest contribution to the anisotropic behaviour⁵. Thus these observations provide us with concrete strategies to increase the hard magnetic volume per particle and thus the integration density by careful interface design¹², ⁵.

3 The Role of Microstructure in Magnetic Shape Memory Alloys

Magnetic shape memory alloys are another important class of functional magnetic materials with a high application potential. Once more, their functionality is intimately linked to structural features on the nanoscale in terms of a hierarchical microstructure spanning different length scales. It adapts the stress, which forms during a structural phase transformation at the interface between the high-temperature high symmetry crystal structure
Figure 2. Left: The atomic arrangement of a L1\textsubscript{0} ordered Fe\textsubscript{80}Pt\textsubscript{67} nanoparticle encapsulated with a monolayer of Au in cross-section (foreground). Blue spheres refer to Fe atoms, brown to Pt and yellow to Au. Right: Magnetocrystalline anisotropy of 147 atom FePt nanoparticles with different caps and interface terminations. Small circles refer to particles with stripped decorations but retained atomic positions and clean nanoparticles, showing the impact of lattice relaxations in the core due to the encapsulation. Adapted from Ref. 12.

(austenite) and a low-temperature low-symmetry structure (martensite) (for a recent discussion and further references, see Refs. 13–16). In the prototype system, the Heusler alloy Ni\textsubscript{2}MnGa, this specific microstructure is responsible for giant strains of up to about 10% which can be obtained reproducibly in moderate magnetic fields\textsuperscript{17,18}. The mechanism is the reversible growth of one martensitic variant (which exhibits a specific orientation relation to the lattice of the austenite) on the expense of another. This is possible due to an extraordinarily high mobility of the mesoscopic twin boundaries separating these variants.

The full phase sequence of Ni-Mn-Ga Heusler alloys is rather complex. At high temperatures, one finds the cubic L2\textsubscript{1} austenite, which proceeds with decreasing temperature to the tetragonal L1\textsubscript{0} low-temperature phase through a variety of modulated structures (referred to as 6M, 10M and 14M), which finally exhibit the MSM features\textsuperscript{19}. To address the complex nature of the phase sequence and the mobility of the respective twin interfaces, we investigated the stress induced martensitic reorientation process by means of ab initio molecular statics calculations. Starting point were twin interfaces of the smallest building block, the unmodulated low temperature L1\textsubscript{0} martensite. The comparison with quasi-static simulations of shear-induced twin boundary motion and calculations of the magnetocrystalline anisotropy energy, shows energy scales encountered in the coherent reorientation processes of twins and magnetisation are not of the same order of magnitude on this length scale\textsuperscript{20,3}. Recent follow up calculations demonstrated that the perfect L1\textsubscript{0} interface plays a decisive role despite its immobility, since it can be introduced at practically no cost in energy\textsuperscript{14,21}. This fosters the formation of a periodic (5\textsuperscript{2}2)\textsubscript{2} nanotwinned superstructure (described by a double stacking of twins of five L1\textsubscript{0} layers oriented in one direction and two in the other) which is identical to the experimentally observed 14M modulated phase and nearly degenerate to the simple, nonmodulated (NM) L1\textsubscript{0} martensite (left side of Fig. 3). According to the concept of Khachaturyan the formation of such a nanotwinned
Figure 3. Illustration of the close analogy between the ordered Ni$_2$MnGa and the disordered Fe-Pd MSM systems from DFT total energy calculations. The upper graphs show the energy along the transformation path characterised by the tetragonal distortion $c/a$. The energies were obtained for nonmodulated (NM) structures at ground state magnetisation $M_0$ and reduced (finite temperature) values. For Ni$_2$MnGa NM $L1_0$ ground state is found at $c/a = 1.26$, while modulated structures (6M and 10M) are metastable for pseudo-tetragonal supercells with $c/a < 1$. Lower energies are found for the monoclinic ($32\bar{1}$)$_2$ and ($52\bar{1}$)$_2$, aka 14M (depicted below) nanotwinned structures consisting of $L1_0$ building blocks which are marked at the corresponding (approximate) tetragonal distortion of the supercell$^{14, 21}$. Straining Fe$_{68}$Pd$_{32}$ beyond the traditional transition path (Bain path) connecting austenite (fcc) and low temperature martensite (bcc), one encounters a second shallow minimum which corresponds to a nanotwinned configuration consisting of tetragonal (fct) building blocks (see below). The orange arrows finally denote the proposed evolution of structures with decreasing temperature, which proceeds analogously in both systems: Cubic austenite is stabilised by magnetic disorder and vibrational entropy. Soft acoustic phonons in [110] direction which originate from a band-Jahn-Teller type instability of the electronic structure of austenite result in modulated and nanotwinned structures. Their sharp interfaces are aligned along [110] and do not impose a significant cost in energy. The final step to the low temperature martensite requires the removal of twin interfaces in a potentially irreversible coarsening process. Partly adapted from Refs. 14, 21, 22.

(“adaptive”) structure minimises the energy of the austenite-martensite interface$^{23, 24}$. The formation of the nanotwinned 14M can evolve downhill in energy through the formation of the 6M sinusoidal modulations (open circles) in $L2_1$ austenite, which are commonly recognised as premartensite. The origin of these modulations is related to an electronic instability. Ni $d$-states close to the Fermi level form extended nesting sheets on the Fermi surface$^{13}$. These sheets can be connected through a wave vector along [110] in reciprocal space, which exactly corresponds to a soft phonon mode observed in experiment$^{25, 26}$. The redistribution of a respective fraction of the degenerate states at the Fermi level gains sufficient energy to stabilise the corresponding modulated structural arrangement. In turn, magnetoelastic coupling is pivotal for the stabilisation of the high-symmetry phase, the
austenite. The collapse of the magnetic moment at elevated temperature reduces the energy difference between austenite and non-modulated martensite, while vibrational entropy eventually stabilises the cubic phase\textsuperscript{27,28}.

If this picture claims generality, we should be able to confirm it in other MSM systems such as ordered Fe\textsubscript{3}Pt and disordered metastable Fe\textsubscript{70}Pd\textsubscript{30}. Both systems exhibit large magnetic field induced strains albeit not as pronounced as Ni\textsubscript{2}MnGa and the latter also fairly close to ambient conditions. On the other side, there are marked differences with respect to the Heusler MSM alloys. The Fe-based systems do not contain main group elements and instead of a body centred average coordination of the atoms in the Heusler case, the austenite is rather face centred. Magnetic field induced strains are encountered in a slightly tetragonal fct phase occurring between fcc austenite and the low-temperature martensite. Nevertheless, essential analogies can be established: Similar to the Ni\textsubscript{2}MnGa prototype, Fe\textsubscript{3}Pt and Fe\textsubscript{70}Pd\textsubscript{30} possess an essentially flat energy surface without a significant energy barrier separating the austenite state from the martensite (see Fig. 3)\textsuperscript{29,4}. Both exhibit a soft acoustic phonon branch in [110] direction in reciprocal space supporting corresponding to shear motion of [110] lattice planes in real space. The low-symmetry atomic arrangement arising from the condensation of the phonon finally leads to tetragonal lattice parameters as seen in experiment. In both, Fe\textsubscript{3}Pt and Fe\textsubscript{70}Pd\textsubscript{30}, the structural changes can be traced back to a redistribution of an enhanced minority spin density of states at the Fermi level\textsuperscript{29,4} – just as reported for the Ni\textsubscript{2}MnGa Heusler system – while the stabilisation of the austenite is once again related to a magneto elastic mechanism. This becomes evident from Fig. 3, where the energetic order of the structural minima of Fe\textsubscript{70}Pd\textsubscript{30} is inverted through magnetic excitations, modelled here by a fixed number of flipped Fe-spins.

The key stone is, however, the verification of adaptive nanotwinning in the Fe-Pd system from both, large scale DFT calculations with 500 transition metal atoms and experiments on coherently strained epitaxial films\textsuperscript{22}. If the conventional transformation path (“Bain path”) starting at the low temperature martensite (bcc/bct, \(c/a \approx \sqrt{1/2}\)) is extended towards the austenite (fcc, \(c/a = 1\)) towards \(c/a > 1\), the simulation box fills with a pattern consisting of slightly tetragonal fct twins oriented in [110] direction, corresponding to the aforementioned soft phonon. In turn, a diffraction analysis of the correspondingly strained epitaxial film indicates a rotational movement of the atoms at the [110] twin boundaries being entirely consistent with the presence of fct twins with \(c/a < 1\)\textsuperscript{31}. With the knowledge of the energy landscape and the geometric relation between simulation box and twins we can confirm that the formation energy of the interface between fct twin is extremely low – in close analogy to Ni\textsubscript{2}MnGa\textsuperscript{22,14}. The consequence is a nanotwinned microstructure, which is, however, not as regular as the 14M in Ni-Mn-Ga. But it is consistent with the conjecture that in any conventional MSM system the highly mobile martensitic twin boundaries, which can be shifted easily in external magnetic fields, are supplied by higher order twin interfaces between extended nanotwinned regions, as previously proven for Ni-Mn-Ga\textsuperscript{32}. In the heuristic search for novel MSM systems\textsuperscript{33,31,34,35}, this analogy finally provides us with a catalog of properties, which a successful candidate should fulfil.

4 Concluding Remarks

Based on large scale first-principles calculations, we were able to obtain a thorough understanding the discussed materials. Our theoretical approach allowed us to identify the mi-
croscopic (electronic) mechanisms behind the specific functionality and helps us to make
concrete suggestions for improved compositions. This applies to examples as different as
hard-magnetic $3d$-$5d$ nanoparticles for recording applications and transition metal alloys
and compounds for magnetic field induced actuation purposes. One main advance of this
project is thus the successful demonstration of how the predictive power of massively par-
allel first-principles calculations can be employed as an independent predictive method for
the heuristic search for better compositions and their characterisation in the development of
nanoscale magnetic materials with novel functionalities. Thanks to the dynamic evolution
of the high-performance computing sector, modern electronic structure methods are now
establishing as an a eye level companion to state-of-the-art experimental characterisation
techniques. For the present examples, the correspondence between experiment and theory
turned out to be very good. This is owed to the fortuitous fact that for the systems under
consideration, standard DFT with local spin density and generalised gradient approxima-
tion to the exchange-correlation functional provided already an excellent description of the
fundamental structural properties. This is – of course – not always the case and must be
considered appropriately, for instance in terms of higher level description of exchange and
correlation. This will certainly charge its price in terms of increased computing power, but
computing capabilities are still advancing with exponential pace.

Acknowledgements

The author would like to thank in particular C. Antoniak, P. Entel, M. Farle, S. Fähler,
S. Hamann, T. Hickel, S. Kaufmann-Weiβ, A. Ludwig and H. Wende for substantial dis-
cussions. Furthermore, gratitude is expressed to the NIC, the JSC and the CCSS (Duisburg-
Essen) for computing time and support. The efforts of Dr. P. Vezolle of IBM and Dr. C.
Henriet of Cray Inc. in optimising the VASP code are gratefully acknowledged. Financial
support was granted by the DFG through of SPP 1239 and SFB 616.

References

6. J. Lyubina, B. Rellinghaus, O. Gutfleisch, and M. Albrecht, vol. 19 of Handbook of
2008.
12. C. Antoniak, M. E. Gruner, M. Spasova, A. V. Trunova, F. M. Römer, A. Warland,
B. Krumme, K. Fauth, S. Sun, P. Entel, M. Farle, and H. Wende, Nature Commun. 2,
528, 2011.

221
First-Principle Calculations on the Phase Evolution of Si Electrodes in Li-Ion Batteries

Jochen Rohrer and Karsten Albe

Technische Universität Darmstadt, Institut für Materialwissenschaft, Fachgebiet Materialmodellierung, Petersenstr. 32, 64287 Darmstadt, Germany
E-mail: {rohrer, albe}@mm.tu-darmstadt.de

We present large-scale calculations on the phase evolution of Li-Si alloys. The calculations combine density-functional-theory calculations with iterative insertion/extraction of Li and statistical sampling via replication as to mimic electrochemical (de)lithiation of Si anodes. Our calculations predict the presence of various two-phase regions. These two-phase regions are believed to be key to degradation as they are accompanied by large inhomogeneous volume changes that can lead to crack formation and eventually pulverisation of the anode.

1 Introduction

Rechargeable Li-ion batteries become increasingly important as energy sources for portable electronic devices and, in particular, for zero-emission electric vehicles. Key challenges for enhancing Li-ion technology are to increase energy density (gravimetric and volumetric capacity), power density, cycle life and open-circuit voltage without decreasing the safety of the batteries. These issues require development of new electrode materials as well as electrolytes. In the following the focus will be on a specific electrode material, namely Silicon, which has high potential for application as anode in Li ion batteries but which also faces issues that so far have prevented commercialisation.

Silicon possesses a Li storage capacity as high as 3600 mAh/g, corresponding to about ten times the capacity of commercially used graphite. This makes Si one of the most desirable anode materials for Li-ion batteries. However, Si suffers from high irreversible capacities, that is, the amount of Li that can be extracted in a cycle is much lower than the amount of Li that was previously inserted. It is widely believed that this behaviour originates from large volume changes upon (de)lithiation (upon full lithiation, the electrode volume increases by 280%) leading to cracking and pulverisation and thus to a very limited lifetime. On this basis, experimental attempts to functionalise Si have largely focused on capacity-limited cycling. During capacity-limited cycling only a fraction (up to one fourth) of the maximum capacity is exploited. Correspondingly, volume changes are significantly reduced. Although capacity-limited cycling has indeed contributed to extending the lifetime of Si anodes, irreversible capacities are still a major issue preventing the commercial use of Si anodes.

In this paper we study the behaviour of Si during insertion and extraction of Li on the atomic scale. In particular we develop a method that employs large-scale density-functional-theory (DFT) calculations combined with iterative insertion/extraction of Li and statistical sampling via replication. Our calculations allow us to theoretically predict the phase evolution of Li-Si alloys during cycling and to identify two-phase regions that may contribute to degradation of Si electrodes during cycling without capacity limitation.
We also use our data to suggest a model that may contribute to understanding degradation in Si anodes during capacity-limited cycling.

The paper is organised as follows. In Sec. 2 we summarise our computational method and introduce a modelling strategy that mimics electrochemical lithiation and delithiation of a Si electrode. Results are reported in Sec. 3 and their discussion follows in Sec. 4. In Sec. 5, we finally summarise our work and give our conclusions.

2 Method

2.1 Structure Generation

In Fig. 1 we illustrate our iterative scheme for generating amorphous Li$_x$Si model configurations (a-Li$_x$Si), representing an evolving Si electrode-particle during cycling. Here, $x$ denotes the ratio between the number of Li and Si atoms in the cell, $x = n_{\text{Li}}/n_{\text{Si}}$. As starting point we use crystalline Si, represented by a cubic unit cell that contains 64 atoms. The iterative step consists of three substeps. First, for a given Li content $x$, five identical replicas of the Li$_x$Si model are created. In each of these replicas, eight additional Li atoms are inserted such as to produce a homogeneous distribution of Li in the simulation cell. For details on this procedure, we refer to Ref. 8. All five replicas are then optimised using ab initio molecular dynamics simulations and subsequent relaxations. Finally, the lowest-energy configuration is chosen as the new Li$_x$Si model which is further lithiated.

In addition to the lithiation cycle described above, we also perform a delithiation cycle. Again five copies of the system are made. In each copy, Li atoms are now removed and five copies of the system are created and eight Li atoms are inserted in each replica. The resulting systems are optimised by ab initio molecular dynamics simulations at 400 °C and subsequent relaxations of the atomic positions and the simulation cell shape and volume. Among the optimised systems, the lowest-energy configuration is chosen for further insertion of Li. Colour coding: yellow balls represent Si atoms; green balls represent Li atoms; pink balls represent Li atoms that are added to a replica.
randomly and the delithiated copies are equilibrated and optimised. The lithiation and delithiation protocols were implemented in python\textsuperscript{9} using the functionality of the atomic simulation environment\textsuperscript{10}.

### 2.2 Computational Method

All calculations are performed using the projector-augmented wave DFT code VASP\textsuperscript{11,12}. The exchange-correlation energy is approximated by the PBE functional\textsuperscript{13}. For Li all three electrons are taken into account; for Si only the 3s and 3p electrons are treated explicitly.

For \textit{ab initio} molecular dynamics we use a plane-wave cutoff of 280 eV, a timestep of 2 fs and a temperature of 400°C (slightly below the temperature above which crystalline Li-Si are known to form); the overall simulation time is 2-6 ps (depending on the cell size) and the Brillouin zone (BZ) is sampled at the Γ point, only. For relaxations of atomic positions, cell shape and cell volume, we use a plane-wave cutoff of 380 eV and a $2 \times 2 \times 2$ Monkhorst-Pack k-point mesh\textsuperscript{14} for the BZ sampling; relaxations are considered as converged as the forces do not exceed 0.03 eV/Å.

### 3 Results

In Fig. 2 we detail the calculated formation energies of Li$_x$Si as a function of the Li content. The case of insertion is shown in the left panel; the right panel corresponds to extraction.

![Figure 2](image-url)  

**Figure 2.** Calculated formation energies of homogeneously lithiated Li$_x$Si structure models (open circles). The formation energy of crystalline Li$_{12}$Si$_4$ is also given for comparison (open square). Solid lines indicate two-phase regions. The corresponding phase-evolution and relevant structure models are shown in Fig. 3.
The formation energy is defined as

\[ \epsilon_f(x) = \epsilon_{Li_xSi} - x \cdot \epsilon_{bcc-Li} - \epsilon_{c-Si}. \]  

Here, \( \epsilon_{Li_xSi} \) is the energy per Li\(_x\)Si unit, \( \epsilon_{bcc-Li} \) is the energy per Li in bcc Li and \( \epsilon_{c-Si} \) is the energy per Si in c-Si; this definition corresponds to the use of bcc-Li and c-Si as reference electrodes.

In Fig. 2, open circles (connected by a black curve) correspond to calculated formation energies of a homogeneously lithiated a-Li\(_x\)Si electrode-particle. For comparison, we have added the formation energy of crystalline Li\(_{1.5}\)Si\(_4\), represented by a square. In addition, several straight lines are drawn. These straight lines connect data points for two distinct values of \( x_a \) and \( x_b \) in such a way that data points at intermediate values of \( x \) are all located above the line. Neglecting interface energies, this implies that the formation energy of the phase-separated Li\(_a\)Si/Li\(_b\)Si system (consisting of both Li\(_a\)Si and Li\(_b\)Si) is lower than the...
formation energy of the homogeneous Li$_x$Si system ($a < x < b$); the phase-separated system is therefore preferred.

The calculated formation energies suggest the phase evolution detailed in Fig. 3. There, we also present structure models for relevant values of $x$. We start with lithiation of a c-Si electrode-particle. Insertion of Li directly results in a two-phase region where c-Si and a-Li$_2$Si coexist, see solid red line in the left panel in Fig. 2. Further Li insertion leads to a homogeneous a-Li$_x$Si one-phase region with $2 < x < 3.75$. We note that in this region an a-Li$_2$Si/c-Li$_{15}$Si two-phase region is in principle thermodynamically favoured. This is indicated by the green dashed line in the left panel in Fig. 2. However, this two-phase region requires crystallisation. For this to happen, the local concentration of Li has to increase from $x = 2$ to $x = 3.75$. This is unlikely due to the curvature of the calculated formation-energy curve. In addition, even if thermal fluctuations may increase the local Li concentration, we expect a critical nucleus-size below which the energy gain of the volume phase will be outweighed by the excess energy due to the interface between the crystalline phase and the amorphous surrounding. Thus we consider an a-Li$_2$Si/c-Li$_{15}$Si$_4$ two-phase region as unlikely. At $x \approx 3.75$, this situation changes. There the slope of the formation-energy curve is close to zero, and therefore, crystallisation of Li$_{15}$Si$_4$ becomes much more likely. In Fig. 2, crystallisation is indicated by the solid grey line. However, due to finite-size effects, the amorphous phase may still be favoured. In this case a second two-phase region is predicted where a-Li$_{3.75}$Si and a-Li$_{4.4}$Si coexist, see solid blue line in Fig. 2.

We now focus on delithiation. The phase evolution upon delithiation is qualitatively independent of the final state of the lithiation cycle. In both cases we first expect a two-phase region, followed by a one-phase region until a-Si is reached. However, the details differ slightly. Starting at a-Li$_{4.4}$Si, we expect an a-Li$_{4.4}$Si/a-Li$_{3.75}$Si two-phase region followed by a homogeneous a-Li$_x$Si one-phase region with $3.75 > x > 0$. Starting at Li$_{15}$Si$_4$, we expect a Li$_{15}$Si$_4$/a-Li$_2$Si two-phase region, followed by a a-Li$_x$Si one-phase region with $2 > x > 0$ (note that this two-phase region, which is absent during insertion, now can occur since amorphisation does not require a critical nucleus size; this is indicated by the solid green line in the right panel of Fig. 2).

Finally we consider (re)insertion into a-Si. This can be achieved by reading the right panel of Fig. 2 from left to right. Lithiation of a-Si proceeds homogeneously up to a Li content of $x \approx 3.75$. From there on, the a-Si electrode-particle behaves identical to a c-Si electrode-particle.

4 Discussions

In the literature, the volume expansion of Si is often pointed out as the main reason for cracking of the electrodes during cycling and thus for the rapid capacity fading$^{2,3}$. Progress in functionalisation of Si anodes has been made by capacity-limited cycling$^{4,5}$. It is often believed that capacity-limited cycling reduces crack-formation due to volume changes. However, even when limiting the capacity to one fourth of the maximum capacity the volume changes are still huge (+70% upon insertion). Our calculated phase evolution strongly encourages capacity-limited cycling, not because of a limited volume expansion, but due to the absence of highly destructive two-phase regions. In fact, homogeneous volume changes are largely irrelevant for cracking; the system simply undergoes a breathing be-
Figure 4. Potential degradation mechanism of a multi-particle Si electrode during capacity-limited cycling. (Top panel) Initially, the anode consists of a collection of amorphous Si particles connected by a binder phase (not shown) and permeated by the electrolyte (white). During first lithiation, only a fraction of the Si particles is accessed (possible reasons for this are discussed in the text). In particular, some particles will be almost unlithiated, while others become fully lithiated leading to crystallisation of Li$_{15}$Si$_4$. During delithiation, the parts of the anode that contain crystalline Li$_{15}$Si$_4$ fracture due to large stresses because of inhomogeneous volume shrinkage (c-Li$_{15}$Si$_4$/a-Li$_2$Si two-phase region). The fully lithiated core of these particles loses contact to the (partially) delithiated outer shell. Since the core is not in contact with the binder it becomes inactive for further Li insertion/extraction. (Bottom panel) In forthcoming cycles, more and more particles become lithiated and thus more and more particles degrade.

We note that currently Si anodes are still problematic for commercial application due to persistent irreversible capacities that occur even during capacity limited cycling. Our calculations indicate that these irreversible capacities are not an intrinsic problem of Si electrode-particles themselves, that is, essentially no Li is expected to be trapped inside a Si particle if it is cycled at limited capacity. However, in reality, an electrode consists of thousands of electrode particles. Capacity limitation can only be achieved for the collection of particles, not for individual particles. In particular the formation of a solid-electrolyte interphase around individual particles may prevent an electrode from being homogeneously lithiated on a global scale. As a result, some particles may be fully lithiated, while others are not lithiated at all. In fully lithiated particles Li$_{15}$Si$_4$ will crystallise and subsequent delithiation can lead to cracking of that anode particle. As a result this particle loses contact and cannot be fully delithiated. Such a scenario could explain persistent irreversible capacities during capacity limited cycling and is depicted in Fig. 4.

---

haviour without (significant) crack formation. Inhomogeneous volume changes that occur within two-phase regions, may, on the other hand, easily lead to cracking and pulverisation. Using a-Si and avoiding crystallisation of Li$_{15}$Si$_4$ (at high capacities), therefore leads to a highly enhanced cycling behaviour of Si.
5 Conclusions

We present large-scale first-principle modelling of amorphous Li-Si alloys relevant for Silicon-based anodes in Li-ion batteries. Our calculated phase evolution is very much in line with experiments\textsuperscript{16–18}. In particular the occurrence of a two-phase region for lithiation of c-Si and for delithiation of Li\textsubscript{15}Si\textsubscript{4} have been reported earlier by experimental means. The a-Li\textsubscript{4}Si\textsubscript{4}/a-Li\textsubscript{1.75}Si two-phase region, on the other hand, has not been discussed in the literature so far. This may be mainly due to the fact that Li content of $x = 4.4$ has been experimentally reached only recently by the use of nanowires\textsuperscript{19}, while the phase content of Si anodes has only been investigated for nanoparticles or thin films.

Concerning degradation of Si anodes we identify the coexistence of Li\textsubscript{15}Si\textsubscript{4} and a-Li\textsubscript{2}Si as a key factor. The large inhomogeneous volume changes are sources for cracking and pulverisation. Capacity-limited cycling, on the other hand, should prevent crystallisation of Li\textsubscript{15}Si\textsubscript{4} and therefore lead to the absence of the main degradation mechanism. However, for multi-particle electrodes, global inhomogeneities may arise, in particular due to a solid-electrolyte interphase. In fact, such inhomogeneities have indeed been reported recently\textsuperscript{20}. This can then lead to crystallisation of Li\textsubscript{15}Si\textsubscript{4} in individual particles and thus to their degradation. Potentially, by alloying Si with impurity atoms\textsuperscript{21}, crystallisation may be strongly suppressed.

Acknowledgements

The authors gratefully acknowledge the SPP 1473 WeNDeLIB of the German research foundation (DFG) and BMBF project “Elektrochemie für Elektromobilität - Verbund Süd” for financial support. The Jülich Supercomputing Centre (Project HDA17) is acknowledged for providing computational resources.

References


http://pubs.acs.org/doi/abs/10.1021/jp401379d

https://wiki.fysik.dtu.dk/ase/


15. C. K. Chan et al., *Surface chemistry and morphology of the solid electrolyte interphase on silicon nanowire lithium-ion battery anodes*, J. Power Sources 189, 1132–1140, 2009.


20. D. Robert et al., *Mechanism of the first lithiation of micrometric and nanometric Silicon particles in Li-ion batteries studied by electron Microscopy*, Oral Poster O48, 6th Lithium Battery Discussion Forum, LiBD 2013, June 16th -21st, Arcachon, France; E. Radvanyi et al., *Study of the failure mechanism of silicon electrodes by using AES, XPS and EIS*, Oral Poster O50, 6th Lithium Battery Discussion Forum, LiBD 2013, June 16th -21st, Arcachon, France.

Direct Excitation of Charge-Transfer States in the ZnPc-C$_{60}$ Donor-Acceptor Complex

Oliver Brügner and Michael Walter

Freiburg Materials Research Center, Stefan-Meier-Str. 21, 79104 Freiburg, Germany
E-mail: Michael.Walter@fmf.uni-freiburg.de

The binding of the prototypical donor-acceptor complex of CuPc and ZnPc with C$_{60}$ is investigated using van der Waals corrected density functional theory. Binding energies are found to be much higher than previously reported, i.e. 0.89 eV and 0.91 eV for CuPc and ZnPc, respectively. The optical absorption spectrum of the complex is analysed on its charge transfer character. We find transitions that have both large oscillator strength and partial charge transfer suggesting that the excitation of these states facilitate the charge transfer process.

1 Introduction

Organic solar cells (OSC) are promising candidates for cheap solar energy harvesting\(^1\). The OSCs still suffer from lower efficiencies as compared to semiconductor cells, but there is a lot of effort to improve in this respect. The possibilities for advance are manifold as many aspects of the OSC function are still poorly understood. The heart of organic solar is the donor-acceptor complex where the photon energy is used to create an excitation (exciton) in the donor molecule D. In a subsequent step, the exciton has to be separated, in that the electron is transferred to the acceptor molecule A. Then both the hole residing on D and/or the electron have to find their way to the electrodes to enable the use of the electrical energy. All these steps influence the efficiency of the solar cell and critically depend on the composition and structure of the donor-acceptor blends used.

The charge separation step requires a close contact of D and A and happens on their interface. Therefore an increase of this interface should enhance the possibility of charge transfer. The maximal contact is obtained for D-A complexes. Chemically linked D-A dyads show low energy excitations in the absorption spectrum that are interpreted as charge-transfer transitions\(^2\). Similar effects have been observed in the absorption and emission spectra of D-A blends in thin films with analogical interpretations\(^3\). The excitation spectrum of dyads\(^4,5\) and D-A complexes\(^6,7\) were studied using time-dependent density functional theory (TD-DFT). Ghosh and Gebauer found that the TD-DFT excitations are at lower energies than the experimental ones\(^5\). Only a completely charge separated state obtained with constraint DFT\(^8\) is found at higher energy in better agreement with experimental excitation energy.

There is a problem with the interpretation of the excitations seen in the experiment as completely charge separated states, however. Transitions from the ground state to a completely charge separated state should have vanishing absorption cross section due to missing spatial overlap. Therefore we expect the excited states to have only partial charge transfer character if they are directly accessible by photon absorption. We therefore take a different viewpoint here, in that we study the excitation spectrum of a D-A complex and search for excited states the provide both partial charge transfer as well as significant oscillator strength.
There are many different donor-acceptor complexes possible. One popular combination is metal-phtalocyanines (MPc) combined with C_{60} derivatives. We choose CuPc and ZnPc complexed with C_{60} as our model system.

2 Methods

We describe the Cu/ZnPc-C_{60} D-A complexes in the gas-phase using density functional theory (DFT). The electron density and the Kohn-Sham wave-functions are calculated within the projector augmented wave method using GPAW. The smooth wave-functions are represented on real space grids with a grid-spacing of \( h = 0.2 \) Å. The molecules are bound via van der Waals interactions that are known to be a difficult task for the usual approximations used in DFT. Gradient corrected functionals (GGAs) that are very successful in the description of chemical bonds, usually completely fail to provide dispersion related attraction, whereas the lower level local density approximation (LDA) shows some, but unphysical contribution. This deficiency can be cured efficiently by a correction that adds the dispersion contributions to the energy obtained by traditional DFT. We use a particular formulation proposed by Tkatchenko and Scheffler (TS09) that is mainly based on calculated values from ab initio theory and corrects the otherwise very successful PBE functional.

The electronic structure calculation is based on the PBE approximation as the TS09 approach influences the total energy only. The analysis of a possible charge transfer character requires the assignment of the electron density to the atoms. This procedure is not unambiguous and different partitioning schemes were proposed in the literature. We use two popular schemes here: the Hirshfeld partitioning, which weights the space regions by the electron density of the free gas-phase atoms and the Bader approach, where the atomic regions are separated by the minima of the electron density. Excited states properties are obtained via spin restricted DFT and time dependent DFT (TD-DFT). The TD-DFT calculations use the linear response formulation to calculate excitation energies, dipole oscillator strengths and to create approximate wave functions.

3 Structure of the Complex

We have studied complexes of CuPc and ZnPc with C_{60} in different conformations that are depicted in Fig. 1 for the example of CuPc (the structures for ZnPc are nearly identical and not shown). The complexes have their energetic ground state in configuration (a) of Fig. 1 with binding energies of 0.89 eV and 0.92 eV for CuPc and ZnPc, respectively. In terms of potential energy, this structure is followed by configurations (b) and (c) that are approximately 0.1 eV higher. Configurations (d) and (e) are again 0.1 eV higher and (f) provides only about 0.3 eV binding energy independent of the central metal atom.

The binding to C_{60} leads to a substantial disturbance of the unperturbed flat structure of CuPc as seen in configurations (a-e) in Fig. 1 in agreement with the finding of Ref. 19. The bowled shape of CuPc can be attributed to the increase of interaction with the C_{60} ball. The inclusion of van der Waals interactions leads to a substantial increase in binding energy as compared to the reported values of 0.23 eV and 0.19 eV for Zn PC using Hartree-Fock and B3LYB, respectively. Accordingly, we found binding energies of \( \approx 0.4 \) eV in the LDA and purely repulsive interaction within PBE.
4 Charge Transfer Excitations

In order to study the charge transfer properties of ground and excited states, we restrict ourselves to ZnPc-C\textsubscript{60} in its energetic ground state structure (Fig. 1(a)). Within the PBE functional, ZnPc is found to have a ionisation potential of 6.51 eV and the electron affinity of C\textsubscript{60} is predicted to be 2.78 eV. The transfer of an electron from well separated ZnPc to C\textsubscript{60} would therefore cost 3.73 eV. Weak van der Waals binding should not influence this energy substantially, such that we do not expect much charge transfer in the ground state of the complex.

This expectation is confirmed by the charging analysis presented in Tab. 1, obtained from Bader and Hirshfeld approaches. Notably, the two schemes are quite different in their assignment of the atomic charges. Nominally the Zn atom is expected to give two electrons to Pc, but the actual charging of the Zn atom is 1.29 $e$ (the electron charge is denoted by $-e$) according to Bader and only 0.4 $e$ according to Hirshfeld partitioning. Bader assigns -1.21 $e$ to each of the nitrogen atoms, while Hirshfeld partitioning finds only 0.14 excess electrons per N. In contrast to this disagreement in the atomic charges, both partitioning schemes agree in that there is negligible charge transfer between ZnPc and...
Table 1. Charges in |e| in the singlet and triplet ground and first excited states of the ZnPc C₆₀ complex.

<table>
<thead>
<tr>
<th></th>
<th>Bader</th>
<th>Hirshfeld</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singlet ground state 0 eV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZnPc</td>
<td>0</td>
<td>-0.02</td>
</tr>
<tr>
<td>Zn</td>
<td>1.29</td>
<td>0.40</td>
</tr>
<tr>
<td>N</td>
<td>-1.21</td>
<td>-0.14</td>
</tr>
<tr>
<td>C₆₀</td>
<td>0</td>
<td>0.02</td>
</tr>
<tr>
<td>Triplet ground state 1.10 eV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZnPc</td>
<td>0.28</td>
<td>0.22</td>
</tr>
<tr>
<td>Zn</td>
<td>1.30</td>
<td>0.40</td>
</tr>
<tr>
<td>N</td>
<td>-1.21</td>
<td>-0.14</td>
</tr>
<tr>
<td>C₆₀</td>
<td>-0.28</td>
<td>-0.22</td>
</tr>
</tbody>
</table>

C₆₀, however.

The first excited state of the complex is expected to be the triply degenerate triplet. These states can be directly accessed by a spin restricted DFT ground state calculation, i.e. by setting the spin projection $S_z$ defined by the difference in majority and minority spin density, to two electrons. We can expect a charge transfer from ZnPc to C₆₀ in these states, as in the Kohn-Sham single particle picture there has to be a transfer of one electron from the ZnPc-HOMO to the C₆₀-LUMO. Such a state is found by spin-restricted DFT to be 1.10 eV higher in energy than the singlet ground state and its charge distribution is also analysed in Tab. 1. The atomic charges are again very different for the two assignment schemes, but total charges are in good agreement with each other. Bader and Hirshfeld assign a charge transfer of 0.28 and 0.22 electrons to C₆₀ in the triplet state, respectively. The charges of Zn and N atoms do not change much as compared to the singlet ground state indicating a delocalisation of the charge over the molecules.

The first excited triplet states analysed above are not directly accessible via photon absorption as the photon does not couple to spin degrees of freedom, however. We would like to have a criterion to decide whether an optical transition can lead directly to charge transfer or not. In the following, we use an approximate scheme to assign each TDDFT-excited state an electron density and in this way enable the charge transfer analysis. We will find that there are indeed excited states that both show charge transfer and non-vanishing optical matrix elements.

The only variable needed in DFT is the electron density and the Kohn-Sham orbitals do not have a connection to reality in principle. It is common practice to view the Kohn-Sham states as electrons living in an effective potential, however. This approach is very successful e.g. for the interpretation of photo-electron spectra. Furthermore, the interpretation of a Slater determinant constructed from the occupied Kohn-Sham states can be considered as an approximate wave-function. The corresponding density and its difference to the ground state density can then be used as criterion for the charge transfer. For technical simplicity, the charge transfer is analysed using Hirshfeld partitioning in the following, but we expect no difference to a Bader analysis due to the experience from above.

HOMO=highest occupied molecular orbital, LUMO=lowest unoccupied molecular orbital
The optical absorption spectrum of an isolated ZnPc molecule is shown in Fig. 2 a). ZnPc has a doubly degenerated first excitation with large oscillator strength of $f$=0.41 at 2.08 eV in excellent agreement with similar calculations using a B3LYB functional\textsuperscript{20,21}. The DFT calculations overestimate the experimental excitation energy of 1.85-1.95 eV, but perfectly describe the experimental oscillator strength of $f$=0.40\textsuperscript{20}. This excitation is embedded in excitations with vanishing oscillator strength, but well separated from higher optically active excitations above 2.7 eV. The excited state spectrum of C\textsubscript{60} is also known to be dense\textsuperscript{22}, but transitions with larger optical strength appear at higher photon energies only as seen in Fig. 2 a).

Fig. 2 b) depicts the optical spectrum of the complex. The excitation at 2.08 eV is broadened due to the presence of the C\textsubscript{60} molecule and consists of more transitions now. A similar broadening has been predicted for the ZnTPP-C\textsubscript{70} system\textsuperscript{5}. There are also a nearly invisible (due to their small oscillator strength) excitations around 0.9 eV. These excitations are the three HOMO-LUMO transitions corresponding to the first excited singlet states
(spin unpolarised linear response TD-DFT calculations reveal singlet excitations only\textsuperscript{23}). Accordingly, the charge analysis in Fig. 2 c) reveals that these low transitions have large charge transfer. These states are hard to excite directly due to their low $f$. There are also states with large $q < 0$ around $\omega = 2$ eV (where $f$ is larger) followed by a bunch of states with opposite charge transfer $q > 0$. The latter correspond to HOMO(C$_{60}$)-LUMO(ZnPc) transitions and therefore represent hole transfer to the C$_{60}$.

To zeroth order approximation, we can assume that the overlap to a completely charge separated state\textsuperscript{5,6} is proportional to the charge transfer $\delta q$ in the excited state already present. The oscillator strength $f$ gives the probability for excitation by photons of the right energy. Therefore the product $\delta q \cdot f$ can give a measure of the combined probability for charge transfer and photon absorption. This quantity is depicted in Fig. 2 d). Indeed, we can find excited states with high OS and large electron transfer character to the C$_{60}$ slightly below 2 eV. According to this simple picture, these states should lead to enhanced charge separation as compared to the strongest transition nearby. Note, that these states are found about 1 eV above the simple HOMO(ZnPc)-LUMO(C$_{60}$) transitions that are usually analysed. Only slightly above there are states with comparable $\delta q \cdot f > 0$, that should preferably overlap with hole transfer to C$_{60}$ and thus be lost for the solar energy conversion.

5 Conclusions

We have analysed complexes between CuPc and ZnPc donor molecules and the C$_{60}$ acceptor molecules. Using dispersion corrected functionals, we have obtained a binding energy of nearly 1 eV for these complexes, much larger than the reported values before.

We also have have investigated the charge transfer in the ground and excited states. While there is large disagreement of the charge state of the single atoms between Bader and Hirshfeld partitioning schemes, the different schemes agree remarkably well in the degree of charge transfer between donor and acceptor molecules. An approximate scheme to obtain charge transfer from TDDFT excited states predicts the existence of excited states that have both non-vanishing oscillator strength and partial charge transfer. These states are good candidates for large overlaps to truly charge separated states in the donor-acceptor complex.

The nature of the electronic structure and excited states is a difficult task for DFT and TD-DFT in particular if charge transfer is involved\textsuperscript{24}. Range separated functionals are popular candidates to improve this situation\textsuperscript{25} and we are currently working on an implementation of such functionals to the GPAW package. This will enable to reexamine to conclusions drawn above.

Finally, an experimental study of an isolated D-A complex could give further insight here. While the creation and detection of such complexes in the gas-phase could be very difficult, it might be possible to study D-A complexes confined in helium droplets\textsuperscript{26} where molecular aggregates have been observed already\textsuperscript{27}.

Acknowledgements

M. W. thanks for useful discussions about the topic with E. v. Hauff and F. Stienkemeier. Computational resources from FZ Jülich and the BWGrid\textsuperscript{28} are gratefully acknowledged.
References


Condensed Matter
“Condensed Matter” is a wide field, encompassing fluids and solids, “soft matter” like polymers, colloids, etc., and “hard matter” like metals, insulating crystals, etc. Condensed matter is the basis for both living species and for all the materials used for machines and devices. Consequently, some aspects of condensed matter are dealt with in other chapters of this Proceedings volume as well, such as Computational Biology and Biophysics, Chemistry, Materials Science, and Computational Soft Matter Science. In the present chapter, the focus will be on problems where quantum-mechanical aspects of condensed matter play a decisive role.

The first article by Anders and Joschev deals with a popular problem of nanoscience, namely the quantum transport through a molecular junction. In corresponding experiments, a complex organic molecule is contacted by two conducting leads, which in the present study were modelled by featureless free electron gases. The molecule is described by a single active molecular level, coupled to vibrational modes of the molecule, in order to create a minimalistic but generic model for this problem. The charge transport through this device then is analysed by a numerical renormalisation group treatment, yielding the influence of a vibrational mode on the differential conductance, both for a particle-hole symmetric model and a model lacking this symmetry.

The article by Assaad et al. presents a Quantum Monte Carlo study of the so-called Kane-Mele-Hubbard model, which is motivated by the recent discovery of graphene (honoured by a Physics Nobel Prize not long ago!) and its exciting properties. This model describes electrons on the honey-comb lattice, with nearest-neighbour hopping, spin-orbit coupling and an on-site repulsion between the electrons. The ground-state phase diagram of this model is established, exhibiting several phases: an antiferromagnetic Mott insulator, a semi-metal and a quantum spin Hall insulator. The semi-metal to insulator transition has been analysed by finite size scaling methods and could be shown to fall in the Gross-Neveu universality class (describing massless Dirac fermions coupled to a vector boson due to the Goldstone mode of the antiferromagnet). Very interesting is also the transition from the quantum spin Hall phase (which is a “topological insulator”) to the $xy$ antiferromagnetic Mott insulator.

The article by Bockstedte et al. deals with the surface structure of thin crystalline hexagonal ice layers grown on metal surfaces, a problem that had been studied by STM experiments, which seemed to be difficult to interpret. Several possibilities for surface
reconstruction (describable as periodic arrangements of H$_2$O admolecules of the surface) are conceivable. Careful density functional studies of this admolecule termination of ice surfaces could shed new light on this problem.

Finally, Lenchuk et al. use also density functional theory to study interfacial properties in metallic alloys, focusing on the segregation of Zr at symmetric grain boundaries in Mo, for Mo-Si-B-Zr alloys. Zr “microalloying” has the (desirable!) effect of strengthening such materials, and the present study could provide a theoretical explanation for this effect.

Thus, the chapter provides an impression for the wide spectrum of problems in the physics of condensed matter that significantly profit from massive computations on high performance supercomputers: from fundamental problems of quantum phase transitions to the mechanisms of mechanical strength of alloys, or the conductance of nanoelectronic devices, or the question how an ice surface looks like on the atomistic scale.
Influence of Vibrational Modes on the Quantum Transport through a Nano-Device

Frithjof Anders and Andre Jovchev

Lehrstuhl für Theoretische Physik II, Technische Universität Dortmund, 44221 Dortmund, Germany
E-mail: {frithjof.anders, andre.jovchev}@tu-dortmund.de

We use the scattering states numerical renormalisation group (SNRG) approach to calculate $I(V)$ and the differential conductance through a single molecular level coupled to a local molecular phonon. The phonon site peaks in the equilibrium spectral function are related to additional maxima in the differential conductance. Non-equilibrium effects lead to significant deviations between a symmetrically coupled junction and a junction in the tunnel regime. The suppression of the current for asymmetric junctions with increasing electron-phonon coupling, the hallmark of the Franck-Condon blockade, is discussed. The renormalisation of the tunnelling rate and the occurrence of a dynamically generated capacitance are extracted from the equilibrium NRG and related to the non-equilibrium results.

1 Introduction

In the quest for size-reduced and possible low-power consuming electronic devices, the proposal of using molecular junctions for electronics has sparked a large interest in understanding the influence of molecular vibrational modes onto the electron charge transfer through a molecule. Interestingly, hysteretic behaviour of the $I(V)$ curves has been reported in several experiments when sweeping the voltage with a finite rate. However, the observed hysterese are non-universal and depend on the sweeping rate. For infinitesimally slow sweeping the effect vanishes. In some cases a sudden drop of the current has been observed with increasing bias voltage which translates into a negative differential conductance. This all has been accounted to configural changes of the molecule emphasising the importance of vibrational couplings in such devices.

Many experimental facts have been gathered in the last two decades but there is still a lack of an accurate theoretical description of all the reported phenomena. An excellent review by Galperin et al. comprehensively summarises the different theoretical approaches and experimental findings. Single molecular transistors (SMT) promise to offer some advantages over their semi-conductor based counterparts. Both types of single-electron transistors can be controlled by a capacitively coupled external gate. The molecular energy scales, however, are larger in SMTs and reproducibly defined by the chemistry of the molecule. In addition, the coupling to vibrational modes enlarges the parameter space and different physics such a phonon-assisted tunnelling, Frank-Condon blockade or the appearance of inelastic steps in the $I(V)$ curve can be observed.

2 Theory of Quantum Transport through a Molecular Junction

2.1 Model

In molecular electronics experiments, a complex organic molecule is contacted by two conducting leads. We have modelled these leads as two symmetric featureless free electron
gases since the mean-free path in the leads is large compared to the spatial dimensions of the device. In general, the molecule can contain several molecular orbitals which are actively participating in the quantum transport. Furthermore, the internal vibrational modes of the molecule are influenced by charging and discharging of the molecule.

In this work we are not targeting an approximate solution for a realistic but very complicated description tailored for a specific molecular junction. We have focused on the most minimalistic model for quantum transport through a molecule\textsuperscript{4,7} which we solve accurately with a state of the art numerical approach. The model consists of a single active molecular level - all others are energetically well separated – whose charge density is coupled to a local Holstein phonon stemming from the dominating vibrational mode of the molecule. An electron can tunnel from and to the molecule from lead $\alpha = L, R$ by the tunnel matrix element $t_\alpha$. In real materials, band features are important but only influence the single-particle properties which can be accounted for in a frequency dependent charge transfer rate $\Gamma_\alpha(\omega)$ which we treat as a constant for simplicity in our simulations. This widely used\textsuperscript{4,7,8} Hamiltonian is defined as

$$H = E_d d^\dagger d + \omega_0 b^\dagger b + \lambda_{ph}(b^\dagger + b) \left( \hat{n}_d - \frac{1}{2} \right) + \sum_{\alpha=L,R} \sum_k \varepsilon_{k\alpha} c^\dagger_{k\alpha} c_{k\alpha}$$

$$+ \sum_{\alpha=L,R} \frac{t_\alpha}{\sqrt{N}} \sum_k \left( d^\dagger c_{k\alpha} + c^\dagger_{k\alpha} d \right)$$

where $d (d^\dagger)$ annihilates (creates) an electron on the device with energy $E_d$, and $c^\dagger_{k\alpha}$ creates an electron in the lead $\alpha$ with energy $\varepsilon_{k\alpha}$. The local charge-transfer rate to each lead $\alpha$ is given by $\Gamma_\alpha = \pi t_\alpha^2 \rho_\alpha(0)$, where $\rho_\alpha(\omega)$ is the density of states of lead $\alpha$. In order to focus only on the influence of the electron-phonon interaction onto the quantum transport, the spin degree of freedom is neglected in order to avoid obstruction of the competition between spin-flip scattering through the device and polaron formation on the device. For $\lambda_{ph} = 0$, the phonon decouples and the Hamiltonian reduces to an exactly solvable resonant level model (RLM).

![Minimal model of a molecule consisting of a single active molecular level at energy $E_d$ coupled to two leads with tunnelling matrix elements $t_L$ and $t_R$. Depending on the local charge configuration $|0\rangle$ or $|1\rangle$, the ground state of a vibrational degree of freedom is shifted. The relative displacement between the two configuration is given by the dimensionless electron phonon coupling $g = \lambda_{ph}/\omega_0$. The phononic excitations for a fixed charge are multiples of the oscillator energy $\omega_0$.](image)
The spinless Anderson-Holstein model is schematically depicted in Fig. 1. Depending on the local charge configuration, the local harmonic oscillator is displaced and the dimensionless distance between the two ground states is given by $g = \lambda_{ph}/\omega_0$. For modelling realistic situations, the restriction to a single phonon and a single electronic level must be lifted. In spite of a lot of theoretical progress⁴ this model has only been accurately solved in equilibrium⁹,¹⁰,⁸, while its non-equilibrium dynamics has only been perturbatively investigated in lowest order of the coupling constants⁴.

Using the Lang-Firsov transformation¹¹,¹², the local part of the model given by the three first terms in Eq. 1 can be solved exactly. This solution describes the formation of a local polaron which decouples from a shifted harmonic oscillator. The corresponding polaronic energy gain is given by $E_p = \lambda_{ph}^2/\omega_0 = g^2\omega_0$. Coupling this local problem to the two leads defines two competing regimes. For $E_p, \omega_0 \ll \Gamma_0 = \Gamma_L + \Gamma_R$, the phonon dynamics is slow and can be treated perturbatively in this adiabatic regime. $E_p, \omega_0 \gg \Gamma_0$ defines the opposite limit: in this anti-adiabatic regime charge fluctuations are suppressed, the electron moves slowly and the phonon defines the large energy scale. The anti-adiabatic regime is relevant for molecular junctions since the tunnelling coupling of a molecule to the leads is usually small compared to the intrinsic energy scales of the molecule. After the Lang-Firsov transformation, the tunnelling term acquires an additional factor $\exp\left[g(b^\dagger - b)\right]$ whose physical meaning is stripping the original electron content from the locally formed polaron. If $\omega_0 \gg \Gamma$, the local phonon remains in its ground states which yields an exponential suppression of the tunnelling coupling and $\Gamma \rightarrow \Gamma_{eff} \approx \Gamma_0 e^{-g^2}$. In a particle-hole asymmetric junction, this leads to a Franck-Condon suppression of the current to avoid the reorganisation of the nuclear positions of the molecule.

While the effective low-energy fixed point Hamiltonian of the equilibrium is always given by an effective resonant level model with $\Gamma_0 \rightarrow \Gamma_{eff}$, at intermediate energies $\omega$, $\Gamma_{eff} < \omega < \omega_0$, the instable fixed point Hamiltonian is given by an effective interacting RLM (IRLM). Integrating out the phonon degrees of freedom has lead to an additional effective repulsive interaction¹³,¹⁴ of the form $U_{eff}(\hat{n}_c-1/2)(\hat{n}_d-1/2)$ where $U_{eff}$ is of the order $U_{eff}/D \approx \pi g^2 \Gamma_{eff}/2\omega_0$. The effective parameters, $\Gamma_{eff}$ and $U_{eff}$, have been extracted directly from the numerical renormalisation group approach⁸,¹⁴ and excellently agree with these leading order analytical estimates. These different fixed points are schematically depicted in Fig. 2.

Figure 2. The renormalisation group fixed points as function of temperature for the anti-adiabatic and the crossover regime. The low-temperature fixed-point is a stable Fermi-liquid. Below the phonon frequency $\omega_0$, the vibrational dynamics is frozen; we find an interacting resonant level model characterised by an dynamically generated effective capacitance between the molecule above $\Gamma_{eff}$. The free orbital fixed point governs the physics at infinitely high temperature.
Therefore, the equilibrium properties of the model are well understood. It has been shown that the anti-adiabatic regime, in which the phonon dynamics is fast compared to the electron dynamics, i.e. $\Gamma_{\text{eff}} < \omega_0$, is in fact extended to $\Gamma_0 \approx \omega_0$ when the polaron shift $E_p > \Gamma_0$, followed by a crossover regime, where $\Gamma_{\text{eff}} \approx \omega_0$. The adiabatic regime requires that $\Gamma_0$ exceeds $E_p^8$.

### 2.2 Scattering States Numerical Renormalisation Group to Quantum Transport

Quantum transport in mesoscopic systems is traditionally described by the Landauer and Büttiker approach using a single-particle scattering theory. For nano-devices the interaction on the devices modifies the transport properties and requires a many-body description. Hershfield has shown that the steady-state limit of the density operator out of equilibrium is given by a Bolzmannian form $\rho = e^{-\beta(H-Y)}/Z$. However, the $Y$-operator is only explicitly known for a non-interacting system in terms of scattering states. The scattering-states numerical renormalisation group approach utilises that (i) the Hamiltonian (1) is exactly diagonalisable in terms of current carrying Lippmann-Schwinger scattering states and (ii) the analytically known form of the non-equilibrium density operator at finite bias. At time $t = 0$, we switch on $\lambda_{\text{ph}}$ and let the density operator evolve to infinitely long times, and project out the steady-state matrix $\rho_{\infty}$

$$
\rho_{\infty} = \lim_{T \to \infty} \frac{1}{T} \int_0^T dt e^{-iHt} \rho_0 e^{iHt}
$$

which describes the current carrying fully interacting system. Technically this is done using the recently developed time-dependent numerical renormalisation group.

### 2.3 Computational Approach

The properties of quantum impurity systems such as defined by $H$ in Eq. 1 can be very accurately calculated using the numerical renormalisation group (NRG). At the heart of this approach is a logarithmic discretisation of the continuous bath, controlled by the discretisation parameter $\Lambda > 1$. Using a Householder transformation, the Hamiltonian is mapped onto a semi-infinite chain, with the impurity coupled to the first chain site. The $N_{\text{th}}$ link along the chain represents an exponentially decreasing energy scale: $D_N \sim \Lambda^{-N/2}$. Using this hierarchy of scales, the sequence of finite-size Hamiltonians $H_N$ for the $N$-site chain is solved iteratively, discarding the high-energy states at the conclusion of each step to maintain a manageable number of states. The reduced basis set of $H_N$ so obtained is expected to faithfully describe the spectrum of the full Hamiltonian on a energy scale of $D_N$. Details can be found in the review by Bulla et al.

The extension of the NRG to real-time dynamics requires two NRG runs in parallel. One NRG run constructs the density operator for the initial preparation of the system, given an Hamiltonian $H^i$. The evolution of the system can be easily calculated given the eigenstates of the final Hamiltonian $H^f$, and the overlap of the initial and final eigenstates, both obtained in the second NRG run. One efficient way of accurately reconstructing the continuum limit is sampling over an number of $N_z$ different discretisations of the lead electron spectrum. For recovering the high energy phonon side peaks in the local electron green function we have used up to $N_z = 512$ configurations. Therefore, the general purpose
supercomputer such as the JUROPA has been an ideal tool for performing such a run very efficiently. Each MPI Thread processes one TD-NRG run comprising of the two coupled NRG runs. All individual spectra are averaged at the very end of the calculation in order to obtain a faithful representation of the full continuum solution. This also is keeping the communication overhead low.

3 Equilibrium Properties: Crossover from the Adiabatic to the Anti-Adiabatic Regime

In order to set the stage for the steady-state non-equilibrium results, we present the effective charge fluctuation scale $\Gamma_{\text{eff}}$ and the dynamically generated Coulomb repulsion $U_{\text{eff}}$ of the instable IRLM fixed point obtained from the equilibrium NRG calculations. These values can be extracted directly from the NRG level flow and matrix elements. We have used $\Lambda = 1.5$, have kept $N_s = 1500$ NRG states after each iteration, and have included the lowest $N_b = 400$ local phonon states. The bandwidth of a lead with constant density of states $\rho(\omega) = \Theta(D - |\omega|)/2D$ was set to $D = 100\Gamma_0$, and $\Gamma_0$ serves as energy scale throughout the paper. We restrict ourselves to the PH-symmetric limit by setting $E_d = 0$.

The results are depicted in Fig. 3. Part (a) shows the evolution of $\Gamma_{\text{eff}}$ as function of $g^2$ while part (b) displays the change of $U_{\text{eff}}$ with increasing electron-phonon coupling. The upper panels show the exacted parameters while in the lower panels the analytical renormalisation factors are divided out. The larger the phonon frequencies compared to $\Gamma_0$ the closer $\Gamma_{\text{eff}}$ approaches $e^{-g^2}\Gamma_0$. For smaller frequencies, the frozen-phonon approximation becomes invalid, and the phonon dynamics yields to a significant deviation between the simple analytical estimate and the true NRG data. Therefore, we have excluded the $\omega_0/\Gamma_0 = 1$ data and depicted only the evolution of $U_{\text{eff}}$ for $\omega_0/\Gamma_0 = 5, 10, 20$. In this regime, we find universality as depicted in the lower panel of Fig. 3(b).

![Figure 3](image_url)

Figure 3. (a) Anti-adiabatic and crossover regime: $\Gamma_{\text{eff}}$ as function of $g^2 = (\lambda_{ph}/\omega_0)^2$ for $\omega_0/\Gamma_0 = 1, 5, 10$, lower panel rescaled $\Gamma_{\text{eff}}$ with $g^2$ vs $g^2$. (b) upper panel: $\omega_0U_{\text{eff}}$ as function of $g^2$ for four different values of $\omega_0/\Gamma_0 = 1, 5, 10, 20$, lower panel: $f(x) = U_{\text{eff}}/U_{\text{ana}}$ vs $\alpha(\omega_0)g^2$ where $U_{\text{ana}}/D = \pi g^2\Gamma_{\text{eff}}/2\omega_0$. We used the scaling factor $\alpha(\omega_0) = 0.85, 1, 1.1$ for $\omega_0/\Gamma_0 = 5, 10, 20$. (taken from Ref. 14)
4 Quantum Transport: Influence of a Vibrational Mode on Differential Conductance

4.1 Particle-Hole Symmetric Model

The current $I(V)$ is shown as a function of the applied source-drain voltage $V$ in Fig. 4(a) for three different ratios of the tunnelling asymmetry $R = \Gamma_L/\Gamma_R$, a phonon frequency $\omega_0 = 2\Gamma_0$ and a fixed electron-phonon coupling $g = 1$. The prefactor $G_0 = e^2\Gamma_L\Gamma_R/(h\Gamma_0^2)$ contains the trivial reduction of the current upon increasing of $R$ and has been divided out. The corresponding differential conductance $G(V) = dI/dV$ calculated by numerically differentiating of the $I(V)$ curve is displayed in panel (b). For a better comparison, we added the equilibrium spectral function. The pinning of the zero-bias conductance peak for $T \to 0$ is easily understood by applying the Friedel sum rule. The narrowing of the central resonance due to the renormalisation of $\Gamma_0 \to \Gamma_{\text{eff}}$ is exemplified in Fig. 3. Qualitatively, $G(V)$ traces the transmission function $T(\omega) = \pi\Gamma_0\rho_d(\omega)$ but quantitatively significant differences are observed. The full width at half maximum (FWHM) of the zero bias $dI/dV$ peak is not simply given by $4\Gamma_{\text{eff}}$ for $R = 1$. Non-equilibrium effects cause a bias dependency of the spectral function: the width of $dI/dV$ is significantly smaller than predicted from a naive replacement of the bias-depended spectral function $\rho_d(\omega, V)$ by its equilibrium value $\rho_d(\omega, V = 0)$ for $R = 1$.

Figure 4. (a) $I(V)$ for the symmetric model, i.e. $E_d = 0$, and three different ratios $R = \Gamma_L/\Gamma_R = 1, 10, 100$. The vibrational parameters are $\omega_0 = \lambda_{ph}/2\Gamma_0$. (b) $dI/dV$ in units of $G_0 = e^2\Gamma_L\Gamma_R/(h\Gamma_0^2)$ as function of $eV$ obtained by numerically differentiating the curves in (a). Additionally, the equilibrium transmission function $T(\omega) = \pi\Gamma_0\rho_d(\omega)$ obtained from the spectral function $\rho_d(\omega)$ has been added for comparison. NRG parameters: $N_s = 2000$, $\Lambda = 2$, $N_b = 35$, $N_z = 64$, $b = 0.15$, $T = 6.5 \times 10^{-3}\Gamma_0$. (taken from Ref. 14)
The additional transmission maxima are clearly visible at finite voltage which are related to phonon-assisted tunnelling increasing significantly the current above a threshold. This threshold position depend on the asymmetry $R$ of the junction. For the extreme tunnelling regime $R \to \infty$, $dI/dV$ approaches the equilibrium transmission function for $T \to 0$.

### 4.2 Particle-Hole Asymmetric Model

For a particle-hole asymmetric regime, the I-V curve and corresponding the differential conductance $G(V)$ is depicted in Fig. 5. We have set the phonon frequency to $\omega_0/\Gamma_0 = 2$, a moderate electron-phonon coupling $g = \lambda_{ph}/\omega_0 = 1$ and $R = 10$.

![Figure 5](image_url)

**Figure 5.** (a) I(V) for the particle-hole asymmetric model, $E_d/\Gamma_0 = -2$, corresponding to $\epsilon = 0$, and an asymmetric junction $R = \Gamma_L/\Gamma_R = 10$. The vibrational parameter are $\omega_0 = 2\Gamma_0$, and $\lambda_{ph} = 2\Gamma_0$. The solid line shows the SNRG current, The black dotted line indicates the $\lambda_{ph} = \epsilon = 0$ current, the blue dashed line current for a shifted level $E_s/\Gamma_0 = -2.5$ in a non-interacting junction (RLM) at $T = 0$. (b) $dI/dV$ in units of $G_0$ as function of $eV$ obtained by numerically differentiating the SNRG curve in (a). NRG parameters: as in Fig. 4. (taken from Ref. 14)

The strong suppression of the current with increasing $\lambda_{ph}$ and fixed $\epsilon$, as seen by comparing the $\lambda_{ph} = 0$ and $\lambda_{ph}/\Gamma_0 = 2$ curves has been interpreted as Franck-Condon blockade physics. Clearly, the leading order effect stems from a polaronic shift $E_p$ of the single-particle level. The SNRG curve tracks the current through a shifted level at $E_d/\Gamma_0 = -2.5$ and $\lambda_{ph} = 0$ rather well for positive voltages as indicated by the red dashed line.
5 Concluding Remarks

We reported on our recent extension of the scattering states NRG to the charge-transport through a molecular junction\textsuperscript{14}. By analysing the low-temperature equilibrium physics, the model can be mapped onto an interacting resonant level model\textsuperscript{8,14} in the extended anti-adiabatic regime. We have extracted the renormalised charge-transfer scale $\Gamma_{\text{eff}}$ and the effective Coulomb interaction $U_{\text{eff}}$ of the instable low-energy intermediate fixed point.

With increasing electron-phonon coupling the reduction of the charge-transfer scale $\Gamma_{\text{eff}}$ yields a narrowing of the zero bias differential conductance peak. In our calculation, the phonon side peaks are clearly visible in $G(V)$ in contrary to a recent QMC study\textsuperscript{23}. Gating the junction away from the particle-hole symmetric point reveals the Franck-Condon blockade physics with increasing electron-phonon coupling.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft under AN 275/6-2, and supercomputer support was provided by the NIC, FZ Jülich under project No. HHB00.

References

Computer simulations allow to experiment with models relevant for the understanding of crystalline materials and answer some of the following fundamental questions. What is the nature of the ground state and how can we characterise it? What are the thermodynamic and dynamical properties? Can we pin down the nature of the (quantum) phase transitions? One of the main present research interests is constituted by intensive simulations of the so called Kane-Mele-Hubbard model. This model contains the physics of graphene, the physics of correlated topological insulators as well as quantum phase transitions between metallic and insulating states. It hence provides an extremely rich playground to address a number of fundamental questions and to ultimately provide approximation-free answers. Here, we will review some of our recent results.

1 Introduction

Complexity is the defining notion behind correlated electron systems: the whole is more than the sum of the individual parts. This emergent complexity gives rise to a rich variety of fascinating phenomena in fields ranging from biological systems to materials. At the same time, the complexity represents a challenging scientific problem: even if we find a way to solve the corresponding equations of motion of the constituents of a many-body system, it is by no means guaranteed that we can understand the emergent collective phenomena.

In the case of crystalline materials, we can start with a quantum-mechanical description of the electronic motion in the periodic potential of the ions in terms of Bloch states. From these states, we can construct the many-body wavefunction by taking into account the Pauli exclusion principle that demands the wave function to be antisymmetric under exchange of identical fermions. These two ingredients, Bloch states and Fermi statistics, underlie band structure theory and produce a host of very interesting states, most notably insulators and metals with completely or partially filled highest occupied band, respectively. As discovered theoretically by Kane and Mele,1 band theory also permits more exotic insulating states referred to as topological insulators. These states, which arise from strong spin-orbit coupling and time-reversal symmetry, are characterised by insulating behaviour in the interior but metallic behaviour at the edge or surface of the sample. Importantly, the characteristic surface states are robust with respect to disorder, and can therefore be observed in experiments.2 In two dimensions, the edge states are helical, that is, the direction of motion of electrons is coupled to the orientation of their spin. In principle, this property and the topological protection make topological insulators a promising candidate for applications in the field of spintronics, which offers the vision of overcoming the problem of heating inherent to current CPU technology.

Our discussion so far has neglected electron-electron interactions. However, as students already learn in introductory courses on solid state physics, the energy scale associated...
with such correlations is very large, typically several electron volts. Nevertheless, in many cases, electronic correlations do not significantly alter the properties of the system – they are irrelevant in the language of Wilson’s renormalisation group. This remarkable fact, which is known as Fermi liquid theory, hinges on the screening of the Coulomb repulsion, and Fermi statistics which drastically limits the phase space available for scattering due to the Coulomb repulsion.

Materials in which Fermi liquid theory does not hold are of particular interest. In such systems, correlations lead to exotic and/or collective states of matter. In one dimension, the concept of the Fermi liquid breaks down and correlations cause a fractionalisation of electrons into independent, collective spin and charge excitations. More generally, Mott insulators are correlated materials which are incorrectly predicted by band theory to be metallic. The insulating behaviour is often attributed to the observed ordering of, for example, the spin degrees of freedom (that is, the formation of a magnetic state). However, a central question in this context is if these two phenomena are always coupled. Remarkably, the answer to this question is no. Mott insulating states with no associated symmetry breaking can be realised in a number of models of correlated electron systems. In two dimensions, these states are yet another example of exotic states of matter which support anionic excitations with fractional statistics. Mott insulators have also received considerable attention because they produce high-temperature superconductivity upon doping with electrons or holes. Recently, the possibility of a quantum spin liquid phase has been hotly debated for the case of the Hubbard model on the honeycomb lattice\cite{3-5}. Motivated by the discovery of topological insulators, there have been numerous studies of the honeycomb Hubbard model with additional spin-orbit coupling, see Ref. 6 for a review. The Mott transition of electrons on the honeycomb lattice has recently been realised experimentally using cold atoms in an optical lattice\cite{7}.

2 The Quantum Monte Carlo Approach

For quantum many-body problems, the Hilbert space grows exponentially with the number of electrons, or the volume of the system. A brute force diagonalisation of the full problem hence requires an exponential effort. The question is whether or not we have to consider all possible states in order to understand the problem. Here, we choose the stochastic approach, and use the auxiliary-field quantum Monte Carlo method\cite{8}. The method is based on a path integral formulation where the interaction is decomposed with the help of a Hubbard-Stratonovich field that mediates the electronic correlations. The integration over the field configurations is carried out with the Monte Carlo method. A generic problem of this approach is the sign problem (the appearance of configurations with negative weights), causing an exponential increase of computer time with volume and inverse temperature to reach a given accuracy. However, for a rather large class of non-trivial models\cite{9-12}, symmetries permit us to avoid the sign problem. For such models, the numerical effort scales as $N^3\beta$ where $\beta$ is the inverse temperature and $N$ the number of electrons. The method used here can be efficiently parallelised and run on modern architectures. Recently, various optimisation schemes, in particular cache optimisation, have allowed us to gain up to an order of magnitude in performance.
3 The Kane-Mele-Hubbard Model

Motivated by the experimental progress with graphene and related systems (such as silicene and artificial graphene), we consider the paradigmatic Hubbard model for strongly correlated electrons on the honeycomb lattice illustrated in Fig. 1. Electrons can hop from site to site with hopping amplitude $t$, and two electrons at the same lattice site experience a repulsion $U$ that captures the essence of Coulomb repulsion. This model has time-reversal, $SU(2)$ spin, and sublattice symmetry, and the non-interacting band structure is that of massless Dirac fermions. Adding a spin-orbit term in the form of a complex second-neighbour hopping $\lambda$ leads to the Kane-Mele-Hubbard model for correlated topological insulators. The spin-orbit term breaks sublattice symmetry and thereby opens a topological mass gap resulting in a quantum spin Hall insulator (a widely used name for two-dimensional topological insulators). At half filling (one electron per lattice site), the above symmetries permit us to apply the auxiliary-field quantum Monte Carlo method without a sign problem, and hence to investigate the role of electronic correlations accurately and without uncontrolled approximations.

4 Selected Results

In the limits of weak and strong coupling, the low-energy physics of the Kane-Mele-Hubbard model can be understood quite easily. The semi-metallic state of Dirac fermions at $\lambda = U = 0$ is stable with respect to correlations and hence expected to survive at small $U > 0$. Similarly, at $U = 0$, the spin-orbit coupling $\lambda$ establishes a gapped quantum spin Hall state, and the bulk physics remains unaffected by small interactions $U$. For strong coupling $U/t \gg 1$, the charge degrees of freedom become frozen (any fluctuation from the
state with one localised electron per site are suppressed by $U$) but the spin degrees of freedom remain active. Using perturbation theory in the small parameter $t/U$, one can derive effective spin models\textsuperscript{17}. In the absence of spin-orbit coupling, the $SU(2)$ spin symmetry leads to a Heisenberg model with antiferromagnetic nearest-neighbour exchange. Since the lattice is non-frustrated, we expect long-range antiferromagnetic order at zero temperature, and a gapless Goldstone mode corresponding to spin-wave excitations. This antiferromagnetic Mott insulator (AFM) state is labelled $xyz$ AFM in Fig. 2. Spin-orbit coupling reduces the spin symmetry from $SU(2)$ to $U(1)$. The resulting spin model is more complex, and includes frustrated interactions in the $z$ direction of spin\textsuperscript{13}. Long-range magnetic order can develop in the transverse direction, as is the case in the $xy$ AFM phase in Fig. 2.

The validity of the above considerations can be verified using numerical simulations. Importantly, such methods also permit to study the intermediate-coupling regime and hence the evolution from small to large $U/t$. Of particular interest is whether or not there is a direct transition between the weak- and strong-coupling phases, or if instead exotic intermediate phases appear. Moreover, if continuous quantum phase transitions occur, it is of great interest to identify their universality class. In the following, we summarise our present understanding of correlation effects in the Kane-Mele-Hubbard model, as it emerges from large-scale quantum Monte Carlo simulations\textsuperscript{3,10,11,14,5}.

4.1 Semi-Metal to Insulator Transition

The following two scenarios can be envisaged for the transition from the semi-metal at weak coupling to the antiferromagnetic Mott insulator at strong coupling. Starting from strong coupling, and noting that the insulator to semi-metal transition is numerically found
Figure 3. Data collapse of the staggered magnetic moment $m$ at $\lambda = 0$, measured for different system sizes $L$. Here we have used a critical value $U_c/t = 3.78^5$, and critical exponents obtained from an $\epsilon$-expansion and Gross-Neveu-Yukawa theory$^{15,16}$. Results are taken from Ref. 5.

to occur at values of $U$ smaller than the electronic bandwidth $W = 3t$, one can construct an effective spin model for the region close to the transition where higher-order ring-exchange terms proliferate with decreasing $U/t^{17}$. This point of view suggests that the melting of magnetic order may be independent of the metal-insulator transition. Quantum Monte Carlo simulations$^1$ suggested the existence of an intermediary spin liquid phase, with a single-particle gap but no long-range magnetic order, that separates the semi-metal from the magnetic insulator. Similar conclusions have been drawn for the related $\pi$-flux model on the square lattice$^{18}$. The results of Ref. 3 have been challenged by more recent studies: Entropy calculations do not favour a degenerate ground state as expected for the $Z_2$ spin liquid$^{19}$, and Ref. 4 demonstrates that the use of significantly larger system sizes leads to an almost complete disappearance of the spin-liquid from the phase diagram.

Alternatively, we can start from the weak-coupling semi-metallic state. Upon increasing $U/t$, we eventually expect a phase transition to an insulating state with antiferromagnetic order$^{20,21,15}$. Because gapless, linear fermionic excitations exist below the transition, the transition is expected to fall into the Gross-Neveu universality class$^{15,16}$. This expectation is based on Gross-Neveu-Yukawa theory$^{15,16,5}$, which describes massless Dirac fermions with a Yukawa coupling to a vector boson describing the bosonic Goldstone mode of the antiferromagnetic state. The fermionic mass (the gap of the insulating state) is generated by the condensation of the bosonic mode (antiferromagnetic order). The theory has an upper critical dimension of $d = 3$ where the Gaussian approximation becomes exact. Within a first-order $\epsilon$-expansion, one can estimate the critical exponents for the present two-dimensional case. These theoretical predictions are confirmed by recent quantum Monte Carlo simulations$^5$. Using the predicted critical exponents, we observe an excellent data collapse [see Fig. 3] and an identical finite-size scaling of both the single-particle
gap and the staggered magnetisation\(^5\). Importantly, Gross-Neveu theory predicts that the long-ranged tail of the Coulomb repulsion, relevant for possible future experiments, will not alter the nature of the transition. Numerical work to confirm this hypothesis is currently in progress.

4.2 Topological Insulator to Antiferromagnetic Mott Transition

The transition from the quantum spin Hall phase to the xy antiferromagnetic Mott insulator is an insulator to insulator transition. Because the onset of magnetic order breaks the time-reversal symmetry that protects the quantum spin Hall state, a direct transition is possible. Since the system is insulating on both sides of the transition, the latter involves only the spin degrees of freedom which order in the xy plane at the critical point. Such a quantum phase transition, involving a \(U(1)\) order parameter in two spatial dimensions, is expected to fall into the same universality class as the three-dimensional classical XY model. Using the quantum Monte Carlo method, we have successfully verified this prediction in Ref. 11.

4.3 Correlation Effects on Edge States

Ref. 14 demonstrates that below the magnetic transition, the ground state of the Kane-Mele-Hubbard model is adiabatically connected to the non-interacting limit \((U/t = 0)\). Hence, for bulk properties we can safely neglect the interaction. This is not necessarily true for the edge state. Edge states are one-dimensional and for the so called zig-zag cut presented in Fig. 1 have a velocity set by \(\lambda\). Therefore, on the edge correlation effects should be measured in units of \(\lambda\) whereas in the bulk in units of the hopping matrix element \(t\). In the small \(\lambda\) and intermediate \(U\) region of the phase diagram, correlation effects turn out to be very important on the edge, but negligible in the bulk. In Ref. 10,22 we propose to

![Figure 4. Single-particle spectral function \(A_\uparrow(k,\omega)\) from projector-CTQMC simulations for \(L = 24, \theta = 80\) and (a) \(U/\lambda = 8\), (b) \(U/\lambda = 40\). Results are taken from Ref. 22.](image-url)
retain correlations only on the edge and neglect them in the bulk. Integrating out the non-interacting bulk produces a one-dimensional effective action which can be studied very efficiently with the so called continuous-time QMC algorithm. This approach enables us to study in detail the physics of correlated helical liquids, and to compare our results to the Luttinger liquid theory of this state of matter. Fig. 4 shows how inelastic spin-flip scattering can reduce the spectral weight of the edge state, thus leading to a suppression of the Drude weight.

5 Conclusions

Quantum Monte Carlo simulations of the Hubbard model with additional spin-orbit coupling allow us to study the interplay between band topology and correlation effects. Our results on finite lattices are free from systematic errors and therefore also serve as benchmarks for approximate theories. The continuous algorithmic progress enables us to address exciting open questions, such as the effect of a long-range Coulomb interaction on the phase diagram and the phase transitions.

Acknowledgements

Parts of this work were done in collaboration with I. Herbut, T. Lang, Z.Y. Meng, A. Muramatsu and S. Wessel. Funding from the DFG under the grant number AS 1209-1, AS120/10-1, and Ho 4489/2-1 (Forschergruppe FOR 1807) is acknowledged. We thank the Jülich Supercomputing Centre for generous allocation of CPU time.

References


Admolecule Structures at the Surface of Ice

Michel Bockstedte¹, Anja Michl¹,², and Manuel Kolb¹,³

¹ Lehrstuhl für Theor. Festkörperphysik, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany
E-mail: bockstedte@physik.uni-erlangen.de

² present address: Hamburg University of Technology, Institute of Advanced Ceramics, 21073 Hamburg, Germany

³ present address: Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands

The (0001) surface of hexagonal ice (Ih) is conceived to be terminated by a proton-ordered bi-
layer. For thin crystalline ice structures grown on metals, however, experiments report evidence
questioning the bilayer-termination. In fact, a detailed STM analysis of ice structures grown
on Cu(111) demonstrated a termination by admolecule structures on top of the bilayer. Using
density functional theory, we analyse admolecule structures from clusters to hexagon adrows
decorated with admolecules. Admolecule structures, indeed, represent low energy surface ter-
nominations, yet, exceeding the surface energy of the proton-ordered bilayer by a small margin.
However, for thin films on metal substrates weak water-metal interaction may energetically
drive the formation of admolecule terminations.

1 Introduction

Ice grown on metal substrates, owing to the flexibility of the water bond network and the
nature of the water-metal interaction, exhibits a wide range of structures including clusters
of varying shape, amorphous, and crystalline islands. Here, focus is put on crystalline
ice. Crystalline hexagonal ice (Ih) consists of mutually bonded ice bilayers stacked along
the (0001) direction. While oxygen occupies hexagonal lattice sites, the orientation of the
water molecules is disordered and only governed by the restrictions of the hydrogen-bond
network. So far the Ih(0001) surface is conceived as being terminated by the bilayer¹,².

In contrast to the bulk, the dominance of proton order at the surface was predicted by
theoretical modelling in the framework of density functional theory³ and Monte Carlo
simulations using empirical force fields⁴. In the ideal case of the Fletcher-striped bilayer
termination, neighbouring rows of surface water molecules with dangling protons (OH
groups) and flat-lying water molecules alternate.

On the other hand, recent experiments⁵,⁶ reported evidence for a termination by water
admolecules rather than by the bilayer. In particular the detailed structural analysis of ice
islands grown on Cu(111) by high-resolution STM-experiments⁶ demonstrated a surface
termination by admolecule structures on top of two bilayers. The structural analysis re-
vealed (2 × 1), (2 × 2) and c(2 × 4) admolecule reconstructions after annealing at T=130 K
and hexagon adrows decorated with admolecules between the rows upon annealing below
145 K. Earlier theoretical investigations of admolecules focused on the monomers⁷ or used
empirical force field⁸ with limited accuracy. Complex admolecule reconstructions so far
have not been addressed.
2 Theoretical Approach towards the Surface Termination of Ice

The aim of the present work is to (i) analyse energetically favourable motifs of admolecule structures guided by structures identified in the STM experiments\(^6\) and (ii) to understand the mechanisms of their energetic stabilisation. Thermodynamic and kinetic aspects shall be discerned by elucidating the consequences of the thermodynamic approach. The basic quantities here are the surface energy \(\gamma\) and correspondingly for water clusters and hexagon adrows the formation energy \(E_f\) per \(\mathrm{H}_2\mathrm{O}\) or unit length

\[
\gamma = \frac{1}{A} (E_{\text{surf}}^{\text{tot}} - n_{\mathrm{H}_2\mathrm{O}} \mu_{\mathrm{H}_2\mathrm{O}}) \quad \text{and} \quad E_f = \frac{1}{n_{\mathrm{H}_2\mathrm{O}}} (E_{\text{cluster}}^{\text{tot}} - n_{\mathrm{H}_2\mathrm{O}} \mu_{\mathrm{H}_2\mathrm{O}})
\]

(1)

where \(A\) is the surface area and \(n_{\mathrm{H}_2\mathrm{O}}\) the number of water molecules contained in the ice structure or surface cluster. \(E_{\text{surf}}^{\text{tot}}\) and \(E_{\text{cluster}}^{\text{tot}}\) are the total energies of the surface structure and the surface water cluster, respectively, as obtained at a first-principles level. \(\mu_{\mathrm{H}_2\mathrm{O}}\) is the chemical potential of water molecules used for the formation of the admolecule terminations or clusters. For the surface of bulk hexagonal ice its value corresponds to that of the bulk. As will be seen below, the description of thin ice films or other situations requires to choose different values of \(\mu_{\mathrm{H}_2\mathrm{O}}\).

Density functional theory sets an excellent framework for the investigation of the structural elements of the ice surface and their energetics\(^3,9\). In the present work admolecule structures are described by slab models building on the bilayer-terminated Fletcher-striped surface as a reference. For simplicity antiferroelectric proton order was assumed along the direction perpendicular to the surface. In order to understand the effect of strain exerted by a metal substrate, compressive strain of 1.7\% corresponding to ice adsorbed on the Cu(111) surface was investigated in addition to unstrained ice. Calculations were performed with the PWscf code\(^{10}\) using the BLYP exchange-correlation functional (for details see Ref.\(^11\)). Test calculations for prototypical structures using the PBE functional confirmed the results.

3 Admolecule Structures on the Bilayer-Terminated Surface

The best approach towards an understanding of admolecule structures on the bilayer-terminated surface is an analysis of small admolecule clusters. The monomer preferentially binds to adjacent surface sites via three hydrogen bonds\(^7,8\). Adsorption sites are located at the centre of six-fold rings or atop a subsurface water and are characterised by 1-2 adjacent water molecules with dangling protons. Stable bonding of dimers and larger water clusters feature a lower surface coordination and mutual bonds among adjacent admolecules. This enables shorter as well as stronger admolecule-admolecule and admolecule-surface bonds than for the water monomers. Admolecule clusters occur in different conformations such as linear and ring-like structures. Examples are displayed in Fig. 1 along with calculated formation energies for monomers, dimers and larger admolecule clusters. The energetics is determined by two interrelated factors: (i) the local proton order at the surface and (ii) the conformation of the admolecules. Variation of the formation energy due to different proton arrangements of up to 30 meV/\(\mathrm{H}_2\mathrm{O}\) was found. For trimers and tetromers linear conformation is favoured over ring-like or other structures. The formation energy per \(\mathrm{H}_2\mathrm{O}\) of linear clusters decreases with increasing length as demonstrated by comparison of the
trimer (72 meV/H$_2$O) and tetramer (61 meV/H$_2$O) with the limiting case of an infinite linear admolecule chain (31 meV/H$_2$O) [cf. Fig. 2(c)]. Compressive strain of 1.7% reduces the formation energy in almost all cases slightly – the dimers and the hexamer being exceptions. Still the slight effects can shift the balance between dimers and trimers.

Periodic admolecule structures formed of such clusters are discussed next. High resolution STM-experiments$^6$ report admolecule structures with nominal periodicity of $(2 \times 1)$, $(2 \times 2)$, and $c(2 \times 4)$ with coverage of 35%, 29%, and 6%, respectively. The dominant $(2 \times 1)$ admolecule structure consists of rows of neighbouring admolecules that are separated by rows of unoccupied surface sites. Nominally 1/4 of the bilayer surface sites are occupied, which prevents a three-fold surface coordination. Bond formation among admolecules takes place and reduces the nominal $(2 \times 1)$ symmetry along the rows. The dimer and trimer models, shown in Fig. 2, match the observed surface occupation and nominal periodicity. Other models with $(2 \times 1)$ periodicity were found to be less favourable. The dimer and trimer models are both admolecule terminations of low surface energy [cf. Fig. 2(d)]. The clear energetic preference for the dimer model by 0.2 meV/Å$^2$ in the unstrained case is lost under compressive strain of 1.7%. Note, that the trimer model consists of trimers with different proton arrangement and hence different formation energy. This suggest that the optimal arrangement should consist of a mixture of dimers and trimers.

Although longer linear clusters possess lower formation energy per H$_2$O, due to larger admolecule occupation of the surface, admolecule structures composed of such clusters do not have a lower surface energy. This is seen for a model of parallel infinite linear chains as shown in Fig. 2(c). In this model all surface sites are occupied by admolecules. Without strain the chain model and the dimer model have equivalent surface energies. Upon compressive strain, however, the dimer and trimer models are energetically preferred over the linear chain model. The surface energy of the investigated admolecule terminations is larger than that of the Fletcher-striped surface by a small margin. However, as will be seen below, in thin films energetic stabilisation takes place.

---

Figure 1. Admolecule clusters on the bilayer-terminated ice surface: geometry of tetramers and a hexamer (a) linear tetramer, (b) star-like tetramer, (c) ring-like tetramer, and (d) hexamer – note the strong distortion due to an optimisation of surface bonds. (e) formation energy of admolecule clusters for unstrained and strained conditions.
Figure 2. Admolecule reconstruction (a) $(2 \times 2)$ dimer model, (b) nominal $(2 \times 1)$ trimer model with the unit cell vectors indicated, and (c) linear chain model; (d) excess surface energy of these admolecule structures.

4 Hexagon Adrows Decorated with Admolecules

The analysis of STM experiments\(^6\) identified hexagon adrows of varying width and distance decorated with admolecules as another motif of the observed admolecule structures. In this section hexagon adrows and their interaction with admolecules is studied as a key element of surface termination. On the Fletcher-striped surface two row orientations with respect to the stripes of dangling OH groups are distinguished: (i) adrows running parallel to the stripes and (ii) intersecting adrows (cf. Fig. 3). Furthermore, hexagon adrows are characterised by two distinct edges: (i) one with the edge water molecules being located atop a surface site (edge E3) and (ii) one with the edge molecules being situated over subsurface sites (edge E2). Fig. 3 shows that edge molecules at E2 strongly relax towards the surface to form bonds with the surface and to obtain three-fold coordination.

The orientation of water molecules in the adrow is conditioned by the surface proton order underneath. Obeying constraints the orientation has to be optimised such as to reduce the electrostatic repulsion among the dangling OH groups or flat-lying molecules. This was done for repeated models of intersecting and parallel adrows with a width of

Figure 3. Hexagon adrows on the bilayer-terminated Fletcher-striped ice surface. (a) Possible orientations of the adrows with respect to the Fletcher stripes of dangling protons. (b) Top view of the intersecting adrow. (c)-(e) side view of intersecting and parallel adrows. The edges E2 and E3 with distinct surface bonding of the edge water molecules are indicated. Formation energy of adrows in the unstrained and strained cases (f) per water molecule and (g) per unit length.
one hexagon as shown in Fig. 3. For the parallel orientation purely O- and H-terminated adrows are most favourable (Hex\(_{O}\) and Hex\(_{H}\), respectively). An alternating one was obtained for the intersecting adrow (Hex\(_{I}\)). These models were extended to adrows with a width of up to 4 hexagons. Calculation of the formation energy per H\(_2\)O [cf. Fig 3(f)] indeed shows that the hexagon adrows have lower formation energy than the admolecule terminations discussed above. The energetically more favourable hexagon adrows are the intersecting ones. With increasing row width the energy decreases rapidly. Considering the hexagon adrow as a stripe cut from the perfect surface bilayer, its formation energy has contributions from the loss of coordination at the edges E2 and E3 and the strain exerted by the relaxation of the edge E2. With increasing width the contribution of the edges tends to zero. Fig. 3(g) demonstrates that the formation energy per unit length therefore reflects a much weaker dependence on the row width. Compressive strain reduces the formation energy to a large extent (~9 meV/H\(_2\)O). The non-negligible effect amounts to 30-40% of the formation energy for the smallest row.

Interaction of admolecules with an adrow follows distinct mechanisms at the edges E2 and E3 owing to different bonding and available surface sites. At the edge E3, edge molecules bind directly to the surface and two adjacent unoccupied surface sites are within second neighbour distance. In the most favourable configuration the admolecule attaches to these sites and the edge molecule as depicted in Fig. 4. The two-fold surface coordination of the admolecule implies that only every second edge water can be decorated in this way. The change of the adrow formation energy induced by the decoration is plotted in Fig. 4. Typically the formation energy increases due to the decoration, although, the amount drops for wider rows. Under compressive strain, however, admolecule decoration can stabilise the adrows Hex\(_{I}\) and Hex\(_{H}\) if the width exceeds 2 or 3 hexagons.

The decoration of the edge E2 is more subtle. There, an admolecules cannot interact with the edge without breaking a surface bond. Single admolecules form one bond with the surface and one with the edge by which, however, energy is gained by the admolecule.

![Figure 4. Hexagon adrows decorated with admolecules. Edge E3: (a) change in the formation energy \(\Delta e_f\) per unit length due to the admolecule. (b) admolecule attached to a two hexagons wide adrow with an H-terminated edge E3. Red and blue balls represent oxygen in the lower and upper position in the hexagon adrow. The green arrow displays the unit vector along the row. Edge E2: (c) formation energy per unit length of adrows decorated with different admolecules at E2 and one admolecule at edge E3. Admolecule configurations at edge E2 with periodicity of 2 unit vectors along the row: (d) intersecting dimer, (e) pair, (f) trimer, and (g) triangle.](image-url)
Interaction with additional admolecules may lead to more stable configurations. Among the tested configurations those shown in Fig. 4 are the energetically most favourable ones. They comprise a dimer attached at an angle, an admolecule pair, a trimmer, and a triangle. Two hexagons wide intersecting and H-terminated parallel adrows (Hex$_{2I}$ and Hex$_{2H}$, respectively) were decorated. Inspired by the STM experiments and due to the conformational constraints a periodicity of two unit vectors along the row was imposed. Rows decorated at edge E2 are characterised by a higher formation energy than the undecorated ones. For Hex$_{2H}$ on unstrained ice the most stable configurations are the admolecule pair and the trimmer, whereas the dimer and the triangle are metastable as compared to admolecule clusters at the surface. At Hex$_{2I}$, in contrast, pair, trimmer, and triangle configurations are equally stable with only the dimer one being metastable. Under compressive strain, however, the energetic hierarchy changes. At the H-terminated row the dimer is stabilised whereas the pair becomes metastable. At the intersecting row both the pair and dimer configuration are metastable. Here, the most stable configurations are the trimmer and triangle ones. The latter configuration can be obtained from the adrow Hex$_{3I}$ by removing every second edge molecule at E2. Decoration of Hex$_{3I}$ at E2 goes along with an increase in energy until the next wider row is completed. A similar argument holds for Hex$_{3H}$. Notably, parallel and intersecting adrows decorated at the edge E2 are energetically equivalent. A clear energetic preference for intersecting adrows is lost. Calculations show that the above conclusions are valid for wider adrows. The STM experiments$^6$ demonstrated that annealing of ice structures on Cu(111) below 145 K leads to the formation of decorated hexagon adrows with an admolecule $(2 \times 1)$ structure in between adrows. Hexagon adrows with a width of two hexagons and separation of 2-6 hexagons are not uncommon motifs. Our findings for the favourable admolecule configurations nicely agree with the experimental analysis.

## 5 Energetic Stabilisation of Admolecule Terminations

The present investigation demonstrated that the admolecule structures and decorated hexagon adrows indeed are low energy terminations of the (0001) surface of hexagonal ice. Nevertheless, the surface energy in all cases exceeds that of the Fletcher-striped termination. The latter marks the lower bound of the surface energy of the bilayer-terminated surface$^{3,4,12}$. Estimates of the effect of proton disorder indicate that the surface energy should fall in a range of 4 meV/Å$^2$ above that of the Fletcher-striped surface. The excess surface energies of admolecule terminations as reported above varies between 0.91 and 2.0 meV/Å and thus is located in this range as well. Nevertheless, the energetic preference for the bilayer-terminated (0001) surface is in agreement with experimental findings for thick ice layers$^{1,5}$. The STM analysis$^6$ of thin crystalline ice structures on Cu(111) demonstrates, on the other hand, the preference for admolecule terminations.

An explanation of the apparent contradiction can be obtained by revisiting the expressions for the surface energy. The general assumption here is that equilibrium can be established with the source of admolecules, although some kinetic process may not have been thermally activated during annealing or growth implying restrictions regarding the formation of some surface terminations. In the previous sections a surface of bulk ice was considered with the reference energy per molecule $\mu_{\text{H}_2\text{O}}$ corresponding to hexagonal ice. In a thin ice film or due to poor crystal quality the value may deviate from the bulk value. For instance for a free standing ice film of two bilayers this deviation $\Delta \mu_{\text{H}_2\text{O}}$ amounts to 90 meV. In a thin film on Cu(111) one ice-vacuum interface is replaced by an ice-metal
interface with the corresponding change in the interface energy. Estimates by DFT calculations\textsuperscript{13} amount to 16 meV/H\textsubscript{2}O. If the excess surface energy per admolecule is lower than $\Delta\mu\text{H}_2\text{O}$, the ice film/structure will break up. Molecules diffuse onto the surface and form admolecule terminations. With increasing number of bilayers in the ice structure the average energy approaches the bulk value and admolecule terminations will become unstable.

The energetic argument is best illustrated by a plot of the excess surface energies of different admolecule terminations vs. $\Delta\mu\text{H}_2\text{O}$ shown in Fig. 5. Here in addition to the models $(2 \times 1)$ admolecule structure also two models of a adrows Hex\textsubscript{2H} decorated with $(2 \times 1)$ admolecule structures inspired by the experimental findings by Mehlhorn and Morgenstern\textsuperscript{6} are considered. The plot clearly demonstrates the stabilisation of admolecule structures for realistic values of $\Delta\mu\text{H}_2\text{O}$. In the experiments an evolution of the surface termination from admolecule via decorated hexagon adrows to pyramidal islands was observed with annealing stages at 130, below 145 K and at 145 K. The present analysis suggests that at the different stages the formation of the subsequently more stable admolecule terminations is kinetically limited. Note, a similar argument based on adsorption energies of water bilayer and 3D clusters helped to understand the water wetability of close-packed metal surface\textsuperscript{14}.

6 Concluding Remarks

In the present work admolecule terminations of the (0001) surface of hexagonal ice were investigated in the framework of density functional theory. The structural analysis revealed the building blocks of admolecule structures, the physics of hexagon adrows and their decoration with admolecules. Such structures constitute low-energy surface terminations. Energetically driven deviation from the proton-ordered bilayer termination is predicted for thin ice films.

Acknowledgements

This work is part of a joint project with the groups of K. Morgenstern (U. Bochum) and U. Bovensiepen (U. Duisburg-Essen) on the Interaction of organic molecules with long-
living solvated electrons at the ice surface funded by the Deutsche Forschungsgemeinschaft (BO1851/3). A generous grant of computer time from the John von Neumann Institute for Computing, Jülich (HER14) is gratefully acknowledged. Complementary calculations were carried out at the HPC-cluster at the RRZE, FAU Erlangen-Nürnberg.

References

Atomistic Modelling of Zr-Segregation in Symmetric Grain Boundaries of Molybdenum

Olena Lenchuk, Jochen Rohrer, and Karsten Albe

Technische Universität Darmstadt, Institut für Materialwissenschaft, Fachgebiet Materialmodellierung, Petersenstr. 32, 64287 Darmstadt, Germany
E-mail: {lenchuk, rohrer, albe}@mm.tu-darmstadt.de

The influence of Zr segregation on the cohesive strength of grain boundaries in molybdenum is studied by means of density-functional theory (DFT) calculations for the case of symmetric tilt and twist boundaries in bi-crystal geometry. First, the driving force for Zr segregation to the grain boundaries is investigated. Secondly, the influence of Zr on the GBs cohesive strength is studied with respect to energy and stress. The results reveal that the presence of Zr leads to reduced cohesive strength of the GBs in Mo and therefore provides suitable conditions for intergranular fracture to occur.

1 Introduction

Improving the efficiency, performance, and cost-effectiveness of energy production will depend critically on our capability to develop substantially improved materials and material systems as well as on our ability to understand and to design their properties. Turbine blades, for example, working inside the hottest part of the gas turbine are exposed to significant environmental and mechanical stresses and thus need to be resistant to creep, fatigue and oxidation. Nickel-based superalloys are traditionally used to sustain the most stressed parts of turbine blades, but have a limited service temperature (around 1000 °C).1, 2

Applying materials with higher melting points in turbines would allow to operate these components with increased gas inlet temperatures and to optimise the thermodynamic behaviour leading to a reduction of fuel consumption, emissions and costs. Hence, a lot of research has focused on Mo-Si-B alloys, since they exceed the melting point of Ni-based superalloys resulting in higher operating temperatures. Their low thermal expansion coefficients and high thermal conductivity are desired for turbine blades to achieve excellent fatigue properties.3–5. Mo-Si-B alloys are of scientific interest also because of their high-temperature strength and creep resistant.3, 6. Besides, a protective borosilicate glass layer is self-forming during the exposure at high temperatures,7, 8 making this material oxidation resistant. However, to meet the requirements for the desired applications, fracture toughness and ductility at ambient temperatures need to be improved1, 9–11.

The desired fracture toughness and ductility of the Mo-Si-B alloys can be achieved by strengthening the grain boundaries (GBs) of the Mo phase. Zr microalloying was identified as a useful approach to improve the ductility and to increase the strength of both types of materials.8, 12, 13. However, the physical origin of this effect is not clear. At this point, total energy calculations based on density-functional theory (DFT) are useful for obtaining a detailed understanding of the energetics on the atomic scale.

Using grants of computer time at the John von Neumann Institute for Computing (NIC) we have performed a study of the alloying effect of Zr in GBs of molybdenum. After a brief overview of the employed computational methods, the effect of Zr alloying on the
cohesive strength of the GBs using energy and stress approaches is studied. We conclude our discussion with a brief summary.

2 Computational Method

Our calculations are carried out in the framework of the density-functional theory (DFT)\textsuperscript{14,15} using the program package \textit{vasp}\textsuperscript{16,17}. The GGA-PBE\textsuperscript{18} functional for describing exchange-correlation interactions is employed and Blöch’s projector augmented wave (PAW) method\textsuperscript{19} to define pseudopotentials. Two different symmetric grain boundaries are studied in bi-crystal geometry:

The twist $\Sigma 5[001]$ grain boundary (GB) is modelled by a slab of 30 layers (150 atoms) that are separated by a 15 Å vacuum layer. The tilt $\Sigma 5(310)[001]$ GB is represented in a periodic supercell of 160 atoms and consists of two equivalent GBs. The calculations are done using a planewave energy-cutoff of 450 eV. A $7 \times 7 \times 1$ k-point sampling for the slab model and a $2 \times 6 \times 8$ k-point sampling for the periodic supercell model is used. Atoms are relaxed while keeping the cell parameters fixed until the maximum force acting on each of them was less than 0.05 eV/Å\textsuperscript{2}.

3 Results

3.1 Solute Segregation

Fig. 1 illustrates the structural models of the (a) twist $\Sigma 5[001]$ and the (b) tilt $\Sigma 5(310)[001]$ model grain boundaries (GBs). These high-angle GBs are constructed by twisting (tilting) the bcc Mo cell by an angle of 36.9° C around the [001] axis. Non-equivalent lattice sites within the GB are marked by “2” and “3”, whereas coincident sites are marked as “1”.

On each non-equivalent lattice site within the GB, we replace one Mo atom by Zr and optimise the structure. By comparing the heat of formation of Zr in the GB relative to Zr in the bulk, we determine a thermodynamic driving force for segregation. The heat of formation is defined as:

$$H_f = E_{GB+y\cdot Zr} - E_{GB} - y \cdot \epsilon_{MoZr} + (1 + 2y) \cdot \epsilon_{Mo}.$$  \hfill (1)

Here, $E_{GB+y\cdot Zr}$ and $E_{GB}$ are the total energies of the GB model with and without Zr, respectively. $MoZr$ is used as a reference state for Zr. It means that the strength of the reservoir for Zr is determined by the chemical potential of Zr in the next stable phase that appears in the Mo-Zr phase diagram\textsuperscript{20}. The total energy of bulk molybdenum per Mo atom is represented by $\epsilon_{Mo}$.

Comparing the heat of formation for Zr occupying a non-equivalent lattice site within the GB relative to a position in the bulk of Mo reveals the existence of a strong driving force for GB segregation. Tab. 1 summarises the corresponding changes in the heat of formation. It should be pointed out that the heat of formation in Mo bulk equals $H_{Mo}^{0Zr} = 0.27$ eV. Thermodynamically stable positions for Zr are “1” for tilt and “2” for twist configurations, respectively, (see Fig. 1). Zr being an oversized additive prefers to occupy sites with a sufficient distance to the nearest neighbours ($\approx 2.8\text{--}3.0$ Å). The other sites at the tilt GB,
namely “2” and “3”, as well as coincidence site (“1”) at the twist bi-crystal, have more dense atomic environment and therefore can not provide enough volume for Zr to accommodate itself.

Our results reveal that Zr shows similar behaviour for both GB orientations and therefore the driving force for segregation seems to be rather insensitive to the GB orientation.

<table>
<thead>
<tr>
<th></th>
<th>Twist $\Sigma_5[001]$</th>
<th></th>
<th>Tilt $\Sigma_5[310][001]$</th>
<th></th>
<th>Mo(001)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zr</td>
<td>+0.15</td>
<td>-0.45</td>
<td>-0.51</td>
<td>-0.17</td>
<td>-0.15</td>
</tr>
</tbody>
</table>

Table 1. Change in the formation enthalpy $H_f$ of Zr segregating at the GB relative to Mo bulk ($H_f^{bulk} = 0.27$ eV). Reduction in energy corresponds to negative values. All numbers are given in eV.

3.2 Effect on GB Cohesive Strength

Zr preferring to segregate to Mo GBs, as defined in the last section, has an impact on the GB cohesive strength changing the bonds strength at the interface. The GB cohesive strength can be quantified by means of work of separation, according to:

$$W_{separation}^{twist}(\infty) = (E_S + E'_S - E_{GB}(d_0))/A.$$  \hspace{1cm} (2)

Here $E_{GB}(d_0)$ is the total energy of the GB (with or without additive) at the equilibrium grain separation $d_0$, $E_S$ and $E'_S$ are the total energies of the two surfaces created by cleaving the GB into two parts. In the case of alloyed GB, the surface with a substitutional additive
Table 2. Work of separation of the pure and alloyed twist $\Sigma 5 \{001\}$ and tilt $\Sigma 5 \{310\} \{001\}$ Mo GBs. Zr is inserted at the GB substitutionally and occupy favourable sites. O is added interstitially. All values have the units of J/m$^2$.

<table>
<thead>
<tr>
<th></th>
<th>Pure GB</th>
<th>1Zr</th>
<th>2Zr</th>
<th>4Zr</th>
<th>8Zr</th>
<th>Zr+O</th>
<th>Zr+2O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twist</td>
<td>4.06</td>
<td>3.82</td>
<td>3.56</td>
<td>-</td>
<td>2.22</td>
<td>3.38</td>
<td>2.98</td>
</tr>
<tr>
<td>Tilt</td>
<td>4.44</td>
<td>4.26</td>
<td>-</td>
<td>3.96</td>
<td>3.80</td>
<td>4.32</td>
<td>3.96</td>
</tr>
</tbody>
</table>

is marked as $S'$ while $S$ represents pure molybdenum surface. In the case where additives are not present, $S'$ is identical to $S$. The total area of the GB in a supercell corresponds to $A$.

Tab. 2 lists the work of separation $W_{sep}$, required to split the pure and alloyed GB in two surfaces with their further structural relaxation. We cleave the twist GB in the centre between grains. The fracture plane is marked as black dashed line in Fig. 2a, while the close-up of the produced surfaces is shown on Fig. 2b, c. By cleaving the tilt GB we construct several possible surfaces and chose one that requires the lowest $W_{sep}$ as energetically more preferable. To split a pure twist Mo GB, one needs to apply 4.06 J/m$^2$ while for pure tilt 4.44 J/m$^2$ is required.

With Zr in the twist GB, the work of separation $W_{sep}$ decreases relative to the pure GB. This reduction is proportional to the amount of Zr atoms and is equal to 0.24 J/m$^2$ ($\sim 6\%$) for one Zr atom and 1.84 J/m$^2$ ($\sim 45\%$) for 8 Zr atoms (in this case, zirconium occupies all favourable sites within the GB). The atomic structure of the fully relaxed pure GB containing 8 Zr atoms is shown on Fig. 2a. The GB is viewed along the $[100]$ axis. The separation between grains for relaxed GB increases by 35\% compared to pure one weakening the bonds through the interface and providing suitable conditions for intergranular fracture to occur.

Segregating to the grain boundary with tilt geometry, Zr shows the same behaviour as for twist but the decay in $W_{sep}$ is less dramatic. The work required to separate the

Figure 2. Atomic structure of the relaxed twist $\Sigma 5 \{001\}$ Mo GB containing 1 Zr atom occupying favourable site at the GB. (b), (c) represent relaxed surfaces created by cleaving the GB. Colour coding: see Fig. 1. The fracture plane is marked as dashed black line.
corresponding GB decreases continually with increasing Zr content. Segregating at the GB, one Zr atom reduces the work of separation by 0.18 J/m$^2$ ($\sim 4\%$), whereas 8 Zr atoms by 0.64 J/m$^2$ ($\sim 14\%$) corresponding to the occupation of all favourable sites at the GB. The decrease in the work required to be done to separate the grain boundary evidencing that the energy released during structural relaxation will facilitate the fracture of the GB$^{21}$. Zr behaves as a weak embrittler of Mo and is not sensitive to the GB orientation. For an oversized segregant the only way to accommodate itself at the GB is to push the grains apart, whereas at a free surface it may protrude into vacuum without significantly disturbing its surrounding atoms$^{22}$. The same picture we see in the present study.

To investigate the effect Zr on the cohesive strength of Mo GBs in the presence of oxygen, we insert one and two oxygen atoms interstitially at different sites, choose energetically more favourable positions and define the work of separation by splitting the boundary. All additives occupy the lowest-energy positions. Independent of the GB orientation, in the case when both impurities are present at the GB the work of separation decreases compare to case when only one of them (either Zr or O) is present. It should be pointed out, that the tilt GB contain more atoms at the interface compared to twist and therefore the obtained results are not as dramatic.

3.3 Stress Approach

The insight how additives can embrittle a solid susceptible to crack growth through cleavage, is offered by stretching the grains apart$^{23}$. In this way, we set bounds on the maximum cohesive strength of molybdenum and show in which manner Zr effects the cohesive strength of the GB leaving the energy contributions aside. We uniformly stretch the twist $\Sigma 5[001]$ Mo GB (pure GB and one containing Zr atom) in the $<001>$ direction along a pair of adjacent planes that create a boundary. We again compute a work of separation $W_{\text{sep}}$ between grains as a function of grains separation $d$ according to:

$$W_{\text{sep}}(d) = (E(d) - E_S - E_{S'})/A,$$  

where $E(d)$ is the total energy of the GB at separation distance $d$, but now in the calculation we do not perform any relaxation of the atoms within the unit cell. The obtained data is fitted to the universal binding-energy relationship for crystals (UBER$^{24}$). The maximum slope of this curve corresponds to the maximum stresses the system can withstand. As soon as Zr enters the system, the GB strength decreases and the slope of the curve is not as steep as in the case of pure Mo GB.

Fig. 3 shows a traction curve obtained from the derivative of binding energy-separation curve. It describes the stresses acting on the faces of the crack. The theoretical cohesive strength of the GB $\sigma_{th}$ corresponds to a maximal stress and equals 27 GPa for pure twist $\Sigma 5[001]$ Mo GB. For the GB alloyed with Zr, $\sigma_{th}$ decreases to 23 GPa.

4 Concluding Remarks

In this contribution, we have presented calculations of Zr segregation at the GBs in molybdenum based on density-functional theory (DFT). Our results reveal the driving force of Zr to segregate to Mo boundaries and its effect on the GBs cohesive strength. Specifically, we
have shown that Zr segregation to the GBs is energetically favourable. On the other hand, not all lattice sites at the GBs are favourable, since Zr being an oversized additive needs enough free volume to accommodate itself within the GBs. The decrease in the cohesive strength of the GBs in the presence of Zr, points to the more complicated effect of Zr on the Mo properties and is the subject of further investigations.

Acknowledgements

The project is supported by the German Research Foundation (DFG) in the frame of the research unit 727 “Beyond Ni-based Superalloys”. The calculations described here were performed with a grant of computer time provided by the John von Neumann Institute for Computing (NIC) in Jülich, which is gratefully acknowledged.

References


Computational
Soft Matter Science
Soft Matter is a rather general term for a huge class of liquids and solids which neither fall into the class of simple liquids nor into the class of hard solids as metals or other anorganic materials are commonly called. As such soft matter includes biological and synthetic materials ranging from oil or biological liquids such as blood to everyday plastics. Thus soft matter, in more technical terms, includes (bio-) polymers, emulsions (e.g. food), surfactants, liquid crystals, gels etc. They are pervasive and affect our daily life in many respects. Soft matter typically is structured on a nanometre scale and does not display long-range order. Consequently analytical theory provides a rather generic, yet very important understanding, while more specific questions rely on extensive computer simulations for a better understanding of systems and processes. Along this line three very interesting contributions are presented in this year’s volume of the NIC Proceedings which deal with rather different topics.

The first contribution by S. Frijters, F. Günther and J. Harting deals with a special kind of emulsions. Emulsions contain at least two different liquids which are immiscible. These two liquids are intermixed and the minority phase is finely dispersed in the majority phase. Close to equal concentration of both components, so called bicontinuous phases (bijel) are also possible. Because of the high surface tension between the incompatible liquids, the two phases segregate into macroscopic domains. To prevent this from happening additional components, which stabilise the surfaces of the droplets, have to be added. One example of such emulsifiers are small solid particles, which aggregate at the surfaces of the droplets, creating so called Pickering emulsions. Frijters et al. systematically study such systems using Lattice Boltzmann (LB) simulations, where the LB simulation of the liquids is coupled to a particle based simulation of the stabilising particles. This approach combines several advantages. The methodology can be parallelised efficiently, which enables a very efficient use of the Blue Gene/Q systems. LB simulations include hydrodynamics interactions which are relevant for intermediate states during structure formation and dynamics in general, as was studied here. This allowed a systematic study of the dynamic transition between bijel and Pickering phases. Finally, anisotropic particle emulsifiers have been investigated too.

The second contribution by D. Winter, P. Chaudhuri and J. Horbach deals with micro rheological properties of simple glass formers. Despite tremendous theoretical and experimental efforts, the rheology of glass forming materials and the glass transition itself are by far not understood. Thus it is still very appropriate to study such problems on the basis of extremely simple model systems. One such model system is a mixture of Yukawa particles, i.e. a mixture of particles interacting with a repulsive screened Coulomb potential. Besides their simplicity, the beauty of these systems comes from the fact that they are quite good models for charge stabilised colloids, which can directly be observed with visible light by
modern experimental techniques. Then these observed caging effects for particle motion or whole trajectories as well as their impact on the surroundings can be investigated and also compared to experiment. Non equilibrium molecular dynamics (NEMD) are the perfect tool to address such problems. Typical trajectories of particle which are pulled - as also could be done by optical tweezer techniques experimentally - show a characteristic step-like motion where the time a particle needs before leaving a cage correlates with the temperature. In another set of simulations, the flow of a glass forming Yukawa liquid close to a solid wall is studied, displaying characteristic shear banding.

Finally M. Müller, D. Sun and U. Welling deal with a polymer physics problem which lies at the heart of many modern technological applications, namely the direction of self-assembly processes of block-copolymers in both bulk and especially on structured surfaces. Such simulations often suffer from the very long timescales needed for structure formation of such systems. Consequently the simulated systems usually are not that large, chains are short and times cannot cover material transport beyond scales that correspond to the order of the chain size. Müller and coworkers employ a new, advanced simulation approach where the individual chain does not interact not any longer explicitly with the individual beads of its neighbours, but with a density field created by them. Though this affects the microscopic dynamics, the overall dynamics of pattern formation is expected to be reproduced on time scales above the diffusion time of the individual polymers, at least for polymers, which are not highly entangled. This makes the simulation extremely fast and is also quite easy to parallelise. By this the structure formation on (patterned) surfaces as well as the structure formation in bulk subject to shear forces or a pressure quench has been studied. The authors clearly demonstrate the relevance of such external parameters on the pattern forming process. Since very long times can be reached, such simulations have all the potential to be employed in parallel to experiments and to guide experimental investigations.

The three above mentioned examples impressively demonstrate the power of modern computer simulation methods for our overarching goal to better understand soft matter systems, independently of the fact whether of man made synthetic or biological origin, and to eventually arrive at better and new materials.
Dynamic Phase Transitions and Time Dependent Structuring Effects in Particle-Stabilised Emulsions

Stefan Frijters\(^1\), Florian Günther\(^1\), and Jens Harting\(^1,2\)

\(^1\) Department of Applied Physics, Eindhoven University of Technology, Den Dolech 2, 5600MB Eindhoven, The Netherlands
\(^2\) Institute for Computational Physics, University of Stuttgart, Allmandring 3, 70569 Stuttgart, Germany

Particle-stabilised emulsions are commonly used in various industrial applications, but their fundamental properties remain poorly understood. Computer simulations are well-suited to probe properties of emulsions that are hard to access experimentally. We use the lattice Boltzmann (LB) method with some extensions to simulate multiple fluids and the suspended particles. We review two examples of our recent research on particle-stabilised emulsions. First, we showcase the possibility of changing an emulsion from a Pickering emulsion to a so-called bijel or vice versa, by changing in time the wettability of the particles. When Pickering emulsions are generated, the droplet size distribution is possibly bimodal, which can be observed in our simulations. Second, we demonstrate the effect of anisotropy of the ellipsoidal particles on the formation of an emulsion. Compared to emulsions stabilised by spherical particles, additional timescales are observed, owing to the extra degrees of freedom introduced by the difference in length of the different axes.

1 Introduction

Particle-stabilised emulsions differ from traditional surfactant-stabilised emulsions in some crucial ways: Firstly, the particles in these mixtures block Ostwald ripening, which allows for long-term stabilisation of these emulsions. Secondly, they can be present in different forms, such as the Pickering emulsion (suspended droplets of the minority phase in the majority phase) or the bijel (two bicontinuous fluid domains of the two phases), which have different characteristic topology and rheology. Various parameters, such as the ratio between the two fluid phases, volume fraction of the stabilising particles and the surface chemistry of the particles can affect which emulsion state forms\(^1,2\). In general, the stabilising particles can be of any shape, but spheres and ellipsoids are of particular interest, for their relative simplicity (spheres) and added degrees of freedom (ellipsoids).

Computer simulations are well-suited to probe properties of emulsions that are hard to access experimentally. We use the lattice Boltzmann (LB) method to simulate the fluid phases in our simulations\(^3\). Furthermore, we use the Shan-Chen multicomponent model to treat the fluid-fluid interactions\(^4\), and we use a variant of Ladd’s method to couple finite-sized particles to the fluids\(^5,6\). Particle-particle interactions and particle dynamics are modelled by a molecular dynamics algorithm. The code has recently been expanded to allow for ellipsoidal particles as well as spherical ones\(^2\). For a more detailed description of our simulation methods the reader is referred to\(^7,8\). The LB method lends itself very well to implementation and execution on massively parallel architectures, and our simulation code...
LB3D is able to efficiently leverage hundreds of thousands of CPUs. A typical simulation of an emulsion might use 8192 to 32768 CPUs for a week on the Blue Gene/Q system “JUQUEEN”. An open-source version of LB3D, providing a subset of the functionality described here, is available at http://mtp.phys.tue.nl/lb3d/ under the LGPL version 3 license.

Here, we show two examples of our recent research on particle-stabilised emulsions. First, we consider the possibility of fostering a transition from a Pickering emulsion to a bijel or vice versa, by changing in time the particle wettability. When Pickering emulsions are generated, the droplet size distribution is possibly bimodal, which can be observed in our simulations. Second, we demonstrate the effect of anisotropy of the ellipsoidal particles. Compared to emulsions stabilised by spherical particles, additional timescales are observed, owing to the extra degrees of freedom introduced by the difference in length of the different axes.

2 Induced Transitions of Particle-Stabilised Emulsions

The work presented here expands on the work done by Jansen et al., who studied the formation of emulsions and the time evolution of their domain sizes. This study is now taken a step further; in particular we consider the evolution of domain sizes in time, combined with the possibility of a fostered transition from a bijel to a Pickering emulsion. Such a triggered transition could have applications in the development of new filter technologies or self-assembled functional materials. Preliminary results suggest that inducing a phase transition is indeed possible, although this has now only been observed for colloidal properties which may be hard to obtain in experiments. This is because sweeping the parameter space is costly, so extreme values are chosen to check whether the looked-for transition can occur at all. Further simulations with less extreme parameters are planned for the near future. We simulate the equilibration of emulsions from their initial state of a random mixture of fluids and particles. The initial conditions determine the phase the emulsion will evolve into. Ignoring other effects, from geometrical arguments one can see that, for the case of an oil-in-water emulsion, hydrophilic particles favour a Pickering emulsion, while hydrophobic particles favour a bijel configuration. Once either phase is reached and the domain sizes are no longer changing in time, the wettability of the stabilising particles is allowed to linearly change in time, changing the particles from being hydrophobic to hydrophilic or vice versa. The wettability of the particles is expressed through the contact angle $\theta$, with $\theta > 90^\circ$ being hydrophobic and $\theta < 90^\circ$ being hydrophilic. Although from the evolution of systems with either wettability one might expect the emulsion to change its phase, there is a large energy barrier associated with the extensive reorganisation of the system. This introduces a strong hysteresis effect, allowing the transition only when the particles are approaching the limit of superhydrophobicity or -hydrophilicity, and not as the particle wettability crosses the border between hydrophobicity and hydrophilicity, as one might expect.

We use the reduced average cluster size to determine if an emulsion is in a bijel state or a Pickering state:

$$I_{av} = \left( \frac{\sum_{i=1}^{n_{max}} i_n n^2}{N_c} - \frac{n_{max}^2}{N_c} \right), \quad (1)$$
where $N_c$ denotes the total number of sites which are oil-dominated, $i_n$ is the number of clusters consisting of $n$ lattice sites, and $n_{\text{max}}$ is the size of the largest such cluster. All units in this article are non-dimensionalised by the lattice Boltzmann lattice constant $\Delta x$ and the corresponding timestep $\Delta t$. By subtracting the contribution of the single largest cluster in this way, we observe a very sharp decrease of this quantity when transitioning from the Pickering to bijel regime. In the case of a bijel, $I_{av} = 0$. Examples of a transition in either direction are shown in Fig. 1. Here, we use a system of $256^3$ lattice sites, and have initialised the initial configuration with random densities of oil and water at each lattice site, such that the oil:water ratio is 0.56. The surface tension of the oil-water interface is $\sigma = 0.014$. The particles are spheres with radius $r = 5.0$ and have a volume concentration of 25%. The wettability of the particles is changed in time in a piecewise linear fashion, following the values presented in the insets. Even though wettability in the second period (circles) has changed from hydrophilic to hydrophobic or vice versa, no transition is observed. When the wettability is taken to extremes, a transition is eventually induced (sharp jump / drop of the reduced average cluster size), and the system stays in its new state even when the wettability is changed to more moderate values afterwards.

![Figure 1](image-url)  
Figure 1. Transition from a bijel state to a Pickering emulsion (top) and vice versa (bottom). The pictures on the left and right show examples of an initial and final state of the system. $I_{av}$ is a measure of the domain sizes in the system and is identically zero for a bijel state. The simulation starts from a random mixture and evolves in the time. The contact angle $\theta$ of the stabilising particles is changed in various time intervals (see insets). Even when the particles are changed from hydrophilic to hydrophobic or vice versa, no phase change occurs until the contact angle is varied to extremes. Returning to more moderate values then retains the current phase.
With the data already in place, it is trivial to consider the size distribution of the domains. For a bijel state, there is only one domain of the minority phase and as such the exercise is not very interesting. However, many droplets are present in a representative equilibrated Pickering emulsion and they show a rather broad distribution. An example is shown in Fig. 2. To get better statistics, the system size has been increased to $512^3$ as compared to the previous simulations, with all other parameters being kept constant. As the droplets in the emulsion are approximately spherical, it is natural to measure their size by the radius a droplet with the same volume would have, if it were perfectly spherical. In the histogram, two peaks are visible, which might be indicative of a bimodal droplet distribution also observed in experiments\textsuperscript{10}. More data is currently being accumulated to support this statement.

![Histogram of the domain size distribution in a Pickering emulsion. Two peaks are visible, which might be indicative of a bimodal droplet distribution also observed in experiments\textsuperscript{10}.](image)

Figure 2. Histogram of the domain size distribution in a Pickering emulsion. Two peaks are visible, which might be indicative of a bimodal droplet distribution also observed in experiments\textsuperscript{10}.

3 Timescales in Emulsions Stabilised by Anisotropic Particles

In this section we discuss the influence of the particle anisotropy on the time dependence of the formation of Pickering emulsions. We use an elongated ellipsoid with an aspect ratio $m = R_\parallel/R_\perp = 2$ ($R_\parallel = 8$ and $R_\perp = 4$) as a realisation of an anisotropic particle. Similar to the previous section we use a system with $V_S = 256^3$ and $\sigma = 0.014$. The volume concentration, contact angle and fluid ratio are chosen as $C = 0.24$, $\theta = 90^\circ$ and $5:2$, respectively. As an alternative to Eq. 1, the average size of fluid domains can be determined by measuring

$$L = \frac{1}{3} \sum_{i=x,y,z} L_i,$$

where $L_i = \frac{2\pi}{\sqrt{\langle k^4_i(t) \rangle}}$ is the average domain size in direction $i$. \(\langle k^4_i(t) \rangle = \sum_k k^i(t) \zeta(k,t)/\sum_k k^4_i(t)\) is the second-order moment of the three-dimensional struc-
Figure 3. Pickering emulsion: (a) Time development of the domain size $L$ (see Eq. 2) for $m = 1$ and $m = 2$. At first view, a steady state is reached after some $10^5$ timesteps. (b) Zoom for $m = 2$: Time development of the domain sizes (second timescale). Particle anisotropy leads to a more complex time development of $L$.

ture function $\varsigma(\vec{k}, t) = \langle 1/\varsigma_n \rangle |\varphi'_k(t)|$. $\varphi' = \tilde{\varphi} - \langle \tilde{\varphi} \rangle$ is the fluctuation of $\tilde{\varphi}$ which is the Fourier transform of the order parameter $\varphi = \rho^r - \rho^b$. The time development of $L$ is shown for different particle aspect ratios ($m = 2$ and $m = 1$) for a Pickering emulsion in Fig. 3(a). After an initial formation of droplets, their growth is driven by Ostwald ripening. Finally, further growth is only possible by the coalescence of the droplets. When two droplets unify, the area coverage fraction of the particles is increased because the surface area of the new droplet is smaller than that of the two smaller droplets before. At some point the area coverage fraction of the particles is high enough to stop further coalescence. The state which is reached at that time is (at least kinetically) stabilised and we obtain a stable emulsion. It seems that $L$ reaches the steady state after a few $10^5$ timesteps for both types of emulsions. However, if we zoom in we see that $L$ develops for a longer time period if the particles have a non-spherical shape. The reason for this phenomenon is the additional degree of freedom coming from the particle orientation. If a particle changes its orientation as compared to the interface or a neighbouring particle this generally changes the interface shape. In this way the domain sizes are influenced, leading to changes of $L$. Fig. 3(b) shows a zoom-in of the time development of $L$ and Pickering emulsions with $m = 2$, respectively. The kink in Fig. 3(b) after about 1.2 million timesteps is due to the coalescence of two droplets of the Pickering emulsion. The decay of $L$ is larger for the case of a bijel due to its larger interface.

In order to understand the properties of emulsions and especially the time development of $L$ for anisotropic particles we investigate the behaviour of some “model” systems. The simplest one is the adsorption of a single particle to a fluid-fluid interface which is treated in detail in Refs. 2, 11. Here we restrict ourselves to a model system which also allows to study the effect of particle interactions, i.e. a particle ensemble at a flat interface. A measure for global ordering effects of the particles is the orientational order parameter $S$. We define the uniaxial order parameter $S^{11.8}$ as

$$S = \frac{1}{2} \langle 3 \cos^2 \vartheta - 1 \rangle,$$

(3)
where \( \langle \rangle \) denotes the averaging over particles. Originally, \( S \) is an order parameter for studying liquid crystals which indicates the phase transition from the isotropic to the anisotropic/nematic phase. Here, the parameter \( S \) is used as a measure for the orientation of the particle ensemble towards the interface. If all particles are oriented orthogonal to the interface one has \( S = S_\perp = 1 \). The orientation of all particles parallel to the interface leads to \( S = S_\parallel = -0.5 \).

The local ordering effects are taken into account by a discretised form of the pair correlation function \( g(r) \) defined as

\[
g(r) = \frac{1}{2\pi g_0 N} \left\langle \sum_{i \neq j} \int_{r-\frac{1}{2}}^{r+\frac{1}{2}} \delta(\vec{r} - r_{ij}) d\vec{r} \right\rangle,
\]

where \( N \) is the number of particles, \( r \) and \( r_{ij} \) are the distance from a reference particle and the distance between the two particle centres of particle \( i \) and \( j \) in units of \( R_\parallel \), respectively, and \( g_0 \) is a normalisation factor chosen such that \( g(r) \to 1 \) for \( r \to \infty \). \( g(r) \) gives a probability to find a particle at a distance \( r \) from a reference particle. \( g(r) \) is a measure for the ordering of the particle centres and ignores the orientation.

The flat interface considered in this section is periodic in two dimensions parallel to the interface and each period has a size of \( A_I = 512^2 \). The system is confined by walls at a distance of 40 lattice nodes from the interface in the third direction. The particle coverage fraction for \( N \) particles adsorbed at the interface is defined as \( \chi(\xi, \vartheta) = \frac{N A_P(\xi, \vartheta)}{A_I} \). \( A_P(\xi, \vartheta) \) is the area which the particle would occupy in a hypothetical flat interface and depends on the distance between the particle centre and the undeformed interface and the particle orientation relative to the flat interface. In the following, we relate the coverage fraction to the case of \( \xi = 0 \) and \( \vartheta = 90^\circ \) (\( \chi_I \)) corresponding to the initial state and the final steady state\(^5\)\(^-\)\(^1\)\(^1\) for \( \vartheta = 90^\circ \). This leads to \( \chi_I = \frac{N A_{P,I}}{A_I} \) with \( A_{P,I} = \pi R_\parallel^2 \). Fig. 4(a) shows the time development of \( S \) for different values of \( \chi_I \) (\( \chi_I \approx 0.0767 \) (squares), \( \chi_I \approx 0.3835 \) (circles), \( \chi_I \approx 0.4602 \) (upward pointing triangles) and \( \chi_I \approx 0.5229 \) (downward pointing triangles)). The parameter \( S \) starts for all values of \( \chi_I \) with a value of \( S_\perp = 1 \), corresponding to the initial configuration. For lower values of \( \chi_I \), \( S \) reaches \( S_\parallel = -0.5 \), corresponding to the case that all particles have flipped towards the interface. For higher values of \( \chi_I \), the final value of the parameter is \( S_{\text{final}} > S_\parallel = -0.5 \). This corresponds to the case where some particles cannot flip completely to the equilibrium orientation because there is not sufficient space.

To characterise the local ordering effects of the particles we utilise the pair correlation function \( g(r) \) defined above. The particles have a contact angle of \( 90^\circ \), so there is no capillary interaction between them in the final state when all of them have flipped completely and the system has reached a steady state. However, there are dipolar interface deformations and thus interactions during the flipping process of the particles. Furthermore, for \( \chi > \chi_c \), capillary interactions occur due to not all particles being able to flip. Even after all particles have rotated towards the interfaces, there are still some capillary waves going through the system, leading to time dependent interactions between the particles.

The rotation of the individual particles towards the interface and the time development of \( S \) are not sufficient to explain the long-time development of \( L \) since those are completed after much shorter timescales. Thus, we study the process of the particles aligning at the interface. Fig. 4(b) shows \( g(r) \) at different moments between 10000 and 1000000
Figure 4. (a) Time development of the order parameter $S(t)$ (see Eq. 3) for $m = 2$, $\theta = 90^\circ$, $\sigma \approx 0.041$ and different values of $\chi_I$. There is an incomplete flip of particles for high values of $\chi_I$. (b) Time development of $g(r)$ for $\chi_I \approx 0.38$. The second peak is more pronounced at later timesteps. The particles reorder and the ordering increases. The reordering process is almost complete after 400000 timesteps. (c) Snapshot of a typical simulated system depicting the alignment of the particles parallel to the fluid-fluid interface.

timesteps. The first peak decreases but at later times the following peaks are more pronounced demonstrating that the degree of ordering increases. After 400000 timesteps this development is almost coming to an end and only few changes in the particle configurations occur. The particles form domains where they align parallel to each other. These domains become larger with time and continue to rearrange. In the case of an emulsion, this realignment effectively influences the maximum interface size that can be stabilised by a given number of particles. As such, the effective average domain size will become smaller in our simulations of emulsions in a fixed volume. This section shows that the presence of particles at an interface leads to two additional timescales in the reordering. The first one is the rotation of the particle towards the interface. The particle rotates towards its final orientation parallel to the interface. For lower values of $\chi_I$ this process does not depend on $\chi_I$ and is not different from the single particle adsorption. For larger values of $\chi_I$ the time required to arrive at its final orientation increases. Hydrodynamic as well as excluded volume effects become more important. Above a critical value not every particle reaches its “final” orientation.
4 Concluding Remarks

In this report we have reviewed recent simulation studies of particle stabilised fluid interfaces. We focused on dynamic transitions between bijel- and Pickering emulsion states and demonstrated that those can be obtained by modifying the contact angle a particle forms with the fluid-fluid interface. In the second part of the report we investigated the impact of particle anisotropy on the formation of stable particle-stabilised emulsions. We have shown that the rotational degrees of freedom of the particles lead to additional timescales in the formation and growth of particle-stabilised fluid domains.

Acknowledgements

Besides support from JSC and IBM experts present at the “First JUQUEEN Porting and Tuning Workshop 2013” we acknowledge computing resources that were granted via NIC grant hss08 and by PRACE project pra082.

References

Non-Linear Response of Glass-Forming Systems
to External Forces

David Winter\textsuperscript{1}, Pinaki Chaudhuri\textsuperscript{1,2}, and Jürgen Horbach\textsuperscript{2}

\textsuperscript{1} Institut für Physik, Johannes Gutenberg-Universität Mainz, Staudingerweg 7, 55099 Mainz, Germany

\textsuperscript{2} Institut für Theoretische Physik II – Weiche Materie, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

E-mail: horbach@thphy.uni-duesseldorf.de

Non-equilibrium molecular dynamics (NEMD) simulation is used to investigate the non-linear response of a glass-forming Yukawa mixture to external fields. As a first problem, single particles are pulled through deeply supercooled liquids by a constant force (non-linear active micro-rheology). The pulled particles exhibit anomalous diffusive transport in the direction parallel to the external force. We show that this anomalous transport is associated with the appearance of long residence times in the cages formed by the particles surrounding the pulled one. In the second problem, the Yukawa mixture in the glass state is confined between two walls and studied under the imposition of a uniform shear stress. Here, shear-banded structures are observed below and around the yield stress of the confined glass.

1 Introduction

The nonlinear response of supercooled liquids and glasses to external fields has been in the focus of recent research revealing a wealth of novel phenomena (see, e.g., Refs. 1–13). The study of these phenomena on a microscopic level may result in a better understanding of the nature of the glass transition\textsuperscript{14, 15}. By applying external fields to glassforming systems below the glass transition temperature \( T_g \), relaxation processes can be studied under well-defined conditions, e.g. by investigating the relaxation of a sheared system to a quiescent glass state after switching off the shear field\textsuperscript{2, 5}. Non-equilibrium molecular dynamics (NEMD) simulations are an appropriate tool to study – on a microscopic level – the non-linear response of glass-forming systems to external fields. In this work, this is demonstrated for a simple glass-forming Yukawa mixture (see below). The following themes are discussed:

- **Non-linear active micro-rheology in the glass** (Sec. 2). Single particles (tracer particles) are pulled by a constant force through an otherwise quiescent fluid. We show that in the direction parallel to the force the tracer exhibits a superdiffusive motion. This superdiffusive behaviour is connected to a hopping motion of the pulled particle from cage to cage.

- **Creep for a system confined between walls applying a constant force to the upper wall** (Sec. 3). Here, the onset of flow is associated with a time scale that diverges with decreasing the magnitude of the applied force (or external stress \( \sigma_0 \) corresponding to this force). This defines the yield stress \( \sigma_d \) which is the minimal stress required to bring the system into a steady-state flow. For \( \sigma_0 < \sigma_d \), we observe a creep flow and we show that this creep flow is associated with the occurrence of shear-banded structures.
2 Non-Linear Active Micro-Rheology

In active micro-rheology, a tracer particle is inserted into a host fluid and one measures the response of the tracer to an external force $f$. For colloidal and biological systems, the latter experiment can be realised, e.g., by using a combination of magnetic tweezers and confocal microscopy\textsuperscript{16, 17}. Of particular interest is active micro-rheology in glass-forming systems since these systems exhibit a strong non-linear response even for relatively small external forces. This has been demonstrated in an experiment of a dense hard-sphere-like colloidal fluid by Habdas \textit{et al.}\textsuperscript{18}.

In the linear response regime (i.e. for sufficiently small $f$), the steady-state velocity $v$ of the tracer particle is proportional to the applied force, $f = \xi v$, with the friction coefficient $\xi$ being independent of $f$. In the latter case, the friction coefficient is exactly related to the equilibrium self-diffusion coefficient $D$ via a fluctuation-dissipation theorem, $D = k_B T / \xi$ (with $k_B$ the Boltzmann constant and $T$ the temperature). As has been shown recently in the framework of a fluctuation theorem\textsuperscript{19}, for glass-forming systems the linear response regime shrinks to a window of very small forces and eventually to zero when approaching to glass transition. Then, the non-linear response of the tracer shows a strong dependence of the friction coefficient on the force $f$. Similar to the phenomenon of shear thinning in macro-rheology, the friction coefficient may decrease by orders of magnitude with increasing $f$. Moreover, the non-linear response of the tracer particle is associated with an anisotropic diffusion dynamics with anomalous diffusion (superdiffusion) in the direction parallel to the external force\textsuperscript{11, 12}.

We use MD simulations to investigate a binary glass-forming soft-sphere mixture where A and B particles interact with each other via screened Coulomb (Yukawa) potentials. The details of the model and the parameters of the potential can be found in

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Snapshots of typical independent trajectories of a pulled A particle for $T = 0.14$ and $f = 1.5$. The trajectories are drawn such that they all start at the same initial point. From Ref. 22.}
\end{figure}
Refs. 1, 2, 11, 12. We just note that we consider an equimolar mixture at the number density \( \rho = 0.68/d^3 \) (with \( d \) the diameter of A particles) where the reduced critical mode coupling temperature is at \( T \approx 0.14 \). The micro-rheological simulations start from fully equilibrated configurations in the temperature range \( 1.0 \leq T \leq 0.14 \). In each run, a single particle is pulled with a constant external force \( \vec{F}_{\text{ext}} = (f, 0, 0) \) in \( x \)-direction, assuming periodic boundary conditions in all three spatial directions. To keep temperature constant, the system is coupled to a dissipative particle dynamics (DPD) thermostat that provides local momentum conservation and Galilean invariance. For most of the simulations, a cubic shape of the simulation box with linear size \( L = 13.3 \, d \) is chosen, corresponding to \( N = 1600 \) particles, i.e. 800 of each type. We note that such relatively small systems are large enough to avoid finite-size effects. This finding is in agreement with recent micro-rheological studies of Schroer and Heuer\textsuperscript{20,21} for a binary Lennard-Jones mixture who even considered systems with only 64 particles. In the production runs, the magnitude of the force, \( f \), was varied in the range \( 0.5 \leq f \leq 30 \) (with \( f \) in units of \( k_B T/d \)). For each temperature and force, about 1000 independent trajectories of pulled particles were generated to obtain data with reasonable statistics.

Fig. 1 shows snapshots of typical trajectories of A tracer particles at the temperature \( T = 0.14 \) and the force \( f = 1.5 \). Note that the behaviour of the B particles is very similar and so we only show results for the A particles in the following. The snapshots indicate that the motion of the pulled particle can be described by a hopping motion from cage to cage in the quasi-frozen host liquid where the cage is formed by nearest-neighbour particles around the pulled particle. Essential for this hopping motion is the time scale separation between the motion of the pulled particle and that of the host liquid.

The diffusion properties can be analysed by the mean-squared displacement (MSD) of a tagged particle. In force direction the MSD is defined by

\[
\langle \Delta x^2(t) \rangle - \langle \Delta x(t) \rangle^2 = \langle (x(t) - x(0))^2 \rangle - \langle (x(t) - x(0)) \rangle^2,
\]

where the second term, \( \langle \Delta x(t) \rangle^2 \), corrects for the drift of the particle due to the external force \( f \). In Fig. 2a, the MSDs are displayed in a double-logarithmic plot for different

![Figure 2](image-url)

Figure 2. a) Mean-squared displacement \( \langle \Delta x^2(t) \rangle - \langle \Delta x(t) \rangle^2 \) for pulled A particle at \( T = 0.14 \). The curves correspond to the forces \( f = 0.0, 0.5, 1.0, 1.5, 2.5, 4.0, 6.0, \) and 10.0 (from right to left). b) Effective exponents \( \alpha \) as a function of \( f \) for different temperatures, as indicated. From Ref. 12.
forces at the temperature $T = 0.14$. Also included is the equilibrium MSD ($f = 0.0$). To allow for a direct comparison with the latter MSD, we have multiplied MSDs in force direction, Eq. 1, by a factor of 3. Compared to the equilibrium case ($f = 0.0$), the dynamics accelerates in the long-time limit with increasing $f$. Even more remarkable is the fact that at long-times and intermediate forces the MSDs exhibit a superlinear behaviour, $(\Delta x^2(t)) - (\Delta x(t))^2 \propto t^\alpha$ with $\alpha > 1$. Only at very large values of $f$ (here, $f = 10.0$) the MSD is again diffusive in the long-time limit. At short times (in the ballistic regime), the curves for $f = 6.0$ and $f = 10.0$ are slightly shifted to the left on the $t$ axis corresponding to a slight drift of the temperature from its assigned value $T = 0.14$. This is due to the fact that for such high forces the thermostating of the DPD dynamics is not efficient anymore thus leading to slight temperature drift. However, this drift does not affect the qualitative behaviour of the high-force MSDs at long times. We emphasise that all the results in Fig. 2a are taken in the steady-state regime. Thus, the superlinear behaviour occurs in the steady-state and is not a transient phenomenon.

Fig. 2b shows the effective exponent $\alpha$ as a function of $f$ for different temperatures. At $T = 0.14$, it first increases from about 1.3 to 1.5 in the interval $0.5 \leq f \leq 2.5$, then it is constant around 1.5 between $f = 2.5$ and $f = 6$, before it decreases to 1.0 for $\alpha > 6$. Note that all the considered values of $f$ at $T = 0.14$ are beyond the linear response regime. At higher temperatures the behaviour of $\alpha$ is qualitatively similar but the effective exponents are significantly lower than for $T = 0.14$. From this, we can conclude that the superlinear (or superdiffusive) behaviour is directly related to the time scale separation between the motion of the pulled tracer particle and that of the surrounding host fluid. A pronounced superdiffusive behaviour is seen if the surrounding host fluid is quasi-frozen on the time scale of the tracer particle.

The time scale separation between the motion of the pulled tracer particle and the quasi-frozen host liquid is associated by a hopping motion of the tracer from cage to cage. This is indicated in Fig. 3 which shows typical trajectories for the force $f = 1.0$. While at high temperatures, say at $T = 0.18$, the trajectory is relatively smooth, at low temperatures one can clearly identify the residence in the cage and the hop to the next cage. Furthermore, the residence time in the cages is rather heterogeneous at low temperatures (cf. the trajectories

![Figure 3](image)

**Figure 3.** a) Typical trajectories, $x(t)$, of pulled A particles at $T = 0.14$ and $f = 1.0$. b) Typical trajectories, $x(t)$, of pulled A particles at $f = 1.0$ and different temperatures, as indicated. From Ref. 12.
at $T = 0.14$ in Fig. 3a). In a recent work$^{11}$, we have determined waiting time distributions $P(\tau)$ from trajectories $x(t)$ such as those shown in Fig. 3. For $T = 0.14$, these distributions can be described by a stretched exponential function, $P(\tau) \propto \exp(-C\tau^\beta)$, with $\beta = 0.5^{11}$. Thus, $P(\tau)$ can be classified as a broad distribution with broad tails towards large values of $\tau$. The latter tails are reflected by the occurrence of very long cage residence times in the trajectories of Fig. 3a.

Many of the features that we have seen in the micro-rheological behaviour in force direction are also found in certain random force field models, proposed by Bouchaud et al.$^{23,24}$ One of these models can be mapped onto a directed walk among traps with a broad release time distribution, given by a master equation for a one-dimensional lattice model with a random distribution of asymmetric transition rates at each lattice site. In this model, superdiffusion is observed if the drift force experienced by the walker is sufficiently large such that the walker sees only very high potential barriers, corresponding to very long times in the tails of the release time distribution. A similar mechanism may also apply in non-linear active micro-rheology of glass-forming systems. It is not clear, of course, to what extent a structural glass can be represented by simple random trap models (at least on a coarse-grained level), but such models can be seen as minimal models showing superdiffusion caused by a force-induced drift motion.

3 Confined Yukawa Mixture Under an Imposed Shear Stress

A macro-rheological experiment that is complementary to active micro-rheology is provided by a stress-controlled Couette flow setup. One considers, e.g., a glass-forming system confined between two walls and then, one pulls one of the walls with a constant force $F_0$, thus imposing a constant shear stress of $\sigma_0 = F_0/A$ (with $A$ the area of the pulled wall). There is an important difference between such a macro-rheological experiment and active micro-rheology. In the latter case, the state of the host liquid keeps essentially unchanged over a broad range of the force $f$ (see above) applied to the tracer particle$^{12}$, whereas in the stress-controlled macro-rheological experiment the state of the entire system is changed. As a consequence, there are phenomena in sheared systems that do not seem to have a counterpart in micro-rheology$^{11}$. One of these phenomena is the existence of a yield stress $\sigma_d$: for a material being in a glass state, the material exhibits only a transition towards steady-state flow if the applied stress exceeds the threshold value $\sigma_d$. For $\sigma_0 < \sigma_d$, creep flow instead of steady-state flow is observed$^{25}$. The spatio-temporal characteristics of this creeping motion are essentially unknown and so we have performed non-equilibrium MD simulations to elucidate the nature of creep flow.

In our simulations, we consider the same soft-sphere mixture as for our micro-rheological studies (see above); now a system of $N = 12800$ particles is considered, placed in a simulation box with dimensions $L_x = 26.66 \, d$, $L_y = 53.32 \, d$ and $L_z = 13.33 \, d$. Glass configurations in confined geometry are generated as follows: We start with $m = 24$ fully equilibrated bulk configurations at $T = 0.2$ (for these systems periodic boundary conditions are used). These configurations are then instantaneously quenched to $T = 0.05$ and then aged for durations of $t_{\text{age}} = 10^3$, $10^4$, $10^5$. After the time $t_{\text{age}}$, particles at $0 < y < 2 \, d$ and $L_y - 2 \, d < y < L_y$ are frozen in, thereby obtaining samples in a glass state that are confined between rough walls. These samples are sheared by pulling the top wall in $+x$-direction with a constant force $F_0$ (corresponding to a constant shear stress $\sigma_0$, see above). Further details of the simulation can be found in Ref. 7.
During the shear simulations, we determine the velocity of the wall, \( v_w(t) \), as a function of time, from which, similar as in a creep experiment, the average effective strain at the wall is obtained via \( \gamma_w(t) = \langle \gamma_\alpha(t) \rangle_e \) with \( \gamma_\alpha(t) = \int_0^t \frac{v_w(t')}{(L_y - 4d)} dt' \) the wall strain of the \( \alpha \)-th sample and \( \langle \cdot \cdot \cdot \rangle_e \) the average over the ensemble of the \( m = 24 \) trajectories. Fig. 4a shows \( \gamma_w(t) \) for three different stresses \( \sigma_0 \) above the yield stress \( \sigma_d \approx 0.085 \) (for the estimate of \( \sigma_d \), see Ref. 7). Similar as in a recent experiment, different time regimes can be identified: (i) an early time regime where the strain approaches an oscillatory part (the latter oscillations are due to an elastic response of the sample); (ii) an intermediate time regime with a superlinear behaviour of the strain; (iii) an asymptotic linear regime corresponding to steady flow at a constant shear rate. Below the yield stress at \( \sigma_0 = 0.08 \) (Fig. 4b), creep flow is manifested by a power law behaviour, \( \gamma_w(t) \propto t^{0.25} \), at long times. For \( \sigma_0 > \sigma_d \) and intermediate time scales, the strain decreases with increasing \( t_{age} \) while it reaches the same asymptotic steady state at long times (as expected). Also for \( \sigma_0 < \sigma_d \) the strain decreases with increasing \( t_{age} \), but the exponent of the power law remains the same for the three considered aging times.

To characterise the nature of the creep flow, we now analyse deformations due to the shear on a local scale. To this end, we compute maps of transverse displacements since they provide a measure of the non-affine motions due to local rearrangements (the details of the calculation can be found in Ref. 7). Fig. 5 shows the evolution of displacement maps for an imposed stress of \( \sigma_0 = 0.085 \), i.e. \( \sigma_0 \approx \sigma_d \). Note that the times at which the maps are shown fall all into the creep flow regime in \( \gamma_w(t) \). As the maps indicate a shear-band forms that remains stable during the entire simulation run. We note that we find the occurrence of similar shear-bands in all the considered samples in the creep flow regime. Thus, the creep
flow is associated with a spatially heterogeneous dynamics. This finding may stimulate theoretical models that predict such heterogeneities in the form of shear-banded structures. In this context, a recent study is very interesting where in the framework of a soft glassy rheology model, a relation between creep flow and shear-banded velocity profiles has been predicted\textsuperscript{26}.

4 \textbf{Summary}

We have done extensive NEMD simulations of a glass-forming soft-sphere mixture to study its non-linear response to external forces. On a microscopic level, our simulations give insight into the mechanisms that lead to anomalous transport as well as inhomogeneous flow patterns.

In the case of active non-linear micro-rheology, a superdiffusive motion of the pulled tracer particle in the direction parallel to the applied force is observed. The tracer performs a hopping motion from cage to cage, associated with a broad distribution of residence times of the tracer in the cages. The motion of the pulled tracer particle is to some extent similar to a directed walk among traps with a broad release time distribution that also shows the occurrence of superdiffusive behaviour.

The second problem considered in this work is a macro-rheological flow setup that is complementary to to active micro-rheology. We have studied the glass-forming Yukawa mixture under constant-stress conditions in order to characterise the creep flow that is observed, if the external stress is smaller than the yield stress. From the analysis of displacement maps we showed that the creep flow takes place in the form of shear-banded structures.

\textbf{Acknowledgments}

We thank Kurt Binder and Peter Virnau for useful discussions. We gratefully acknowledge financial support by the German DFG, projects No. SFB TR 6/A5 and No. FOR 1394/P8 (HO 2231/7-1). Allocation of computer time at the NIC Jülich is gratefully acknowledged.
References

22. A. Winkler, D. Winter, P. Chaudhuri, A. Statt, P. Virnau, J. Horbach, and K. Binder, submitted to EPJ-ST.
Directing the Assembly of Block Copolymers:
A Single-Chain-in-Mean-Field Simulation Study

Marcus Müller, Dewen Sun, and Ulrich Welling
Institut für Theoretische Physik, Georg-August-Universität,
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany
E-mail: mmueller@theorie.physik.uni-goettingen.de

The self-assembly of copolymer materials has attracted abiding interest for their ability of forming structures on the nanometre scale. By virtue of the minuscule free-energy differences between competing structures, however, many defects form and the ordering kinetics is protracted. Hence it is important to predict the kinetics of self-assembly and direct the structure formation. We discuss two strategies for directing the self-assembly using (i) static electric fields and (ii) kinetically trapping after a rapid pressure quench, and illustrate their potential by large-scale computer simulation of a soft, coarse-grained model.

1 Introduction

Diblock copolymers are long flexible macromolecules that comprise two blocks of incompatible monomer species \( A \) and \( B \). Since the different segments repel each other they tend to segregate into domains. The connectivity of the two blocks, however, prevents macroscopic phase separation and, instead, these amphiphilic molecules form periodic microphases. The characteristic domain spacing, \( D_0 \), is dictated by the mean-squared end-to-end distance, \( R_{e0} \), of the macromolecules and can be experimentally controlled by their molecular weight. Typically, \( D_0 \) ranges from 10 nm to 100 nm, and these microphases have found ample applications as templates for fabricating electronic circuits or memory devices or as scaffolds for catalysts, filtration sieves or 3D photonic crystals\(^1\).

The equilibrium phase behaviour in the bulk is well understood in terms of the interfacial tension, which favours large domains, and the entropic penalty associated with the stretching of the conformations as the molecules fill space uniformly in a melt\(^1\). The free-energy differences between morphologies, however, is minuscule – on the order of \( 10^{-3} k_B T \) per molecule. Hence the driving forces to anneal defects is very small and the ordering kinetics is protracted. Moreover, there are many metastable minima in the free-energy landscape that can be categorised according to the scaling of the free-energy with spatial dimension – alternate periodic morphologies, grain boundaries, disclination lines and point defects. In fact, the free-energy landscape of these soft materials has been likened to that of glasses\(^2\).

Consequently, different strategies for directing the assembly of block copolymers have been devised including topographically structured supports (graphoepitaxy\(^3\)), chemically patterned substrates (chemoepitaxy\(^4\)), electric fields\(^5,6\) or shear\(^7\). In this proceedings contribution, we illustrate large-scale simulations of a soft, coarse-grained polymer model to explore two strategies for directing the kinetics of self-assembly: (a) an electric field and (b) a rapid pressure quench.
2 Model and Computational Technique

In order to investigate the kinetics of structure formation in block copolymer materials on the time scale of minutes and length scale of hundreds of nanometres the use of a coarse-grained representation is mandatory. Therefore we use a top-down approach representing the universal features of the collective ordering kinetics by a particle-based model that only retains the relevant interactions – the connectivity of the segments along the macromolecular backbone and the non-bonded interactions that drive the microphase separation. We consider \( n \) polymers comprised of \( N = 32 \) segments in a simulation cell. \( fN \) of the segments are of type \( A \) and the others belong to the \( B \) species. The connectivity of the macromolecules is modelled by a bead-spring Hamiltonian,

\[
H_b = \frac{3(N-1)k_BT}{2R_{e0}^2} \sum_{i=1}^{n} \sum_{s=2}^{N} \left[ r_i(s) - r_i(s-1) \right]^2,
\]

where \( r_i(s) \) denotes the spatial position of the \( s \)th segment on polymer \( i \). The non-bonded interactions take the form

\[
H_{nb} = \int \frac{dr}{R_{e0}^3} f(\phi_A, \phi_B)
\]

where \( \phi_A \) and \( \phi_B \) denotes the local density of \( A \) and \( B \) segments, respectively. The latter can be computed from the explicit particle coordinates via a collocation grid. For the first application – symmetric diblock copolymers in an electric field – the non-bonded interactions are comprised of the standard field-theoretic Hamiltonian

\[
f_{nb} = \sqrt{\bar{N}} k_BT \left\{ \frac{\kappa N}{2} (\phi_A + \phi_B - 1)^2 - \frac{\chi N}{4} (\phi_A - \phi_B)^2 \right\},
\]

where the first term describes the near-incompressibility of the melt (with \( \kappa \) being proportional to the inverse isothermal compressibility) and the second term captures the free-energy of mixing \( A \) and \( B \) (with \( \chi \) and \( \bar{N} \) being the Flory-Huggins parameter and the invariant degree of polymerisation) and, additionally, a contribution due to the electric field,

\[
f_{el} = -\frac{R_{e0}^3}{2} \epsilon(\phi_A, \phi_B) E^2
\]

where \( \epsilon \) is the composition-dependent dielectric constant. We use a simple constitutive relation \( \epsilon = \frac{\epsilon_A \phi_A + \epsilon_B \phi_B}{\phi_A + \phi_B} \). The electric field satisfies Laplace’s equation

\[
\nabla \cdot \left[ \epsilon(\phi_A(r), \phi_B(r)) \mathbf{E}(r) \right] = 0
\]

For the second application – structure formation in response to a rapid pressure quench – we utilise a fourth-order virial expansion \( f_{nb} \) to model the equation-of-state of the compressible, binary mixture and the diblocks are asymmetric, \( f = 1/8 \).

The polymer conformations are updated by Monte-Carlo simulations using the strong bonded forces to bias the local diffusive motion of segments. In order to exploit the separation between the strong bonded and weak but computationally expensive non-bonded forces, we use the Single-Chain-in-Mean-Field algorithm, which temporarily approximates the non-bonded interactions by external fields that are frequently computed from the instantaneous, local density created by the particles. This quasi-instantaneous field approximation is controlled by a parameter that depends on the discretisation of the chain contour \( N \) and the spacing of the collocation grid and it can be conceived as the Monte-Carlo analog of the multiple-time-step algorithms in molecular dynamics simulation.

Since the polymer chains move independently between the updates of the external fields, this algorithm is straightforwardly implemented on parallel computers by scattering the polymer configurations across the processors independently of their spatial position. In the presence of an electric field, this computational algorithm is particularly useful because
the Laplace equation, Eq. 2, is only used when the external fields are updated, i.e., typically after all polymer segments had on average the chance to be moved once (1 MC step).

In order to solve the Laplace equation we discretise the electric field by a finite difference scheme on the same collocation grid as the particle densities, obtaining a linear set of equations for the electric potential $\psi$ defined by $E = -\nabla \psi$. Since densities and potential evolve only little between subsequent updates of the fields, we iteratively solve the set of equations by a parallel conjugate-gradient solver using a spatial domain decomposition.

3 Simulation Results

3.1 Directing the Assembly by an Electric Field

Electric fields have been utilised experimentally to align block copolymers\textsuperscript{5,6} but the mechanism of alignment and its kinetics is not completely understood. The advantage of an electric field in directing copolymer self-assembly is based on its long-range character. Moreover, whereas a homogeneous electric field retains the degeneracy of grains that only differ by a rotation around the field axis, inhomogeneous fields allow for 3D orientation control.

Fig. 1 shows the assembly of a symmetric block copolymer in a concentric wedge of radii, $R_1$ and $R_2$, and height $H$, in the absence of an electric field. The inner and outer concentric surfaces prefer one segment species. This graphoepitaxy locally directs the morphology into a concentric morphology but the aligning effect of the confinement does not propagate far into the volume; it is restricted to a few lamellae. Although concentric lamellae are thermodynamically stable for the larger thickness, even extended simulation runs, lasting $20\tau_0$ (with $\tau_0 = Re_0^2 / D$ and $D$ being the self-diffusion coefficient in the disordered state), do not result in defect-free ordering; defects remain and the detailed structure depends on the commensurability of $H$ with the lamellar spacing, $D_0$.

If we start in a perfect concentric phase and switch on an electric field, the AB interfaces will align parallel to $E$ and we induce a transition from the concentric to the radial morphology as illustrated in Fig. 2. This demonstrates the favourable orientation effect

![Figure 1. Morphology of a symmetric block copolymer in a cylindrical segment of height $H = 1D_0$ (left) and $H = \frac{3}{4}D_0$ (right), $R_1 = 30D_0$, $R_2 = 40D_0$, without E-field. Circular lamellae form due to a contact potential at the inner and outer rings but defects remain in the middle.](image-url)
of the E-field that pervades the volume. In this particular geometry the inhomogeneity of the field not only orients the bulk lamellar morphology but additionally aids in mitigating defects of the radial phase due to the gradual variation of the lamellar thickness with radial position. Due to the inhomogeneity the applied potential can be used to precisely adjust the range of the ordering effect.

The early kinetics of the transition from the concentric to the radial phase, depicted in panels b and c of Fig. 2, proceeds via an undulation instability of the innermost AB interface that is exposed to the highest field. The observation that the transition does not involve a localised nucleation event but, instead, the new radial morphology forms everywhere at the inner ring, indicates that the electric field is strong enough to render the concentric phase absolutely unstable and allows for rapid reordering. A detailed analysis of this interfacial instability reveals that the growth rate is proportional to the square of the wave vector, $q$, because the change of the local interface position involves the diffusive transport but the maximum growth rate occurs at $q_{\text{max}}D_0 \approx 2\pi$ and is rather independent from the field strengths.

### 3.2 Directing the Assembly by a Rapid Pressure Quench

Complex network morphologies in block copolymers have attracted abiding interests due to their beneficial physical attributes for emerging technological applications ranging from materials for fuel cell to templates for 3D photonic crystals. In linear diblock copolymer melts, however, more than 80\% of the phase diagram is occupied by the lamellar, hexagonal, BCC, and CPS phases at $\chi N = 100$. These classical morphologies have no multiple domains that continuously percolate across the whole system in all three dimensions. Traditional approaches try to alleviate the packing frustration of the flexible...
Figure 3. Rapid quench structure formation from a stable BCC morphology at \( f_{\text{start}} = 0.15 \) to \( f_{\text{end}} = 0.5 \). Since the change of the thermodynamic state on the time scale \( \sim 10^{-3} \) s is fast compared to the molecular relaxation, \( \tau_0 \), the system reaches the free-energy landscape (blue curve) of the new thermodynamic state, \( f_{\text{end}} \), in the BCC morphology (star \( \star \)) that is highly unstable. The spontaneous structure formation after this rapid quench becomes quickly trapped in a metastable R8F morphology on the time scale of \( \sim 10 \) s. Only after a much longer time, \( \sim 10^3 \) s, the thermodynamically stable lamellar phase eventually nucleates via a thermal fluctuation, and the system reaches equilibrium.

diblock copolymers in network morphologies in order to stabilise them with respect to competing classical morphologies. One strategy consists in exploiting chain-length polydispersity\(^{12,13}\), and blending copolymers with homopolymers or solvents\(^{14-16}\).

We have begun to pursue a different approach. Rather than fine-tuning the copolymer material in order to make the desired network structure the thermodynamic equilibrium state, our strategy consists in reproducibly directing the kinetics of structure formation to be temporarily trapped in the desired network morphology. This “process-controlled” assembly, illustrated in Fig. 3, provides a potential avenue for reproducibly fabricating metastable morphologies without the need for synthesizing new macromolecular materials or blending existing ones. To this end, it is necessary (i) to control the generation of well-defined, highly unstable states and (ii) to design the unstable state such that the ensuing spontaneous kinetics of structure formation reaches the desired metastable morphology. The former is feasible in macromolecular systems because the growth rate of spinodal modes of an unstable state is on the order \( 1/\tau_0 \), where the single-chain relaxation time \( \tau_0 \) is much larger than the time scale of altering a thermodynamic control parameter, e.g., pressure or temperature.

The design of a suitable process requires the prediction of the spontaneous structure formation from an unstable state, which in the present example is generated by a rapid pressure quench (affine deformation). We have investigated such a deep pressure quench by computer simulation of our compressible, coarse-grained model. The equation of state
Figure 4. MFEP from the unstable BCC morphology to the metastable R8F network morphology and the conserved transformation path to the expanded BCC state. Snapshots illustrate the initial BCC morphology, the metastable R8F and BCC morphologies as well as the metastable R8F structure along the MFEP towards the BCC morphology at \( s = 0.67 \). The results demonstrate that the local conservation of density alters the directly accessible metastable structure from R8F to BCC.

is chosen such that the minority block \( A \) is supercritical and the majority block \( B \) is nearly incompressible. At high pressure a BCC structure forms, whereas at lower pressure the volume fraction of \( A \) increases and a lamellar phase is observed. Thus the pressure quench mimics an alchemical transformation from a highly asymmetric to a nearly symmetric diblock copolymer.

In the kinetic simulations the system becomes trapped in a metastable network morphology – the reverse single 8-fold (R8F) phase – after a pressure jump from the BCC morphology. This R8F morphology involves the formation of a continuous \( A \) domain inside the \( B \) matrix, which is facilitated by the high pressure inside the spherical \( A \) domains immediately after the pressure quench (“evaporation”) and by the stretching of the \( B \) blocks, which contract and pull the short \( A \) fragments into the \( B \) matrix. If, however, we use a field-theoretic umbrella potential to equilibrate the chain conformations with the density fields of the BCC structure after the quench, then we only observe an expansion of the spherical \( A \) domains, highlighting the importance of the stretched conformations after the rapid expansion.

Using field-theoretic umbrella sampling\(^{17}\) and the on-the-fly string method\(^{18}\) we have studied the minimum free-energy path (MFEP). As shown in Fig. 4 we observe a direct path from the unstable BCC morphology to the R8F state – in contradiction to the kinetic results because the MFEP assumes the chain conformations be in equilibrium with the instantaneous density and should predict the swelling of the \( A \) domains of the BCC morphology. The formation of the R8F along the MFEP occurs because the system reduces its free energy by non-locally increasing the \( A \) concentration inside the \( B \) matrix for there
is no local conservation of the density imposed. If we include this constraint\textsuperscript{19}, however, we obtained the expansion of the BCC morphology in agreement with the simulations. The results, compiled in Tab. 1, highlight the importance of (a) non-equilibrium molecular conformations in the starting state and (b) the local conservation of density. Since the molecular conformations evolve on the same time scale as the morphology, there is no time-scale separation between the slowest Rouse mode(s) of the macromolecular conformations and the morphology. In fact the metastable state is reached around one Rouse time. Thus a description in terms of the density fields alone is insufficient.

4 Concluding Remarks

We have studied the directed assembly of block copolymers by molecular simulations of a soft, coarse-grained model. The use of efficient coarse-grained models, optimised algorithms (e.g., SCMF simulations and an efficient solver for the electrostatic problem), and advanced free-energy techniques (e.g., field-theoretic umbrella sampling and string method) contribute to provide simultaneous insights into the structure, thermodynamics and kinetics of these fascinating materials. Our simulations may help to guide experiments but they also point to deficiencies of the current theoretical description.

Acknowledgements

We thank Yoav Tsori and Kostas Ch. Daoulas for stimulating discussions. Financial support has been provided by the DFG under grants Mu 1674/9-2 and Mu 1674/11, and we have benefited from a generous grant of computer time at the Jülich Supercomputing Centre.

References


Earth and Environment
Throughout the last decades, computational methods have become an essential tool, being on eye level with theoretical and experimental approaches in the process of scientific discovery. High performance computing, as a core enabling technology, allows to tackle problem classes which seem unreachable only a few years ago. In the field of Earth and Environment, High Performance computing meets a field of high relevance to society and economy. A particular challenge in this field is the immense spectrum of spatial and temporal scales which are involved in processes affecting Earth and Environment. With ongoing research, we become more and more aware of the fact, that the different spheres, i.e. the geosphere, atmosphere and hydrosphere are closely coupled. A volcano eruption, having its origin deep in the Earth, can significantly influence the climate. On the other hand, the presence of an atmosphere, early during the evolution might have influenced the further thermal evolution of the Earth’s interior. Time scales involved range from million of years for geological events to milliseconds as typical for some fluctuations in the Earth’s magnetic field. The Earth’s magnetic field, with its origin deep in the centre of the Earth, protects us against energetic particles from the sun and is today viewed as a necessary prerequisite for the habitability of a planet. State of the art computer models allow for investigating such phenomena. Of course “realistic” models can often not be achieved in this field, even employing today’s technology. However, often the models serve to understand key mechanisms by testing several hypothesis and such lead to a much deeper understanding of the sometimes very remote phenomena.

The first contribution by Kurzmann et al. describes a quite practical application. The exploration of the subsurface underground receives increasing attention. Substantial energy sources are stored in this part of our planet and it may also become important for the storage of man made waste. Usually information about the subsurface is retrieved from seismic waves, but often only simple characteristics of the waves are analysed, essentially the echoes of the waves are recorded. Here, Kurzmann et al. propose a by far more developed inversion method, allowing for taking into account information from the full waveform. This inversion technique allows for resolutions which go far beyond procedures like traveltime tomography which are essentially based on ray theory. The authors demonstrate the power of their approach and show at the same time that their procedures can easily exhaust today’s available computing power.

In order to develop an understanding of processes in the deep Earth, the behaviour of materials under extreme conditions must be explored. The proposal by Jahn et al. aims at this ambitious goal. Atomic structures of geomaterials vary significantly from metallic iron, deep in the Earth’s core, to silicates and oxides in the Earth’s mantle and crust. In the presence of fluids ionic and molecular species do coexist. High pressure fluids can have
a strong influence of dynamic processes within the Earth. For example, they can strongly influence the viscosity of the mantle material and locally facilitate creeping flows in the mantle. Another topic concerns the investigation of the silicate melts. Material properties of silicate melts, again especially the viscosity, play a key-role for geodynamical processes on all scales. Ranging from the thermochemical evolution of a magma chamber and it’s implications for the formation of igneous deposits to the evolution of the Earth’s mantle on a global scale. Here first-principle modelling techniques are employed which are based on quantum mechanic treatment of the electronic structure of the materials. Differently from experimental approaches they can principally deal with the whole range of pressures and temperatures, however, similar to the previous example are still limited by computer resources.

Convection is clearly the main mechanism behind the dynamics of our planet. It drives the circulation of the atmosphere, also the large scale motion in the ocean is driven by convective flows. Further, the tectonic activity on Earth and the generation of the magnetic field is ultimately driven by convection. Boundary layers are always of key importance for the dynamics of a convecting system and this is also reflected in the contribution of Mellado et al.. The dynamics of the planetary boundary layer is in the focus of that contribution. The planetary boundary layer is the lowest layer of the atmosphere, being in contact with the surface of the planet. Turbulent mixing is an essential phenomenon taking place within the boundary layer. The authors employ direct numerical simulation of a Boussinesq fluid to better understand entrainment in the boundary layer and also the turbulent collapse of the boundary structure. Besides an improved understanding of the role of the planetary boundary in the meteorological context, one can also hope for deeper insight into the phenomenon of turbulence, resulting from these efforts.

Last, but not least the contribution of Truhetz et al. directly addresses the issue of climate change. They operate so called Regional Climate Models (RCM), driven by the General circulation models GCM, in order to better answer question of regional scale. Their models provide a much finer resolution than the GCM’s and thus they may be able to predict extreme events on a local spatial scale. As a connection to the previous contribution it seems noteworthy that the regional models belong to a class of so called convection permitting models. It is recognised that deep moist convection plays a major role in extreme precipitation events and this type of convection becomes resolvable in the approach of Truhetz et al.. In my view, one can be truly curious, how this will influence the Climate models in the Alpine region.
3D Acoustic Full Waveform Tomography

André Kurzmann, Simone Butzer, Stefan Jetschny, Anna Przebindowska, Sven Heider, Lisa Groos, Martin Schäfer, and Thomas Bohlen

Geophysical Institute, Karlsruhe Institute of Technology, 76187 Karlsruhe, Germany
E-mail: andre.kurzmann@kit.edu

In this work, a massively parallelised implementation of 3D acoustic full waveform tomography (FWT) has been developed. Nowadays, FWT emerges as a powerful method in seismic imaging. It is able to reconstruct high-resolution subsurface models from seismic data recorded at the earth’s surface. Almost all computational efforts of this inverse method are required for modelling of seismic wavefields. Our C implementation extensively exploits the possibilities of the message passing interface (MPI). The parallelisation comprises domain decomposition in finite-difference modelling and simultaneous computation of several modelings. Thus, network communication is reduced significantly and the implementation shows a nearly perfect scaling behaviour. Furthermore, we apply 3D FWT to a synthetic reflection experiment involving a complex subsurface model within a marine environment. Starting from a simple model assumption, the FWT recovers a satisfactory velocity model from seismic data.

1 Introduction

Living in a time where natural resources are scarce and precious, the number of underground constructions is increasing, and the storage of waste and other materials in the earth is necessary, it is important to find accurate ways of mapping geological structures in the earth’s interior. A considerable amount of money and effort has been spent on the field of reflection seismology, the science of collecting echoes and transforming them into images of the subsurface. Conventional seismic imaging methods applied in reflection seismology utilise a small portion of the information of the echoes we obtain from the subsurface. Most methods analyse the arrival time of the echoes or specific signal amplitudes only.

In this project we further develop a new seismic imaging technique that uses the full information content of the seismic recordings. Each echo or reflection from geological discontinuities in the earth is used to unscramble the earth structure. For an improved estimation of subsurface parameters we thereby exploit the full information contained in the seismic waveforms. Full waveform tomography (FWT) is a cutting-edge inverse method that accounts for the full seismic waveform recorded over a broad range of frequencies and apertures. It iteratively retrieves multi-parameter models of the subsurface by solving the full wave equations each time. It allows for a mapping of structures on spatial scales down to less than the seismic wavelength, hence providing a tremendous improvement of resolution compared to traveltime tomography based on ray-theory. Especially in exploration geophysics and earthquakes seismology the interest in FWT is increasing continuously.

The FWT is split into two major directions, the time- and frequency-domain FWT (Virieux et al.\(^8\)). Although first implementations of the FWT in the 1980s were conducted in the time-domain by Tarantola\(^8\) and Mora\(^8\), the frequency-domain version of FWT developed in the 1990s by Pratt\(^5\) has now emerged as an efficient imaging tool. In our work we concentrate on the implementation of the former method and its application to seis-
mic problems. It comprises two- and three-dimensional modelling of acoustic and elastic wavefields in the time-domain. Its main advantage is the efficient parallelisation by domain decomposition (Bohlen\textsuperscript{1}) and shot parallelisation (Kurzmann \textit{et al.}\textsuperscript{2,3}) leading to a significant speedup on parallel computers.

Figure 1. Top: The Marmousi P-wave velocity model\textsuperscript{7} consisting of a marine environment with a seawater layer and complex geological structures. The markers illustrate the seismic source (air gun, red) and exemplary receivers (hydrophones, green) dragged by the research vessel, e.g., the Ramform of Petroleum Geo-Services (image). The acoustic wavefield (emitted by the source, shown in blue/magenta) represents an exemplary snapshot after 1.5 seconds of wave propagation. Bottom: A synthetic pressure data set for the source location shown above. The seismic recordings correspond to the exemplary receivers. The arrows indicate the relation between seismic measurement based on real subsurface structures (observation) and the inverse problem where we have to find a subsurface model which explains the observed data.
2 Motivation

The main part of the FWT consists of seismic modelling which is a straightforward method, i.e., synthetic seismic data are computed by the numerical solution of the acoustic or elastic wave equation. Modelling is essential for simulations of wave propagation and the computation of “observed” data sets based on model assumptions of the subsurface. In contrast, the FWT is an inverse method with the aim to recover the underlying model by explaining the observed seismic data (Fig. 1).

The acoustic FWT is limited to modelling of compressional body waves (P-waves). Hence, we concentrate on the processing of pressure data with the purpose of application to marine seismic data. In 2D marine seismics, a seismic source and a streamer are towed behind the seismic vessel. The pressure pulse generated by the source travels through the water layer, down through the earth and back to the surface. The streamer detects this reflected energy using pressure sensitive devices called hydrophones. The research ship travels in a straight line at a constant speed, dragging sources and streamer with a length of several kilometres. The survey collects data at individual, evenly spaced, source points (in seismics: referred to as “shots”) along the survey lines to provide the section through the earth’s crust. The depth and the type of geological target for a given survey have an influence on the streamer length and the vessel speed. Deep surveys require longer times for reflection to return from greater depths and respectively longer streamers to record these events. Complex seismic surveys involve high demands on imaging techniques. For example, the imaging of deep and large geological structures requires expensive seismic modelling. That is, we need both huge subsurface models and numerous seismic sources to ensure a sufficient illumination of shallow and deep structures. Due to the high demands of FWT with respect to resource consumption and computational time, its feasibility requires high-performance supercomputing.

3 Description of Methods and Algorithms

Full waveform tomography aims to find the optimal subsurface model by iteratively minimising the misfit between observed and modelled data. That is, this model has to explain the observed seismic data. The implementation is written in C and is based on the conjugate gradient method as described by Tarantola and Mora.

3.1 Seismic Modelling

Seismic modelling is the fundamental part of full waveform tomography and requires nearly all the computation time. In dependence of the field of application and the complexity of the subsurface, the wave-propagation physics for an underlying subsurface model has to be described by an appropriate wave equation. In our work, we use the acoustic approximation of the wave equation, which is applicable to marine seismic experiments. However, a high complexity of subsurface structures (see Sec. 4) requires the solution of the 3D wave equation involving 3D models (applications are referred to as 3D FWT). The numerical implementation of the wave equations consists of a time-domain finite-difference (FD) time-stepping method in Cartesian coordinates. In detail, the FD-scheme solves either the pressure formulation or the velocity-pressure formulation by utilising pressure and/or
particle-velocity wavefields. Due to finite model sizes, the wave equations are expanded by perfectly matched layer terms (PML) to avoid artificial boundary reflections.

3.2 Full Waveform Tomography

The solution of the inverse problem comprises several steps shown in Fig. 2. The method is initialised by the choice of a 2D or 3D initial parameter model. Seismic velocities or mass density are assigned to the model \( m_0 \) at the first iteration. The initial model can be estimated from \textit{a priori} information or computed by conventional imaging methods. For each source of the acquisition geometry, seismic modelling is applied, i.e., the wavefield is emitted by the source and propagates forward across the medium. A time series of this wavefield has to be stored in memory with respect to the whole volume. Synthetic seismic data is recorded at the receivers and the difference of observed and synthetic data is calculated – resulting in residuals. For each source, the residual wavefield is back-propagated from the receivers to the source position. The cross-correlation of forward-

![Diagram of Full Waveform Tomography](image)

Figure 2. General scheme of the time-domain FWT: As described by Tarantola\(^6\), the model is improved by using an iterative gradient method.
and back-propagated wavefields yields shot-specific steepest-descent gradients $\delta m_n$. The computation of the global gradient $\delta m_n$ for the entire acquisition geometry is given by the summation of all gradients $\delta m_n$. An optimised gradient $\tilde{\delta} m_n$ is computed by subsequent preconditioning and application of the conjugate gradient method. The update of the model parameter is the final step of a FWT iteration. The gradient $d_n$ has to be scaled by an optimal step length $\mu_n$ to get a proper model update at iteration step $n$. The estimation of $\mu_n$ is performed at each iteration and requires additional modellings.

3.3 Parallelisation and Performance

The success of a tomography depends on a sufficient illumination of the subsurface. Thus, several shots, such as $(10, \ldots, 100)$, and many receivers are necessary. Due to the separate wavefield simulation of each shot, modellings require most of the entire computation time. The huge computational efforts can be handled by a massive parallelisation. Our FWT implementation offers two types of parallelisation. The conventional strategy is the domain decomposition, where the model is subdivided into several subdomains and computations are distributed to numerous CPU cores (Fig. 3). In order to ensure an accurate calculation of the wave equation, additional padding layers have to be located around the model. At each time they are exchanged by MPI communication. Due to an increasing communication overhead with increasing number of cores, modellings cannot benefit from domain decomposition with a high number of very small subdomains. On condition of enough memory, the domain decomposition should be restricted to the cores of one node and combined with shot parallelisation (Fig. 3). This second possibility allows the simultaneous modelling for several shots, which are performed independently. In an optimal case the simulations can be distributed to all available nodes. The reduction of inter-node network communication results in a significant decrease of computational time.

Apart from speed, we have to consider the memory requirements of our 3D acoustic FWT implementation representing a pure time-domain method. The imaging condition

![Figure 3. Parallelisation strategies on a cluster computer (2 nodes, 4 cores "C#" each). The model is divided into subdomains with padding layers to exchange wavefields. Pure domain decomposition requires intra-node and inter-node communication (arrows). Shot parallelisation distributes simulations among the nodes and omits inter-node communication.](image-url)
of FWT is expressed by a cross-correlation of forward-propagated and back-propagated pressure wavefields (see Sec. 3.2 and Fig. 2). Therefore, a 3D FWT has to store a time series of forward-propagated 3D wavefield snapshots. We exploit the Nyquist criterion to reduce the number of mandatory 3D wavefield snapshots. In other words, it is not necessary to store snapshots at every finite-difference time step. This ensures both a sufficiently accurate sampling of the time series and a feasible memory consumption.

4 3D Acoustic FWT in a Marine Environment

The evaluation of marine seismic data is a typical field of application for a 3D acoustic FWT. In the following, we show the successful application of our 3D FWT code to a synthetic 3D reflection experiment (see also Kurzmann). A similar test configuration is utilised to perform a speedup benchmark.

4.1 Reflection Seismic Experiment: Synthetic Example

Due to the demands in marine reflection seismics of recovering vast subsurface structures, we generated a 3D velocity model based on the 2D Marmousi model (cp. Versteeg), which resembles typical geological structures in a marine environment (Fig. 4a). The model is used as real model of the subsurface and covers a volume with the edge lengths \( (9.6 \times 4.8 \times 3.52) \) km. Based on it, the “recorded” data is computed. For this and for all seismic modellings, we need a finite-difference discretisation which comprises \( 960 \times 480 \times 352 \) grid points and \( 4600 \) time steps to obtain a recording length of 6 s. The reflection acquisition geometry is composed of 60 homogeneously distributed air guns (generating a source signal with a peak frequency of 9 Hz) and 1002 hydrophones (along

Figure 4. 3D acoustic experiment: (a) shows the true velocity model with seawater layer and complex geological structures. (b) depicts the initial model for FWT and the acquisition geometry at the sea surface (sources in red and receivers in green). (c) illustrates the reconstructed velocity model.
nine receiver lines). The initial model for FWT is given by a simple 1D gradient which represents an appropriate assumption in case of lacking a priori information (Fig. 4b).

The time-domain FWT approach is able to recover the velocity model at high resolution (Fig. 4c). Even deep high-velocity structures – composed of salt – are recovered sufficiently. In particular, a small-scale low-velocity zone within the complex fault structure bearing hydrocarbons, such as oil and gas, is recovered perfectly (at a depth of \( \approx 1.2 \) km in Fig. 4c). The time-domain FWT found a velocity model which provides a nearly perfect explanation of the observed data (not shown). In other words, for the given initial velocity model and acquisition geometry, it is impossible to obtain further improvements of the inversion result. Either we have to provide a better initial model or we have to consider more seismic data by increasing the number of sources.

The FWT was performed on 7680 CPU cores of JUROPA and required a total memory of 10.2 TB for wavefield storage (1.35 GB / core). The inversion result was obtained after 173 iterations with 12.3 minutes / iteration. The runtime amounts to 35.5 hours, which is equivalent to 273 000 core hours or 218 % of the monthly quota of project HKS01.

4.2 Reflection Seismic Experiment: Speedup Benchmark

We investigated the scaling behaviour of the 3D acoustic FWT code by performing a speedup benchmark on JUROPA. Therefore, we used a smaller version of the 3D model introduced in Sec. 4.1. The new model covers a volume with the edge lengths \((8 \times 3.2 \times 3.2)\) km (not shown), the finite-difference discretisation comprises \(800 \times 320 \times 320\) grid points and the acquisition geometry consists of 50 sources and 1206 receivers. All other parameters are adopted from Sec. 4.1. According to Sec. 3.3, the implementation uses both domain decomposition and shot parallelisation. The number of cores ranges within \(N_{\text{core}} = 160, \ldots, 8000\). That is, in case of pure domain decomposition the model domain is divided into \(160, \ldots, 8000\) subdomains. The combination with shot parallelisation employs a domain decomposition using 160 cores per source. The parallel wavefield simulation of all 50 sources involves 8000 cores. During forward-propagation, the pressure wavefield is stored at every 6th time step (766 3D snapshots). In dependence

![Figure 5. Speedup benchmark of a 3D experiment for different parallelisation strategies using 160, 320, 800, 1600, 4000 and 8000 cores on JUROPA. (a) shows the memory consumption for pure domain decomposition and the combination with shot parallelisation. (b) compares the computational times for one FWT iteration. (c) illustrates the scaling behaviour – the number of iterations computed within one hour – and compares with the perfect linear speedup (based on the computation with 160 cores).](image)
of the parallelisation strategy, the load varies over several orders of magnitude. The following paragraph only considers the memory required for wavefield storage. In particular, we need additional memory (order of magnitude: 100 MB/core) to store seismic data or 3D models, such as gradients or velocity models.

On the one hand, the memory consumption of the pure domain decomposition ranges from 1.46 GB/core ($N_{\text{core}} = 160$) down to 29.9 MB/core $N_{\text{core}} = 8000$, while it is constant in case of shot parallelisation (1.46 GB/core). On the other hand, the total memory usage – shown in Fig. 5a – is either constant (domain decomposition, 234 GB) or scales with the number of sources (shot parallelisation, 234 GB,..., 11.4 TB). Due to the restriction of MPI communication to blocks of 160 cores each, the shot parallelisation exhibits a nearly perfect scaling behaviour. In contrast, the domain decomposition suffers from an extensive communication overhead and shows a significant performance collapse at a very high number of cores. Figs. 5b,c illustrate both the computation time required for a complete FWT iteration and the corresponding scaling behaviour expressed by the number of iterations per hour.

5 Concluding Remarks

We have implemented 3D acoustic full waveform tomography (FWT) which benefits from efficient time-domain seismic modelling. We combine domain decomposition in finite-difference modelling with the demands of FWT, that is, we apply the shot parallelisation which allows a simultaneous computation of several modellings by reducing inter-node network communication. In a speedup benchmark – involving up to 8000 CPU cores – our FWT implementation shows a nearly perfect scaling behaviour. The 3D FWT has been applied successfully to a synthetic experiment. However, even this medium-sized problem clarifies, that a 3D FWT is able to exhaust computational resources of high-performance supercomputers with ease. In particular, the feasibility of 3D time-domain FWT is limited by the available amount of random access memory.

Acknowledgements

This work is supported by the Verbundnetz Gas AG and the sponsors of the Wave Inversion Technology consortium (WIT). The FWT computations were performed on the supercomputer JUROPA at Jülich Supercomputing Centre.

References


First-Principles View on Element and Isotope Cycles in the Earth’s Interior

Sandro Jahn, Volker Haigis, Piotr M. Kowalski, Daniela Künzel, Georg Spiekermann, and Johannes Wagner

GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany
E-mail: jahn@gfz-potsdam.de

Throughout Earth’s history, chemical elements and their isotopes have been continuously recycled and redistributed. The related chemical and transport processes are especially efficient in the presence of melts or fluids. We use molecular dynamics simulation techniques to investigate the thermodynamics of fluid- or melt-rock interactions and develop predictive models of equilibrium trace element partitioning and stable isotope fractionation between coexisting phases. This approach does not only require an accurate description of atomic interactions but also reliable structure models for the constituting minerals as well as for the melt or fluid phase at relevant pressure and temperature conditions of the Earth’s interior. On top of classical and ab initio molecular dynamics simulations used for a statistical sampling of fluid and melt structures we develop and employ advanced numerical methods to validate our models against experiments and to derive element and isotope partition coefficients.

1 Introduction

Minerals and rocks contain unique information about their geological history, which may date back to the early days of our planet billions of years ago. With modern analytical methods, the chemical and isotopic composition of a sample can be measured with high precision even for elements with very low concentration. Yet, it remains challenging to link this chemical record to the corresponding mineral- or rock-forming event(s). The most efficient geological processes involve at least some amount of liquid or supercritical fluid. In these cases element transport can be fast and proceed e.g. via dissolution of one mineral and precipitation of a new one. For a better understanding of geological processes we need to quantify their driving forces that are usually governed by thermodynamics as well as their mechanisms and rates.

Element partition and isotope fractionation coefficients between minerals, melts and fluids have been studied experimentally for many years. Such investigations are still challenging since they need to be performed at the often extreme conditions of pressure and temperature prevalent in the Earth’s interior. Moreover, measured thermodynamic properties do not provide any information about the molecular scale mechanisms that are responsible for the observed behaviour. Thus, structural and spectroscopic investigations in situ at high pressures and temperatures are required to shed light onto the chemical interactions that determine the atomic structure of materials. However, in most cases the latter experiments provide only fingerprint information about the sample and especially for disordered phases such as fluids and melts it is impossible to derive directly a three-dimensional structure model without ambiguity.

First-principles modelling techniques that are based on a quantum-mechanical treatment of the electronic structure of materials are unique in the sense that they provide simultaneous access to the atomic structure and dynamics and to thermodynamic properties.
Furthermore, they are often accurate and predictive over the whole range of pressures and temperatures and equally applicable to all types of Earth materials. The main limitation of such an \textit{ab initio} modelling approach is the huge computational demand, which often requires the use of supercomputer facilities. Still, compromises have to be made for the number of atoms in the simulation box (typically up to a few hundred), which puts constraints on the chemical complexity that can be modelled. The introduction of sensible approximations and the development of efficient numerical algorithms for a specific problem can help to open new perspectives on geological processes. Here, we summarise some of our recent work on the structure and dynamics of silicate melts and aqueous fluids and present efficient methods to predict the equilibrium fractionation of isotopes and trace elements between minerals, melts and fluids.

2 Methods

Electronic structure calculations are performed in the framework of density-functional theory (DFT) using a planewave basis set as implemented in the CPMD code\textsuperscript{1}. The exchange-correlation functional is treated in the generalised gradient approximation (GGA) and norm-conserving pseudopotentials were employed. The planewave cutoff is varied between 70-80 Ry for \textit{ab initio} molecular dynamics simulations and 140 Ry for calculations of vibrational frequencies. Car-Parrinello or Born-Oppenheimer molecular dynamics simulations are performed with time steps of 0.1 and 0.5 fs, respectively. System sizes are typically in the order of 100 to 400 atoms. Temperatures for the aqueous fluids are varied between 300 K and 1000 K. For silica-rich fluids, temperatures above 2000 K are needed to allow for breaking and formation of Si-O bonds\textsuperscript{2}. Force constants and vibrational frequencies are computed using the finite atomic displacement scheme after relaxing atomic positions to minimise the potential energy for a given structure. \textit{Ab initio} molecular dynamics simulations are computationally very demanding and require the use of high performance supercomputers such as JUQUEEN, where the CPMD code shows reasonable scaling up to \texttt{bg.size=1024} (16384 cores). For the calculation of frequencies, snapshots from the molecular dynamics simulations are chosen. The structural optimisation is efficiently performed on JUROPA.

For our explorative study of the effect of melt composition and structure on the mineral-melt trace element partition coefficients we perform classical molecular dynamics simulations using a polarisable ion potential\textsuperscript{3} in conjunction with the method of thermodynamic integration as implemented in the CP2K code\textsuperscript{4}. The interaction potentials for the Y-Ca-Al-Si-O system are optimised by fitting forces, ionic dipoles and stress tensors predicted by the classical force field to corresponding DFT data for a number of reference structures\textsuperscript{3}. The molecular dynamics simulation cells representing the silicate melts contain typically between 1000 and 2000 ions including a single ion of the trace element, i.e. Y$^{3+}$ in the present case. Thermodynamic integration is used to evaluate the equilibrium constant of an exchange reaction between two phases whereby the Y$^{3+}$ cation (trace element) is exchanged with an Al$^{3+}$ cation (main element). CP2K simulations using the classical potential show a good scaling behaviour on JUROPA of up to at least 16 nodes (128 cores).
3 Structure and Vibrational Properties of Silicate Melts and Supercritical Fluids

Before extracting thermodynamic properties from the simulations or making predictions of properties under experimentally unexplored conditions the model approach should be validated. In the case of fluids and melts at elevated temperatures this may be done by comparison to experimental data. Classical molecular dynamics results could also be compared to reference ab initio data. The structure factor $S(Q)$ of a melt or fluid that is measurable by neutron or x-ray diffraction and that provides information about the short- and intermediate-range atomic order is readily computed from averaging the particle positions along a molecular dynamics trajectory. Since $S(Q)$ is composed of contributions from partial correlations that are weighted with the concentrations and the scattering cross-sections of the individual elements, it is mainly representative of the major and/strongly scattering elements. Thus, diffraction is usually not very sensitive to minor components such as solutes in aqueous solutions or trace elements in melts.

Figure 1. Computed (left) and experimental (right) x-ray Raman spectra of liquid and supercritical water. Significant changes in the spectra with temperature are caused by variations in the molecular structure of water especially with respect to the hydrogen-bond network. The structure model obtained from ab initio molecular dynamics simulations seems to capture the essential changes observed experimentally (modified after Ref. 6).

Therefore, site-specific spectroscopies, such as x-ray absorption or Raman spectroscopy are needed. The comparison of results from molecular dynamics simulations with such experimental spectra is more complicated. For complex chemical systems, the interpretation of the measured spectra is not straight forward and therefore it seems preferable to compute theoretical spectra. X-ray absorption (or x-ray Raman) spectra of fluids are derived from individual snapshots of the molecular dynamics simulation by calculation of the electronic excitation spectrum at the corresponding absorption edge. For example,
the theoretical and the experimental x-ray Raman spectra of pure water evolve in a very similar way when the sample is heated from ambient and supercritical conditions\(^6\) (Fig. 1).

The calculation of theoretical Raman spectra of disordered materials at high temperatures is formally more challenging as they are related to temporal changes of the sample polarisability. The application of density-functional perturbation theory would require relaxation of all atomic positions, which is not only computationally demanding but may also lead to significant structural changes in the fluid phase. To avoid such ambiguities, we have developed a projection method to extract quasi-normal modes of molecular units such as \(\text{SiO}_4\) or \(\text{H}_2\text{O}\) in aqueous fluids or silicate glasses\(^7\). Spectra of these individual modes are very useful for the interpretation of Raman spectra of complex systems and especially help to assign modes of the measured spectra. The major drawback of this approach is that it does not provide any information about the Raman cross-sections, i.e. relative scattering intensities of individual modes. We therefore explore the possibility to derive Raman spectra directly from the time correlation function of the polarisability tensor using a molecular dynamics trajectory\(^8\) or to parametrise a bond polarisability model.

![Figure 2. Snapshots from ab initio molecular dynamics simulations of an \(\text{H}_2\text{O}-\text{SiO}_2\) fluid at temperatures of 2400 K (left) and 3000 K (right) and a pressure of about 4 GPa\(^2\). Si(yellow)-O(red) bonds and O-H(white) groups are shown as cylinders. The number of molecular \(\text{H}_2\text{O}\) (shown as balls and sticks) increases towards lower pressures and temperatures, which eventually leads to unmixing into an \(\text{SiO}_2\) bearing aqueous fluid and a hydrous silica melt. The release of volatiles from magmatic melts during decompression is an important cause for explosive volcanism.](image)

With the confidence that the predicted structure models for melts or fluids are reasonable, which is verified by comparing experimental and modelled spectral fingerprints of the studied system, we can now extend our predictive modelling approach to experimentally unexplored conditions. For instance, Fig. 2 shows two snapshots from \textit{ab initio} molecular dynamics simulations of a fluid with 32 \(\text{SiO}_2\) and 32 \(\text{H}_2\text{O}\) molecules at temperatures well above 2000 K and a pressure of about 4 GPa\(^2\). While at these extreme conditions a single fluid phase is thermodynamically stable, silica-bearing aqueous fluid and hydrous silica melt are expected to phase separate at lower temperatures and pressures below about 0.4 GPa. As a precursor of this unmixing, we observe a significant increase in the number of molecular \(\text{H}_2\text{O}\) when the temperature in the simulation is lowered from 3000 K to 2400 K\(^2\).
At even lower temperatures, the frequency of Si-O bond breaking or formation events becomes slower than the timescale accessible to molecular dynamics simulations\(^2\). Thus, advanced sampling methods will have to be used in the future to explore silicate-bearing model melts and fluids under geologically relevant conditions.

### 4 Prediction of Equilibrium Stable Isotope Fractionation

Stable isotopes are used as tracers of geochemical processes in the Earth’s interior. Assuming that the chemical interactions are unchanged when replacing one isotopic atom by another one of the same element, isotope fractionation between two phases is dominated by the mass effect on the vibrational frequencies of the system, which results in a small difference in the free energy. If the coordination environments of the isotopic element are different between two phases, the heavy isotope preferentially fractionates into the stronger bonded and usually less coordinated site. Isotope fractionation is a small effect and resulting concentration differences are measured in units of per mille.

While the stable isotope fractionation between molecules and crystals has been accurately predicted from electronic structure calculations for many systems\(^9\), explicit simulations of mineral-melt or mineral-fluid isotope fractionation coefficients at elevated pressures and temperatures are still rare. The two major challenges for a quantitative \textit{ab initio} prediction of isotope fractionation between minerals and melts or fluids are (1) the necessity to sample the structure of the disordered phase with sufficient statistical accuracy (which could be achieved by running long molecular dynamics simulations) and (2) to compute efficiently the reduced partition function ratio that defines the fractionation factor \(\beta\) between the isotopic atom in the phase of interest and the reference monatomic gas in the harmonic approximation\(^10\)

\[
\beta = \prod_i \frac{u^h_i}{u^l_i} \exp \left[ \frac{u^l_i - u^h_i}{2} \right] \frac{1 - \exp(-u^l_i)}{1 - \exp(-u^h_i)}
\]

where \(l\) and \(h\) denote the light and the heavy isotope and \(u_i = \hbar \nu_i / k_B T\) with \(h\) and \(k_B\) being Planck’s and Boltzmann’s constants. \(T\) and \(\nu_i\) are the temperature and the vibrational frequency of the \(i\)th degree of freedom. In a periodic system with \(N\) atoms in the simulation cell, there are \(3N\) degrees of freedom. The harmonic frequencies of a crystal are usually derived from a normal mode analysis that is performed on a structure with all atomic positions relaxed to local minima of the potential energy surface. However, this concept is less suited for high temperature fluids and melts due to the huge (often unaffordable) computational cost and possible changes in the local atomic structure of these systems during the relaxation procedure, which was already mentioned above in the context of computing vibrational spectra.

Major progress in this field was achieved recently when we presented a new approximate scheme for the prediction of isotope fractionation coefficients at high temperatures\(^11\). Instead of computing the full normal mode spectrum we only relaxed the isotopic atom or its nearest neighbour environment and derived three pseudo-frequencies from the force constants acting on the isotopic atom. This procedure reduces the computational cost for evaluation of the \(\beta\)-factor by at least two orders of magnitude for systems of about 100 atoms. Further, our new pseudofrequency approach performs much better than the well-known Bigeleisen and Mayer approximation\(^10\), which is demonstrated in Fig. 3 for two
molecules, $\text{H}_3\text{BO}_3$ and $\text{H}_4\text{BO}_4^-$. A similar good agreement between the full normal mode and our approximate approach is also achieved for condensed phases\cite{11}. So far, we have not only quantitatively reproduced experimental data of Li and B isotope fractionation between minerals and aqueous fluids but we could also provide new insights into relations between the molecular structure of a phase and its isotope fractionation behaviour and help to resolve some inconsistencies between experimental data and isotopic signatures of natural samples\cite{11,12}.

5 Melt Composition Effect on Trace Element Partitioning

Theoretical predictions of element partition coefficients between different phases is formally more complex than those of isotope fractionation. When one chemical element is replaced by another one the chemical interactions are affected. If ions of different charge are concerned, a charge compensating mechanism is needed, e.g. a coupled substitution or electronic charge transfer. On the other hand, the differences in free energy are usually much larger and element concentration ratios, i.e. partition coefficients, can be up to several orders of magnitude. A very successful approach to describe trace element partitioning between minerals and melts is the lattice strain model\cite{13} that expresses the partitioning behaviour in terms of an ideal ionic radius for the substitution site, Young's modulus of the host crystal and a baseline parameter. However, this model assumes that changes in the melt structure do not influence mineral-melt partitioning, which often is not the case\cite{14}.

In a pilot study we used classical molecular dynamics simulations to study changes in the local atomic environment of an $\text{Y}^{3+}$ ion in different silicate melts\cite{3}. A significant decrease of $\text{Y}$ coordination towards less polymerised silicate melts was observed (Fig. 4).
Using the method of thermodynamic integration, we derived exchange coefficients of Y and Al between two distinct melts. Assuming that Al is a major component in silicate melts that has a partition coefficient close to one, we estimated the change Y trace element partition coefficient as a function of melt composition. As a result, Y strongly fractionates into more depolymerised melts. Variations of the mineral-melt partition coefficient with different melts can be as large as two orders of magnitude, which is consistent with experimental results\textsuperscript{14}. In the near future we should be able to make such predictions using \textit{ab initio} methods.

6 Concluding Remarks

First-principles modelling approaches have become increasingly powerful tools to improve our understanding of the molecular structure and dynamics of melts and fluids at extreme pressures and temperatures. They provide unprecedented insight into the mechanisms by which elements and isotopes are distributed in the Earth’s interior during complex geological processes. This progress has been achieved mainly by feasibility of performing \textit{ab initio} molecular dynamics simulations of complex disordered systems on high performance supercomputers such as JUQUEEN and by the development and application of more efficient computational methodology.
Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft in the framework of the Emmy-Noether program, project JA1469/4-1. A large part of the simulations was performed on JUGENE, JUQUEEN and JUROPA at Jülich Supercomputing Centre using allocated computing time of NIC project HPO15.

References

4. CP2K, http://cp2k.berlios.de/
Direct Numerical Simulation of Turbulent Mixing in the Planetary Boundary Layer

Juan Pedro Mellado, Cedrick Ansorge, Alberto de Lózar, Jade R. Garcia, Tomas Keitzl, and Chiel van Heerwaarden

Max Planck Institute for Meteorology, Bundesstr. 53, 20146 Hamburg, Germany
E-mail: {juan-pedro.mellado, cedrick.ansorge, alberto.de-lozar, jade.garcia, thomas.keitzl, chiel.vanheerwaarden}@mpimet.mpg.de

Small-scale turbulence determines some relevant properties of the planetary boundary layer, like the way it grows into the upper troposphere or the details of the near-surface region. However, there are still significant inaccuracies in the parametrisation of these small-scale processes in weather and climate models. These inaccuracies are often due to a lack of understanding of the processes themselves, and not only to the way we represent their effect on a larger scale. This project combines high-performance computing with direct numerical simulation to gain new insight into long-standing, critical aspects of turbulent mixing in the planetary boundary layer and to reduce thereby those inaccuracies. In particular, we investigate open questions regarding (i) how the turbulent convective boundary layer grows into the non-turbulent troposphere (ii) the mechanisms of reduction, intermittency and collapse of turbulence in the stable boundary layer and (iii) turbulent mixing at the stratocumulus cloud-top and the role of local phenomena like evaporative cooling or radiative processes. This paper describes some aspects of these problems.

1 Introduction

The planetary boundary layer (PBL) is the lowest part of the atmosphere, the part that is in contact with the surface and that feels the cycle of night and day. Its most important property is that it is generally turbulent, and mixing on small spatial and temporal scales can be very important. Many aspects of small-scale mixing are poorly understood, but can have a great impact on the performance of weather and climate models. These are, for instance, how turbulence interacts with the surface, or the exact mechanisms that control the evolution of the PBL height. Our research addresses some of these open issues. The aim of this paper is to introduce some of these problems and describe the steps that we have taken to solve them.

As a tool, we use direct numerical simulation (DNS). DNS consists in solving as accurately as needed the partial differential equations governing the evolution of the mass, momentum and energy of a fluid or a fluids mixture – the Navier-Stokes equations. In a turbulent state, we need to account for the very broad range of spatio-temporal scales embracing almost all of the non-linearly interacting, three-dimensional modes. Hence, high-performance computing is required.

Significant contributions to the understanding of turbulence and turbulent mixing have been made by means of DNS during the last decades, though mainly in fundamental aspects and engineering problems. Geophysical systems are characterised by buoyancy effects and strong inhomogeneities, and supercomputers were, so far, not powerful enough to handle the amount of data necessary both in space and time. This situation has changed in the last decade. For instance, the range of scale separation between the large, energy
containing eddies and the small, dissipative motions that we can achieve with the IBM Blue Gene/Q supercomputer (JUQUEEN) at the Jülich Supercomputing Centre is already wide enough to retain, in addition, the gravity waves due to the stable stratification. Thus, we can investigate, with the required accuracy, the turbulence-wave interaction that characterises – and complicates – mixing in difference regions of the PBL.

In the following sections, we elaborate further these ideas using three different problems. First, we consider the convectively driven PBL and discuss the turbulent mixing and entrainment at the PBL top. Then, we address the wind-driven PBL and we investigate what happens when stable stratification increases, e.g., by cooling the surface. Last, we describe how cloud processes at the PBL top modify the local dynamics and how, despite the complexity of the problem, we can still use DNS in combination with high-performance computing to gain new knowledge, useful to improve weather and climate models.

2 Formulation

The atmosphere is described in terms of the velocity vector \( \mathbf{v}(x,t) \) and the buoyancy \( b(x,t) \) acting in the vertical direction \( Oz \). The small magnitude of the velocities found in the PBL and the small size of the domains considered in these DNS studies compared to the scale height of the atmosphere justifies an incompressible approach. Density variations smaller than 5% allows for the application of the Boussinesq limit. The governing equations are

\[
\begin{align*}
\frac{\partial \mathbf{v}}{\partial t} + \nabla \cdot (\mathbf{v} \otimes \mathbf{v}) + 2\Omega \mathbf{k} \times (\mathbf{v} - G\mathbf{i}) &= -\nabla p + \nu \nabla^2 \mathbf{v} + b \mathbf{k} \\
\nabla \cdot \mathbf{v} &= 0 \\
\frac{\partial \chi}{\partial t} + \nabla \cdot (\mathbf{v} \chi) &= \kappa \nabla^2 \chi \\
b &= b^*(\chi)
\end{align*}
\]

(1)

The kinematic viscosity is \( \nu \), \( \kappa \) is a scalar diffusivity, \( p \) is a modified kinematic pressure, \( \Omega \) is the angular velocity of the non-inertial system, \( G \) the corresponding geostrophic wind, and \( \mathbf{i} \) and \( \mathbf{k} \) are the unit vectors along \( Ox \) and \( Oz \), respectively. The field \( \chi \) represents transported scalars (e.g., the enthalpy, the total water content or some combination thereof). The buoyancy function \( b^*(\chi) \) needs to be specified depending on the problem: In the simplest case \( b^*(\chi) = \chi \), e.g., when the buoyancy is due to variation in the temperature), but more complex thermodynamics can be prescribed, e.g., to account for latent heat effects in cloud mixing processes.

The geometry consists of a horizontal layer of fluid and the system is statistically homogeneous inside horizontal planes. Periodicity is used as lateral boundary conditions, which is acceptable as long as the domain size is large compared to the auto-correlation lengths of the turbulent fields. Additional constraints are imposed depending on the particular configuration and are discussed in the corresponding section below.

High-order spectral-like compact finite-differences are used to discretise Eq. 1 on a structured grid, along with a low-storage fourth-order Runge-Kutta scheme for time advancement. The discrete solenoidal constraint is satisfied down to machine accuracy combining a Fourier decomposition along the horizontal directions with a factorisation of the resulting set of equations along the vertical coordinate.
In terms of implementation, the code is written in Fortran90 and uses hybrid parallelisation: MPI for the domain decomposition in $O_x$ and $O_y$, and OpenMP within each node for the main loops. The linear systems are solved very efficiently using LU decompositions, and the code employs the level 1 BLAS routines from the IBM ESSL library and profits from SIMD vectorisation. The library MPI-IO is used for parallel reading and writing.

3 Entrainment in the Convective Boundary Layer

One of the simplest PBL regimes is that of a dry, shear-free convective boundary layer (CBL) forced with a constant surface buoyancy flux $B_0$ and growing into a linearly stratified fluid with a constant buoyancy gradient $N^2$. This situation is commonly found during the daytime when the sun heats the surface. As the CBL thickens in time, tropospheric air from above is incorporated into the turbulent region below (Fig. 1). This process is called entrainment and it occurs in a finite size region called entrainment zone. Within this zone, the turbulence from the CBL is fighting against the stably stratified troposphere above. The CBL growth rate (or, similarly, the entrainment rate) can be well predicted, but other properties, like the entrainment-zone thickness or associated buoyancy scales, cannot. These details become relevant in cases of added complexity, like cloud formation or transition into a wind sheared CBL. Hence, different extensions of the basic models have been proposed in the literature, but results are still inconclusive\(^6\).\(^7\). This still-stand is partly due to the lack of accurate data, which motivates our research based on DNS.

Our results have already shown that, even in this most simple of circumstances, the entrainment zone comprises in reality two different sub-layers, instead of just one as assumed in previous analysis\(^8\). There is one upper sub-layer characterised by the top of the penetrating thermals alternating with the relatively smooth regions in between. This upper sub-layer is where the troposphere opposes most strongly the advance of turbulence. There is a second sub-layer beneath characterised by the troughs between those crests.

![Figure 1. Vertical cross-section showing the logarithm of the magnitude of the temperature gradient field in the convective boundary layer (CBL). The magnitude increases according to the sequence black-blue-yellow. The contrast and sometimes sharp transition between the smooth, gentle oscillations in the upper troposphere (upper part of the figure) and the turbulent motion below marks the entrainment zone, a region about 100 m deep, characterised by an alternating sequence of crests and troughs, and where tropospheric air is incorporated in the CBL (entrainment). The grid size is $5120 \times 5120 \times 840$ (only part of the domain is shown).](image-url)
and the thermal cores, both turbulent and relatively well mixed (Fig. 1). This dual-layer structure helps explaining the existing difficulties to parametrise the vertical structure of the entrainment zone, without having to resort to gravity-wave radiation, which was often found in previous analyse to play a minor role. In addition, we have provided an explicit parametrisation of this vertical structure in terms of the external parameters of the problem $B_0$ and $N$, and as a function of the state of development of the CBL, measured either in terms of elapsed time or in terms of the CBL height.

In turn, this newly found structure implies a new understanding of the equations describing entrainment-zone properties, like the entrainment rate. As mentioned before, the rate itself is well predicted, but the contribution from different terms to this rate was unclear. The accuracy of the DNS data has shown that, contrary to previous assumptions, the contribution from the finite size of the entrainment zone enters indirectly through a mean buoyancy term, instead of directly through a shape or deformation term.

We have also investigated the role of surface heterogeneity in the form of an alternating pattern of square patches with large and small heat fluxes. Starting from the limit in which the patch separation distance is large compared to the CBL height, so that we have individual CBLs on top of each hot patch, we have investigated the merging of these individual CBLs as the CBL height grows. The simulations are long enough to cover (i) the formation of a peak in kinetic energy corresponding to the “optimal” heterogeneity size with the strongest secondary circulations and (ii) the subsequent transition into a horizontally homogeneous CBL. This long time data has allowed us to explain the scaling of the transition times and their properties by means of the effect of the pattern geometry on the spectral characteristics of the flow.

4 Turbulence Collapse in the Stable Boundary Layer

Another common PBL regime is a wind-driven boundary layer in the presence of stable stratification, for instance, a wind current with velocity $G$ and a buoyancy difference $b_0$ between the free troposphere and the surface. Such a situation develops for instance at high latitudes when a mass of air is advected over relatively cold sea surface or over ice. Most weather models, including climate models, rely on flux parametrisations providing the lower boundary conditions and vertical diffusivities in the vicinity of the lower boundary. However, the correct prescription of those properties remains a challenge, in particular under very stable stratification, when turbulence become intermittent and can even collapse, and much of this difficulty stems from the lack of understanding of small-scale turbulence processes. We need data free from the uncertainty introduced by the different turbulence closure models, and this motivates a line of research based on DNS.

Some aspects of this stable boundary layer (SBL) can be investigated using an Ekman flow over a flat plate with a constant temperature as a simplified model (Fig. 2). One of the two controlling parameters quantifies the relative strength of stratification and we can modify it to cover different weather conditions. Thereby, we have shown that DNS reproduces all of the stages of turbulence reduction, intermittency and collapse observed to occur in nature as the stable stratification increases. Importantly, these results demonstrate that there is no need for a surface disturbance to create those transitions.

The second step has been the characterisation of the different stages or regimes in terms of a reduced set of bulk properties, like mean velocity and intensity of turbulence fluctu-
Figure 2. Vertical cross-section showing the logarithm of the magnitude of the temperature gradient field in the weakly stratified stable boundary layer (SBL). The magnitude increases according to the sequence black-grey-white. The wind blows from left to right. The lower panel is a zoom-in to the grey shaded rectangle in the upper one, to show the flow details. There is an evident vertical layering with very high gradients close to the surface. Confined bands of high gradients detach regularly from the surface and reach into the well-mixed region. Above this well-mixed region, we find a large-scale pattern of alternating turbulent bulges and areas of non-turbulent fluid entrained from above. Simulation performed using $3072 \times 6144 \times 512$ grid points (only part of the domain is shown).

ations. In particular, we have assessed the standard Monin-Obukhov theory used in the models and we have quantified deviations thereof, in particular, the strength of stratification at which this theory is no longer valid. As stratification is increased, the organisation of the flow in the outer regions of the boundary layer can penetrate all the way down to the near-wall region, where turbulence is mainly produced, playing thereby a crucial role in turbulence collapse. We have shown that we can partition the flow into turbulent and non-turbulent regions, and we have learned that the amount of kinetic energy that is associated with the non-turbulent fluctuations can become comparable to the contribution from the turbulent ones if the stratification is strong enough. This redistribution of energy can be associated with gravity waves because of the background stable stratification. Although this picture has been previously proposed, we have provided a methodology to quantify this partition and to perform the corresponding conditional statistical analysis. The next step is to incorporate this data into new, better parametrisations.

5 Entrainment at the Stratocumulus Cloud-Top

The small-scale features of the entrainment zone discussed in Sec. 3 become even more important in the presence of stratocumulus clouds (Fig. 3). The reason is as follows. First, stratocumulus strongly influence the radiation energy balance of the earth system because of their relatively large earth coverage ($\approx 20\%$). Thus, climate models need to predict accurately enough when and where these clouds form and disappear. However, the current uncertainty in these predictions is still of order one$^{12,13}$. Last, this uncertainty is traced back to errors in the representation of the small-scale turbulence at the cloud boundary that
mixes the cloud with the dry tropospheric air. New computational capabilities allow us to investigate some aspects of this process.

Cloud processes at the cloud-top boundary alter the local dynamics inside the entrainment zone compared to the dry case (Sec. 3). We focus on two processes: the cooling caused by the evaporation of droplets as the cloud mixes with the dry air above it, and on the cooling caused by the long-wave radiation from the cloud. The former results from mixing at scales of millimetres, the latter is associated to scales of meters. These two processes can cool the upper regions of the cloud strongly enough to render those layers heavier than the environment, and a convective instability can develop: cold parcels of cloudy air plummet down into the core of the cloud deck, promoting turbulence and reinforcing entrainment across the cloud boundary. It turns out that these local processes often become dominant in the stratocumulus-topped boundary layer, so that significant errors in their representation translate into significant errors in the representation of the PBL as a whole.

In our research, we use DNS to solve accurately idealised problems in which we can switch on and off those processes selectively, learning thereby how important these processes really are for the local turbulence properties and in particular for the entrainment rates (see Fig. 2). In this way, we have learned that evaporative cooling alone is too weak to explain the cloud dynamics observed in field measurements, closing thereby a long-standing debate on this issue. This result is independent of the droplet dynamics at a microscopic level, but a result of the bulk effect of evaporative cooling\textsuperscript{15}. In order to become important, evaporative cooling needs the interaction with other local, external processes that strengthen the mixing between the cloud and the upper tropospheric air. One of such processes is a mean wind vertical shear\textsuperscript{16}. Another example would be radiative cool-
Figure 4. Turbulence growth into the upper troposphere due to different cloud-top processes inside the entrainment zone of the stratocumulus-topped boundary layer (Fig. 3). Evaporation alone (blue line) cannot explain the measurements (black line), whereas radiation and, in particular, wind shear can. This result demonstrates the importance of the last two processes in the turbulence growth.

Actually, radiative cooling alone, without shear or evaporative cooling, can explain almost half of the entrainment-rate measurements\(^{17}\). We are further continuing this line of research towards a complete parametrisation of this cloud-top mixing layer.

6 Concluding Remarks and Outlook

We have presented a series of results to explain how our line of research improves our current understanding of small-scale turbulent mixing in the planetary boundary layer. In some cases, our research has led to new insights in the mechanisms that drive the planetary boundary layer; in some other cases, it has led to closing long-standing debates. The ability to simulate accurately the small-scale motions inside the planetary boundary layer, which is key for these results, relies on high-performance computing and it was impossible merely 10 years ago; the possibilities that will emerge in the coming decades are opening new frontiers in weather and climate research.

Acknowledgements

Support from the Max Planck Society through its Max Planck Research Groups program is gratefully acknowledged. Partial financial support was provided by the Deutsche Forschungsgemeinschaft, SPP 1276 Metström program. The authors gratefully acknowledge the Gauss Centre for Supercomputing (GCS) for providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS share of the supercomputer JUQUEEN at Jülich Supercomputing Centre (JSC).
References

Cutting the Edge of Regional Climate Models: Highly Resolved Climate Simulations in the Alpine Region

Heimo Truhetz, Andreas Prein, András Csáki, and Andreas Gobiet

Wegener Center for Climate and Global Change (WEGC), University of Graz, A-8010 Graz, Austria
E-mail: heimo.truhetz@uni-graz.at

Since their introduction in the late 1980s, regional climate models (RCMs) are widely applied in climate research and climate change impact studies. They are used to provide climate change data on model grids with \( \sim 10 \) km grid spacing or above. The project described pursues multiple objectives: (1) pushing RCMs forward to smaller (convection-resolving) scales, (2) build-up of knowledge and understanding for regional climate processes and their changes due to climate warming, and (3) generating top quality highly resolved climate simulations for the Alpine region and beyond. This article gives a brief status report on the on-going project work and provides an outlook on planned future actions.

1 Motivation

The impact of climate change represents a major challenge for human society in the 21st century.\(^1\) Potential characteristics of climate change like increased frequency of weather and climate extremes and climate change impacts such as droughts, floods, exacerbation of diseases, and adverse conditions for water supply, agriculture, tourism and other sectors\(^1\) threaten to cause enormous economic and social costs\(^2\). In order to avoid such costs in an early stage, the United Nations Framework Convention on Climate Change (UNFCCC) formulated an objective already back in 1992 “[…] to achieve stabilisation of greenhouse gas concentrations in the atmosphere that would prevent dangerous anthropogenic interference with the climate system”\(^3\). On the ground of this, the European Council formulated a climate target in 1996: “The Council believes that the global average temperature should not exceed \(2^\circ\)C above pre-industrial level […]”\(^4\) and the International Panel on Climate Change (IPCC) started to prepare greenhouse gas emission scenarios for estimating the behaviour of the climate system under future greenhouse gas concentrations via climate models.

In order to evaluate climate change risks and to develop mitigation and adaptation strategies decision makers as well as the general public need detailed information on future climate and its potential impacts. Depending on the application, the level of detail of required climate information looks differently\(^5\). For instance, in hydrological impact studies Maraun et al.\(^5\) define these needs as the correct representation of (1) intensities, (2) temporal variability, (3) spatial variability, and (4) consistency between different local-scale meteorological variables. Although global climate models (GCMs) have been developed over the last decades (e.g. the GCM ensemble of the Coupled Model Intercomparison Project (CMIP) was introduced by the World Climate Research Programme (WCRP) in 1995), GCMs are not able to fulfil these requirements\(^5\) due to their limited spatial resolution (horizontal grid spacings are larger than 100 km). In particular, GCMs cannot resolve circulation patterns leading to hydrological extreme events\(^6\) and hence GCMs need to be further downscaled for a better representation of such processes.
One of the most prominent ways to derive detailed information of climate change on local scale is to run regional climate models (RCMs), which are driven by global climate models (GCMs) in turn. This approach was introduced by Giorgi and Bates and has successfully been applied in well-known large scale projects like PRUDENCE, its successor ENSEMBLES or the “North American Regional Climate Change Assessment Program” (NARCCAP). At that time, RCMs generated climate scenarios with horizontal resolutions from 50 km to 25 km grid spacing. Since these projects were based on multi-model ensemble simulations, estimations of regional climate change uncertainties became available for the first time. Within national climate research programmes even higher resolutions have been achieved on the sub-continental scale (e.g. 10 km grid spacing in the Alpine region).

Due to the general progress in computing technology, higher resolutions became feasible also on continental scales. Nowadays, a total of 35 climate simulations with ~12 km grid spacing are currently conducted by 26 collaborating universities and institutes within the EURO-CORDEX initiative. EURO-CORDEX, which is coordinated by the Climate Service Center (CSC), Hamburg, and the Wegener Center for Global and Climate Change (WEGC), is the European branch of WCRP’s Coordinated Regional Downscaling Experiment (CORDEX).

While the application of RCMs with a grid spacing on the 10 km scale and above became internationally accepted, a further increase of the resolution towards convection-permitting scales (horizontal grid-spacing <4 km) is not straightforward: (1) RCMs (originally developed for coarser resolutions) need significant reconsideration, because relevant processes on formerly unresolved (parameterised) scales become resolved, (2) highly resolved land surface information is necessary as input for the lower boundary conditions, (3) highly resolved observational data for model evaluation only exist in exceptional cases (e.g. special observation campaigns) and hence alternative reference data have to be found (e.g. data from nowcasting systems), and (4) the demand of computational resources increases exponentially when grid spacing is reduced. A further difficulty lies in model evaluation: at smaller scales, one has to deal with the double penalty problem (i.e. shifts in time/space between modelled and observed processes which are limiting the applicability of traditional error statistics). In other words, the detection of the added value at such scales becomes complicated and asks for special statistical methods and highly resolved reference data. Nonetheless, these so-called convection-permitting climate simulations (CPSCs) are becoming more and more established (e.g. Hohenegger et al., Meissner et al.), because they have two major advantages: (1) deep moist convection, which is the most important process in the majority of extreme precipitation events, becomes resolved and (2) the representation of orography and surface fields, which interact closely with precipitation, is improved.

Our activities on the high performance computing (HPC) systems of the Jülich Supercomputing Centre (JSC), entitled “Cutting the Edge of Regional Climate Models: Highly Resolved Climate Simulations in the Alpine Region”, supports the build-up of knowledge and understanding for regional climate processes in the complex Alpine region and their changes due to climate warming. In particular, these themes are associated with the following goals:
• Advancing climate modelling techniques for local (convection resolving) scales (1 km to 3 km grid spacing) and analysing scale-dependent model performance.

• Investigating atmospheric conditions and related model resolved indicators for extreme events and regional-scale feedback processes.

• Generating high-resolution climate change scenarios for the Alpine region and beyond.

• Analysing and quantifying all uncertainty components of high-resolution Alpine climate scenarios.

2 Methods and Technical Aspects

Within our first computing time period (11/2009 - 10/2010) we began to use the HPC systems of JSC. We installed the Mesoscale Model of the Fifth Generation (MM5) as well as the COSMO model in Climate Mode (CCLM), a RCM that evolved from the “Lokal Modell (LM)” model of the German Weather Service (DWD) in 2002, on JUROPA. Meanwhile, we also have the successor of MM5, the Weather Research and Forecasting (WRF) model, in operational mode.

All three models are non-hydrostatic limited-area models of the atmosphere. They are using a non-hydrostatic approach to solve the governing equations, which allows them to be operated on small scales (<10 km grid spacing) and in complex terrain. Thereby, the variables of state of the atmosphere (three wind components, air pressure, temperature, and humidity) as well as cloud parameters, and turbulent kinetic energy are integrated in time. This allows the models to capture the full 3D dynamics, but also limits their applicability due to the amount of needed computing resources. Following the Courant-Friedrichs-Lewy (CFL) criterion the time-step depends on the grid spacing. In order to reduce the run time to an acceptable amount, the models are parallelised on distributed memory systems. By doing so, the model domain, which consists of a specified number of grid points, is split into several sub-domains in a way that each computing core has the same number of grid cells to process. Each core then processes the full model equations for its sub-domain. At each time step (or integration step) the current state (i.e. the values of all variables that are calculated) at adjacent sub-domains has to be exchanged within the cores. Hence, the run time of the model is a function of two limiting parameters: the number of cores and the network.

The number of cores and the influence of the network traffic depends on the problem size (i.e. the number of grid cells in the model). For instance, a usual simulation with a resolution of 10 km grid spacing covering the Alpine region consists of 125 × 111 × 32 = 440,000 grid cells (32 model levels). The observed run time of this simulation is decreasingly reduced the more cores are engaged. At a number of 128 cores no further run time reduction is achieved due to the increase of network traffic. The same behaviour in principle can be observed when the grid spacing is reduced to 3 km (300 × 200 × 32 = 180,000 grid cells). If the grid spacing of a given model domain is halved, the model’s run time needs to be multiplied by 8 (a factor of 4 for doubling the horizontal grid cells and a factor of 2 for halving the time-step). For instance, if the Alpine region should be simulated with 3 km grid spacing instead of 12 km (which is a factor of 4), the model run time will be 64 times the 12 km simulation.
1′920′000 grid cells). However, in that case the speedup factor is much larger (at least twice as large) in general, because the inter-procedural communication between the cores has a smaller impact on the overall run time. Hence, the more cores are engaged, the faster the network needs to be to make efficient use of the cores. Or in other words, there exists an optimal model configuration (in terms of cores and grid cells) for an optimised usage of a specific hardware given.

Depending on the number of grid cells, we typically use up to 256 cores on JUROPA in our simulations. For instance, for a 150-year climate simulation (period 1951 to 2100) with 50 km grid spacing for the European continent (EURO-CORDEX) we use 96 cores. That is enough to have the simulation finished within 50 days. Of course CPCSs take significantly longer: currently a 22 years simulation with 3 km grid spacing covering the Alpine region is running on 256 cores. It will take ∼200 days in total.

3 Results so Far

Thanks to JUROPA, WEGC was able to contribute to both, advancements in climate modelling techniques for local (convection-resolving) scales and generating high-resolution climate change scenarios for the Alpine region and beyond.

First systematic sensitivity studies on convection permitting scales (which will probably be the scale of the next generation climate simulation) have been conducted on JUROPA during the research project “Non-hydrostatic Climate Modelling, Part I (NHCM-1)”, funded by the Austrian Science Fund (FWF) (project number P19619-N10) and its embedded international model-inter-comparison campaign for RCMs, called “Local Climate Model Intercomparison Project, Phase 1” (LocMIP-1)\textsuperscript{15}. By means of three RCMs (CCLM, MM5, and WRF) operated with 10 km, 3 km, and 1 km grid spacing in two different test areas in the Eastern Alps, the performance of various model components, in particular physical parameterisations, were investigated. Sensitivity simulations in the field of highly resolved (12 km, 3 km grid spacing) regional climate modelling have also been conducted on JUROPA within the EU-FP7 project “Assessing Climatic Change and Impacts on the Quality and Quantity of Water (ACQWA)” (http://www.acqwa.ch) (project number 212250) to investigate the applicability of convection-permitting simulations for investigating small-scale hydrological extreme events (flash floods)\textsuperscript{22}.

Important results from that work can be summarised as follows: Diurnal cycles of convective precipitation (onset and maximum, cf. Fig. 1), small-scale extreme precipitation rates, the spatial distribution of precipitation on hourly time scales, and the size and/or shape of precipitation objects are improved in the Alpine region due to the high (convection-permitting) resolution with a grid spacing <4 km. In addition, near surface fields of temperature, wind (speed and direction), and relative humidity are captured in a more realistic way. This is the condensed output of three PhD thesis\textsuperscript{23−25}, two peer-reviewed papers\textsuperscript{15,26}, and numerous reports and presentations on international conferences. Browsing through these publications, it can also be preliminarily concluded, that convection-permitting simulations will not necessarily improve biases and (spatial and temporal) correlation coefficients. However, a more accurate implementation of turbulence raises expectations of improved model performance. Especially, for CCLM for which a new turbulence scheme is currently under development\textsuperscript{27} at the German Weather Service (DWD). Nonetheless, due to the resolved treatment of deep convection, CPCSs are sup-
posed to be suited well enough to allow (at least) process-oriented investigations of small scale processes and their respond to climate change.

Figure 1. Example diurnal cycle of the average MM5 JJA precipitation. The solid blue line depicts the total precipitation of the 10 km simulation, the red line the parameterised (convective scheme) part, and the orange line depicts the resolved part. The dashed blue line shows the diurnal cycle of the corresponded 3 km CPCS and the dotted blue line shows the precipitation of a 3 km simulation with a 10 km orography. The black line represents the observed precipitation and the grey back-ground depicts its daily variability. Adapted from Prein et al.15.

Concerning the generation of regional climate simulations, we contributed to the Austrian projects reclip:century11 and reclip:century II, both funded by the Austrian Climate Research Programme (ACRP) (project numbers A760437 and A963768), by means of a climate simulation (period 1960 to 2100) with 10 km grid spacing covering the Alpine region. In addition, we conducted two climate simulations (period 1950 to 2100) within the recently finished ACRP project “Next Generation Regional Climate Scenarios for the Greater Alpine Region (ReCliS:NG)” (project number B068694): one reaching the “2 °C target” and one with less ambitious mitigation measures. Both simulations were based on novel IPCC greenhouse gas scenarios, the so-called Representative Concentration Pathways (RCPs).28 This extended the ensemble of highly resolved climate simulations in the Alpine region and provided local climate information to the climate impact community. In addition, first EURO-CORDEX simulations for the European continent were conducted on JUROPA.

4 Outlook

Since we have started our activities on JUROPA, global fossil CO₂ emissions are lying constantly high in the upper half of the IPCC emission scenarios29 and have reached his-
torical peaks in series over the last couple of years: 9.17 Gt/y in 2010, 9.46 GtC/y in 2011, and 9.67 GtC/y in 2012. In the face of such high CO₂ emissions the question “What happens, if the 2 °C target is missed?” is more than just of academic interest. The situation becomes even more dramatic, because climate warming is assumed to provoke meteorological extreme events. This is related to the Clausius-Clapeyron relationship and its dependency from climate warming (known as Super Clausius-Clapeyron Scaling effect).

In order to find answers to current research questions, WEGC follows its strategy of conducting (1) research-oriented simulations for model development and investigating climate processes in the Alpine region and (2) application-oriented simulations providing state-of-the-art climate change simulations.

In the upcoming years, we will focus on the CPCSs’ ability to capture the shielding effect of the Alps massif within the FWF project NHCM-2 (project number P 24758-N29), the successor of NHCM-1. NHCM-2 has started in January 2013 and will last for three years. It focuses on error components and their characteristics with respect to orographically forced climate processes. Several model components (e.g., boundary-layer and turbulence, radiation, and convection schemes) will be systematically perturbed in order to contribute to model development and provide well-balanced model configurations for CPCSs of the next generation. NHCM-2 is a cooperation with CSC Hamburg and it is also supported by the German Weather Service (DWD) and the Leuphana University, Germany.

In addition we are planning to investigate climate change effects in a severely changing climate (way beyond the 2 °C target) and its impact on extreme precipitation events in the Alpine region. Due to the models’ capabilities, we will focus more on process-oriented investigations of extreme precipitation events by means of CPCSs. Our work will aim at (1) investigating synoptic and small scale climate change implications assuming the RCP8.5 scenario, which is associated with a global warming of +4.9 °C by the end of the 21st century, and (2) developing new highly resolved (3 km and 12 km grid spacing) climate change scenarios based on RCP8.5 for the Alpine region and Europe to complement EURO-CORDEX and the Austrian pool of climate simulations.

Acknowledgements

The simulations were performed with a total grant of computing time of 2.3 Mio coreh on JUROPA since 11/2009. We are very thankful to our ex-colleagues, N. K. Awan, A. Leuprecht, and M. Suklitsch, for their scientific, technical as well as organisational contributions.

References

2. International Climate Change Taskforce, Meeting the climate challenge (Institute for public Policy Research, London, UK; Center for American Progress, Washington, DC, USA; Australian Institute, Canberra, Australia, 2005).

21. R. Courant, K. Friedrichs, and H. Lewy, über die partiellen differenzengleichungen der mathematischen physik, Mathematische Annalen (in German) 1, 32–74, 1928.


23. N. K. Awan, Performance of state of the art high resolution regional climate models in the european alpine region, Ph.D. Thesis, Wegener Center for Climate and Global Change, University of Graz, Austria, 2011.


Computer Science and Numerical Mathematics
In this chapter the main topics concern the parallelisation of multigrid methods, adaptive meshes and load balancing.

In the contribution of M. Bolten about “Evaluation of a multigrid solver for 3-level Toeplitz and circulant matrices on Blue Gene/Q” the main focus lies on the parallel implementation of a multigrid method for special matrices. Weak and strong scaling results are presented. Depending on the structure of the matrices the implementation is based on a 3D or torus topology and maps the problem to the processor topology.

The scavenging process of a 2-stroke engine, including the combustion processes, is considered in the paper “Solving the reactive compressible Navier Stokes equations in a moving domain” by R. Klöfkorn, M. Nolte. The underlying mathematical model consists of the compressible Navier-Stokes equations in a time-dependant domain, due to the moving piston in the cylinder. For the spatial discretisation a third order discontinuous Galerkin approach with a limiter based stabilisation for the convective terms is used. Dynamically adaptive meshes increase the efficiency and the necessary dynamic load balancing is taken into account. The strong scaling of the MPI implementation shows an efficiency of more than 70% for 65356 cores on JUGE. This work was funded by the Baden-Württemberg-Stiftung.

In the paper “Scalability tuning of the load balancing and coupling framework FD4” by M. Lieber, W. E. Nagel, H. Mix the coupling of the well established COSMO code for atmospheric flows and spectral cloud micro physics model SPECS are considered. This model takes also water droplets, frozen particles and insoluble particles into account. Since these particles appear with a great spatially and temporally varying presence, dynamic load balancing is necessary, which is done by the FD4 part of the code. In this contribution a new algorithm for coupling the data of COSMO and SPECS and the organisation of the dynamic load balancing is presented. The final results show that the overhead of the dynamic load balancing can be considerably reduced. This work was funded by the German Research Foundation (DFG).
Evaluation of a Multigrid Solver for 3-Level Toeplitz and Circulant Matrices on Blue Gene/Q

Matthias Bolten
Bergische Universität Wuppertal, Fachbereich C – Mathematik und Naturwissenschaften,
Gaußstraße 20, 42119 Wuppertal, Germany
E-mail: bolten@math.uni-wuppertal.de

Multigrid methods are optimal solvers for a broad range of applications, including certain banded Toeplitz and circulant matrices. We present a highly scalable parallel implementation of a solver for such systems as well as the underlying theoretical foundations and evaluate it on the Blue Gene/Q architecture. Strong scaling results were obtained on up to 1 rack, i.e., 16384 cores, weak scaling results on up to 4 racks. The method and its implementation show a perfect weak scaling behaviour and the strong scaling results show that the application of the method is feasible even for small system sizes.

1 Introduction

In many applications in computational science and engineering the solution of a linear system

\[ Ax = b, \]

where \( A \in \mathbb{C}^{n \times n} \) and \( x, b \in \mathbb{C}^n \) is sought for. Due to the cost of direct solvers often iterative solvers like Krylov subspace methods are used to solve these systems. Their convergence rate depends on the condition number of the system matrix that in many applications grows when the system size is increased, resulting in a higher iteration count that is necessary to solve the system up to the desired accuracy. Multilevel solvers like multigrid methods often do not pose this problem but rather show a convergence rate that is independent of the system size. For this reason multigrid methods are used in many different applications and the demand for scalable parallel implementations is very high. Different implementations exist, e.g., the multigrid solvers in the hypre package\(^1,2\) or the Trilinos package\(^3\). Most of these solvers focus on the solution of linear systems arising from the discretisation of partial differential equations. While the approaches used in these solvers work very well for these problems and other systems with similar properties, e.g., Laplacians on graphs, there are other problems for which plain multigrid methods, including algebraic multigrid, will not work well. Examples include the discretisation of integral operators like they arise in signal and image processing. However, it is possible to construct multigrid methods sharing the same fundamental ideas as the aforementioned methods for problems that are not like discretisations of partial differential equations. This includes methods for Toeplitz matrices and matrices from matrix algebras like the circulant matrix algebra, where the so-called generating symbol possesses certain properties.

Multigrid goes back to Brakhage\(^4\), Fedorenko\(^5,6\) and Bakhvalov\(^7\) and later to Brandt\(^8\) and Hackbusch\(^9,10\). The idea was then extended to more general systems of equations, resulting in algebraic multigrid (AMG)\(^11\) and aggregation-based multigrid\(^12,13\). Based on
the same ideas as classical AMG, multigrid methods for Toeplitz matrices have been developed by various authors\textsuperscript{14–16}, methods for circulant matrices were derived by Serra and co-workers\textsuperscript{17, 18}.

Being an efficient solver for problems involving various PDEs, multigrid methods have been parallelised for many years. Parallel multigrid methods have been used in many different applications, including computational fluid dynamics in 2d\textsuperscript{19} and 3d\textsuperscript{20}, or groundwater flow simulation\textsuperscript{21}. Today, parallel multigrid solvers are included in packages like the aforementioned hypre\textsuperscript{1, 2}. A survey of parallelisation techniques can be found in the article by Chow et al.\textsuperscript{22}.

Before describing the evaluated implementation, we will briefly review the theoretical foundations of multigrid in general and multigrid for Toeplitz and circulant matrices.

### 2 Multigrid Methods for Toeplitz Matrices and Circulant Matrices

A Toeplitz matrix \( T_n \in \mathbb{C}^{n \times n} \) is a matrix with constant entries on the diagonals, i.e., \( T_n \) is of the form

\[
T_n = \begin{pmatrix}
    t_0 & t_{-1} & \cdots & t_{-n+1} \\
    t_1 & t_0 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & \ddots \\
    t_{n-1} & t_{n-2} & \cdots & t_0
\end{pmatrix}.
\]

A Toeplitz matrix is closely related to its generating symbol \( f : \mathbb{R} \rightarrow \mathbb{C} \), a \( 2\pi \)-periodic function given by

\[
f(x) = \sum_{j=-\infty}^{\infty} t_j e^{i2\pi jx}.
\]

The generating symbol induces a sequence \( \{T_n(f)\}_{n=1}^{\infty} \) of Toeplitz matrices \( T_n(f) \), where the \( T_n(f) \) are band matrices in the case that \( f \) is a trigonometric polynomial.

Circulant matrices are the “periodic cousins” of Toeplitz matrices. For a circulant matrix we have \( t_{-k} = t_{n-k} \), \( k = 1, 2, \ldots \), i.e.,

\[
C_n = \begin{pmatrix}
    t_0 & t_{n-1} & \cdots & t_1 \\
    t_1 & t_0 & \ddots & \vdots \\
    \vdots & \ddots & \ddots & \ddots \\
    t_{n-1} & t_{n-2} & \cdots & t_0
\end{pmatrix},
\]

further \( C_n \) is diagonalised by the Fourier matrix \( F_n \) and the eigenvalues are given by a sampling of the generating symbol, so in contrast to the Toeplitz case the circulant matrices form a matrix algebra. Like above, a generating symbol \( f \) induces a sequence \( \{C_n(f)\}_{n=1}^{\infty} \) of circulant matrices \( C_n(f) \).

Both concepts can be easily extended to the block case: A \( d \)-level Toeplitz or circulant matrix is a Toeplitz/circulant block matrix that has \( (d-1) \)-level Toeplitz/circulant blocks. The generating symbol of such a matrix is a \( d \)-variate function. In the case \( d = 2 \) these matrices are often referred to as block Toeplitz Toeplitz block (BTTB) or block circulant circulant block (BCCB) matrices, here we deal with 3-level matrices.

If stationary iterative methods like Jacobi are used to solve a linear system, the convergence rate degrades as the system becomes more ill-conditioned. This is the case in many applications, e.g. when a PDE is discretised using a finer mesh-spacing. The reason for this
Algorithm 1 Multigrid cycle $x_{n_i} = MG_i(x_{n_i}, b_{n_i})$

$$
x_{n_i} \leftarrow S_i^{ν_1}(x_{n_i}, b_{n_i}) \\
r_{n_i} \leftarrow b_{n_i} - A_i x_{n_i} \\
e_{n_{i+1}} \leftarrow R_i r_{n_i} \\
\text{if } i + 1 = l_{\max} \text{ then} \\
e_{n_{l_{\max}}} \leftarrow A_{l_{\max}}^{-1} r_{n_{l_{\max}}} \\
\text{else} \\
f \text{or } j = 1, \ldots, ξ \text{ do} \\
e_{n_{i+1}} \leftarrow MG_{i+1}(e_{n_{i+1}}, r_{n_{i+1}}) \\
\text{end for} \\
\text{end if} \\
e_{n_i} \leftarrow P_i e_{n_{i+1}} \\
x_{n_i} \leftarrow x_{n_i} + e_{n_i} \\
x_{n_i} \leftarrow \tilde{S}_i^{ν_2}(x_{n_i}, b_{n_i})
$$

is that error components belonging to large eigenvalues are damped efficiently, while error components belonging to small eigenvalues get damped slowly, so Jacobi is a “smoother”. This motivates the introduction of another, coarser grid, as the error is smooth and well-represented on this coarser grid. The residual $r = b - Ax$ is computed on the finer grid, transferred to the coarser grid and the residual equation $Ae = r$ is solved on the coarser level. After transferring the error back to the finer grid, the current approximate solution can be updated. Recursive application of this two-grid idea yields a multigrid method.

To construct a multigrid method to solve Eq. 1 where now $A$ is a 3-level Toeplitz or circulant matrix, denote the system on the finest level by $A_0 = A$. Further denote the multi-indices of the system sizes on each level by $n_i < n_{i-1}$, $i = 0, \ldots, l_{\max}$. Here $l_{\max}$ is the maximum number of levels used, the number of unknowns on each level is then given by $N_i = \prod_{d=1}^{l_{\max}} (n_i)$. To transfer quantities from one level to another restriction operators $R_i : \mathbb{C}^{N_i} \rightarrow \mathbb{C}^{N_{i+1}}$, $i = 0, \ldots, l_{\max} - 1$ and prolongation operators $P_i : \mathbb{C}^{N_{i+1}} \rightarrow \mathbb{C}^{N_i}$, $i = 0, \ldots, l_{\max} - 1$ have to be defined, furthermore a hierarchy of operators $A_i \in \mathbb{C}^{N_i \times N_i}$, $i = 1, \ldots, l_{\max}$ is needed. After choosing appropriate smoothers $S_i$ and $\tilde{S}_i$ and the numbers of smoothing steps $ν_1$ and $ν_2$ on each level, the multigrid method $MG_i$ is given by Algorithm 1.

Now, to construct a multigrid method for Toeplitz and circulant matrices, let $A_i$ be Toeplitz or circulant and let $f_i$ be the generating symbol of $A_i$. Assume that $f_i \geq 0$ with a single isolated zero of even order and thus that $A_i$ is positive (semi-)definite and assume that in the Toeplitz case each component of $n_0$ is a power of 2 minus 1, and in the circulant case that each component of $n_0$ is a power of 2. These choices guarantee that the resulting coarse grid matrix will again be Toeplitz respectively circulant.

The common choice for the pre- and postsmoothing iteration is relaxed Richardson, so the pre- and postsmoother are given by

$$S_i(x_{n_i}, b_{n_i}) = \left( I - ω_i A_i \right) x_{n_i} + ω_i b_{n_i},$$

with the option of using different relaxation parameters $ω_i$ and $\tilde{ω}_i$. For a proper choice of
relaxation parameters these fulfil the pre- and postsmoothing property\(^1\).

To reduce a quantity from a fine to a coarse level cut matrices \(K_{n_i} \in \mathbb{C}^{n_{i+1} \times n_i}\) are introduced, these are given by

\[
K_{n_i} = \begin{bmatrix}
0 & 1 & 0 \\
1 & 0 & \ddots \\
& \ddots & 1 \\
& & & 1 & 0
\end{bmatrix}.
\]

in the Toeplitz case and result in skipping every other component starting with the first one when a vector is transferred to the coarse level, the cut matrix in the circulant case misses the first zero column. In the 3-level case the cut matrix is defined by the Kronecker product

\[
K_{n_i} = K_{(n_i)_1} \otimes K_{(n_i)_2} \otimes K_{(n_i)_3}.
\]

Given a generating symbol \(p_i\) the prolongation operator is defined as

\[
P_i = C_{n_i}(p_i)K_{n_i}^T
\]

in the circulant case and similar in the Toeplitz case. If the order of the zero \(x^0\) of the generating symbol is \(2q\), \(p_i\) is usually being chosen as

\[
p_i(x) = c \cdot \prod_{j=1}^{d}(\cos(x^0_j) + \cos(x_j))^q.
\]

3 Parallel Implementation

As the underlying structure is either a cuboid or a 3d torus the whole implementation is based on either a 3d cartesian mesh or torus processor topology. The components of the vectors are mapped to the corresponding processor such that the number of components in each direction is as evenly distributed amongst the processors, as possible. To implement a multigrid method only matrix vector multiplications with matrices \(A := A(f)\), where \(A = \mathcal{F}\) or \(A = \mathcal{C}\), and restriction and prolongation of vectors is needed. The implementation is targeting banded matrices and as the 3-level matrices under investigation can be written as sums of tensor products of 1-level matrices, we only describe the one level case, here.

For a 1-level circulant matrix with fixed bandwidth \(m\), independent of the system size \(n\), to calculate the \(i\)-th entry of the matrix-vector product we only need the information of the entries that have indices from \(\mathcal{I}_{i,m,n}\), where

\[
\mathcal{I}_{i,m,n} = \{(i - m) \mod n, \ldots, (i - 1) \mod n, i, (i + 1) \mod n, \ldots, (i + m) \mod n\},
\]

i.e., we have

\[
(Ax)_i = \sum_{j \in \mathcal{I}_{i,m,n}} a_{i,j}x_j.
\]

In order to evaluate this product in parallel it is favourable to have as much of this information stored locally as possible. Therefore we choose to distribute the vector over the
processors block-wise, i.e., when we have \( p \) processors the \( i \)-th processor gets the components ranging from \((i - 1)\lceil n/p \rceil \) to \( \min\{i\lceil n/p \rceil , n\} \). Using this distribution the processors are logically arranged in a 1d torus and they only have to exchange components with \( \lceil m/\min\{\lceil n/p \rceil , n - (p - 1)\lceil n/p \rceil \}\rceil \) neighbours in a one-dimensional torus. Obviously for the Toeplitz case a 3-dimensional mesh is sufficient.

Going down to the coarser levels, the locality of a variable on the fine level determines on which processor the coarse level variable will be located. The variables on the coarse level are located on the same processor as their fine grid counterpart. This leads to a structured communication scheme on the lower levels. Starting with a 3-dimensional torus we have communication with the next neighbours holding the \( m \) needed components, as long as all processors still have variables to treat. At some point, namely when the number of processors in one direction is bigger than the number of unknowns, we will have idle processors, which do not hold any variable on that level anymore. These processors then have to be ignored when the communication takes place. Technically we tackle this issue by storing the neighbourhood information on each level. In the initialisation step processors ask the neighbours of the previous level which neighbour they should use on this level. The asked processor answers this question with its own id, if it still has to do work, or with its own neighbour. This scheme requires that only every second processor may become idle per level. This is guaranteed if the unknowns are equally distributed on the finest level at the beginning.

The whole method is implemented in C99 using MPI as communication library.

4 Results

As our implementation uses a 3d-torus or a 3d-mesh it is particularly well-suited for IBM’s Blue Gene architecture. While the code was originally developed with Blue Gene/L and Blue Gene/P in mind, we now evaluated the scalability of the implementation on the Blue Gene/Q system JUQUEEN at the Jülich Supercomputing Centre. For our tests we chose

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\text{32x32x32 unknowns per core} & \text{16 ranks per node} & \text{32 ranks per node} & \text{64 ranks per node} \\
32 & 64 & 128 & 256 & 512 & 1024 & 2048 & 4096 & \\
\text{time [s]} & & & & & & & & \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\text{64x64x64 unknowns per core} & \text{16 ranks per node} & \text{32 ranks per node} & \text{64 ranks per node} \\
32 & 64 & 128 & 256 & 512 & 1024 & 2048 & 4096 & \\
\text{time [s]} & & & & & & & & \\
\end{array}
\]

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\text{128x128x128 unknowns per core} & \text{16 ranks per node} & \text{32 ranks per node} & \text{64 ranks per node} \\
32 & 64 & 128 & 256 & 512 & 1024 & 2048 & 4096 & \\
\text{time [s]} & & & & & & & & \\
\end{array}
\]

Figure 1. Weak scaling results obtained on the Blue Gene/Q system JUQUEEN at the Jülich Supercomputing Centre. While the method shows a perfect weak scaling behaviour, for these system sizes the use of additional ranks does not yield an improvement.
Figure 2. Strong scaling results obtained on the Blue Gene/Q system JUQUEEN at the Jülich Supercomputing Centre. For large system sizes the use of additional ranks is beneficial.

Toeplitz matrices with the generating symbol given by

\[ f(x, y, z) = 24 - 4(\cos(x) + \cos(y) + \cos(z)) - \\
4(\cos(x) \cos(y) + \cos(x) \cos(z) + \cos(y) \cos(z)), \]  

related to a 4th-order compact discretisation of the Laplacian that also alters the right hand side\textsuperscript{23}. This symbol has a single isolated zero of order 2 at the origin, so it fulfils the
requirements of the theory. As the solver is based on the matrix, only, no rediscretisation is
used and the special treatment of the right hand side does not have to be taken into account
on the coarser levels, a benefit of the algebraic formulation of the method.

The weak scaling results are shown in Fig. 1. We have tested the implementation with
approximately $32 \times 32 \times 32$, $64 \times 64 \times 64$, and $128 \times 128 \times 128$ unknowns per core.
As the tested systems were Toeplitz systems some processors have one variable less in
some directions as the overall system size is a power of two minus one. For the current
implementation the use of additional ranks on each node of the Blue Gene/Q system does
not improve the overall performance but rather the overall time is increased.

Additionally, the strong scaling of the implementation was evaluated. In Fig. 2 the
results for systems with $63^3$, $127^3$, $255^3$, and $512^3$ unknowns are depicted. For the small
system with $63^3$ unknowns using 1024 nodes, i.e., $32 \cdot 32$ nodes, still results in a speedup of $\approx 2.9$. We like to emphasise that for this system size each core only has at most 16
unknowns on the finest level. For larger system sizes the speedup is much better, as ex-
pected. Additionally, for the larger systems the use of additional ranks to make use of the
multi-threaded architecture is beneficial.

5 Conclusion

We tested a parallel implementation of a multigrid method for 3-level Toeplitz and circulant
matrices. The implementation uses a 3d mesh or torus topology, depending on the type of
matrix, and maps the problem to the processor topology. The presented results show a very
good scaling behaviour. For larger systems, the use of additional ranks is beneficial, so
we expect that a hybrid parallelisation will be beneficial on the Blue Gene/Q architecture
compared to the current implementation

References

1. R. D. Falgout, J. E. Jones, and U. M. Yang, “The design and implementation of hypre,
a library of parallel high performance preconditioners”, in: Numerical Solution of
Partial Differential Equations on Parallel Computers, A. M. Bruaset and A. Tveito,
(Eds.), vol. 51 of Lecture Notes in Computational Science and Engineering, chapter 8,
2. R. D. Falgout and U. M. Yang, hypre: a Library of High Performance Precondition-
ers, in: Computational Science - ICCS 2002 Part III, P. M. A. Sloot, C. J. K. Tan., and
J. J. Dongarra, (Eds.), vol. 2331 of Lecture Notes in Computer Science, pp. 632–641,
3. M. Heroux, R. Bartlett, V. H. R. Hoekstra, J. Hu, T. Kolda, R. Lehoucq, K. Long,
R. Pawlowski, E. Phipps, A. Salinger, H. Thornquist, R. Tuminaro, J. Willenbring,
4. H. Brakhage, Über die numerische Behandlung von Integralgleichungen nach der
5. R. P. Fedorenko, A relaxation method for solving elliptic difference equations, USSR
Solving the Reactive Compressible Navier-Stokes Equations in a Moving Domain

Robert Klöfkorn\(^1\) and Martin Nolte\(^2\)

\(^1\) National Center for Atmospheric Research, Boulder, CO, USA
E-mail: robertk@ucar.edu
Supported by the DOE BER Program under award DE-SC0006959.

\(^2\) Abteilung für Angewandte Mathematik, Universität Freiburg, Freiburg, Germany
E-mail: nolte@mathematik.uni-freiburg.de
Supported by the Baden-Württemberg Stiftung under contract HPC-11.

In this paper we describe the reactive compressible Navier-Stokes equations in a moving domain. Our application is the simulation of the scavenging process of a 2-stroke engine including the combustion process. The implementation of the solver is based on the Discontinuous Galerkin method and the software framework DUNE. Preliminary numerical results include the combustion of methane triggered by the compression of the fluid.

1 Introduction

In this project the continued development of an interface library for solving partial differential equations especially evolution equations is the main aspect.

These systems are used in many different fields of applications from astrophysics to industrial problems to model the evolution of physical quantities \(U\), e.g., density, momentum, and energy. A very general form of these equations is

\[
\partial_t U(x,t) + \nabla \cdot (F(U(x,t), x, t) - A(U(x,t), x, t)\nabla U(x,t)) = S(U(x,t), x, t). \tag{1}
\]

This is an advection-diffusion equation with a source term that represents chemical reactions.

In this paper we describe the reactive compressible Navier-Stokes equations in a moving domain. Our application is the simulation of the scavenging process of a 2-stroke engine including the combustion process. We will briefly describe the mathematical model, the numerical discretisation, the software, some performance results, and preliminary numerical results. A more detailed simulation will be reported in a follow-up paper.

Our main goal is the further development of a general framework for solving evolution equations based on modern software design techniques. Here, efficiency and reuse of code are major aspects in the design process. Since an important part of our work is the design and the improvement of numerical methods, the underlying code should be easy to apply to different test cases, model equations in one, two, or three space dimensions, and must allow for easy comparison of new methods with established ones. Also, due to the complexity of the key applications mentioned above, parallelisation combined with local grid adaptation is a central ingredient for the development of efficient numerical schemes. To be able to concentrate on the considered applications and on the design of new numerical schemes, aspects such as dynamic load balancing in a distributed computational environment should not be a concern to the researcher working on the project; the software library we are
developing should handle all these aspects in a transparent manner. Furthermore, progress made in this project contributes to usability of DUNE for all DUNE users, for example, the implementation of a general approach for handling moving grids that was done in this project.

## 2 Mathematical Model

We consider the evolution of a gas mixture of \( n \) species and denote by \( \vec{\chi} \in \mathbb{R}^n \left[ \frac{mol}{m^3} \right] \) the vector of concentrations. Given the diagonal matrix \( W \in \mathbb{R}^{n \times n} = \text{diag}(w_1, \ldots, w_n) \) consisting of the molar weights \( w_i \left[ \frac{kg}{mol} \right] \) of the \( n \) individual species, the vector of mass densities \( \vec{\rho} \in \mathbb{R}^n \left[ \frac{kg}{m^3} \right] \) is given by \( \vec{\rho} = W \vec{\chi} \).

For the fluid dynamics, we only consider the evolution of the gas mixture. To this end, we introduce the total mass density \( \bar{\rho} = \sum_i \rho_i \), the total momentum density \( \bar{\rho} \vec{v} \left[ \frac{kg}{m^2 s} \right] \) with the fluid velocity \( \vec{v} \left[ \frac{m}{s} \right] \), and the total energy density \( \bar{\rho} E \left[ \frac{J}{m^3} \right] \).

To model the evolution of the gas mixture in the piston we consider the compressible Navier-Stokes equations in a time-dependent domain \( \Omega_t \subset \mathbb{R}^d \), \( \mathcal{Q}_T := \{(x, t) \mid x \in \Omega_t, t \in [0, T]\} \),

\[
\begin{align*}
\partial_t \bar{\rho} + \nabla \cdot (\bar{\rho} \vec{v}) &= 0, \\
\partial_t \bar{\rho} \vec{v} + \nabla \cdot (\bar{\rho} \vec{v} \otimes \vec{v} + p \mathbb{I}) - \nabla \cdot \tau &= 0, \\
\partial_t \bar{\rho} E + \nabla \cdot (\bar{\rho} E \vec{v} + \rho \vec{v}) - \nabla \cdot (\tau \vec{v} + \vec{q}) &= 0, \\
\partial_t \vec{\rho} + \nabla \cdot (\vec{\rho} \otimes \vec{v}) - \nabla \cdot j &= W \dot{\vec{\chi}} \end{align*}
\]

in \( \mathcal{Q}_T \) (2)

with pressure \( p \), viscous stress tensor \( \tau \), heat flux \( \vec{q} \), chemical diffusion flux \( j \), and chemical source term \( \vec{\dot{\chi}} \). \( \mathbb{I} \) denotes the identity matrix.

For the pressure \( p \), we assume the ideal gas law

\[ p = \bar{\chi} R T = \bar{\rho} R_s T, \]

where \( \bar{\chi} = \sum_i \chi_i \) is the total concentration, \( R \left[ \frac{j}{mol \cdot K} \right] \) is the universal gas constant, and \( T \left[ K \right] \) is the temperature of the gas mixture. The specific gas constant \( R_s \left[ \frac{j}{kg \cdot K} \right] \) is given by \( R_s = \frac{R}{\bar{\omega}} \), where the mean molar mass \( \bar{\omega} \left[ \frac{kg}{mol} \right] \) is defined by the relation \( \bar{\rho} = \bar{\omega} \bar{\chi} \).

The temperature is related to the specific internal energy density \( e = E - \frac{1}{2} |\vec{v}|^2 \) by the relation

\[ e = c_v T + \vec{h}^0 \cdot \bar{\chi}, \]

where \( c_v \left[ \frac{j}{kg \cdot K} \right] \) denotes the specific heat capacity of the gas mixture at constant volume and \( \vec{h}^0 \left[ \frac{j}{mol} \right] \) denotes the vector of standard enthalpies of the gas mixture. We obtain the specific heat capacities \( c_v \) and \( c_p \) (specific heat capacity at constant pressure) from the relations \( R_s = c_p - c_v \) and \( \gamma = \frac{c_p}{c_v} \), where \( \gamma \approx 1.4 \) is the adiabatic exponent.

Putting all these relations together, we obtain the following formula for the pressure:

\[ p = (\gamma - 1) \bar{\rho} \left( E - \frac{1}{2} |\vec{v}|^2 - \vec{h}^0 \cdot \bar{\chi} \right). \]

The viscous stress tensor is given by

\[ \tau = 2 \mu D \vec{v} - \frac{2}{3} \mu (\nabla \cdot \vec{v}) \mathbb{I}, \]
where \( \mu \left[ \frac{kg}{m \cdot s} \right] \) is the dynamic viscosity and \( D \vec{v} = \frac{1}{2} (\nabla \vec{v} + \nabla^T \vec{v}) \) is the symmetric velocity gradient. For the dynamic viscosity \( \mu \) we use the Sutherland model introduced in Ref. 18

\[
\mu = \mu_0 \left( \frac{T}{T_0} \right)^2 \frac{T_0 + C}{T + C},
\]

where \( T_0 \) is a reference temperature, \( \mu_0 \) is the dynamic viscosity at temperature \( T_0 \), and \( C \left[ K \right] \) is the Sutherland constant. For the sake of simplicity we use the values \( T_0 = 291.15 \, K \), \( \mu_0 = 1.827 \times 10^{-5} \, \frac{kg}{m \cdot s} \), and \( C = 120 \, K \) for air.

The heat flux \( \vec{q} \) is given by

\[
\vec{q} = \lambda \nabla T,
\]

where \( \lambda \) is the thermal conductivity of the gas mixture. For the sake of simplicity, we assume the Prandtl number \( Pr = \frac{c_p}{\mu \lambda} \) to be constantly 0.72. This ensures that the thermal conductivity still depends on the gas mixture through the specific heat capacity \( c_p \).

The chemical diffusion flux \( \vec{j} = (\vec{j}_1, \ldots, \vec{j}_n) \) is given by

\[
\vec{j}_k = D_{mix} \left( \nabla \rho_k - \rho_k \frac{\vec{\chi}}{\chi} + \rho_k \left( 1 - \frac{w_k}{\vec{w}} \right) \nabla p \right),
\]

where \( D_{mix} \) is the averaged mixture diffusion coefficient of the \( k \)-th species. For a detailed representation of \( D_{mix} \), we refer to Ref. 10.

The chemical source term \( \vec{\dot{\chi}} \) is determined from the reaction mechanism consisting of \( N \) elementary forward reactions. It is given by

\[
\vec{\dot{\chi}} = (\nu'' - \nu') \vec{r},
\]

where the matrices \( \nu', \nu'' \in \mathbb{R}^{n \times N} \) contain the stoichiometric coefficients of the chemical reactions and \( \vec{r} = (r_1, \ldots, r_N) \left[ \frac{mol}{m^3 \cdot s} \right] \) denotes the vector of reaction rates, given by

\[
r_i = k_i \prod_{j=1}^{n} \chi_j^{\nu_{ij}''},
\]

The reaction rate coefficients \( k_i \) are related to the temperature \( T \) through a modified three-parameter Arrhenius law

\[
k_i = A_i T^{B_i} \exp \left( \frac{-E_{act,i}}{RT} \right).
\]

### 3 Spatial and Temporal Discretisation

For the spatial discretisation of the mathematical model Eq. 2 we use a Discontinuous Galerkin (DG) approach. For convection dominated flows a limiter-based stabilisation technique is used\(^1\). The viscosity terms are discretised using the Compact Discontinuous Galerkin 2 method (CDG2)\(^4\). Both methods developed by the authors are especially tailored to the mathematical model described in Sec. 2. A comparison of the CDG2 method with other DG methods in terms of performance and stability for the compressible Navier-Stokes equations is included in Ref. 4. A comparison of our DG solver with COSMO, the operational weather forecast model of the German Weather Service (DWD) is carried out in Ref. 3 for test cases related to atmospheric flow.
As basis functions for the DG space we use piecewise quadratic polynomials leading to a third order scheme in regions where the solution is smooth. The discretisation of the flow in a moving domain requires the following concatenation of discrete spatial operators: the inverse mass operator $M_t^{-1}$, the limiter $\Pi$, and the DG operator $L_{DG}$. The overall spatial discrete operator is then given by

$$L := L_{DG} \circ \Pi \circ M_t^{-1}.$$ 

Given the spatial discretisation the following ODE has to be solved:

$$\partial_t mU = L[mU],$$

where $mU$ denotes the conservative variables including the mass which is necessary to obtain a mass conservative scheme when the change of domain is included.

The time integration is done by Strong Stability Preserving Runge-Kutta methods of explicit, implicit or semi-implicit type. The implicit and semi-implicit solvers are based on Jacobian-free Newton-Krylov methods where a GMRES solver is used for the linear solves. Further details can be found in Ref. 16. The implementation of the spatial discrete operator $L$ is carried in a matrix-free fashion. This simplifies the implementation of the non-linear spatial operators and has a moderate memory consumption. On the contrary preconditioning of linear solvers for implicit time discretisations is very difficult and a currently very active field of research for these kind of problems (cf. Ref. 7). In this work we use a semi-implicit time integration where the viscosity and source terms are treated implicitly while the advection is treated explicitly. Our preconditioning is based on the inverse of the chemical source term.

4 Technical Description of the CFD solver

In this section we will briefly describe the technical details of the implementation of our CFD solver, such as the software framework DUNE, the parallelisation strategy, and the implementation of the moving grid.

4.1 The DUNE Interface Library

All the design restrictions mentioned in the introduction require the use of modern software design techniques. These strategies have already been successfully applied during the development of the DUNE-GRID interface\textsuperscript{1,2}, an interface library allowing generic access to the grid structure. Such a grid forms the basis of most numerical schemes for solving partial differential equations. Due to the use of modern code design techniques in C++ such as static polymorphism, numerical methods can be implemented efficiently. Comparisons with a direct implementation using the same underlying grid structure show that the cost of the interface is very small\textsuperscript{6}. This overhead seems reasonable if set against the advantage of using an unified interface. Numerical methods can thus be implemented generically for many different grid structures (locally adapted, hexahedral/tetrahedral or distributed grids). Parallelisation is straightforward and does not require a major redesign of the code.

In addition to the grid interface module DUNE-GRID, DUNE also provides a module for the solution of large linear systems on parallel computers (called DUNE-ISTL). Recently, a new module is being developed providing interfaces and implementations of finite element spaces (called DUNE-LOCALFUNCTIONS).

356
4.2 The DUNE-ALUGRID Module

DUNE-ALUGRID is an implementation of the DUNE grid interface providing unstructured grids for parallel computations. This grid implementation combines both, non-conforming local grid adaption for hexahedral and tetrahedral grids in 3d and distributed memory parallelisation based on domain decomposition. For time dependent problems dynamic balancing of work load (based on the Zoltan library\(^8\)) is essential for an efficient parallelisation. Furthermore, checkpointing of the adaptive grid structures is supported by DUNE-ALUGRID. Nevertheless, utilisation of DUNE-ALUGRID on many-core systems including dynamic load balancing is limited by the choice of domain decomposition which is based on the coarsest mesh. Recently, we have been able to relax this problem by employing a hybrid parallelisation strategy\(^{15}\).

4.3 The DUNE-FEM and DUNE-FEM-DG Modules

The DUNE-FEM package is the central part of our numerical methods library. The goal is to follow the mathematical concepts behind numerical schemes as closely as possible in the software design process. Thus function spaces, discrete function spaces, operators mapping function spaces to function spaces, and discrete operators realising mappings between discrete function spaces are the main parts of the interface. Combined with the usual mathematical operations on the discrete operators, this allows us to develop numerical methods for complex systems of equations based on the implementation of a few basic numerical schemes. Especially using the ideas of Discontinuous Galerkin methods combined with Runge-Kutta methods we can solve complex evolution equations of the form (1). More details can be found in Refs. 5, 12, 11, 4.

The DUNE-FEM-DG package contains the implementation of our DG based CFD solver. For solving compressible flow problems such as the Euler or Navier-Stokes equations the package contains a state-of-the-art DG discretisation accompanied with a limiter-based stabilisation (cf. Ref. 11) and a variation of diffusion operators (cf. Ref. 4).

4.4 Moving Grids in DUNE – the Meta Grid Approach

For realistic simulations we have to include the movement of the piston. Unlike other approaches, where the boundary conditions are modified to account for the moving wall, e.g. Ref. 9, we would like the moving wall boundary to coincide with the boundary of our computational domain at all time steps to avoid a complicated handling of boundary conditions. The generic DUNE grid interface allows to implement layers of grid implementations that add features to existing grid without changing the original code. These grid layers are called meta grids. Here, we use the meta grid approach to place the boundary of the computational domain exactly at the location of the moving boundary. Cells that are outside of the domain because the moving boundary has passed them will be filtered using a filtering meta grid, e.g., FilteredGrid contained in DUNE-METAGRID. For cells that now contain the moving boundary we apply a simple cut cell approach by shrinking these cells such that the cell boundary coincides with the moving wall boundary. This can be achieved with the GeometryGrid approach which allows to replace the geometry of the original grid. This stacking of meta grids is presented in Fig. 1.
For the discretisation of Eq. 2 in the moving domain, we use an Arbitrary Lagrangian-Eulerian (ALE) formulation of the problem (see, e.g. Ref. 13). We would like to point out that the meta grid approach allows for the exact representation of the time-dependent domain $\Omega_t$. In combination with the ALE approach, this method conserves mass up to machine precision. This is not the case for the results presented in Ref. 9 where a mass loss of the order $O(h)$ is observed. For a higher order discretisation this not acceptable.

5 Performance of the Parallel CFD Solver

In this section we try to briefly describe the implemented features of our CFD solver in order to make efficient use of the target hardware. The code implements modern HPC concepts such as hybrid parallelisation techniques based on MPI/pthreads including asynchronous communication (overlap of computation and communication), automated code generation for high floating point performance, dynamic load balancing for efficient adaptive computations, and data I/O and checkpointing based on SIONlib\textsuperscript{14}. In the following we briefly describe the achieved results. A detailed presentation of the performance results can be found in Ref. 15.

**Strong scaling of the MPI parallelisation:** The implemented DG solver shows a strong scaling with an efficiency of above 70% on up to 65,356 cores of the JUGENE with very few degrees of freedom per core ratio. This is crucial for our application, because the adequate solution of the stiff chemical source term requires many time steps.

**Shared memory parallelisation:** Our shared memory parallelisation is based on POSIX threads. As for the MPI parallelisation, we use domain decomposition to decompose the computational domain assigned to one MPI task by using graph partitioning methods. Each partition is assigned to one thread. The matrix-free implementation of the spatial discretisation allows for a straightforward shared memory parallelisation and since only a small part of the code had to be changed in order to work with shared memory parallelism. Here, we avoid any synchronisation of data by computing fluxes at thread domain boundaries twice, as it is done in the distributed parallelisation. The DG solver is able to achieve strong scaling results above 85% for as few as 2,500 degrees of freedom per task.
Efficiency in terms of peak performance: The performance of our DG solver with respect to floating point operations per second (FLOPS) has been studied carefully. We apply an automated code generation to performance critical numerical kernels which leads to high rates of floating point performance. On an Intel Core i7 (Nehalem) CPU Q740 @ 1.73 Ghz we are able to achieve more than 25% of theoretical peak performance.

Data I/O and checkpointing: Data I/O and checkpointing is implemented in the code and fully operational. During a computation checkpoints are written to allow to restart the simulation in case of hardware failure. The same methodology is used to generate output data for later post processing and visualisation purposes. The actual I/O is done using the SIONlib software package to write one or multiple files allowing for very good I/O performance\textsuperscript{14}. Byte order is respected and data generated on the Blue Gene architecture can be post processed on different architectures.

6 Preliminary Numerical Results

In this section we report a simple numerical example that shows that the implemented CFD solver works as expected. We use the simplified methane combustion mechanism from Ref. 17 which consists of 5 species and 3 reactions. Note that more complicated mechanisms can be easily incorporated into the CFD solver.

To test the overall setup we chose a cylindrical domain with radius 0.025m and a height of 0.075m, which corresponds to the dimensions of the piston of the 2-stroke engine from Fig. 2. The piston movement is modelled by a sinus function such that the piston is initially at \( z = 0 \text{m} \) \((t = 0 \text{s})\) and moves to \( z = 0.06 \text{m} \) at \( t = 1/320 \text{s} \) and back to \( z = 0 \text{m} \) at \( t = 1/160 \text{s} \). Due to the compression of the gas we expect the combustion reaction to be triggered at or shortly after the highest point of the piston.

Initially the piston is at the lowest position (\( z = 0 \text{m} \)). As initial data we choose constant values throughout the domain, i.e. zero velocity, a pressure \( p = 303900 \text{Pa} \) and constant values for the species concentrations as follows:

\[
\text{CH}_4 = 1.6 \text{mol/m}^3, \quad \text{O}_2 = 16 \text{mol/m}^3, \quad \text{CO} = 10^{-7} \text{mol/m}^3, \\
\text{CO}_2 = 10 \text{mol/m}^3, \quad \text{H}_2\text{O} = 6 \text{mol/m}^3.
\]

The conservative initial conditions are computed using the relations given in Sec. 2. Except for the moving wall the boundary conditions are chosen to be reflective. At the moving wall we added a moving wall boundary incorporating the velocity of the wall, see Ref. 9.

Figure 2. Moving piston in a 2-stroke engine grid with 52,500 hexahedrons divided into 8 sub domains.
Figure 3. Initial grid and partitioning for 256 nodes (left) and pressure/velocity at $t = 0.001$ (middle) and $t = 0.0026$ (right).

The piston movement creates waves in the fluid which at the highest point trigger the combustion of the methane gas since the temperature has increased to allow such combustion. The left part of Fig. 3 shows the computational grid with 81,920 cells divided into 256 partitions (corresponding to one node on JUQUEEN used with 16 pthreads). Fig. 3 also shows the pressure distribution and the velocity field during later stages in the simulation. Fig. 4 shows the temperature distribution in the early stages of the simulation. One can see that the temperature increases due to the compression as well as shock waves arising from the wall movement. At the current state of development the flow solver is very well integrated into the moving grid setup. The transport of the species works as expected. However, it is not yet possible to simulate the reaction in a reasonable amount of time since the time stepping is not very well suited for dealing with the very stiff reaction source terms. This problem will be addressed next.

Figure 4. Temperature distribution for $t = 0.0002, ..., 0.0004$ for left to right. For each picture the colour bar was changed to visualise the position of the shock wave.
7 Conclusions and Outlook

We briefly described our recent activities of developing a state-of-the-art CFD solver based on the DG method. The solver is capable to solve a very complex advection-diffusion-reaction system in a domain with moving boundaries. Modern techniques of HPC have been considered to make this solver run efficiently on today’s many core systems. A semi-implicit time stepping was considered to avoid small time steps resulting from the viscous terms.

Future work will focus on the improvement of the time integration to be able to treat the very stiff reaction terms. Furthermore, the simulation of the full system should consider the complex 2-stroke engine geometry including local grid adaptivity which hopefully will reduce the computational cost of solving this system.

Acknowledgements

The authors gratefully thank Tobias Malkmus for helpful comments and discussions. Furthermore, the authors acknowledge the Gauss Centre for Supercomputing (GCS) for providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS share of the supercomputer JUQUEEN at Jülich Supercomputing Centre (JSC) and also the high-performance computing support from Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR’s Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Robert Klöfkorn also acknowledges the Cluster of Excellence Simulation Technology for former support of this work. The 2-stroke engine geometry was kindly provided by the STIHL AG & Co. KG.

References


Scalability Tuning of the Load Balancing and Coupling Framework FD4

Matthias Lieber, Wolfgang E. Nagel, and Hartmut Mix

Center for Information Services and High Performance Computing (ZIH), Technische Universität Dresden, 01062 Dresden, Germany
E-mail: {matthias.lieber, wolfgang.nagel, hartmut.mix}@tu-dresden.de

In this paper, we discuss the scalability tuning of the HPC software framework FD4. The framework provides dynamic load balancing and model coupling for multiphase and multiphysics simulations. We first investigate scalability bottlenecks using the Vampir performance analysis tool-set on the Blue Gene/Q system at JSC. Then, we describe and evaluate our optimised algorithms: A new hierarchical 1D partitioning algorithm for SFC-based dynamic load balancing and a new method to organise the coupling metadata. The final scalability benchmark shows that the overhead of FD4 has been reduced by a factor of 3.3 at 262 144 ranks, which leads to a considerably increased scalability of the whole application.

1 Introduction

Driven by the growing capacity of supercomputers and the increasing knowledge about the underlying processes, simulation models become more and more complex. For example in the atmospheric sciences, models are coupled to incorporate more elements of the geosphere in the simulations, e.g. climate models or earth systems models. Additionally, more complex descriptions of individual phenomena are included in the models. Running such model systems efficiently on large scale supercomputers is a challenging task and requires highly tuned methods for data decomposition, communication, load balancing, and model coupling, which should be implemented preferably in an independent and reusable software framework. In this paper, we analyse scalability bottlenecks of the framework FD4 and describe our improvements. The work has been carried out on the Blue Gene/P and Blue Gene/Q systems at Jülich Supercomputing Centre. The architecture of these systems offers an excellent platform to prepare applications and algorithms for the Exascale era, where the number of nodes is expected to grow even further.

FD4 (Four-Dimensional Distributed Dynamic Data structures) has been developed for dynamic load balancing in the model system COSMO-SPECS, which describes the interaction between aerosols, clouds, and precipitation with a high level of detail. It consists of the COSMO model, a non-hydrostatic limited-area atmospheric model, and the spectral cloud microphysics model SPECS. SPECS introduces 11 new variables to describe three types of hydrometeors (water droplets, frozen particles, and insoluble particles). These variables are each discretised into 66 size classes, leading to a high amount of data per grid cell. Since SPECS computational costs are very high and strongly depend on the spatially and temporally varying presence of hydrometeors, a new version called COSMO-SPECS+FD4 with dynamic load balancing has been developed: SPECS has been separated from the static partitioning of COSMO and decomposed in three dimensions into small, rectangular blocks transparently managed by FD4, see Fig. 1. FD4 provides dynamic load balancing of the blocks, communication between neighbour blocks located on
different MPI processes, and coupling to the static partitioning of the atmospheric model. After first scalability experiences with COSMO-SPECS+FD4, we observed that much attention needs to be paid to the load balancing and coupling methods in order to prevent an increasing overhead that impedes any benefit of the load balancing.

2 Scalability Bottleneck Analysis of FD4

In this section, we discuss the scalability bottlenecks we detected in FD4. For a detailed analysis, we used the Vampir performance analysis tool-set. The VampirTrace performance monitor records application events from all processes, like function calls and MPI messages, along with their time stamp to an OTF trace file. This data is then displayed with the Vampir trace visualiser by means of several charts like timelines or statistics. For our analysis, we performed runs with 32,768 processes on Blue Gene/Q. We recorded MPI events and certain functions of COSMO-SPECS+FD4. Additionally, we added user-defined counters, e.g. to observe the number of FD4 blocks per process over time. To reduce the trace file size, only a few time steps (e.g. the last 3 out of 180) have been recorded. This selective instrumentation caused a runtime overhead of less than 1% with additional trace writing time of approx. 90 s. In spite of the large number of application processes, the compressed trace files have a total size of approx. 4 GiB only.

We ran COSMO-SPECS+FD4 with an artificial benchmark scenario: An initial heat bubble leads to the formation of a precipitating cumulus cloud after 30 min simulated time. This basic problem can be replicated arbitrarily along the horizontal axes to scale the grid size. Note, that the benchmark does not perform any I/O of model data. For the bottleneck analysis, we used a horizontal grid of 512 × 256 cells with 48 vertical layers containing 128 replicated clouds, resulting in extremely small COSMO partitions of 2 × 2 × 48 cells.

2.1 Analysis of Dynamic Load Balancing

Since the workload of the blocks varies over time, COSMO-SPECS+FD4 regularly performs dynamic load balancing. The workload (i.e. weight) of a block is determined by measuring the time of the microphysics computations. FD4 uses a Hilbert space-filling curve (SFC) for dynamic load balancing. The SFC linearises the 3D block decomposition of the grid and, thus, reduces the partitioning problem from three to one dimension. Then, the remaining 1D partitioning problem is to find consecutive partitions in the block vector with a minimal maximum load among the partitions (i.e. bottleneck). Current published algorithms are either fast, parallel heuristics or serial exact methods.
Fig. 2(a) shows a Vampir timeline chart for a 1% subset of the 32,768 processes. The time interval is approx. one time step, which starts at ca. 0.6 s and ends at ca. 2.4 s. Within a time step, first COSMO is computed (green), then SPECS performs four micro time steps (blue and purple), and finally dynamic load balancing is carried out. All MPI communication is shown in red. In Fig. 2(a), a parallel heuristic is used for partitioning. However, the achieved load balance is rather poor leading to large waiting time after SPECS.

Fig. 2(b) shows the same time step when using an exact partitioning method (adapted from Pınar and Aykanat’s exact bisection algorithm\(^1\))\(^5\). While the load balance is clearly better, the partitioning algorithm requires substantial serial computation time on rank 0, as shown in the upper part of the chart (orange). This leads to large waiting time in all other ranks (red). Additionally, all block weights need to be transferred to rank 0 first, which causes high network traffic and non-scalable memory consumption. Overall, the benefit of the exact method is reduced strongly due to the great overhead at large scale.

As Fig. 2(c) shows, we solved this problem with a new algorithm that is scalable and provides high load balance at the same time. The algorithm combines the approaches of the heuristic and the exact method. We briefly describe the algorithm in Sec. 3.1.
2.2 Analysis of Model Coupling

FD4 separates the decomposition of SPECS from the COSMO model. Consequently, each process owns one dynamic partition of SPECS that does not necessarily overlap with the process’ static COSMO partition. Since data needs to be transferred between both models each time step, model coupling techniques are required\(^\text{16}\), i.e. firstly map the partitionings to find communication partners (handshaking) and secondly transfer the data. In general, handshaking needs to be done only once at initialisation for static partitions. But in applications with dynamic load balancing, handshaking needs to be repeated every time the partitioning is changed and, thus, becomes critical to performance. A naive implementation of handshaking would, for example, search the list of COSMO partitions for each local FD4 block to find overlaps. Of course, this algorithm is not scalable at all. We optimised the handshaking by performing this matching with the smallest bounding box containing all local blocks. This approach strongly reduces the handshaking time in most cases. However, as shown in Fig. 2(d), we observed that at certain time steps a very small number of (obviously arbitrary) processes require a larger handshaking time. In this example, two ranks consumed around 66 ms for handshaking (i.e. FD4 without MPI), while most of the processes required 15 ms only. Of course, this imbalance leads to waiting time in the coupling data transfer and at subsequent synchronisation points. We found out, that the slow processes have disconnected FD4 partitions with parts far apart in the block decomposition. This leads to a large bounding box and, thus, increased handshaking time.

3 Improving Scalability of FD4

In this section, we briefly describe the algorithms we developed to improve the scalability of dynamic load balancing and model coupling in FD4 and evaluate the improvement.

3.1 Improvement of Dynamic Load Balancing

We developed a novel hierarchical 1D partitioning algorithm that combines the scalability of the parallel heuristic and the high quality of the exact method. Our algorithm starts with a parallel heuristic, but creates only \( G < P \) partitions, where \( P \) is the total number of processes. For sufficiently small \( G \), this will create almost optimal partitionings\(^{14}\). Each of the \( G \) coarse partitions is assigned a group of processes. In the second phase, each group independently partitions their coarse partition using an exact method. This way, we reduce the communication costs for gathering the block weights and the computation costs for the exact method, since it is run for a subset of blocks and processes only. Because no global vector of block weights is assembled, memory consumption is reduced in comparison to running the exact method for all processes. For a full description of the method, we refer to Ref. 4.

The group count \( G \) is like a slide control and tunes the influence of the heuristic versus the exact method. Therefore, we performed a series of runs on 65 536 processes with a horizontal grid of 512 \( \times \) 512 cells and varied \( G \) from 16 to 8192. The grid was decomposed into 786 432 FD4 blocks for balancing SPECS. Fig. 3 shows the total runtime of COSMO-SPECS+FD4 divided into main components (without initialisation). Additionally, three important dynamic load balancing metrics are presented: runtime of the load balancing,
average load balance (defined as average process load divided by maximum load), and the average amount of migration per time step. Note, that dynamic load balancing is carried out every time step in this benchmark, i.e. 180 times. The results show, that our new method is almost as fast as the heuristic in terms of partitioning calculation, but achieves the optimal average load balance. At the optimal group count $G = 256$, the runtime of the application is reduced by more than 10% compared to the existing partitioning methods. The increased partitioning calculation time at $G = 512$ is a reproducible artifact resulting from some collectives on subcommunicators when distributing the result to all ranks. The results for average load balance and migrated tasks illustrate the slide control feature of $G$.

In practice, performing load balancing at every time step generates noticeable overhead. To reduce the number of load balancing invocations, FD4 is able to decide automatically if load balancing is beneficial. In this auto-mode, FD4 weighs the time lost due to imbalance against the time required for load balancing. Both times are measured at runtime and a history from the last 4 load balancing invocations is kept. Using the auto-mode, the execution time of the benchmark in Fig. 3 is reduced further to 234.9 s.

### 3.2 Improvement of Model Coupling

Handshaking with disconnected FD4 partitions that cause large bounding boxes has been tuned by replacing the simple bounding box with a cluster of a limited number of boxes. The clustering leads to small additional costs, but effectively solves the load imbalance issue when performing handshaking.

However, the total costs for handshaking are still very high at large scale, because every partition of the model coupled to FD4 (i.e. COSMO) has to be checked for overlap with the local blocks. We developed a general method to dramatically reduce the search space by spatially decomposing the data that describes the coupled partitions (i.e. position, size, variables, owner). The so-called meta data subdomains (MDSDs) are rectangular subsets of the domain and cover multiple FD4 blocks. They contain the descriptions of all partitions that begin within the region of the MDSD, i.e. each partition is assigned to
a unique MDSD. For handshaking, only the MDSDs that overlap with the FD4 blocks and additionally the MDSDs at the spatially next lower coordinates need to be evaluated, given that no coupled partition is larger than the MDSDs (see Fig. 4). Since the MDSD decomposition is straightforward, the necessary MDSDs can be identified in constant time. Thus, the effort for handshaking does not depend on the total number of coupled partitions. The advantage of the MDSD concept is that arbitrary partition structures are supported, i.e. more complex geometries than rectangular partitions as in COSMO.

We evaluated the improvement achieved by the clustering and the meta data subdomains by running the COSMO-SPECS+FD4 benchmark on 131 072 ranks with a horizontal grid of 1024 × 512 cells. Fig. 5 shows that the calculation part of coupling (i.e. handshaking) is strongly accelerated with increasing number of MDSDs. Without the optimisations, the imbalanced handshaking (see Sec. 2.2) leads to waiting time in COSMO and FD4 load balancing, which increases the respective runtimes.
3.3 Scalability Benchmark

We evaluated the scalability of COSMO-SPECS+FD4 with a strong scaling benchmark using a horizontal grid size of $1024 \times 1024$ cells and 3,145,728 FD4 blocks. Fig. 6 compares the results for the original version and the version incorporating the optimisations described above. In both versions, the auto-mode to trigger dynamic load balancing was used. The speed-up of the original version from 16,384 to 262,144 ranks is 11.8, while a speed-up of 14.0 is achieved with the optimised version. At 262,144 processes, the overhead of FD4 has been reduced by a factor of 3.3, not considering the side-effects of handshaking imbalance.

4 Conclusion and Outlook

We discussed scalability optimisations for the HPC software framework FD4 that combines dynamic load balancing and model coupling. Scalability bottlenecks were analysed using the Vampir performance analysis tools-set and optimisations were developed, implemented, and tested in FD4. The results show exemplarily, that complex coupled simulation models with dynamic workload behaviour can be run efficiently on large-scale HPC systems. FD4 is available as open source\ref{17}.

Future scalability improvements include the elimination of global meta data to reduce the memory consumption and global communication overheads. This would benefit the dynamic load balancing, since no global partitioning information needs to be distributed to all ranks, as well as the model coupling, since no global coupling meta data is stored on a single rank. Therefore, the meta data subdomains can be used to store local meta data only and dynamically redistribute the meta data after load balancing.

Acknowledgements

We thank the Jülich Supercomputing Centre, Germany, for access to JUGENE and JUQUEEN and the German Weather Service (Deutscher Wetterdienst) for providing the COSMO model. Furthermore, we want to thank Verena Grützun, Ralf Wolke, and Oswald Knoth for their support regarding the COSMO-SPECS model. This work was funded by the German Research Foundation (DFG), grant No. NA 711/2-1.
References

Fluid Mechanics

Nikolaus A. Adams

Institute of Aerodynamics and Fluid Mechanics, Technische Universität München, Bolzmannstraße 15, 85748 Garching, Germany
E-mail: Nikolaus.Adams@tum.de

The variety of subjects reported in this section reflects the significance of fundamental fluid mechanics research for the scientific landscape in Germany. It also reflects the impact high-performance scientific computing has on the further development of research in different areas of fluid mechanics.

The contribution of Thomas Boeck, Dmitry Krasnov, Saskia Tympel, and Oleg Zikanov, of the Institute of Thermodynamics and Fluid Mechanics at Technische Universität Ilmenau, investigates by large-scale simulations the “Transitional and Turbulent Magnetohydrodynamic Flows in Uniform and Non-Uniform Magnetic Fields” and thus addresses a fundamental fluid mechanics problem with high relevance to geophysical or metallurgical research. The presence of a magnetic field induces electromagnetic boundary layers that play an important role in flow transition to turbulence. It is established that transition can be controlled by special magnetic-field properties. The high Reynolds number of liquid metals renders direct numerical simulations as particularly challenging. The numerical approach relies on standard discretisation schemes, high computational efficiency is facilitated by the simple geometric structure of the problem and by the use of efficient libraries.

The contribution of Bernhard Vowinckel, Tobias Kempe, and Jochen Fröhlich of the Institut für Strömungsmechanik of Technische Universität Dresden, investigates the “Highly-resolved Numerical Simulations of Bed Load Transport in a Turbulent Open Channel Flow”, also addresses a fundamental fluid mechanics problem. Similarly as with the previous paper a rather simple generic shear-flow geometry is considered. However, complexity is introduced by the presence of solid particles that modify the developing turbulent flow. Of particular concern in this paper is the effect of wall roughness, which is accomplished by immobilising wall particles. The numerical model relies on a standard staggered-grid incompressible Navier-Stokes solver, particles are modelled by an immersed-boundary method and tracked in time. Again, the simple geometry in connection with a well-developed flow solution algorithm, and a careful implementation of the particle integration scheme allow for high computational efficiency.

The contribution of Gregor Gassner, Manuel Torrilhon, Andrea Beck, Sophie Knechtel, and Thomas Boelemann of the Institute for Aerodynamics and Gasdynamics, Universität Stuttgart, and of the Center for Computational Engineering Science, RWTH Aachen, on “Comparison of Navier-Stokes-Fourier Equation and Grad’s Moment Equation Solutions for Turbulence”, focuses on a comparison of different physical models for the numerical simulation of turbulent flows. This focus addresses the need of considering alternative physical and numerical models in terms of their inherent computational efficiency. The fact that the three different physical models share a similar algorithm structure is exploited...
to come up with an efficient parallel implementation. The paper addresses the questions which physical models results in which modelling error when compared with the accepted continuum-flow model.

The contribution of Andreas Fiolitakis, Christian Eberle, Antonio Filosa, Juliane Prause, and Berthold Noll of the German Aerospace Center (DLR), Institute of Combustion Technology, has the objective of “Numerical Investigation of Reacting and Non-Reacting Flows in Gas Turbine Related Configurations”. The underlying numerical method is based on a pressure-based approach to solving the compressible Navier-Stokes equations in the low-Mach-number limit, as is appropriate for combustion-chamber flows in gas turbines. Due to its similarity with the fully compressible solution method DLR tau it inherits grid structure and flexibility on dealing with complex geometries. Standard turbulence models and turbulent combustion models are employed. Within the focus of performance analyses was the numerical particle model to solve the probability-density transport equation.

The contribution of Jiajia Zhou, Aleksey V. Belyaev, Evgeny S. Asmolov, Olga I. Vinogradova, and Friederike Schmid of the Institut für Physik, Johannes Gutenberg-Universität Mainz, the DWI, RWTH Aachen, and several cooperating Institutes from Russia, focuses on the “Computer Simulation of Flow Past Superhydrophobic Striped Surfaces”. Improvement of microfluidic mixing is an important subject in bottom-up synthesis which is considered as one of the promising future technologies in chemical process engineering. Surface structuring is one approach to improve microfluidic mixing and is analysed by numerical simulation in this paper. As microfluidic simulations need to take into account explicitly mesoscopic effects, but as also microscopic approaches such as molecular dynamics are infeasible for this purpose, the authors resort to the dissipative-particle dynamics method. Computational power essentially is used for performing a wide range of parameter variations with each individual simulation being of moderate size.
Transitional and Turbulent Magnetohydrodynamic Flows in Uniform and Non-Uniform Magnetic Fields

Thomas Boeck, Dmitry Krasnov, Saskia Tympel, and Oleg Zikanov

1 Institute of Thermodynamics and Fluid Mechanics, Technische Universität Ilmenau, P. O. Box, 100565, 98684 Ilmenau, Germany
E-mail: {thomas.boeck, dmitry.krasnov, saskia.tympel}@tu-ilmenau.de

2 Dept. of Mechanical Engineering, University of Michigan-Dearborn, 4901 Evergreen Road, Dearborn, MI 48128-1491, USA
E-mail: zikanov@umich.edu

We investigate the effects of static magnetic fields on wall-bounded flows of liquid metals using direct numerical simulations. Such flows are modified by the Lorentz forces arising from the electromagnetic induction of eddy currents. In a homogeneous field we find that electromagnetic boundary layers at the walls parallel to the magnetic field are essential for transition to turbulence. A strongly inhomogeneous dipole field can act as a magnetic obstacle and cause transition to turbulence in its wake.

1 Introduction

Electromagnetic induction causes eddy currents and a resulting Lorentz force when a conducting liquid such as a liquid metal or semiconductor flows in the presence of a static magnetic field. The velocity distribution in such a magnetohydrodynamic (MHD) flow can thereby be significantly modified. For this reason, magnetic fields are used for flow control in metallurgical processes such as casting or crystal growth. A considerable advantage is the absence of mechanical contact with the typically hot and aggressive liquid which avoids contamination of the melt. Since liquid metals are also opaque one cannot apply optical flow measurement methods in MHD. Numerical simulations are therefore an essential tool for the analysis and prediction of liquid-metal MHD flows.

Our work is concerned with fundamental phenomena and basic mechanisms in liquid-metal MHD flows. For this purpose we use direct numerical simulations with specialised codes on structured grids. The typical flow domains are therefore fairly simple. Our focus is on straight ducts of rectangular cross-section and straight pipes of round cross-section.

For MHD flows in uniform fields we are interested in the transition from laminar to turbulent flow and in properties of developed turbulent flow. Transition in MHD duct flow was among the first problems studied when MHD came into existence as a branch of fluid dynamics, but could only be quantified by measurements of pressure loss. Today numerical simulations provide information on the complete velocity field and can thereby reveal phenomena such as localised turbulent spots or near-wall vortical flow structures sustaining turbulence. Thanks to the significant advances in computing performance in the past years, we have been able to make significant progress in the understanding of transitional and turbulent MHD ducts flows. Our results are of interest in the general context of subcritical transition to turbulence in shear flows and also from the perspective of wall turbulence.
We also study MHD flows in non-uniform fields by simulations. The motivation is provided by a novel technique for contactless flow measurement in liquid metals called Lorentz force velocimetry (LFV), which is based on electromagnetic induction. By Newton’s third law, the induced Lorentz force on the flow past an exterior magnet is accompanied by an equal but opposite force on the magnet system. Measurement of this force on the magnet can be used to determine the velocity. When the magnetic field pervades only a fraction of the flow, the braking Lorentz force modifies the flow in a similar way to a solid obstacle such as a cylinder. We investigate this so-called magnetic obstacle effect for small magnets that can be approximated by point dipoles. By contrast to a uniform field, the localised field promotes instabilities and turbulence in the wake of the magnetic obstacle.

2 Mathematical Model and Numerical Method

The mathematical model for liquid-metal MHD flows comprises the incompressible Navier-Stokes equations and the induction equation in the so-called quasi-static approximation, which applies at low magnetic Reynolds numbers typical of liquid-metal flows in the laboratory. The induced magnetic field is weak in this limit, and can be neglected in comparison with the imposed magnetic field. Both eddy currents and Lorentz force then depend linearly on the velocity field. The eddy currents are determined by Ohm’s law for a moving electric conductor in combination with a Poisson equation for the induced electric potential to ensure charge conservation. The walls of the flow domain are insulating, i.e. eddy currents are confined within the flow.

For simulations we use an in-house code based on the finite-difference method. It is second-order accurate in space and time and uses a standard projection procedure to satisfy incompressibility. Time integration is explicit. Spatial discretisation is done in a highly conservative way following the principles developed in Ref. 2 for incompressible flows and extended in Ref. 3 to the case of MHD flows in the quasi-static limit. The discretisation is conducted directly on the non-uniform grid with collocate arrangement. Details are provided in Ref. 4.

In the case of periodic streamwise boundary conditions, the solution of Poisson problems for electric potential and pressure uses the Fast Fourier Transform (FFT) applied in the periodic $x$-coordinate, whereby it reduces to a series of two-dimensional (2d) separable elliptic problems for coefficients of the Fourier expansion. Each of these 2d problems is discretised by central differences and solved by the cyclic reduction method, a part of the software package Fishpack. For in- and outflow conditions we use an expansion in cosine functions rather than in full Fourier series.

For the effective use of the modern parallel computers both MPI (distributed memory) and Open MP (shared memory) parallelisation interfaces are employed in the code. In the real space, the computational domain is divided along the spanwise $y$-direction into slabs so that each slab is tackled by an MPI process. In the Fourier space the domain is divided along the wavenumbers in streamwise direction $x$. Therefore, a transpose operation from $y$-to $x$-decomposition and back is necessary, which causes most of the MPI communication. In addition, each MPI task can start several Open MP threads. This hybrid parallelisation makes the code very flexible, such that an optimal combination of MPI tasks and Open MP threads can be found for the best performance. The code is implemented with Fortran 90 programming language and can be run on virtually any number of CPUs and grid points.
3 Uniform Magnetic Fields

Laminar MHD flow in a duct with a transverse magnetic field $B$ is characterised by the presence of a flat velocity distribution and electromagnetic boundary layers where the induced eddy currents concentrate. These layers are the so-called Hartmann layers at the walls perpendicular to the magnetic field and side layers (or Shercliff layers) at the walls parallel to the magnetic field. Their thickness is characterised by the Hartmann number defined as

$$Ha = BL\sqrt{\sigma/\rho \nu}.$$

Here $\rho$, $\nu$ and $\sigma$ are the density, kinematic viscosity and electric conductivity of the liquid metal. The length $L$ is a characteristic length, which we choose as half the side length of the duct. The Hartmann layers have a thickness that is proportional to $1/Ha$ whereas the side layers have a thickness proportional to $1/\sqrt{Ha}$. These layers are essential for transition to turbulence and the properties of MHD wall turbulence. Turbulent flow occurs when the mean velocity $U$ is large, i.e. the Reynolds number

$$Re = U L/\nu$$

is sufficiently high to destabilise these layers.

3.1 Transition in MHD Duct Flow

In order to understand the transition process in detail we have reproduced the mercury experiments by Hartmann in our simulations. In these experiments the flow enters a rectangular duct or pipe with a cross-section of a few millimetres and a transverse magnetic field, and the pressure difference between two stations along the pipe is measured. The mean velocity $U$ is fixed and the magnetic induction $B$ and thereby the Hartmann number can be adjusted. We have first conducted simulations with periodic boundary conditions at a fixed Reynolds number. The Hartmann number $Ha$ was increased until the initially turbulent flow becomes fully laminar. This does not happen at once but in distinct stages. Turbulent fluctuations are first eliminated from the bulk and the Hartmann layers at the walls perpendicular to the magnetic field. The side layers remain turbulent for higher $Ha$. The details of the relaminarisation in the side layers depend strongly on the streamwise size of the computational domain. It was essential for our studies to use long domains (up to about 100 side lengths). We could identify the co-existence of laminar and turbulent regions in different streamwise sections of the duct, which are also localised in the side layers. The turbulent regions can either exist in a quasi-stable manner, or grow in size over time. Such states are also known in other transitional shear flows such as hydrodynamic pipe flows. The peculiarity in our magnetohydrodynamic flows is the localisation of these turbulent zones in the side layers. The quasi-stable states called puffs may either appear isolated on one side, in staggered patterns, or in symmetric patterns. Similar states are also observed in simulations of magnetohydrodynamic pipe flow. The friction coefficients measured by Hartmann are in good agreement with our direct numerical simulations. However, they provide no indication of the existence of turbulent zones as their impact on the total friction is fairly low.

Recently, we have performed more realistic transition simulations at $Re = 3000$ using in- and outflow conditions. They confirm the general conclusions from the simulations.
with periodic boundary conditions. Fig. 1 shows a typical spatial development of the flow in this case. Turbulent inlet conditions are prescribed by running a concurrent simulation of duct turbulence on a periodic domain and copying its velocity field on a cross-sectional plane to the inlet plane of the non-periodic duct. The magnetic field jumps from zero to a finite value at the first streamwise position marked as A. Downstream of this position the turbulent fluctuations are reduced by the magnetic damping and both Hartmann layers and the bulk of the duct become laminar. At $Ha = 12$ turbulence is still maintained in extended zones located at the walls parallel to the field. At $Ha = 13$ we observe the appearance of relatively stable puffs of well-defined length with a tendency to arrange in a staggered pattern. At $Ha = 14$ the magnetic damping is already too strong to sustain turbulence. The puffs that are still generated occasionally die out quickly and the flow is essentially laminar already at the first position marked B.

The friction coefficient has again been measured in order to compare with the experiments. Fig. 2 shows that the friction coefficient agrees well between experiments and periodic and non-periodic simulations. Upon increasing $Ha$ the friction is initially reduced because of the magnetic damping of turbulence. For $Ha > 15$ it increases linearly with $Ha$ because friction in the laminar Hartmann layers becomes the dominant contribution. Relaminarisation occurs close to the minimum of the friction coefficient. In experiments this has typically been found for a parameter

$$R = Re / Ha \approx 200.$$
Figure 2. Friction coefficient $f$ vs. Hartmann number $Ha$ for the spatial transition simulations illustrated in Fig. 1. Experimental data$^1$ and results of simulations$^5$ in a periodic domain are shown for comparison.

3.2 Fully Developed Turbulent MHD Duct Flow

The properties of turbulent MHD duct flow at high Reynolds numbers have been explored by a series of simulations conducted at $Re = 100000$ with periodic boundary conditions$^7$. In this case one expects a stronger anisotropy of turbulence before relaminarisation occurs near $R \approx 200$ because the ratio $Ha^2/Re$ of Lorentz forces to inertial forces is higher. We have again increased the Hartmann number in steps and observed the transformation of the velocity distribution. Between $Ha = 100$ and $Ha = 200$ the Hartmann layers become laminar and turbulence disappears near the centre of the duct as can be seen in Fig. 3. For even larger $Ha$ turbulence is only sustained in the side layers and eventually disappears at $Ha = 400$. With exception of the Hartmann layers, the mean velocity distribution becomes effectively uniform along the magnetic field. Anisotropy of the velocity field is less pronounced for small-scale turbulent structures near the wall but becomes significant at $Ha = 100$ and $Ha = 200$.

Figure 3. Results of duct flow simulations at $Re = 10^5$. Instantaneous distributions of streamwise velocity on a plane $x = \text{const.}$ are shown. The contour levels are the same in all plots ranging from 0 (blue) to 1.25 (red), the magnetic field is applied in the vertical z-direction.
Figure 4. Snapshot at $Re = 10^7$ and $Ha = 350$. Isosurfaces $-0.01\lambda_2^{\text{rms}}$ (yellow) and $-0.005\lambda_2^{\text{rms}}$ (cyan) of the second eigenvalue $\lambda_2$ of the tensor $S^2 + \Omega^2$ are shown to visualise vortices. Also shown are the streamlines of velocity fluctuations in the mid-plane $z = 0$. Numerical resolution is $10^9$ grid points.

larger distances from the wall. This is shown by Fig. 4, which demonstrates the appearance of large-scale vortices aligned with the magnetic field. They are least susceptible to magnetic damping.

The mean velocity distribution in the side layers is also different from normal wall turbulence. We observe a combined linear-logarithmic dependence of streamwise velocity on the distance from the side wall. The linear contribution originates from the magnetic damping since its slope scales approximately with the inverse Joule damping time $\tau$. On the basis of our simulation data we are currently developing a simple mixing-length model for such turbulent side layers. Another interesting possibility concerns the existence of localised turbulent spots close to relaminarisation at $Re \approx 200$. Our simulations have been conducted with a relatively short streamwise domain length. It may turn out that such structures exist also in the strongly anisotropic case at high $Ha$ if longer simulation domains are used. The numerical effort to detect them is going to be substantial and will certainly exceed the $3 \times 10^5$ core hours required for the low Reynolds number simulations $5$.

A further open question is the role of the parameter $R$. It characterises the stability of the Hartmann layers since it is the effective Reynolds number based on the Hartmann layer thickness. Its significance for transition in the side layers remains to be explained by theory.
4 Nonuniform Magnetic Fields

We have also investigated the flow transformation in a duct by a strong localised magnetic field causing a magnetic obstacle effect. The work is motivated by the envisaged use of localised fields for local velocity measurement by Lorentz force velocimetry (LFV). In our study the magnetic field is represented by a point dipole in a certain distance from the duct. Another geometry parameter is the orientation of the magnetic dipole moment. Because of the inhomogeneous field, the Hartmann number is based on the maximum induction $B_{\text{max}}$ within the flow domain. In Ref. 8 we have performed a systematic investigation of the parameter influences including $Re$ and $Ha$ for laminar flow upstream of the dipole.

For distances of the order of the duct side length, the magnetic obstacle effect leads to a wake pattern with vortical structures. For the spanwise orientation of the magnetic dipole, this pattern becomes time-dependent at fairly moderate values of $Re$, whereas it remains stationary for streamwise and vertical orientations. Fig. 5 shows a snapshot of the flow and its complex wake structure in this case. The reasons for the stability in the other cases are not fully understood. A preliminary analysis using ideas from the stability of one-dimensional shear flows makes it plausible that the spanwise orientation is most likely to induce instability. Time-dependence appears as a periodic vortex shedding. In contrast to the cylinder wake, the flow is fully three-dimensional and, therefore, far more difficult to analyse than the Karman vortex street.
5 Conclusions

Our simulations have revealed interesting new phenomena in wall-bounded MHD flows. The results on transition in a uniform field can serve as a starting point for more sophisticated analyses with methods from the theory of dynamical systems and statistical physics. For inhomogeneous fields the next goal is the exploration of the magnetic obstacle effect for a turbulent inflow. We are also developing a numerical method for simulations with the full induction equation in the duct geometry to remove the constraints imposed by the quasi-static approximation.

Acknowledgements

We are grateful to the DFG and the US NSF for financial support, also in the framework of the DFG Research Training Group “Lorentz Force Velocimetry and Lorentz Force Eddy Current Testing”. We acknowledge the generous support by the John von Neumann Institute and the Research Centre Jülich through several grants of computing time on the JUROPA cluster.

References

Highly-Resolved Numerical Simulations of Bed Load Transport in a Turbulent Open Channel Flow

Bernhard Vowinckel, Tobias Kempe, and Jochen Fröhlich

Institut für Strömungsmechanik, Technische Universität Dresden, 01062 Dresden, Germany
E-mail: {bernhard.vowinckel, tobias.kempe, jochen.froehlich}@tu-dresden.de

The paper presents highly-resolved simulations of a turbulent open channel flow laden with up to 27000 spherical particles that are conveyed across a rough bed. The rough bed consists of fixed spherical particles of the same size arranged in hexagonal ordering. The mobility of the particles constituting the bed load is around its critical value, so that particles are constantly colliding either with each other or the fixed bed. The bulk Reynolds number of the flow is 2941, the flow depth is 9 particle diameters. The numerical method employs an Euler-Lagrange approach with an Immersed Boundary Method to resolve each individual particle with 22 grid points per diameter. Five different runs were carried out to investigate the role of mass loading and particle density, which are the key parameters of the problem. The computational domain is very large to allow particles to create bed form structures with high fidelity. The observed particle structures differ substantially between the runs by their scales in time and space. The results are in agreement with experimental evidence observed at higher Reynolds number.

1 Introduction

Prediction of turbulence-induced erosion and near-bed transport of sediment particles in water is important for many processes in environmental engineering, such as the protection of bridge piers, the development and management of aquatic habitats, the design of sewerage and pipeline systems, to mention but a few. Beyond its relation to sediment transport, the results of the present study are relevant as well for numerous industrial applications, particularly in the field of process technology, where solid particles are transported by a carrier flow.

Traditional methods for the prediction of sediment transport are empirical and based on averaged bulk quantities. Shields introduced the ratio of shear stress to gravity as a characteristic number which today is know as Shields number. He then developed an empirical relation describing the critical value of this number, beyond which particles are eroded, as a function of the particle Reynolds number. This type of relation is still common for estimating bed stability and simple empirical formulae of this kind are employed for bed load transport in the framework of statistical turbulence modelling (RANS). Due to their statistical concept, the predictive power of these formulae is low, because homogeneity of the sediment is postulated. The particle structures created by the interaction between mobile sediment and turbulent flow, however, are diverse and controlled by many factors. Despite the long history of research on this topic a detailed understanding of sediment stability and the physical mechanisms involved in sediment transport close to the threshold of mobilisation is still missing due to the lack of highly-resolved data under controlled flow conditions.

Recently, sufficient computational resources have become available to conduct Direct Numerical Simulations (DNS) of such phenomena. In particular, the Immersed Boundary
Method (IBM) has proven to be a valuable tool to tackle this problem. Computational studies of bed load transport, however, are still rare as they lead to very costly simulations\textsuperscript{3–5}. The simulations carried out so far, however, either lacked physically realistic modelling, spatial resolution, or appropriate domain size to provide truly reliable physical information.

The presented study reports on DNS of mobile sediment transported over a rough wall, extending the previous work of the present authors\textsuperscript{6} by using a larger domain. The parameter values of the disperse phase were chosen similar to experiments reported in the literature\textsuperscript{7,8}. The highly-resolved simulations provide detailed and physically reliable instantaneous as well as statistical information on bed load transport and cover parameter ranges so far not reached.

2 Numerical Method

The study was conducted using the in-house code PRIME (Phase Resolved sIMulation Environment), which solves the unsteady three-dimensional Navier-Stokes equations for incompressible fluids

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = \frac{1}{\rho_f} \nabla p + \nu_f \nabla^2 \mathbf{u}$$

$$\nabla \cdot \mathbf{u} = 0$$

(1)

on a regular Cartesian staggered grid, where $\mathbf{u}$ is the velocity vector, $t$ time, $\rho_f$ the fluid density, $p$ pressure, and $\nu_f$ the constant kinematic viscosity. The disperse phase is represented by an enhanced immersed boundary method\textsuperscript{9}. For particle contact, the adaptive collision model (ACM)\textsuperscript{10} was used. It accounts for all relevant mechanisms that have to be modelled during the collision process: short-range lubrication forces, normal forces during surface contact, and tangential forces due to friction between particles. It was validated in great detail for single collisions. Previous investigations showed that for bed load transport the collision model is important for the formation of large-scale structures\textsuperscript{11} and suggested that the ACM yields realistic results that are in agreement with experiments\textsuperscript{6}.

3 Computational Setup

An open channel flow is considered with a computational domain $24H \times (H + H_{sed}) \times 12H$ in streamwise, vertical, and spanwise direction, respectively, with $H$ the height of the water level above the fixed sediment. The sediment bed is constituted of two layers of fixed spheres of diameter $D$, which are arranged in a hexagonal pattern of thickness $H_{sed}$. This defines the case $Fix$ in Tab. 1. The origin of the vertical coordinate $y$ is set to $H_{sed}$ throughout. Periodic boundary conditions are applied in $x -$ and $z -$ direction and a free-slip condition at the top. At the bottom of the domain and at the particle surfaces, a no-slip condition is imposed. The bulk Reynolds number based on the channel height $H$ and the bulk velocity of the fluid is barely above the threshold for turbulent flow of an unladen channel with smooth walls. The resulting friction Reynolds number is $Re_\tau = 193$ and the particle Reynolds number is $D^+ = \frac{D u_c}{\nu_f} = 21$. The resolution of the equidistant, Cartesian grid is set to $D/\Delta_x = 22.2$ to guarantee proper resolution of the viscous effects. This results in a total amount of $1.4 \cdot 10^9$ grid cells, which to the knowledge of the authors is the largest domain employed so far for this kind of problem.
In the reference run (case: \textit{Ref}) the upper layer of the sediment bed was changed from immobile to mobile. To provide an un-biased initial condition, these particles were released at slightly higher positions. The Shields number, describing the mobility of the particles was varied by varying the relative submerged density \( \rho^\prime = (\rho_p - \rho_f)/\rho_f \) with \( \rho_p \) the particle density starting with a value slightly above the threshold of initiation of motion\(^1\). Subsequently, additional simulations with varying mass loading (cases \textit{FewPart} and case \textit{ManyPart}) and mobility (cases \textit{LowSh} and \textit{HighSh}) were carried out to elucidate the effect of these parameters on bed load transport. The specific values of the six simulations are assembled in Tab. 1. After initialisation, the simulations were run until an equilibrium between erosion and deposition was obtained\(^1\). For each simulation, statistical data was gathered over more than 240 bulk units \( \tau_b = H/U_b \).

<table>
<thead>
<tr>
<th>Case</th>
<th>( N_{p,\text{fixed}} )</th>
<th>( N_{p,\text{mobile}} )</th>
<th>( \rho^\prime )</th>
<th>( \rho^\prime/\rho_{\text{crit}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fix</td>
<td>27000</td>
<td>0</td>
<td>( \infty )</td>
<td>0.00</td>
</tr>
<tr>
<td>Ref</td>
<td>13500</td>
<td>13500</td>
<td>0.116</td>
<td>1.18</td>
</tr>
<tr>
<td>FewPart</td>
<td>13500</td>
<td>6750</td>
<td>0.116</td>
<td>1.18</td>
</tr>
<tr>
<td>ManyPart</td>
<td>13500</td>
<td>27000</td>
<td>0.116</td>
<td>1.18</td>
</tr>
<tr>
<td>LowSh</td>
<td>13500</td>
<td>13500</td>
<td>0.182</td>
<td>0.75</td>
</tr>
<tr>
<td>HighSh</td>
<td>13500</td>
<td>13500</td>
<td>0.074</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Table 1. Specific values of the simulations.

4 Results

4.1 Fluid–Particle Interaction

The five simulations of particle-laden flows were analysed by means of suitable statistical tools. In particular, the wall-normal profiles of the velocity and the fluctuations of both the fluid and the disperse phase were compared to the unladen flow of case \textit{Fix}\(^3\), the hydraulic resistance of the particles on the flow was addressed in terms of the porosity as a ratio of solid phase over fluid phase\(^3\), and the correlation between the interaction between the fluid and the dispersed phase was investigated by an analysis of the flow structures and the developed particle clusters\(^4\). In the following, a concise summary is given.

Case \textit{Ref}

The reference run (Fig. 1) produces two dunes in the computation domain with their major axis in spanwise direction. The dunes are roughly sinusoidally shaped and travel on a layer of principally mobile but resting particles usually termed inactive. This is similar to patterns observed in experiments\(^7\). The distance between the two dunes is \( 12H \) on average. The layer of inactive particles shows an increased porosity, because more than 26% of the mobile particles are moving. This allows a significant flow in the interstitial at \( y < 0 \). On top of this inactive layer, the dunes form a bed load layer with a thickness of more than \( 2D \). This enhances the turbulent fluctuations in this regions substantially. The large scale fluid structures adapt to the streamwise length scales of the clusters, which stabilises the observed pattern.
Figure 1. Instantaneous particle distribution of case Ref. Iso-surfaces of fluid fluctuations in blue $u'/U_b = -0.3$, particles in yellow: fixed, white: white = $|u_p| < 1.5u_T$, black = $|u_p| > 1.5u_T$.

Case FewPart

Reducing the mass loading results in inactive particle structures oriented in streamwise direction (Fig. 2). These so-called ridges elongate the coherent fluid structures in the streamwise direction and generate a secondary flow in the troughs. Fast travelling particles move in the troughs and occasionally erode the ridges at different locations. As the fixed bed is not completely covered by inactive particles, spatial inhomogeneity in the spanwise direction develops that allows a significant flow with high turbulent intensity in the region $y < 0$. The relative amount of moving particles is reduced in comparison to Ref, as the inactive ridges are self-stabilising through hiding and shading mechanisms induced by the spatial inhomogeneity. The ridges extend over $12H$ in the streamwise direction on average and move very slowly. This results in a separation of the morphological time scales of the particle clusters from the fluid time scales.

Figure 2. Same as Fig. 1, but for case FewPart.
Case ManyPart

A higher mass loading enhances the development of the dune-like structures observed in case Ref. With the fluid mass flux kept constant, the fluid is displaced upwards by the particles which results in an accelerated flow in the outer region of the channel. The dunes travel across a closed layer of resting particles with a spacing of roughly $6H$, although the dunes are not as pronounced as in case Ref. The coherent fluid structures increase in size as the turbulent fluctuations are strongly enhanced by the large amount of mobile particles.

Case LowSh

Due to the low mobility of the particles of case LowSh, almost all mobile particles settle onto the fixed particles and form a closed plane bed. Only a small percentage of particles is eroded. These particles travel individually across the closed bed and are randomly distributed. The lack of hiding and shading mechanisms leads to higher translational particle velocities. This dilute regime does not have a large impact on the fluid flow. Hence, the
turbulent structures resemble the large-scale behaviour observed for unladen flows over rough walls with a high submergence\(^8\).

**Case HighSh**

A higher mobility of the particles has a somewhat similar impact on the flow as observed in case ManyPart. Although the mass loading is lower, the percentage of eroded particles is larger than observed in Ref, which enhances the development of dune-like structures. The dunes are less pronounced in this case compared to Ref and ManyPart. The larger amount of mobile particles accelerates the outer flow and enhances the turbulent fluctuations.

### 4.2 Modification of Fluid Statistics

The effect of the particles on the mean fluid properties can be analysed by the porosity of the disperse phase \(\phi_{AT}\). It is a measure to quantify the amount of fluid displaced by the disperse phase\(^{15}\). Similar to the porosity, the intrinsic average of a fluid property \(\bar{\theta}\) is evaluated and denoted by angular brackets \(\langle \theta \rangle\), i.e. only the part of the domain occupied by fluid is accounted for in the averaging procedure. As expected, the stronger the decrease of the porosity in the near-wall region, the slower the fluid in that region (Fig. 6 a and b). This deceleration is compensated by an accelerated outer flow. The turbulent fluctuations of the flow are enhanced by the mobile particles in all cases (Fig. 6 c and d). Obviously, the Reynolds shear stress \(\langle u'v' \rangle\) increases with the number of moving particles. The stresses of the streamwise component, \(\langle u'u' \rangle\), of case FewPart and ManyPart exceed the maximum of case Ref peaking at \(y = -0.1H\) and \(y = 0.2H\) respectively. While the resting particles of the cases Ref, LowSh, and HighSh form a layer that is seen by the fluid as a homogeneous plane bed, the ridges of case FewPart and the dense dunes of case ManyPart introduce heterogeneity and act as roughness elements. For LowSh and HighSh the relative increase of \(\langle u'u' \rangle\) is quite small compared to the other three stresses displayed while significant enhancement is reported for the Reynolds stresses for all five cases.
Figure 6. Wall-normal profiles of several statistical quantities describing the continuous phase: a) porosity of the sediment bed, b) mean fluid velocity (streamwise component) c) mean fluid fluctuations (streamwise component), and d) Reynolds shear stress.

5 Concluding Remarks

Highly-resolved simulations of five particle-laden flows across an idealised sediment bed were carried out to elucidate the impact of mass loading and mobility on bed load transport. The resulting fluid structures and particle patterns were investigated by suitable statistical tools, to describe the key mechanisms of the interaction between the continuous and the disperse phase.

Acknowledgements

The present work is funded by the German Research Foundation (DFG) via the project FR 1593/5-1. The authors gratefully acknowledge the Gauss Centre for Supercomputing and its staff for providing computing time and helpful advice during a Large Scale Project on the supercomputer JUQUEEN at Jülich Supercomputing Centre.
References


Comparison of Navier-Stokes-Fourier Equation and Grad’s Moment Equation Solutions for Turbulence

Gregor Gassner¹, Manuel Torrilhon², Andrea Beck³,
Sophie Knechtel¹, and Thomas Bolemann¹

¹ Institute for Aerodynamics and Gasdynamics, Universität Stuttgart, 70569 Stuttgart, Germany
E-mail: gassner@iag.uni-stuttgart.de

² Center for Computational Engineering Science, RWTH Aachen,
Schinkelstr. 2, D-52062 Aachen, Germany

In this work we consider and compare different gasdynamic models for the simulation of turbulent flows. We choose three different continuum models based on the full Boltzmann equations, namely the Navier-Stokes-Fourier equations, the Grad 13 equations and the regularised Grad 13 equations. For the numerical experiment a massively parallel discontinuous Galerkin framework is developed tailored to the Blue Gene/Q architecture and used for direct numerical simulation of a canonical turbulent flow with all three models.

1 Introduction

The Boltzmann equation for the density distribution of particles in a fluid is the most fundamental equation of non-equilibrium gas kinetics and thus fluid motion. Due to its generality however, the numerical solution of the Boltzmann equation is only feasible for systems with low total particle count, e.g. rarefied gases. For cases with higher particle counts, the equilibrium assumption is often valid, which allows the use of the well known equations of gasdynamics, the Navier-Stokes-Fourier system. In between these two regimes and overlapping them lies a physical regime where the validity of the equilibrium assumption is not clear-cut. In here, the numerical solution of the Boltzmann equation is still too computationally demanding, while the Navier-Stokes-Fourier system loses its validity. One way of tackling this problem is finding better mathematical descriptions of this regime, i.e. different sets of equations, often termed the “extended gasdynamics models”.

The goal of this work is to take a first significant step towards the evaluation of turbulent solutions for these extended gasdynamics models, including the Grad 13 and the regularised Grad 13 (R13)¹,² equations. These equations are higher order approximations of the Boltzmann equation with respect to the Knudsen number Kn, see e.g., Chapman and Cowling³. The Knudsen number is defined as the ratio of the mean free path $\lambda$ and a characteristic physical length scale $L$ of the problem

$$Kn = \frac{\lambda}{L}.$$  \hspace{1cm} (1)

If the Knudsen number is small, i.e. $Kn << 1$, the continuum assumption is valid. In this case, the compressible Navier-Stokes-Fourier (NSF) equations are used for the approximation of compressible turbulent flows. If we relate the NSF equations to the Boltzmann equation, we can interpret them as a first order approximation with respect to the Knudsen number.
number. Thus, solutions of the NSF equations are $O(Kn)$ accurate with respect to Boltzmann. Obviously, only for small values of $Kn$, we can expect the NSF equations to cover all the physical aspects and phenomena included in the Boltzmann equation.

There are two typical flow scenarios with large values of $Kn$:

- If $\lambda$ is large, i.e. the mean free path of molecules is comparable to the typical length scale of the problem. This is for instance the case for atmospheric re-entry vehicles.

- If $L$ is small, i.e. the typical physical length scale of the problem is in the range of the free path. This is typically termed micro fluid dynamics and is immanent in a lot of technical devices, as for instance in the cooling of microprocessors used e.g. in the supercomputer JUQUEEN.

In this work, the focus is on the second class of problems, i.e. on the effect of small length scales $L$. In contrast to classical micro fluid dynamics, where the length scale is related to the physical length of a technical device, we focus in this work on small scale structures at high local Knudsen number existent in the flow itself.

We presume that due to the massive range of spatial (and temporal) scales in turbulent flows, the local Knudsen number for the turbulent small scale physics is orders of magnitude larger than the Knudsen number for the large scales. Thus, accompanying the well-known turbulent scale cascade from large eddies to Kolmogorov’s smallest features exists a range of characteristic Knudsen numbers, associated with the local length scale, where the smallest physical feature experience the highest Knudsen number. Thus, we raise the question whether the equilibrium-based Navier-Stokes-Fourier flow model is appropriate to describe the small scale turbulence physics. Exact analytical turbulent solutions are not available and accurate approximations of turbulent solutions of the full Boltzmann equations are out of scope with the currently available computing resources. To still get an answer whether the Knudsen number range has an effect on the small scale physics, instead of the full Boltzmann equations we consider continuum models with higher Knudsen approximation order than the NSF system: Grad 13 is a second order model $O(Kn^2)$, R13 is a third order model $O(Kn^3)$, see Chapman and Cowling.

Based on these continuum flow models of extended gas dynamics, we use direct numerical simulation to generate accurate turbulent solutions for all the different models.

Direct numerical simulation means resolution of all spatial and temporal scales. However, a characteristics of turbulent flow is the wide range of scales in time and space. Thus, a full resolution of turbulent flows is a challenging task for the simulation framework. The essential ingredient to such a simulation is access to massive computing resources as provided by the Forschungszentrum Jülich, namely the Blue Gene/Q system JUQUEEN. From these direct numerical simulation results, we can then investigate the question whether the NSF and extended gasdynamics equations agree at the small scale physics of turbulence and use this as a starting point for investigations into the applicability range of the different equation systems.

## 2 Model, Method and Code

All the physical models used can be formulated in a similar mathematical form:

$$U_t + \nabla \cdot F(U) - \nabla \cdot F_v(U, \nabla U) = S(U).$$  \hspace{1cm} (2)
The vector $U$ contains the conserved variables mass, momentum and energy in the case of Navier-Stokes-Fourier. In the case of the extended models Grad 13 and R13, additional variables for stress and heat flux are introduced. $F(U)$ denotes the advection flux matrix, $F_v(U, \nabla U)$ denotes the viscous flux matrix and $S(U)$ a source term. For Navier-Stokes-Fourier we have $S = 0$, whereas the source term is non-linearly dependent on the solution for the extended models Grad 13 and R13. As a special case, we have for the Grad 13 model that the viscous fluxes are zero, i.e. $F_v = 0$, which means that it can be discretised similar to a pure hyperbolic problem.

To rule out numerical artifacts in our results, we carry out direct numerical simulations (DNS) of a canonical turbulent test case for all the different models. As we are interested in the small scale physics, we not only need to resolve most of the small scales in a turbulent flow but even have to capture them accurately to ensure that the experimental observations are driven by the actual physics of the model and not by the numerical approximation. In addition, we avoid the influence of boundary conditions on the small structures by choosing a triply-periodic flow of homogeneous isotropic turbulence, the well-studied Taylor-Green vortex. The Taylor-Green vortex flow is characterised by the non-linear interaction of large scale vortices and the subsequent development of the characteristic turbulence energy spectrum with a large range of scales. Fig. 1 depicts results from a simulation showing the temporal development from large ordered scales to turbulent chaos with small scale interaction. The Taylor-Green vortex has been studied extensively in literature and serves as a workhorse for investigations into the physics of vortex dynamics as well as the faithful numerical representation of turbulence. It is particularly well-suited for the proposed investigations for two main reasons: 1) Due to the temporal development of the scales and the transition from laminar flow to turbulence, it allows us to cover different local Knudsen number regimes (from moderate to high) in one run. 2) Since no production of turbulent kinetic energy at the large scale occurs, once the large scales have broken down, the flow behaviour is completely determined by the dissipative scales. Since we are interested in investigating effects in that range, we can thus expect these effects to be detectable and not masked by the (order of magnitude larger) physics on the large scales.

For the spatial discretisation of this problem we choose a discontinuous Galerkin approach, see e.g. Kopriva. The discontinuous Galerkin method is a finite element type discretisation with ideas from finite volume methods incorporated to address stability is-
sues when simulating advection dominated (hyperbolic) problems. We use a special variant of the method tailored to the problem we want to solve. We restrict ourselves to hexahedral and Cartesian meshes and choose a high order nodal tensor product basis with polynomial degree $N$ to represent the solution inside each element. The degrees of freedom are determined by a variational formulation, where the inner products are replaced by Gaussian quadrature. To optimise the number of necessary operations, the nodal basis is collocated with the discrete quadrature rule resulting in the so-called discontinuous Galerkin collocation spectral element variant, see e.g. Kopriva\textsuperscript{5} and Hindenlang \textit{et al.}\textsuperscript{4} and references therein for more details.

The time discretisation is purely explicit with a standard low storage five stage fourth order accurate Runge-Kutta method. The single most important motivation for choosing the Runge-Kutta discontinuous Galerkin approach is the inherent parallelism of the resulting discretisation and the resulting access to massively parallel super computing architectures such as the Blue Gene/Q.

This discretisation is implemented in the Fortran 95 code “STRUKTI”. The code has been developed from scratch with a strong emphasis on high performance computing (low memory consumption, high parallel efficiency). The code is flexible in the sense that it allows easy implementation of new physical continuum models, such as the Navier-Stokes-Fourier, Grad 13 and R13 equations. Based on the discontinuous Galerkin method, the code allows to choose the polynomial order of the approximation $N$ as a parameter. Typically, we are interested in high order computations with polynomial degrees $N > 5$. These high order variants of the method are discretisations with low dispersion and dissipation error\textsuperscript{8}, which is important when simulating a multiscale problem. An additional important side effect of using high order methods is that for higher polynomial degrees, the data layout is much more compact, which allows to make efficient use of the available memory bandwidth. Furthermore, the amount of operations scales non-linearly with the degree $N$ resulting in a better ratio of operations per memory access and operations per communication data\textsuperscript{4}. The code STRUKTI is parallelised based on MPI. In the design of the parallelisation, the following aspects where incorporated:

- No overhead for MPI domain faces. We split the DGSEM operator in such a way that no parts of the operators are computed twice at the MPI interface. Thus, when increasing the amount of processors, i.e. the communication surface, no extra operations are introduced. This is important as we are interested in massively parallel computations with maximum amount of possible domain decomposition for a given grid.

- Communication latency hiding. The DGSEM operator is split in such a way that parts needing no data from the neighbour processors are computed during the non-blocking MPI communication. Depending on the given architecture and the ratio of computing bandwidth and communication bandwidth, it is possible to completely hide the communication.

- Perfect load balancing. As only structured grids are used, the domain partition is done automatically by STRUKTI yielding always perfect load balancing in the case of a fully periodic cube domain, as considered in this work.
3 Scaling and Results

To evaluate the scaling of STRUKTI on JUQUEEN, we introduce a performance index (PID) which is computed as

\[
PID \left[ \frac{\mu s}{DOF} \right] := \frac{\text{Wallclock Time} \times \text{Number of MPI Ranks}}{\text{Number of Degrees of Freedom} \times \text{Number of Timesteps}}
\]

and basically represents a normalised CPU time needed for one degree of freedom (DOF) to perform one time step. For a discontinuous Galerkin discretisation the number of DOF, or number of total grid points is computed as \( \text{Number of Elements} \times (N + 1)^3 \). For a perfect scaling across different computational configurations would mean that the PID is constant in all cases.

Fig. 2 shows some experimental results on the influence of the computational load (DOF on a processor) and the hyper threading on the PID. In the left part of the figure, we fix the total number of MPI ranks and choose different configurations of number of computational cores and threads per core. As can be clearly seen, the PID depends quite strongly on the specific processor load. It is interesting to note that by decreasing the load on a processor the PID gets smaller and thus STRUKTI gets faster. This can be clearly related to caching effects. Furthermore, it can be observed that the influence switching from one thread to two threads (by keeping the total number of ranks constant) is only small on the PID (red and green curve in the left figure). This basically means that the two threads speed up the computation by a factor of two. In contrast, increasing the threads to the maximum of four with constant total ranks (i.e. only 1/4 the number of cores) increases the PID nearly by a factor of two, which means that there is no real benefit anymore for STRUKTI using hyper threading with four threads. The optimum is clearly two threads. The behaviour in the case of two threads is also investigated for different number of total ranks, plotted in the right part of Fig. 2. As can be seen, the general behaviour is quite similar when increasing the number of cores by a factor of four, except in the case of very low load on a processor, where some scaling effects can be seen.

![Figure 2. Left: Influence of hyper threading and of the degrees of freedom per processor (load) on the performance index (PID) of STRUKTI. Right: Additional influence of the number of ranks on the PID.](image-url)
The strong scaling behaviour of STRUKTI is investigated in different configurations and illustrated in the left part of Fig. 3. This plots shows strong scaling results for different total problem sizes from 8 up to 32,768 MPI ranks with two threads per core. The first curve (red) represents the extreme case with a total of 32,768 elements. This means that for the largest computation with 32,768 MPI ranks in the strong scaling experiment only one element is left on the processor. Even in this extreme case, STRUKTI on JUQUEEN retains a very good 85% strong scaling speed up. By increasing the total load for the experiment (number of grid cells) one effectively increases the load on a processor. In this specific case, a slight increase of the total number of elements by a factor of 8 (from 32,768 up 262,144) has an interesting effect on the scaling (green curve): If we compare for this scaling experiments the behaviour of the PID depending on the corresponding processor load (green curve in right plot of Fig. 3), one can see that strong caching effects occur thus resulting in a super-linear speed up of the computation reaching over 120% strong scaling. By increasing the total problem size by a factor of four (blue curves) one can see in the right plot that there is also a caching effect, although not as strong as in the green case consequently resulting in a slightly worse super-linear strong scaling. However, as the total amount of elements (and thus potential MPI domains) is larger, one can use more MPI processor for the computation and we found out that the blue curve behaves similar to the green curve when reaching lower loads per processor, i.e. higher total number of processors.

This simulation framework is used to investigate the influence of the model on the simulation of flows with small scales. A highly resolved direct numerical simulation with all models was performed for the Taylor-Green-Vortex with Mach number $Ma = 0.1$ and Reynolds number $Re = 100$. For each computation two different resolutions were used for the fully periodic $[0; 2\pi)^3$ computational cube domain. The medium resolution is $24^3$ grid cells with polynomial degree $N = 7$ resulting in $192^3$ grid points which equals roughly to seven million degrees of freedom per variable. The fine resolution is refined by a factor of two in each direction resulting in $364^3$ grid nodes, which is roughly 56 million degrees of freedom per variable. Each model is computed with both resolutions to compare the simulation results and make sure that no influence from the numerical discretisation
augments the outcome and a possible conclusion. All computations are performed with maximum number of MPI ranks, in case of the medium resolution we used two threads and 6,912 cores resulting in $24^3$ MPI ranks and in case of the fine computation we used 55,296 cores resulting in $48^3$ MPI ranks, i.e. always one element on a processor. The PID for Navier-Stokes-Fourier and Grad 13 was in the range of approximately 55\textmu s, whereas the R13 variant is about a factor of three slower compared to the other models with respect to this index.

Fig. 4 shows the resulting dissipation rate of the kinetic energy over time. The dissipation rate is a measure of the small scales and thus an appropriate quantity for our investigations. As mentioned above, the range of scales in the Taylor-Green-Vortex problem increases over time until a maximum dissipation (at about $t = 5$ in this case) is reached. The left part of the figure shows that at first glance all models yield the same result. However, when looking more closely, especially in the part of maximum dissipation, i.e. small scale dominant regime, one can see a difference between the models (right plot of the figure). To further investigate this difference, we have computed the absolute difference between the dissipation rates and plotted these curves in Fig. 5. The left part of the plot shows the difference between the Navier-Stokes-Fourier simulation and the extended gasdynamics results. It is clear the the difference increases with the increasing range of scales and maxes out at about $t = 5$ where the maximum dissipation rate occurs. The higher the range of scales and the smaller the smallest scales, the higher the difference between the models. In the right part of the figure, we also have a comparison between the two extended gasdynamics models. Note that these are both higher order approximations to the Boltzmann equation with respect to the Knudsen number. The difference is about two order of magnitudes smaller compared to the left plot and again grows over time until the maximum is reached. All these results so far support the initial hypothesis that the different gasdynamics models behave differently for small scale physics driven fluid flows. Although the difference is small, it is related purely to the underlying model as the results did not change for the medium and fine resolution configuration. To the authors’ best knowledge, these

![Figure 4. Left: Plot of the Taylor-Green-Vortex kinetic energy dissipation rate for all different gasdynamics models over time. Right: Zoomed view in the range of highest small scale dissipation.](image-url)
Figure 5. Left: Plot of the difference of kinetic energy dissipation rate for the extended gasdynamics models in comparison to the Navier-Stokes-Fourier result. Right: Plot of the difference of the kinetic energy dissipation of the two extended gasdynamic models.

are the first investigations of this kind and further test runs with higher Reynolds numbers are underway.

Concluding, the massively parallel computing architecture of JUQUEEN is ideally suited for our application and allowed us to make a first step into investigating the influence of different continuum models on the small scale behaviour of turbulent fluid flow. The first findings are encouraging and more complex (computationally and physically) simulations are underway.

References

Numerical Investigation of Reacting and Non-Reacting Flows in Gas Turbine Related Configurations

Andreas Fiolitakis, Christian Eberle, Antonio Filosa, Juliane Prause, and Berthold Noll

German Aerospace Center (DLR), Institute of Combustion Technology, 70569 Stuttgart, Germany
E-mail: {andreas.fiolitakis, christian.eberle, antonio.filosa, juliane.prause, berthold.noll}@dlr.de

The present work addresses the investigation of reacting flows in gas turbine related configurations by means of Computational Fluid Dynamics (CFD). Different combustion problems are discussed, among them autoignition processes, soot formation, and combustion in a generic bluff body burner. The numerical methods used here are unstructured finite-volume techniques, where conservation equations for momentum and energy are solved along with balance equations for combustion species and turbulence model quantities. Also stochastic methods, namely the transported probability density function (TPDF) method, are used to study combustion processes. To solve the TPDF model equations a hybrid finite-volume/Lagrangian Monte-Carlo method is applied.

1 Introduction

Although the use of combustion as a source of heat and power is one of the oldest technologies in the world, its theoretical understanding today is still an area of research. Limited availability of fossil fuels and an increased awareness of the environmental problems, i.e. air pollution through nitric oxides (NOₓ) and soot as well as emissions of green-house gases like carbon-dioxide (CO₂), have been the main driving forces for the development of new combustion technologies. This involves both, the improvement of existing and the development of new burner technologies. At the same time, safety concerns like flame stability and control of ignition-processes must be met. The fulfilment of these requirements in a technological framework, i.e. the wide availability of energy at acceptable cost, is a difficult task faced in todays engineering. Technical combustion takes place in turbulent flows which are necessary to get sufficient mixing rates of reactants and to accomplish the required fuel-to-air mixing ratio for combustion. Through these mixing processes, turbulence interacts with the chemistry of different chemical components (species). In order to predict and understand these complex physical processes by means of calculations, Computational Fluid Dynamics (CFD) has become an important tool in the field of combustion engineering and research. At the Institute of Combustion Technology at the German Aerospace Center (DLR) one of the main areas of research is the development of numerical methods and physical models applicable in the CFD of reacting and non-reacting flows in gas turbine related configurations. In the present work, both the development of methods and the application to combustion problems are addressed. Sec. 2 provides a brief overview on the methods employed as well as on their scalability on JU-ROPa. Results are presented in Sec. 3, where the Subsec. 3.1 and 3.2 examine reliability problems in combustion. In Sec. 3.1 computational models for the prediction of unwanted autoignition in reheat burners are investigated, whereas Sec. 3.2 deals with the unsteady heat release in a generic bluff-body flame burner. This burner emulates the conditions in
todays aero-engines, where unsteady heat transfer is responsible for high thermal loads on
the mechanical structure. Pollutant formation is discussed in Sec. 3.3. Soot is still a big
problem in aero-engines and responsible not only for a significant amount of pollution,
but also for high thermal loads on the combustor through heat radiating soot. The com-
putations presented in Sec. 3.3 continue previous work at DLR and apply the outcomes to
the computation of a lifted, non-premixed, turbulent ethylene-air flame. Experimental data
for soot concentration and temperature are available for model validation. In Sec. 3.4 the
application of a stochastic method for combustion is presented. This method is known as
transported probability density function (TPDF) method, which is capable to accurately
describe the interaction of turbulence and chemistry. Here, the applicability of the method
to a premixed model combustor is investigated.

2 Methodology

For the studies presented in this work the DLR THETA code is used. It provides a large
number of physical models along with appropriate numerical solution methods which are
described next.

2.1 Finite-Volume Method

The DLR THETA code is an incompressible solver version of the DLR TAU code. Using
unstructured finite-volume methods, the conservation equations for momentum and energy
are solved numerically along with SIMPLE or a projection method. Latter pressure cor-
rections methods (which have been implemented initially for purely incompressible flows),
have been extended in their formulation for the treatment of “mildly compressible” flows
at low Mach numbers. This enables the calculation of acoustic wave propagation. To this
end, THETA provides also acoustic impedance boundary conditions. In addition a variety
of model equations for turbulence are available. These models cover the range from simple
two-equation turbulence models over hybrid Scale-Adaptive Simulations (SAS) techniques
to Large-Eddy-Simulation (LES) models (i.e. the WALE9 and Smagorinsky10 model). For
the modelling of turbulent combustion various combustion models are available. This in-
volve finite rate chemistry models, where closure for turbulence chemistry interaction
may be provided by an assumed probability density function (PDF) approach in temper-
ature and composition as well as steady flamelet-PDF models for the modelling of non-
premixed combustion. In addition to the combustion models a detailed kinetic model for
soot formation is available. This model is based on a sectional approach and has been
developed during the last decade at DLR1,2,5.

The finite-volume method is programmed in ISO C (C89). A domain based parallelis-
isation of the finite-volume solver is realised by means of the MPI paradigm and allows to
run the code on parallel systems. A benchmark on JUROPA for the finite-volume solver
is given in Fig. 1, where the calculation discussed in Sec. 3.3 is presented. A mesh of a
size of 3.1 million grid points is used here together with a reaction scheme and soot model
which results in 71 unknowns per grid point. Thus, the problem involves 220 million
degrees of freedom (DoF). It turns out that for such large problems the code may scale up
to 512 cores. This is mainly due to the fact that the most expensive computational step is
the implicit treatment of chemistry which can make up to 80 % of the computational time.
At the same time the solution of chemistry does not require any MPI communication, since its evaluation takes place only pointwise.

2.2 Lagrangian Monte-Carlo Method

An additional combustion model which has been recently introduced to THETA is the TPDF method\textsuperscript{11} which solves a transport equation for the one-point one-time PDF of the thermochemical variables in a turbulent reacting flow. By using this PDF all higher moments of the thermochemical variables can be computed. Compared to the aforementioned combustion models, the transported PDF method offers the distinctive advantage, that the effect of chemical reactions appears in closed form. Turbulence chemistry interaction is therefore modelled accurately.

The transport equation for the PDF, however, is an equation with a high dimensionality. Therefore, a Lagrangian Monte-Carlo is used to solve the PDF transport equation, where a discrete representation of the PDF is given by a set of stochastic particles. The evolution of these stochastic particles is governed by a set of stochastic differential equations in physical space as well as in the composition space\textsuperscript{11}. Thus, the Monte-Carlo solver is essentially a solver for stochastic differential equations. Flow velocities as well as turbulent time and length scales are provided by the finite-volume solver outlined in the previous subsection. The finite-volume solver receives at each iteration from the Monte-Carlo method the expectations of density and dynamic viscosity and provides in turn the flow velocities and turbulent time and length scales to the Monte-Carlo method. A benchmark for the test case discussed in Subsec. 3.4 is given in Fig. 2. Here a testcase with 200,000 grid cells and 32 particles per cell is considered. The reaction mechanism used involves 21 unknowns which results in a total of 134 million degrees of freedom. For this problem size a scalability of the code up to 512 cores is observed. In contrast to the parallelisation of the finite-volume solver, which utilises a domain decomposition, particle partitioning is used in the Monte-Carlo method. Here, the stochastic particles are distributed on different processes. The domain, however is the same on all processes. The advantage of this approach is that a good load balancing between the processes is accomplished automatically. The disadvantage is the large communication overhead when the domain size increases. To reduce the communication overhead, the initially purely MPI parallelisation has been extended to a hybrid MPI-OpenMP parallelisation using the tools provided on
JUROPA. In this case, MPI communication occurs only between different computational nodes. On each node the data are processed by a shared memory parallelisation. The benefit is illustrated in Fig. 3 for a methane-air jet flame which has been computed on JUROPA. The hybrid MPI-OpenMP approach provides still a significant speed up beyond 96 cores whereas no speed up is found for the purely MPI based parallelisation.

3 Results

In the next subsections major outcomes of this work are presented. First autoignition processes in a generic mixing zone of a reheat combustor are discussed in Subsec. 3.1, followed by an analysis of a bluff-body burner in Subsec. 3.2. Soot prediction is the subject of Subsec. 3.3. Subsec. 3.4 presents results of the TPDF method.

3.1 Autoignition in Reheat Combustor

Low emissions in gas turbine combustion are achieved by preliminary mixing of fuel and air upstream of a combustion chamber. However, the time required to obtain sufficient mixing upstream of the combustor is limited by the autoignition time of the explosive mixture. Premature autoignition in the mixing section needs to be avoided to ensure a reliable operation. To provide a deeper understanding of the processes yielding to autoignition, the mixing zone of a generic reheat combustor depicted in Fig. 4 is studied experimentally and
Problematic are the random occurrences of autoignition events in the mixing section; the prediction of those by means of CFD is the prime objective here. To this end, SAS studies in combination with a finite-rate chemistry combustion model and a detailed reaction mechanism are performed. A representative result for the temperature distribution in an autoignition event is given in Fig. 5, where red colour indicates areas of high temperature. Depicted here is the region downstream the fuel injection within the mixing section, where a pocket of hot gas forms which yields to autoignition. A comparison to experimental data shows that the SAS model in combination with a suited finite-rate chemistry mechanism is capable to predict these phenomena accurately.

3.2 Investigation of Bluff-Body Flame Burner

Bluff-bodies can be used to anchor the flame and to help to maintain continual combustion. A bluff-body flame burner consists of any immersed object whose downstream shape is blunt enough to create a recirculation zone in the stream behind the object\(^8\). A low-speed region filled with combustion products is created there which help to stabilise the flame. Such flames are used frequently in academia to mimic the situation in aero-engine combustors. In Fig. 6 the temperature distribution in the bluff-body burner studied here is shown, where red colour indicates areas of high temperature. The wedge-shaped bluff-body is clearly visible in the computational domain. Air entering from the left boundary passes the wedge and is mixed with fuel (in this case methane) which is ejected from holes on the wedge surface into the air stream. Behind the wedge the reaction zone is clearly visible as the high temperature region. This region of the burner builds a Rijke tube which induces thermo-acoustic oscillations in the burner\(^6\). Such effects occur in practice in combustors and have to be avoided since they can result in an unsteady heat load on the walls and subsequently in a fatigue of the wall material. The burner provides therefore a simplified framework to study this technically relevant problem both experimentally and numerically. Computational results of the global averaged heat load on the wall as function of combustion-air ratio are given in Fig. 7. These results are obtained using SAS and LES methods.
3.3 LES of a Turbulent, Non-Premixed, Sooting Flame

The reduction of soot is urgently required for several combustion processes. The understanding of soot is, however, still incomplete and its kinetic modelling is an active field of research. Soot models vary in their degree of complexity starting from simple, two equation soot models and ending up with large models with several hundred components. The DLR soot model is a compromise between accuracy and computational cost and has been validated for a variety of problems\textsuperscript{1,2,5}. The present work aims to combine this soot model with LES turbulence models for the calculation of a turbulent sooting jet flame. The flame is fuelled with a jet of ethylene which is surrounded by a coflow of air. Figs. 8 and 9 give an impression of the highly unsteady temperature field and the pockets of soot in the flame. For the present flame experimental data are available to validate the accuracy of the calculation. A comparison of the computed average soot distribution is given in Figs. 10 and 11 where good agreement between calculation and experiment is found. Due to the high spatial resolution and the complexity of soot modelling, the computational cost for obtaining these results is very high. The scalability of THETA for the present test case, which is
demonstrated in Fig. 1, allows to complete such calculations on massively parallel systems within a few months if up to 512 cores are used.

3.4 Transported PDF Modelling of Premixed Burner

The TPDF method outlined in Sec. 2 has been validated during the last years and improved in terms of models and scalability. Recent work demonstrates its capability to account for heat losses in confined flows. The present work applies this methodology in the computation of a generic single nozzle FLOX® (Flameless Oxidation) burner, which has been investigated experimentally. The burner is designed to emulate the conditions in modern low-emission, premixed FLOX® combustors. It consists of a single nozzle at the bottom of a combustion chamber with rectangular cross section as shown in Fig. 12. Through this nozzle a lean mixture of methane and air at an equivalence ratio of 0.7 is injected into the combustion chamber with rectangular cross section as shown in Fig. 12. Through this nozzle a lean mixture of methane and air at an equivalence ratio of 0.7 is injected into the combustor. Due to the high momentum of the jet and the off-centre placement of the nozzle (cf. Fig. 12) a large backflow pattern is established in the combustor as the computed streamlines in Fig. 13 show. This backflow pattern transports hot combustion products back to the nozzle as demonstrated in Fig. 14 where a high temperature region stretches from $X/D = 5$ to $X/D = 20$. These combustion products help to ignite the incoming fuel-air mixture thus stabilising the whole flame. The calculated maximum temperature of roughly 1570 K at $X/D = 27$ (see Fig. 14) is significantly lower than the adiabatic equilibrium temperature of 2064 K. This indicates that heat losses are of paramount importance in the burner, which is confirmed by experimental data. The computations presented here can reproduce this effect of wall heat losses very well.

Figure 14. Temperature distribution in centre plane of generic FLOX®.
4 Summary and Conclusions

The computations presented in this work cover a wide range of combustion problems encountered in modern combustion research. This involves reliability issues (e.g. autoignition and unsteady heat transfer due to acoustic instabilities), soot formation and studies of modern combustion systems like the FLOX® concept. It is found that to model these systems accurately requires computationally expensive models like LES and TPDF methods. Using the high performance computing resources provided by the Jülich Supercomputing Centre it is possible to improve the scalability of the codes and to run such computations in a reasonable time. Up to now a scalability of up to 512 cores is achieved for problems involving at least 220 DoF which gives an option for treating even larger problems in the future.

References

Computer Simulation of Flow Past Superhydrophobic Striped Surfaces

Jiajia Zhou1, Aleksey V. Belyaev2,5, Evgeny S. Asmolov2-3,4, Olga I. Vinogradova2,5,6, and Friederike Schmid1

1 Institut für Physik, Johannes Gutenberg-Universität Mainz, Staudingerweg 7, 55099 Mainz, Germany
E-mail: {zhou, friederike.schmid}@uni-mainz.de

2 A. N. Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Science, 31 Leninsky Prospect, 119991 Moscow, Russia

3 Central Aero-Hydrodynamic Institute, 140180 Zhukovsky, Moscow region, Russia

4 Institute of Mechanics, M.V. Lomonosov Moscow State University, 119991 Moscow, Russia

5 Department of Physics, M.V. Lomonosov Moscow State University, 119991 Moscow, Russia
E-mail: oivinograd@yahoo.com

6 DWI, RWTH Aachen, Forckenbeckstraße 50, 52056 Aachen, Germany

We report results of Dissipative Particle Dynamics simulations for an anisotropic flow past superhydrophobic striped surfaces. We systematically vary several important parameters, such as the flow direction, the area fraction of the gas sectors, and the local slip length of the gas/liquid interface. We compare the simulation results with the numerical solution to the Navier-Stokes equations, analytic expressions, and asymptotic formulas in the limits where the local slip-length is small and large in comparison to the texture periodicity. The simulation results allow one to adjust surface properties to optimise transverse phenomena and passive microfluidic mixing. Our approach could be helpful to rational design of superhydrophobic surfaces.

1 Introduction

Fluid modelling from micrometer to nanometre scale not only is a fundamental problem in fluid mechanics, but also plays an important role in designing modern micro- and nanofluidic devices1,2. These devices have wide applications in different fields such as development of inkjet printheads for xerography, lab-on-a-chip technology, and most importantly, the manipulation and separation of DNA in molecular biology. Two assumptions which are often taken for granted in studies of macroscopic fluid phenomena, the no-slip boundary condition and the homogeneity of the surface, become doubtful in small length scales. The extension to partial-slip boundary condition and heterogeneous/anisotropic surfaces provide ample opportunities for simulations and applications.

Textured surfaces play a major role in microfluidics due to the high surface to volume ratio which greatly enhances the fluid-surface interaction. One example is the superhydrophobic Cassie surface, where no-slip area and partial-slip area arrange in a striped pattern. The partial-slip region consists of trapped gas, which is stabilised by a rough wall texture and leads to a huge slip length3 (see Fig. 1(a)). These surfaces are known to be self-cleaning and cause droplets to roll (rather than slide) under gravity and rebound (rather than spread) upon impact. Besides patterned superhydrophobic materials are im-
important in context of fluid dynamics and their superlubricating properties. In particular, superhydrophobic heterogeneous surfaces in the Cassie state exhibit very low friction, and the resulting drag reduction is associated with the large slippage of liquids. Therefore, the ability to manipulate and control flow properties by varying the surface pattern has promoted interests to quantify the effects of surface heterogeneity and anisotropy.

In general it is difficult to quantify the flow past heterogeneous surfaces. However, analytical results can often be obtained by using an effective slip boundary condition, \( b_{\text{eff}} \), at the imaginary smooth homogeneous, but generally anisotropic surface\(^4\). For an anisotropic texture, the effective slip depends on the direction of the flow and is a tensor, \( b_{\text{eff}} \equiv \{ b_{ij}^{\text{eff}} \} \) represented by a symmetric, positive definite \( 2 \times 2 \) matrix\(^5\)

\[
b_{\text{eff}} = S_{\Theta} \begin{pmatrix} b_{\text{eff}}^\parallel & 0 \\ 0 & b_{\text{eff}}^\perp \end{pmatrix} S_{-\Theta},
\]

diagonalised by a rotation

\[
S_{\Theta} = \begin{pmatrix} \cos \Theta & \sin \Theta \\ -\sin \Theta & \cos \Theta \end{pmatrix}.
\]

The tensorial formalism allows one to calculate the effective slip in any direction given by an angle \( \Theta \), provided two eigenvalues of the slip-length tensor, \( b_{\text{eff}}^\parallel (\Theta = 0) \) and \( b_{\text{eff}}^\perp (\Theta = \pi/2) \), are known. For all anisotropic surfaces, the eigenvalues \( b_{\text{eff}}^\parallel \) and \( b_{\text{eff}}^\perp \) correspond to the fastest (greatest forward slip) and slowest (least forward slip) orthogonal directions.

In this report, we present Dissipative Particle Dynamics (DPD) simulations for a flow over superhydrophobic striped surfaces and compare our results with numerical solutions to Navier-Stokes equations\(^6,7\) and some analytical theories\(^8-11\).

2 Theory

We consider a creeping flow along a planar surface, and a Cartesian coordinate system \((x, y, z)\) (Fig. 1(a)). The origin of coordinates is placed at the flat interface. The local slip-length profile is alternating no-slip (area fraction \( \phi_1 \)) and partial slip (area fraction \( \phi_2 \) and slip length \( b \)) regions, with the periodicity \( L \) (Fig. 1(b)). Our system is based on the limit
of a thick channel or a single interface, so that the velocity profile sufficiently far above the surface may be considered as a linear shear flow.

When the partial-slip region has a small slip length in comparison to the periodicity, \( b \ll L \), it has been predicted that the effective slip-length tensor \( b_{\text{eff}} \) becomes isotropic, regardless of the type of surface textures\(^8,12\),

\[
\frac{b_{\parallel \text{eff}}}{L} \simeq b \phi_2. \tag{3}
\]

In the opposite limit where the gas sector becomes perfect slip, \( b \gg L \), the eigenvalues of the effective slip-length tensor can be derived\(^13,14\)

\[
\frac{b_{\parallel \text{eff}}}{L} = \frac{2 b_{\perp \text{eff}}}{L} \simeq \frac{1}{\pi} \ln \left[ \sec \left( \frac{\pi \phi_2}{2} \right) \right]. \tag{4}
\]

For more general cases of arbitrary value of \( b \), Belyaev and Vinogradova\(^9\) suggested approximate expressions for the effective slip,

\[
\frac{b_{\parallel \text{eff}}}{L} \simeq \frac{1}{\pi} \ln \left[ \sec \left( \frac{\pi \phi_2}{2} \right) \right], \tag{5}
\]

\[
\frac{b_{\perp \text{eff}}}{L} \simeq \frac{1}{2 \pi} \ln \left[ \sec \left( \frac{\pi \phi_2}{2} \right) \right], \tag{6}
\]

These formulas are accurate over a large range of the parameters\(^9\) and recover the right asymptotic in the large \( b \) limit (Eq. 4).

A detailed analysis of the isotropic slip-length, Eq. 3, reveals that it can only apply for surfaces with a continuous slip profile\(^11\), so its validity for the striped surface, which has step-like jump of the local slip at the stripe edges, is questionable. Recently, we derived a new asymptotic formula for weakly slipping surfaces (\( b \ll L \))

\[
\frac{b_{\parallel \text{eff}}}{L} \simeq \varepsilon \phi_2 + \frac{2 \varepsilon^2}{\pi} \left\{ \ln \left[ \frac{\pi \varepsilon}{\sin (\pi \phi_2)} \right] - \gamma \right\}, \tag{7}
\]

\[
\frac{b_{\perp \text{eff}}}{L} \simeq \varepsilon \phi_2 + \frac{4 \varepsilon^2}{\pi} \left\{ \ln \left[ \frac{2 \pi \varepsilon}{\sin (\pi \phi_2)} \right] - \gamma \right\}, \tag{8}
\]

where \( \varepsilon = b/L \) and \( \gamma = 0.5772157 \ldots \) is Euler’s constant.

### 3 Simulation Method

We apply Dissipative Particle Dynamics (DPD) method\(^15-17\) to simulate the flow near striped superhydrophobic surfaces. The DPD method is an established coarse-grained, momentum-conserving method for mesoscale fluid simulations, which naturally includes thermal fluctuations. More specifically, we use a DPD version without conservative interactions\(^18\). The hydrodynamic boundary conditions are implemented using the tunable-slip method\(^19\), which models the fluid-surface interaction using an effective friction force, combined with an appropriate thermostat.
The simulations are carried out using the open source simulation package ESPResSo\textsuperscript{20}. Modifications have been made to incorporate the patterned surfaces. The simulation starts with randomly distributed particles inside the channel, and the flow is induced by applying a body force to all particles. A small body force is used to ensure the flow velocity near the wall is small, in order to avoid the effect of finite Reynolds number in simulations. The large system size leads to an increased simulation time (over $10^6$ time steps) for the flow to reach a steady state. The flow velocity is small in comparison with the thermal fluctuation; thus the measurement also requires many time steps to obtain enough statistics. In this work, velocity profiles are averaged over $10^5$ time steps. A fit to the plane Poiseuille flow then gives the effective slip length\textsuperscript{7}. The error bars are obtained by six independent simulation runs with different initialisation.

Based on the values of the velocities close to the surface, we estimate the characteristic Reynolds number of our system to be of $O(10)$, which is larger than those in real microfluidic devices. Thus, inertia effects may become important in simulations, and the Stokes equation is not strictly valid. This leads to a slight reduction of our simulation results for the effective slip in the transverse direction, but the flow in the longitudinal direction shall not be affected. To reach more realistic Reynolds numbers, we would need to reduce the body force by orders of magnitude. This would reduce the average flow velocity significantly, and the necessary simulation time to gather data with sufficiently good statistics will then increase prohibitively.

4 Results and Discussion

In this section, we present DPD simulation results and compare them with numerical calculations and analytic formulas in Sec. 2.

We start with varying $\Theta$ in a system where the no-slip and partial-slip areas are equal,
\(\phi_2 = 0.5\), and the slip length is in the intermediate region \(b/L = 1.0\). Fig. 2 shows the results for the effective downstream slip lengths, \(b^{(x)}_{\text{eff}}\). Also shown are theoretical curves calculated using Eq. 1, which can be explicitly written as

\[
b^{(x)}_{\text{eff}} = b^{\parallel}_{\text{eff}} \cos^2 \Theta + b^{\perp}_{\text{eff}} \sin^2 \Theta.
\]

(9)

Here, the eigenvalues of the slip-length tensor are obtained by numerical method and from Eqs. 5, 6. The simulation data are in good agreement with theoretical predictions, confirming the anisotropy of the flow and the validity of the concept of a tensorial slip for striped surfaces.

Figure 3. The eigenvalues of the effective slip-length tensor (symbols) as a function of gas-sector fraction \(\phi_2\) for (a) \(b/L = 1.0\) and (b) \(b/L = 0.034\). The numerical results are shown as solid curves. Also shown are asymptotic formulas Eq. 3 (dashed), Eq. 7 (dotted), and the analytic expressions Eqs. 5, 6 (dot-dashed).
Next we examine the effect of varying the fraction of slippery gas/liquid interface, \( \phi_2 \). Fig. 3(a) shows the eigenvalues, \( b_{\parallel}^{\text{eff}} \) and \( b_{\perp}^{\text{eff}} \), of the slip-length tensor as a function of \( \phi_2 \), for a pattern with \( b/L = 1.0 \). The results clearly demonstrate that the gas fraction \( \phi_2 \) is the main factor determining the value of effective slip, which significantly increases with the fraction of the slippery gas sectors. The theoretical curves match the simulation data very nicely.

We further examine the weakly slippery surfaces with \( b/L = 0.034 \) and present the

![Figure 4](image_url)

Figure 4. The eigenvalues of the effective slip-length tensor (symbols) as a function of local slip length \( b \) at the gas/liquid interface. The top figure (a) are results for large values of \( b \) and \( \phi_2 = 0.5 \), and the bottom figure (b) are results for weakly slipping surface and \( \phi_2 = 0.3 \). The numerical results are shown as solid curves. Also shown are asymptotic formulas Eqs. 3,4 (dashed), Eq. 7 (dotted), and the analytic expressions Eqs. 5,6 (dot-dashed).
longitudinal slip length $b_{\parallel}^{\text{eff}}$, in Fig. 3(b). Here we compare several different theoretical predictions: the isotropic formula Eq. 3, the new asymptotic Eq. 7, and the analytical expression Eq. 5. The results for the transverse slip length $b_{\perp}^{\text{eff}}$ are similar to those presented in Fig. 3(b), so we do not show them here. The values for the transverse component are smaller than those for the longitudinal component, indicating that the flow is anisotropic.

Finally, we present the eigenvalues of the effective slip-length tensor as a function of the slip length $b$ at gas/liquid interface. The results for large values of $b$ are shown in Fig. 4(a). The analytic formulas Eqs. 5,6 are shown in dot-dashed lines, which are accurate over a large range of the $b$ value. The effective slip lengths saturate when the slip length $b$ is much greater than the stripe periodicity $L$, and reach the asymptotic values predicted by Eq. 4.

The simulation results and several theoretical predictions for small values of $b$ are shown in Fig. 4(b). Again only the results for the longitudinal component is shown here, while the transverse component has a similar dependence. The surface-averaged slip, predicted by Eq. 3, is well above the exact values of the longitudinal effective slip. The analytical expressions Eq. 5, on the other hand, underestimates the effective slip. The newly developed formulas, Eq. 7, on the other hand, gives the correct asymptotic behaviour in the limit of very small $b/L$.

Acknowledgements

This research was supported by the DFG through SFB-TR6 and SFB 985, and by the RAS through its priority program “Assembly and Investigation of Macromolecular Structures of New Generations”. The simulations were carried out using computational resources at the John von Neumann Institute for Computing (NIC Jülich).

References


Computational Plasma Physics
Computational Plasma Physics:
Amplification of Ultra-Short Light Pulses,
Particle Acceleration, and Radiation Sources

Karl-Heinz Spatschek
Institut für Theoretische Physik, Heinrich-Heine-Universität Düsseldorf, D-40225 Düsseldorf
E-mail: khs@tp1.uni-duesseldorf.de

Laser-matter-interaction and plasma astrophysics are in the centre of the present NIC high-performance computational plasma studies. Plasma astrophysics appears in this volume under a separate chapter. The present chapter concentrates on new developments in laser-matter interaction. The two projects being presented after this introductory remarks are typical examples of projects which lead to new insights into laser-matter-interaction processes via computational physics. They represent two of the most fascinating areas of the last years: Amplification of ultra-short light pulses by collective modes in plasmas on the one side, and particle acceleration as well as radiation generation in plasma-based accelerators on the other side.

The pioneering work on compression of optical pulses by the chirped pulse amplification (CPA) technique opened a new field of laser physics with respect to ultra-short laser pulses. Very high pulse powers, however, present a practical difficulty to CPA techniques, and therefore the use of plasma as amplifying medium was proposed\(^1\). Laser pulse amplification by Raman or Brillouin scattering in plasmas was suggested before in the linear (parametric) regime. However, nonlinear effects become very important when the laser pulses are short and intense\(^2\). The nonlinear plasma processes may become very beneficial for ultra-short pulse generation. As a result, it is expected that nonlinear parametric amplification in plasmas may lead to high-intensity laser beams even up to the exawatt-zetawatt regime\(^3\).

Both, nonlinear Raman as well as Brillouin scattering are candidates for pumping energy into short seed pulses. The initial pulse shaping during Raman seed amplification enters, after an interesting “start-up period”\(^4\), an important nonlinear stage. Within the latter, known as the self-similar region, the seed amplitude grows and seed width narrows. Both tendencies are highly wanted for pumping a seed to high intensities.

Besides scattering also pulse filamentation and particle trapping may occur leading to operational limits for seed amplification. Generally, modulation, filamentation, and focusing can be caused by ponderomotive force and Joule heating. Since the temporal evolution of these processes relies on ion motion, it will not be important for very short pulses. However, a third process, namely relativistic mass variation, can act quickly for incident laser light of sufficiently high intensity and should be included into the modelling. Filamentation of the pump and/or seed pulse due to relativistic mass variation is on the focus in the first paper. From simple plane wave estimates an upper limit for the pump pulse duration and hence the amplifier length follows. The estimated time to avoid seed filamentation will be quite small, i.e of the order of ps. This would cause severe problems for the technical acceptance. Thus, it is important to find whether short pulses also suffer from filamentation. A detailed investigation of this process is the aim of the first paper which shows that configurations are possible where filamentation will not occur for the leading pulse.
The second paper makes use of ultra-short intense laser pulses, and considers plasma wakefield acceleration and subsequent radiation generation. When focused into a plasma, a short laser pulse expels electrons from its propagation path by virtue of the electromagnetic fields it carries. This creates an electron-depleted cavity behind the driver. Its restoring forces pull back the removed electrons and thereby set up collective plasma oscillations. These form a co-propagating wakefield behind the driver which can be used for particle acceleration. The important point is that the allowed electric field gradients in plasmas are by many orders of magnitude larger than the breakdown limit in conventional accelerators. Thus, acceleration of GeV electron beams becomes possible over distances of centimetres.

In the paper by Grebenyuk et al., two methods for controlled injection of electrons into a plasma-wakefield accelerator are discussed, namely magnetic self-interaction and density down-ramp injection. The beams are analysed to assess their applicability for the generation of X-ray radiation from the wakefield itself and to gauge their usefulness as free electron laser (FEL) drivers. Controlled magnetic self-injection is shown to deliver beams which radiate at well-defined frequencies based on their intrinsic betatron motion inside the transversely focusing plasma fields. Density down-ramp injection produces beams with small emittances and strongly correlated longitudinal phase space with brightness exceeding that of FEL drivers. The authors come to the conclusion that controlled electron injection has the potential for a dramatic miniaturisation of a broad range of accelerator-based photon sources.

References

Laser Pulse Amplification by Raman and Brillouin Scattering towards Multi-Petawatt Level

Götz Lehmann
Institut für Theoretische Physik I, Heinrich-Heine Universität Düsseldorf, Germany
E-mail: goetz@tp1.uni-duesseldorf.de

Current solid-state amplification technology is limited by optical damage threshold issues when it comes to generation of ultra-short laser pulses with peak power far beyond several peta-watt. Laser pulses in the exa-watt regime will give access to enhanced second radiation sources, but also to more fundamental physics, e.g. nonlinear quantum electrodynamics. To circumvent damage-threshold issues, the use of plasma as amplification medium has been proposed. A long pump laser pulse will be backscattered off a plasma oscillation into a short seed pulse. The plasma oscillation can be a high-frequency electron Langmuir wave (Raman scattering) or a low-frequency ion oscillation (Brillouin scattering). We study both amplification schemes using three-wave coupling models, derived from a fluid description of the plasma, but also by kinetic particle-in-cell (PIC) simulations. After identifying the principal amplification dynamics in one spatial dimension, we present results from a newly developed two-dimensional three-wave interaction code.

1 Introduction

Reaching the highest optical field intensities and achieving the shortest pulse duration has always been one of the frontiers of laser physics. Currently the largest lasers in the world reach intensities of $10^{23}$ W/cm$^2$ and pulse durations on the order of few fs. When exposed to such strong fields, electrons will experience a relativistic increase of mass. When laser pulses interact with plasma, the electrons will undergo a collective motion which will induce currents. These currents will act as field sources and influence the incoming light. This is the regime of relativistic optics. The currently available lasers are on the verge of entering the regime of ultra-relativistic optics, where ions become relativistic. Already today a lot of effort is dedicated to study e.g. the possibilities to probe nonlinear QED effects by photon-matter or photon-photon interaction using even larger lasers. Amongst the predicted effects are pair-creation from vacuum and vacuum birefringence. High intensity at the same time usually means short pulse duration (Mourou-Tajima conjecture), thus these very intense fields may only last for atto-seconds or even shorter. The generation of the most intense laser radiation today is based on the chirped pulse amplification (CPA) technique or on optical parametric amplification (OPCPA). CPA allows reaching maximum intensities of the order of $10^{23}$ W/cm$^2$ and pulse durations of some tens of femto-seconds. The generation of even shorter high-intensity laser pulses, only consisting of few laser cycles, requires broad-banded amplification, provided by OPCPA. Both techniques rely on solid-state optical components, which operate close to their damage thresholds. Thus, to achieve even higher intensities, the optical components have to become larger in order to spatially and temporally stretch the pulses to keep the intensity below the damage threshold. All this is related to substantially increasing experimental effort and costs. To achieve even higher laser intensities plasma-based amplification schemes are actively discussed.
Underdense plasmas are insensitive to optical damage and support a broad range of frequencies, thus allow for intense and short pulses. The Extreme Light Infrastructure (ELI)\textsuperscript{3} might become the largest laser facility build based on the conventional CPA and OPCPA techniques. The next step might be to use a combination of amplification and compression schemes based on plasma processes\textsuperscript{4}, termed C3 (for Cascaded Conversion Compression). The efforts to develop such a laser system and at the same time discuss physics that can be studied using these intensities are bundled in the IZEST (International Center for Exa- and Zetawatt Science and Technology) initiative\textsuperscript{5}.

In our project on JUROPA we run simulations to study the pulse amplification schemes based on Raman and Brillouin scattering. We make use of particle-in-cell (PIC), as well as fluid-like three-wave interaction simulations.

This manuscript is organised as follows. In Sec. 2 we give an outline of the underlying physics. Sec. 3 gives an insight into results obtained from two-dimensional simulations on JUROPA. Sec. 4 describes new parallel codes we developed on JUROPA to simulate three-wave models. The manuscript concludes with a short summary and outlook.

2 Basic Principle of Raman and Brillouin Amplification

The basic mechanism for seed pulse amplification in plasma is the following. A long pump pulse will collide head-on with a counter-propagating short seed pulse in underdense plasma. Both pulses have in general different frequencies, $\omega_{\text{pump}}$ and $\omega_{\text{seed}}$, respectively. Once the pulses overlap, they will produce a beat-signal at $\omega_p = \omega_{\text{pump}} - \omega_{\text{seed}}$. When the beat-frequency is equal to the frequency of a plasma oscillation, resonant backscattering of the pump into the seed will occur, i.e. the seed pulse will be amplified at the cost of the pump. The plasma mode can either be a high-frequency electron Langmuir wave (stimulated Raman scattering, SRS) or a low-frequency ion wave (stimulated Brillouin scattering, SBS).

In order to generate a resonant beat-signal, the two laser pulses will have to be detuned from each other. The amount of detuning is different for both proposed amplification processes. For Raman amplification, the two pulses have to be detuned by the electron plasma frequency $\omega_{\text{pe}} = (4\pi n e^2/m_e)^{1/2}$. Typical plasma densities $n$ are on the order of $1/1000n_c - 1/100n_c$, where $n_c = \omega_{\text{pe}}^2 m_e/4\pi e^2$ is the critical plasma density. For a 800 nm pump pulse the seed pulse will thus have to be detuned by up to 100 nm. Efficiently shifting the central wave-length by this amount is experimentally challenging. On the other hand, since the plasma frequency is large, the wave-length of the plasma wave is short. Seed pulses with durations shorter than several pico-seconds would not efficiently interact with ion acoustic waves. However, for sufficiently large pump amplitudes (around $10^{14}$ W/cm$^2$ for plasma densities on the order of 0.01$n_c$), we enter the regime of strongly coupled (sc) Brillouin scattering. In this regime the ion mode becomes a pump-driven quasi-mode and is no longer a plasma eigenmode. The frequency $\omega_{\text{sc}}$ is much larger than $\omega_{\text{IAW}}$. Still, the ion frequency is not comparable to the electron plasma frequency, but sufficient
for laser pulses with durations on the order of 50 fs to interact with the ion wave. For Brillouin amplification we may have to start with longer seed pulses compared to Raman amplification, but since $\omega_{sc} \ll \omega_{pe}$, the required detuning between pump and seed pulse is significantly smaller. Experimentally no detuning might be required at all, since the natural band-width of the seed pulse is much larger than the detuning.

All the above is motivated from monochromatic wave theory for the linear regime, where the seed pulse intensity is assumed to be smaller than the pump pulse intensity. Once the seed has grown to the same amplitude as the pump we enter the nonlinear regime. It turns out that the flow of energy from pump to seed is maintained in the nonlinear regime and the seed can become completely depleted by the pump. The seed pulse will undergo a self-similar evolution in this phase. While the maximum amplitude is growing, the pulse also compresses in longitudinal direction.

One way to describe the interaction between pump, seed and plasma wave mathematically is to derive reduced envelope models. These models consist of three nonlinearly coupled partial differential equations. In the case of Raman scattering the system is of the form

$$\left(\frac{\partial}{\partial t} + v_1 \frac{\partial}{\partial x} - i \kappa_1^2 \nabla^2 \right) E_p = -N^* E_s + i \left( \chi_{11}|E_p|^2 + \chi_{12}|E_s|^2 \right) E_p,$$

$$\left(\frac{\partial}{\partial t} + v_2 \frac{\partial}{\partial x} - i \kappa_2^2 \nabla^2 \right) E_s = N E_p + i \left( \chi_{21}|E_p|^2 + \chi_{22}|E_s|^2 \right) E_s,$$

$$\frac{\partial N}{\partial t} = E_s E_p^*.$$

Here $E_p, E_s$ and $N$ are the complex envelopes of the electric fields of pump, seed and the plasma density, respectively. Although the system of Eqs. 1-3 does not contain kinetic effects, like wave-breaking or Landau damping, it allows a good insight into the nonlinear phase of the amplification. The according system for Brillouin amplification is similar to Eqs. 1-3, but contains a second order equation for the plasma density, which is due to the fact that the sc-Brillouin plasma wave is a driven mode and no plasma eigenmode.

By reducing the system of Eqs. 1-3 to a (spatially) one-dimensional model, one can find that the seed evolution during nonlinear Raman amplification can become self-similar. In the case of initially very short, $\delta$-like seed pulses the self-similar solutions are closely related to the $\pi$-pulse solutions of the sine-Gordon equation. Analytic theory predicts a linear growth of the maximum seed amplitude with amplification distance. This behaviour can be verified in simulations. Initially longer seed pulses however will first undergo a re-shaping of their envelope in the linear amplification phase. This then leads to reduced growth in the nonlinear phase. Using simulations of the three-wave model and validating the results with simulations from PIC and Vlasov codes, we could demonstrate this behaviour.

The nonlinear phase of Brillouin amplification also shows a self-similar behaviour. Again an analytic scaling can be derived from the according three-wave model, showing that the maximum seed amplitude scales as $t^{3/4}$, where $t$ is the amplification time. Simulations of the three-wave model, however, showed, that again only initially very short seed pulses will evolve according to the analytic scaling. For longer seed pulses the growth in the nonlinear phase is reduced to $t^{3/4}$, where $\delta < 3/4$ depends on the initial seed duration. The change in the exponent can be interpreted as a varying seeding power parameter in the family of self-similar solutions.
3 **Amplification in Two-Dimensional Geometry**

To achieve a large uptake of energy by the seed and still remain at non-relativistic field amplitudes large beam diameters are needed. For pump intensities of $10^{15}$ W/cm$^2$ beam diameters of the order of a few hundred micron are possible using a TW class pump laser. A 10 ps pump pulse will carry about 1 J of energy. For such wide and long laser beams transversal instabilities may play an important role. The growth-rate for transversal filamentation scales with the square of the magnitude of the electric field of the pulse. For a pump pulse of intensity $10^{15}$ W/cm$^2$ filamentation will occur on time-scales of several hundreds of ps. The seed pulse, while initially smaller in amplitude than the pump, can be amplified to such intensities that filamentation may occur already within a few ps$^{11,12}$.

Not only instabilities will affect the multi-dimensional amplification characteristics. Due to the transversal profile of the pump beam, the centre of the seed pulse will experience more amplification than the wings towards the outside. This influence of the pump profile on the transversal seed structure is important in the context of subsequent focusing of the amplified seed pulse. Focusing will be done by a plasma mirror and the design of the mirror may depend on the final structure of the seed pulse.

To investigate the general characteristics of Raman amplification we simulate pulse amplification in two-dimensional (2D) geometry. We make use of the Epoch PIC code, developed by T. Arber and colleagues at the University of Warwick, and our own three-wave model code which implements the system of Eqs. 1-3 in 2D, see Sec. 4. While the PIC code presents the most consistent way to simulate the interaction, it is also the most computationally demanding, limiting the simulation time to a few ps. The simulation of the three-wave model consumes much less time, but may not contain all relevant effects. At low plasma densities and high pump intensities wave-breaking will influence the amplification efficiency. Hence we may use the three-wave code to simulate scenarios in which wave-breaking will not play a dominant role. To study the influence of wave-breaking we have to run additional PIC simulations.

![Figure 1](image-url)

**Figure 1.** Evolution of a 50 fs seed pulse with initial intensity of $10^{15}$ W/cm$^2$ (corresponding to $E_p = 0.03$ in dimensionless units) during Raman amplification with a $10^{16}$ W/cm$^2$ pump pulse with FWHM of 375 µm. The seed pulse is shown at times $t = 0, 2, 4, 8$ and 9 ps.
Fig. 1 shows results from a simulation of the three-wave model Eqs. 1-3. A 50 fs seed pulse with initial intensity of $10^{15}$ W/cm$^2$ is pumped by a pulse with the same intensity for 9 ps. The plasma density is $n = 0.01 n_c$ and the pump has a wave-length $\lambda_0 = 800$ nm and the seed propagates from right to left. The pump has a Gaussian profile in transversal direction with FWHM of 375 $\mu$m. The figure shows results at times $t = 0, 2, 4, 6, 8$ and 9 ps. At $t = 0$ the dimensionless amplitude of the seed is 0.03, which corresponds to $10^{15}$ W/cm$^2$. After 9 ps of interaction we find that the seed pulse amplitude has increased by a factor of more than 20, while the pulse still is short. The pump pulse becomes almost completely depleted by the seed.

The detailed structure of the seed pulse after 7 ps of amplification is shown in Fig. 2. Let us first consider the left frame of Fig. 2. The pump pulse (not shown) has a Gaussian transversal profile $E_p \sim \exp(-y^2/\sigma^2)$. The central part of the seed is amplified the most and runs ahead of the wings, i.e. the intensity front becomes curved. We also observe a modulation of the seed amplitude in propagation direction. The main part of the seed energy is located in the first two oscillations of the field envelope. To verify that the curvature of the intensity front is due to the transversal pump profile, we repeat the simulation with a super-Gaussian pump profile $E_p \sim \exp(-y^{20}/\sigma^{20})$. The results are shown in the right frame of Fig. 2. The intensity front of the pulse stays flat and all parts are amplified by the same amount. Thus, even though the Rayleigh-lengths for the laser pulses are very large compared to the amplification distance, we expect the seed pulse to have a curved intensity front in a real experiment, which will use a Gaussian transversal pump profile.

4 Code Performance on JUROPA

As mentioned in Sec. 3 we use two different kinds of codes in our project. For PIC simulations we use the Epoch code, developed by T. Arber and his group at the University
of Warwick. The code is an open source project and shows good scaling behaviour on JUROPA up to a few thousand cores\textsuperscript{13}. The second approach we take is to simulate the three-wave model Eqs. 1-3 and its analogon for the Brillouin case. These systems are the basis on which the amplification process itself and additional effects like transversal filamentation, weakly-relativistic self- and cross-interaction, dispersion and defocusing can be studied analytically as well as numerically.

Typically we need to simulate interaction times on the order of several ps and distances of a few mm with transversal dimensions of one mm. To study the full process using PIC simulations is too demanding. To simulate amplification over more than one mm, we have to propagate the pump beam over several ps. On this time-scale noise from the PIC method starts seeding instabilities, which then grow and invalidate the results. PIC simulations of shorter intervals help us to understand the principal influence of e.g. wave-breaking or Landau damping. To keep a low noise level, we have to use a high grid resolution and many particles. For simulations of about 1 mm interaction length we commonly use 40k $\times$ 5k grid points and a total of about $10^{10}$ macro-particles. We find that the Epoch code scales up to around 4k cores for our problem. Typical run-time for one simulation is 24h, i.e. we consume up to 100k core-h for one simulation. A co-moving frame including radiating boundary conditions for pump injection is not yet implemented in the Epoch code, but could help to reduce the amount of time required.

Since PIC simulations consume too much computer time to perform many simulations, we developed new codes which simulate the three-wave interaction models for Raman and Brillouin amplification. The codes we developed simulate Eqs. 1-3 (or the according, similar system for Brillouin amplification) using finite-differences. In spatial direction we use second-order finite differences, for temporal integration we find that Crank-Nicholson and high-order Runge-Kutta schemes give good results.

Our implementations make use of the PETSc library\textsuperscript{14}. The library provides a flexible way to implement our problem without having to decide about e.g. the temporal integration scheme or the root-finding method before runtime. This allows to easily check different combinations of pre-conditioning methods, sparse matrix solvers and temporal integration schemes. All parallel communication is handled by the library internally. For data storage we use the HDF5 file format, which is directly supported by PETSc.

Our simulations typically require a resolution of 2500 $\times$ 2000 grid points. We carry out calculations in the frame of reference co-moving with the seed pulse. In this way we can simulate interactions for several tens of ps with a reasonable number of grid points. Fig. 3 shows the hard-scaling performance for our Raman code in the case of a grid resolution of 2500 $\times$ 2000. Up to 1024 cores we find perfect scaling behaviour, compared to the performance on 128 cores. The code would scale further for larger grid dimensions, but these are currently not required.

A typical simulation with this three-wave model code for one parameter set requires about 1h on 1024 cores of JUROPA, i.e. about 1k core-h. Our PIC simulations, which cover a shorter interaction time and distance require about 50k-100k core-h per run. Using the three-wave model code we can easily run many parameters and study all effects in detail. However, to confirm the results from the reduced model we also have to run PIC simulations to compare to.
5 Summary and Outlook

We have studied Raman and Brillouin backward scattering in the context of short laser pulse amplification. Both processes have the potential to become a central technology in the next generation of laser pulse amplifiers. Using JUROPA we are able to study the multidimensional characteristics of these processes, where many questions concerning the interplay of transversal dynamics and amplification are still open. Eventhough we developed an efficient scheme to simulate the basic phenomena using reduced three-wave coupling models we still require a large parallel machine to perform parameter studies. Future plans are to port the codes to full three-dimensional geometry and to study the mixing of Raman and Brillouin modes predicted and observed in simulations very recently\textsuperscript{15,16}.

Acknowledgements

The work was done under the auspices of SFB-TR18 of the Deutsche Forschungsgemeinschaft. The simulations were performed with computer resources on JUROPA granted by the NIC at Forschungszentrum Jülich under project HDD10.

References


Radiation Generation in Plasma-Based Accelerators with Controlled Electron Injection

J. Grebenyuk1, J. Vieira2, T. Mehrling1, J. L. Martins2, A. Martinez de la Ossa1, R. A. Fonseca2, L. O. Silva2, J. Osterhoff1

1 Deutsches Elektronen Synchron DESY, Notkestr. 85, 22607 Hamburg, Germany
E-mail: jens.osterhoff@desy.de

2 Instituto Superior Técnico, IST, Avenida Rovisco Pais 1, 1049-001 Lisbon, Portugal
E-mail: jorge.vieira@ist.utl.pt

We present three-dimensional particle-in-cell simulations of plasma wakefield acceleration with controlled injection of electrons using the fully relativistic, massively parallel code OSIRIS1. Two methods for particle injection are considered: magnetic-field induced injection and injection from a gradient in plasma density. Our studies show that both techniques may provide electron beams with the potential to drive compact radiation sources. Magnetic-field injection allows for controlled off-axis beam placement. Consequently, the beam undergoes betatron oscillations and emits radiation at well defined frequencies. Density down-ramp injection enables the creation of beams with small transverse emittances and large currents. These properties correspond to brightnesses up to one order of magnitude higher than provided by the free-electron-laser-quality driver for the plasma-acceleration process. Therefore, this scheme may act as an effective beam-brightness up-converter.

1 Introduction

Plasma wakefield acceleration is a fast-developing novel acceleration technology which allows to sustain extreme electric-field gradients exceeding 10 GV/m, many orders of magnitude above the breakdown limit in conventional accelerators. This enables the acceleration of GeV electron beams on distances of centimetres2–4. The accelerating structure in plasma may be driven by an ultra-short intense laser pulse (laser-plasma wakefield acceleration or LWFA5) or by a high-current-density relativistic particle beam (beam-driven plasma acceleration or PWFA6,7). When focused into a plasma, a laser pulse or a particle beam expels electrons from its propagation path by virtue of the electromagnetic fields it carries. This creates an electron-depleted cavity behind the driver. Its restoring forces pull back the removed electrons and thereby set up collective plasma oscillations. These form a co-propagating wakefield behind the driver, which features the strong longitudinal electrical fields mentioned before. Those fields can be used for the acceleration of charged particles when inserted into the wakefield structure at a properly chosen phase. Different techniques for injection exist such as wave breaking8,9, density transitions10–12, magnetic fields13,14, ionisation of a dopant atomic species15–18 or utilising colliding laser pulses19–21. Over the last decade, the quality of the beams generated in plasma-based accelerators has been significantly improved22,23, but still requires further advancement in quality, tunability, and reproducibility to make them viable alternatives to beams from conventional accelerators.

One of the most promising applications of plasma accelerators constitutes radiation generation. Electron beams from plasma-based sources have already shown to be able to generate XUV photons from undulators24 and hard X-rays from the plasma itself25,26. A
central enterprise of the plasma-acceleration community is to realise the creation of coherent radiation from undulators in a free-electron-laser (FEL) process, in which electrons oscillate in periodic magnetic structures, self-bunch, and coherently emit narrow-band radiation in a selectable part of the electromagnetic spectrum ranging from the visible to hard X-rays. Another possibility is to utilise the intrinsic transverse beam motion in a plasma wakefield as a radiation source. The wake exhibits a focusing potential which leads to transverse oscillations of the trapped electrons, called betatron motion, causing the emission of radiation typically in the X-ray domain. The oscillations usually occur in the wiggler regime enabling broad photon spectra. A third option for radiation generation is Thomson scattering. In this process a laser beam is scattered off the produced electron beam and Doppler-shifted into another frequency range.

For all of these applications to work in a reliable and tunable fashion, it is of paramount importance to obtain control over the parameters of the accelerated electron beam. Therefore, the phase-space population during the injection process needs to be controlled. This allows to obtain beams with tailored parameters in energy spread, charge, current, and emittance. In addition, precise manipulation of the initial placement of the beam in transverse and longitudinal direction controls the betatron emission process during acceleration.

The before-mentioned injection techniques offer different levels of control of which we want to explore two in the paper at hand: magnetic-field induced injection, and density down-ramp injection. These methods are examined by means of three-dimensional (3D) particle-in-cell (PIC) simulations, which were carried out using the fully relativistic, massively parallel code OSIRIS. In addition we will discuss the applicability of the generated beams for driving compact radiation sources.

2 Controlled Magnetic Self-Injection and Radiation Generation

The mechanism of controlled magnetic self-injection relies on static magnetic fields oriented perpendicularly to the direction of propagation of a laser or particle bunch driver. This scheme can strongly relax trapping thresholds, controlling the transverse direction and angular region where injection occurs. By controlling transverse self-trapping, this scheme may also be used to control/enhance betatron radiation in plasma based acceleration. We performed 3D PIC simulations of magnetised laser wakefield acceleration. The simulation used a moving window that moves at the speed of light, with dimensions $24 \times 24 \times 12$ ($c/\omega_p$)$^3$ ($\omega_p$ is the plasma frequency), divided into 480 $\times$ 480 $\times$ 1200 cells with 2 electrons per cell. A linearly polarised laser pulse with central frequency, $\omega_0$, given by $\omega_0/\omega_p = 20$, was used with normalised peak vector potential $a_0 = qA_0/(me_c) = 3$ (where $m_e$ and $e$ are the electron mass and charge respectively), a duration $\tau_{\text{FWHM}} = 2\sqrt{\omega_0/\omega_p}$, and transverse spot size $W_0 = c\tau_{\text{FWHM}}$. A parabolic plasma density profile is also used to ensure that the laser is guided without diffracting during propagation. A static external B-field pointing in the positive y-direction was considered, rising with $B_{y}^{\text{ext}} = \omega_c/\omega_p = 0.6\sin^2[\pi z/(20c/\omega_p)]$, constant and equal to $B_{y}^{\text{ext}} = 0.6$ for $L_{\text{flat}} = 40 c/\omega_p$, and dropping back to zero with $B_{y}^{\text{ext}} = 0.6\sin^2[\pi z/(20c/\omega_p)]$. Note that phases inside the $\sin^2$ term are such that the B-field profile is continuous.

Main simulation results are shown in Fig. 1 which illustrates key mechanisms of injection assisted by external B-fields in LWFAs. Fig. 1 shows the electron plasma density in frames (a), (c), and (e), superimposed by the laser pulse in orange. After the plasma bubble
Figure 1. 3D OSIRIS simulation results illustrating the magnetic self-injection mechanisms. (a), (c), and (e) show the electron plasma density in grey, the self-trapped particles in blue, and the laser pulse envelope in red colours at $t = 110\omega_p$, $t = 126\omega_p$, and $t = 159\omega_p$. (b), (d), and (f) show the corresponding $p_l - \xi$ phase-space. The magnetic field leads to off-axis self-injection in a narrow angular region. The inset in (f) represents the transverse momentum phase-space of the self-injected electron bunch residing within the bubble. The B-field profile is schematically represented on the top of the figure. The laser driver moves from left to right as indicated by the arrow.

is formed, electrons start to be injected in the uniform regions of the external magnetic field where $x > 0$. The magnetic field causes plasma electrons to rotate anti-clockwise. For $x > 0$, electrons reach the axis with larger longitudinal momentum $p_l$. This relaxes the conditions for trapping, which can only occur when the longitudinal electron velocity $v_p$ approaches the phase velocity of the bubble $v_\phi$. For $x < 0$, electrons reach the axis with lower $p_l$, and are are lost to the surrounding plasma. A conservative estimate (neglecting wakefields) for the threshold B-field for injection can be estimated yielding $B[T] = 32\sqrt{n_0[10^{16}\text{ cm}^{-3}]/[2\sin(\theta)\sqrt{a_0}}$, where $\theta$ is the angle between the plane of the electron trajectory with the B-field. As a result, injection occurs off-axis (for $x > 0$), and in a well defined azimuthal region. In addition, $B[T] \lesssim 32\sqrt{n_0[10^{16}\text{ cm}^{-3}]}$ so that the wakefield is nearly undisturbed by the B-fields. A stronger injection burst also occurs in the B-field down-ramp regions, within similar angular range to the uniform B-field section (Fig. 1c). For $x > 0$, when the B-field lowers the plasma wavelength increases and the wake phase velocity lowers, thereby greatly relaxing trapping thresholds. For $x < 0$, $v_p > c$, and trapping is fully suppressed. The resulting phase space at $t = 126c/\omega_p$ is shown in Fig. 1d. After the magnetised plasma region, the magnetically injected electron bunch is clearly detached from the back of the plasma wave, leading to a quasi-monoenergetic electron bunch with $\sim 10\%$ energy spread. The magnetically injected electron bunch right after the B-field is shown at $t = 159 c/\omega_p$ in Fig. 1e, and the corresponding phase-space in Fig. 1f.

3D OSIRIS simulations show that magnetically injected electrons could emit betatron radiation at few (due to small wiggler parameter, close to undulator) and well defined harmonics in comparison to current experiments. To explore this possibility we performed
magnetised plasma wakefield acceleration simulations relevant for experiments, and analysed the resulting radiation spectrum using a post-processing radiation code. These simulations considered a 30 GeV electron bunch with a bi-Gaussian density profile with radius $\sigma_\perp = 0.17 \, c/\omega_p$, length $\sigma_\xi = 1.95 \, c/\omega_p$, and peak density $n_b/n_0 = 19$. Corresponding magnetically injected electron trajectories are shown in Fig. 2a. To retrieve the radiation spectrum, a random sample of the self-injected electrons was post-processed using the radiation code JRad. Fig. 2b shows the radiation spectrum in a transverse central line of a virtual detector placed at a distance $5100 \, c/\omega_p$ from the exit of the plasma. Fig. 2b reveals that radiation is emitted at well defined frequencies, which are clearer at larger angles or larger $|x|$. The white dashed lines, which correspond to the theoretical estimates for the emitted harmonics for the parameter of the magnetically injected electron bunch, are in good agreement with numerical simulations.

3 High-Brightness Beams from Density Down-Ramp Injection

Injection on a density down-ramp (DDR) allows to relax the requirements to achieve wave breaking, and thus, enables particle injection. DDR was shown to be a tunable method for delivering beams in LWFA schemes, and has been proposed for planned PWFA experiments. The DDR injection mechanism enables control over the phase-space population by tailoring the longitudinal plasma-density profile, and in this way to obtain electron bunches with the desired parameters. In case of a plasma-density profile $n(z)$ which is changing along the wake-propagation direction $z$, the plasma-oscillation wavenumber $k_p$ also depends on $z$: $k_p(z) = \sqrt{\frac{\sigma^2}{m_e n(z)}},$ where $m_e$ is the electron mass. Therefore, the wake phase has an additional $z$-dependence $\psi = k_p(z)(z - ct)$ with the
speed of light $c$ and time $t$. Thus, assuming a fixed wake phase with $d\psi/dt = 0$, the wake phase velocity $v_p$ can be calculated according to\textsuperscript{35}

$$v_p(z) = c \left(1 + \frac{\zeta}{k_p(z)} \frac{dk_p}{dz}\right). \tag{1}$$

Here, $\zeta = z - ct$ is a co-moving coordinate. The change in wave number along $z$ can be written as $\frac{dk_p}{dz} = \frac{k_p}{2\pi} \frac{dn}{dz}$. For $\frac{dn}{dz} < 0$, i.e. a negative plasma-density gradient, $v_p$ decreases and the wake slows down. This effect relaxes the condition for longitudinal wave-breaking, i.e. reduces the velocity which plasma electrons have to achieve to be trapped, and hence, enables particle injection\textsuperscript{10,36}. Simultaneously, the injected electrons experience a phase shift. The plasma wavelength $\lambda_p$ is inversely proportional to $\sqrt{n}$, causing an increase of $\lambda_p$ in the density downslope. The resulting phase shift can be expressed assuming no phase-change from acceleration during propagation down the ramp\textsuperscript{35} as $\Delta\psi = \psi_0 \sqrt{1 - \frac{n_0}{n_1}}$, with $\psi_0$ being the initial electron phase and $n_0$ and $n_1$ the initial and final plasma density of the ramp, respectively.

We performed 3D PIC OSIRIS simulations of density-ramp injection for PWFA with the driver-beam parameters corresponding to beams from the Flash free-electron laser at DESY. The driver was approximated by a Gaussian distribution in space assuming the following parameters: longitudinal rms size $\sigma_z = 8.4\ \mu m$, transverse rms size $\sigma_{x,y} = 5\ \mu m$, an integrated charge of 180 pC corresponding to a peak current of 2.5 kA, an energy of 1 GeV with a relative energy spread of 1%, and a normalised transverse emittance of $\epsilon_{n(x,y)} = 1\ \mu m$. The plasma-density profile used in the simulations is depicted in Fig. 3. An initial increase from vacuum to a density of $n_0 = 10 \times 10^{17}\ \text{cm}^{-3}$ is followed by a 100 $\mu m$ plateau of constant density and a linear decrease to density $n_1 = 1 \times 10^{17}\ \text{cm}^{-3}$ over a length of $L_{\text{ramp}} = 60\ \mu m$. Following this down-ramp for triggering injection, a plateau of constant density $n_1$ allows for particle acceleration to high energies. $n_1 = 10^{17}\ \text{cm}^{-3}$ corresponds to a plasma skin-depth of $k_p^{-1} = 16.8\ \mu m$ and wavelength of $\lambda_p = 106\ \mu m$. 

Figure 3. Schematic representation of the plasma-density profile as a function of the longitudinal coordinate, $n(z)$. An electron bunch is injected on the down-ramp transition between $z_0$ and $z_1$. Subsequently, its energy is boosted at constant density $n_1$ with $z \geq z_1$. 

431
The numerical investigation was conducted in a simulation frame with a moving window of box dimensions $20 \times 15 \times 15 \frac{k_{-1}}{}$ with a cell size of $0.04 \frac{k_{-1}}{}$. 8 particles per cell were used for both, driver and plasma.

Fig. 4 shows plasma and driver charge densities after 5.4 mm of propagation distance. An electron bunch is injected into the first wake period owing to the integrated density down-ramp, and is further accelerated at constant density with field gradients of 10 GV/m until depletion of the driver beam occurs. After 14.5 cm of propagation through plasma the witness bunch has gained 1.5 GeV. Its projected relative energy spread is 11% originating from the phase spread at injection time, the rms sizes are $\sigma_z = 5.1 \mu m$, $\sigma_x = 0.5 \mu m$, $\sigma_y = 0.45 \mu m$ with a total charge of 32 pC. The projected normalised transverse emittances are $\epsilon_x = 0.58 \mu m$, $\epsilon_y = 0.40 \mu m$. An analysis of the process shows that normalised emittance is small initially after trapping, $\epsilon_{x,y} = 0.06 \mu m$, but then grows rapidly caused by betatron-phase mixing until it saturates\(^3\).

The beam energy spectrum, longitudinal phase-space properties, and sliced emittance and current are depicted in Fig. 5. The longitudinal phase space $[z,p_z]$ is strongly correlated owing a strongly varying accelerating field strength along the length of the beam. Therefore, the uncorrelated energy spread is below 0.5% while the absolute energy spread exceeds 10%. Moreover, the sliced normalised transverse emittance is found to not exceed $0.2 \mu m$ with a sliced current exhibiting an almost flat-top profile of 0.7 kA. From these values the beam brightness exceeds $0.35 \times 10^{17}$ A m$^{-2}$rad$^{-2}$ and is up to one order of magnitude higher than in modern FEL beams\(^3\). Since FEL performance is dependent on beam brightness, controlled injection on a density ramp may provide beams with the potential to drive the next-generation compact, high-brightness FEL source.
4 Conclusions

In this paper, two methods for controlled injection of electrons into a plasma-wakefield accelerator are presented and investigated by means of 3D PIC simulations. The resulting beams are analysed to assess their applicability for the generation of X-ray radiation from the wakefield itself and to gauge their usefulness as FEL drivers. Controlled magnetic self-injection is shown to deliver beams which radiate at well-defined frequencies based on their intrinsic betatron motion inside the transversely focusing plasma fields. Density down-ramp injection produces beams with small emittances and strongly correlated longitudinal phase space with brightness exceeding that of FEL drivers. In summary, such beams show the potential to allow for a dramatic miniaturisation of a broad range of accelerator-based photon sources.

5 Acknowledgements

We acknowledge the grant of computing time by the Jülich Supercomputing Centre on JUQUEEN under Project No. HHH09. We would like to thank DESY IT for their support concerning data storage at DESY. We acknowledge Alexander von Humboldt Foundation for financial support.

References

NIC Series Volume 32
**NIC Symposium 2006 - Proceedings**
1 - 2 March 2006, Jülich, Germany
edited by G. Münster, D. Wolf, M. Kremer (2006), iv, 365 pages
ISBN: 3-00-017351-X

NIC Series Volume 33
**Parallel Computing: Current & Future Issues of High-End Computing**
Proceedings of the International Conference ParCo 2005
ISBN: 3-00-017352-8

NIC Series Volume 34
**From Computational Biophysics to Systems Biology 2006**
Proceedings

NIC Series Volume 35
**Dreistufig parallele Software zur Parameteroptimierung von Support-Vektor-Maschinen mit kostensensitiven Gütemaßen**
by T. Eitrich (2007), xiv, 231 pages
ISBN: 978-3-9810843-1-3

NIC Series Volume 36
**From Computational Biophysics to Systems Biology (CBSB07)**
Proceedings
ISBN: 978-3-9810843-2-0

NIC Series Volume 37
ParCo 2007 Conference, 4 - 7 September 2007
ISBN: 978-3-9810843-3-7

NIC Series Volume 38
ISBN: 978-3-9810843-4-4
NIC Series Volume 39
**NIC Symposium 2008 - Proceedings**
20 - 21 February 2008, Jülich, Germany
edited by G. Münster, D. Wolf, M. Kremer (2008), iv, 358 pages
ISBN: 978-3-9810843-5-1

NIC Series Volume 40
**From Computational Biophysics to Systems Biology (CBSB08) Proceedings**
ISBN: 978-3-9810843-6-8

NIC Series Volume 41
**Multigrid methods for structured grids and their application in particle simulation**
by M. Bolten (2008), viii, 132 pages
ISBN: 978-3-9810843-7-5

NIC Series Volume 42
**Multiscale Simulation Methods in Molecular Sciences - Lecture Notes**
Winter School, 2 - 6 March 2009, Forschungszentrum Jülich
edited by J. Grotendorst, N. Attig, S. Blügel, D. Marx (2009), vi, 576 pages
ISBN: 978-3-9810843-8-2

NIC Series Volume 43
**Towards the Confirmation of QCD on the Lattice**
Improved Actions and Algorithms
by S. F. Krieg (2008), vi, 89 pages
ISBN: 978-3-9810843-9-9

NIC Series Volume 44
**NIC Symposium 2010 – Proceedings**
24 - 25 February 2010 | Jülich, Germany
edited by G. Münster, D. Wolf, M. Kremer (2012), v, 395 pages
ISBN: 978-3-89336-757-3

NIC Series Volume 45
**NIC Symposium 2012 – Proceedings**
25 Years HLRZ / NIC
7 - 8 February 2012 | Jülich, Germany
edited by K. Binder, G. Münster, M. Kremer (2012), v, 400 pages
ISBN: 978-3-89336-758-0
NIC Series Volume 46
Hybrid Particle-Continuum Methods in Computational Materials Physics – Proceedings
4 - 7 March 2013 | Jülich, Germany
ISBN: 978-3-89336-849-5

NIC Series Volume 47
NIC Symposium 2014 – Proceedings
12 - 13 February 2014 | Jülich, Germany
edited by K. Binder, G. Münster, M. Kremer (2014), vi, 434 pages
ISBN: 978-3-89336-933-1
The John von Neumann Institute for Computing (NIC) was established in 1998 by Forschungszentrum Jülich and Deutsches Elektronen-Synchrotron DESY to support the supercomputer-oriented simulation sciences. In 2006, GSI Helmholtzzentrum für Schwerionenforschung joined NIC as a contract partner.

The core task of NIC is the peer-reviewed allocation of supercomputing resources to computational science projects in Germany and Europe. The NIC partners also support supercomputer-aided research in science and engineering through a three-way strategy:

- Provision of supercomputing resources for projects in science, research, and industry.
- Supercomputer-oriented research and development by research groups in selected fields of physics and natural sciences.
- Education and training in all areas of supercomputing by symposia, workshops, summer schools, seminars, courses, and guest programmes for scientists and students.

The NIC Symposium is held biennially to give an overview on activities and results obtained by the NIC projects in the last two years. The contributions for this seventh NIC Symposium are from projects that have been supported by the supercomputers JUROPA, JUGENE, and JUQUEEN in Jülich. They cover selected topics in the fields of Astrophysics, Biophysics, Chemistry, Elementary Particle Physics, Materials Science, Condensed Matter, Computational Soft Matter Science, Earth and Environmental Research, Computer Science, Fluid Mechanics, and Plasma Physics.