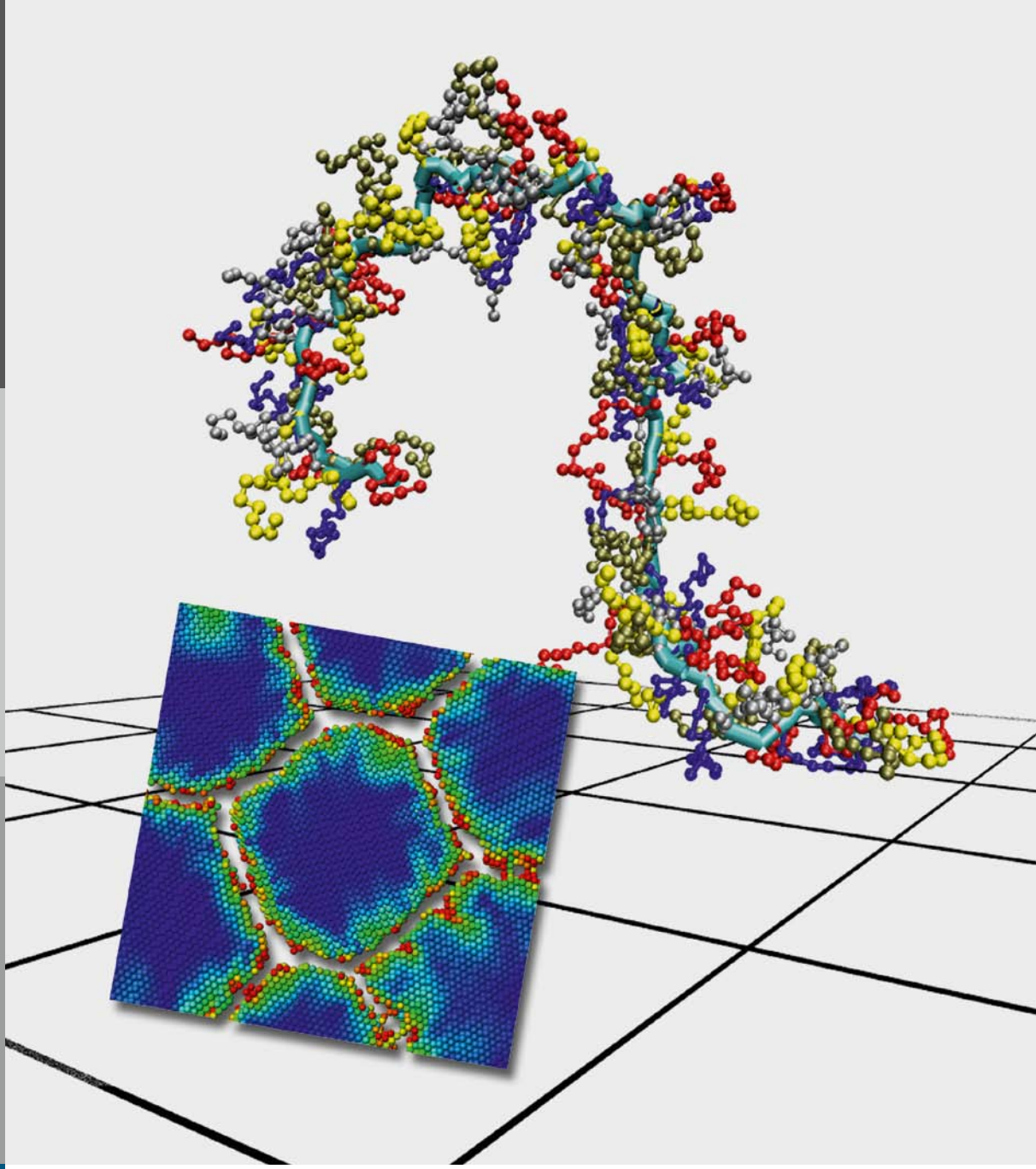


John von Neumann Institute for Computing
NIC Symposium 2010

Proceedings, 24 - 25 February 2010 | Jülich, Germany

G. Münster, D. Wolf, M. Kremer (Editors)

NIC

JÜLICH
FORSCHUNGSZENTRUM

Mitglied der Helmholtz-Gemeinschaft

Schriften des Forschungszentrums Jülich
IAS Series Volume 3

Forschungszentrum Jülich GmbH
Jülich Supercomputing Centre (JSC)
John von Neumann Institute for Computing (NIC)

NIC Symposium 2010

Proceedings, 24 – 25 February 2010 I Jülich, Germany

edited by G. Münster, D. Wolf, M. Kremer

Schriften des Forschungszentrums Jülich
IAS Series Volume 3

ISSN 1868-8489 ISBN 978-3-89336-606-4

Bibliographic information published by the Deutsche Nationalbibliothek.
Die 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 Distributor:	Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag D-52425 Jülich phone: +49 2461 61-5368 · fax: +49 2461 61-6103 e-mail: zb-publikation@fz-juelich.de Internet: http://www.fz-juelich.de/zb
Cover Design:	Grafische Medien, Forschungszentrum Jülich GmbH
Printer:	Grafische Medien, Forschungszentrum Jülich GmbH
Copyright:	Forschungszentrum Jülich 2010

We thank K. Albe (Darmstadt) and H.-P. Hsu (Mainz) for the pictures used in the cover design.

Schriften des Forschungszentrums Jülich
IAS Series Volume 3

ISSN 1868-8489
ISBN 978-3-89336-606-4

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

Persistent Identifier: urn.nbn:de:0001-2010020108
Resolving URL: http://www.persistent-identifier.de/?link=610

Preface

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

Dietrich Wolf
Fakultät für Physik, Universität Duisburg-Essen, 47048 Duisburg, Germany
E-mail: dietrich.wolf@uni-due.de

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

Over the last 10 years the proceedings of the biennial NIC-Symposia have given a fascinating account of supercomputer based research at its best. At the John von Neumann Institute for Computing (NIC) computing time has always been granted according to peer reviews of the project proposals, provided by an internationally composed peer review board. A second peer review of the project *results* obtained in a two years period leads to the selection presented in the proceedings.

This well tried procedure has been adopted by the Gauss Centre for Supercomputing (GCS), an alliance of the three German national supercomputing centres in Jülich, Garching and Stuttgart. By coordinating their interests, the GCS was able to acquire the first European computer with a peak performance exceeding 1 PetaFlop/s. It was installed by the Jülich Supercomputer Centre JSC in 2009 and dubbed JUGENE (an allusion to Jülich IBM Blue Gene/P).

As a consequence, the NIC peer review board at its fall meeting in 2009 was expanded by delegates from the GCS-institutions, and decided about computing time allocation on the GCS-computer JUGENE as well as on the NIC-computer JUROPA (replacement for JUMP). These machines hold rank 4 respectively rank 13 on the Top 500 list of the world's most powerful computers. The expanded peer review board selected projects with particular demands of compute power as *Gauss projects*. This large-scale project status has been given to projects by Jansen and Peters, presented in this volume, and on the recent meeting to Fodor, Schierholz, Katz and Harting.

The formation of an internationally competitive European supercomputing infrastructure is the goal of the European consortium PRACE (Partnership for Advanced Computing in Europe), the German member of which is the GCS. In this context new modes of computing time allocation on the European level are at the horizon. Hopefully the peer review procedure of the NIC will serve as a model there, too. Only the quality of a research proposal should matter, and the research communities should compete freely among each other with their demands for computing power rather than being regulated by quota.

The users of the NIC and GCS computing facilities in Jülich and Zeuthen come from universities and research institutions mostly in Germany, but to a growing extent also from other European countries. The diversity of research subjects is enormous, as reflected in the contents of these proceedings. This requires a great deal of flexibility and open mindedness from the NIC peer review board. External expertise is called in for judgement whenever needed. If a new user community keeps growing that had so far used the NIC facilities less frequently, it will be represented by a new member in the peer review board.

Projects that stand out as particularly excellent are nominated "John von Neumann Excellence Project" by the review board. In 2008 this award was given to Dominik Marx for his project on prebiotic chemistry. In 2009 it was given to a project about the dynamics of galaxy clusters by Stefan Gottlöber and to Zoltan Fodor's ab initio calculation of hadron masses. Fodor's project attracted particular interest, being appreciated by Nobel prize winner Wilczek in *Nature* and selected by *Science* as one of the 10 most important discoveries in 2008.

The NIC is a cooperation between FZJ (Forschungszentrum Jülich), DESY (Zeuthen) and GSI (Darmstadt), all three members of the Helmholtz Association of German Research Centers. We like to mention here two assets of this cooperation. The first one, QPACE, is an initiative together with the universities of Regensburg and Wuppertal in Germany, and Ferrara and Milano in Italy. Its ambitious aim is to develop the most energy efficient computer in the world, foremost for lattice QCD-calculations.

The second asset of the cooperation of the NIC partner institutions is that each of them funds a research group with extensive guest programs. The idea is to appoint scientists from universities for a limited time as directors of the groups, a concept that has proved to be very successful. Presently a research group on statistical physics and many-particle physics in Jülich is led by Peter Nielaba, a group on elementary particle theory at DESY in Zeuthen by Rainer Sommer, and GSI is establishing a research group on hadron physics and QCD guided by Frithjof Karsch.

We thank the referees for selecting the contributions for this book. Most of them are also members of the NIC peer review board. We also thank the authors for representing the NIC with their contributions. Special thanks go to the Jülich Supercomputing Centre JSC and to DESY for providing the high quality computing environments and user service, which were vital to the successful realisation of the projects presented in this book. Thanks go also to Ms. Martina Kamps, who compiled the contributions and produced the present high quality book, and to Ms. Sabine Höfler-Thierfeldt, who took care of the communication with the printing services at Forschungszentrum Jülich. Already here we want to thank Walter Nadler and Ms. Bettina Scheid for their help in organizing the 5th NIC-Symposium.

Jülich, February 2010

G. Münster D. Wolf M. Kremer

Contents

Entering the Petaflop-Era – New Developments in Supercomputing
N. Attig, F. Berberich, U. Detert, N. Eicker, Th. Eickermann, P. Gibbon, W. Gürich, W. Homberg, A. Illich, S. Rinke, M. Stephan, K. Wolkersdorfer, Th. Lippert **1**

The NIC Research Groups

Quark Mass Dependence of the QCD Equation of State
F. Karsch, W. Söldner **13**

Numerical Studies of Structures and Phases in (Nano-) Systems in Reduced Geometry
D. Mutter, Ch. Schieback, J. Neder, F. Bürzle, K. Franzrahe, A. Geng, P. Nielaba **21**

Towards Precision B-Physics from Non-Perturbative Heavy Quark Effective Theory
ALPHA Collaboration: M. Della Morte, J. Heitger, R. Sommer **31**

Astrophysics

Introduction by P. L. Biermann **41**

Clusters of Galaxies in the MareNostrum Universe
St. Gottlöber, G. Yepes, J. Forero-Romero, V. Turchaninov **45**

Unravelling the Mechanism how Stars Explode
A. Marek, B. Müller, H.-Th. Janka **53**

Protoplanetary Discs in Young Dense Clusters
S. Pfalzner, Ch. Olczak, Th. Kaczmarek, M. Steinhausen **61**

Computational Biology and Biophysics

Introduction by U. H. E. Hansmann **69**

Structure and Dynamics of Large Biological Molecules: ATP-Binding Cassette (ABC) Transporters
J.-H. Lin, J. Akola, R. O. Jones **71**

The Inhibition of Hydrogenase by Dioxygen
G. La Penna **79**

Protein Folding and Structure Prediction at the Simulation Laboratory Biology
J. H. Meinke, S. Mohanty, O. Zimmermann **87**

Chemistry
Introduction by W. Klopper **95**

Ionic Liquids
St. Zahn, B. Kirchner **97**

Atomistic Simulations on Polymeric Nitrogen
J. Kotakoski **105**

Prebiotic Peptide Synthesis on Blue Gene Platforms at "Iron-Sulfur-World" Conditions
N. N. Nair, E. Schreiner, D. Marx **111**

Calculation of ^{31}P and ^{183}W NMR Chemical Shifts and Nuclear Spin Spin Coupling Constants in Phosphinidenoid and Other Transition Metal Complexes by the ZORA DFT Method
G. von Frantzius, R. Streubel **119**

Elementary Particle Physics
Introduction by R. Kenway **129**

Dynamical Overlap Fermions with Small Pion Masses and Topology
N. Cundy, I. Koutsou, Th. Lippert, A. Schäfer **131**

Lattice QCD with Fat Clover Fermions: *Ab Initio* Determination of Flavour Non-Singlet Mesons and Baryons
Z. Fodor **139**

QCD Simulations with Light, Strange and Charm Dynamical Quark Flavours
K. Jansen **147**

Tuning Lattice QCD to Petascale on Blue Gene/P
St. Krieg, Th. Lippert **155**

Materials Science
Introduction by R. O. Jones **165**

Deformation Processes in Nanostructured Metals and Alloys
K. Albe, A. Stukowski, J. Schäfer, Y. Ritter, D. Farkas **167**

Anisotropic Magnetic Interactions in Low Dimensions
G. Bihlmayer, M. Heide, Y. Mokrousov, A. Thiess, St. Heinze, St. Blügel **175**

Pattern Formation: From the Macro- to the Nanoscale
E. A. Brener, M. Fleck, C. Hüter, H. Müller-Krumbhaar, D. Pilipenko, D. Temkin, R. Spatschek **183**

First Principles Approach for Vibrational Relaxation of Adsorbates on Semiconductor Surfaces
S. Sakong, P. Kratzer **191**

Determination of Structure and Electronic Properties of Free, Supported and Ligand Protected Metal Clusters by Density Functional Theory
M. Walter, M. Moseler **199**

Magnetic Properties of Co and Ni Multilayers on Diamond Surfaces
B. Stärk, P. Krüger, J. Pollmann **207**

Condensed Matter
Introduction by K. Binder and A. Muramatsu **215**

Quantum Transport Through a Nano-Device: A Numerical Renormalization Group Approach to Open Quantum Systems
F. B. Anders **219**

Van der Waals Interaction of Molecules on Surfaces from First Principles
P. Lazić, N. Atodiresei, V. Caciuc, St. Blügel **227**

On the Potential of Semiconductor Quantum Dots as Active Media for Future Semiconductor Lasers
M. Lorke, P. Gartner, F. Jahnke, W. W. Chow **235**

Simulation of Magnetic Friction
M. P. Magiera, D. E. Wolf **243**

Three-Body Strongly Correlated Polar Molecules in Optical Lattices
L. Bonnes, St. Wessel **251**

Computational Soft Matter Science
Introduction by K. Kremer **259**

Phase Transitions and Relaxation Processes in Macromolecular Systems: The Case of Bottle-Brush Polymers
H.-P. Hsu, W. Paul, P. E. Theodorakis, K. Binder **263**

Studying Directed Assembly of Diblock Copolymers on Patterned Substrates by Soft, Coarse-Grained Models
M. Müller, K. Ch. Daoulas **271**

Membrane-Protein Interactions in Lipid Bilayers: Molecular Simulation versus Elastic Theory
B. West, F. Schmid **279**

Semidilute Polymer Solutions under Shear Flow
G. Sutmann, C.-C. Huang, R. G. Winkler, G. Gompper **287**

Earth and Environment
Introduction by U. Hansen **295**

IMACCO: Virtual Institute for Inverse Modelling of the Atmospheric Chemical Composition
H. Elbern, A. Strunk, L. Nieradzik, N. Goris, J. Schwinger, K. Kasradze **297**

Turbulent Rotating Rayleigh-Bénard Convection
St. Stellmach, U. Hansen **305**

Adaptive Optimization of GOCE Gravity Field Modeling
W.-D. Schuh, J. M. Brockmann, B. Kargoll, I. Krasbutter **313**

Computer Science and Numerical Mathematics
Introduction by M. Griebel **321**

Threshold Determination for Reliable Quantum Error Correction
B. Trieu, R. Kennedy, M. Richter **323**

Optimization of Packing Problems with Heuristic Algorithms Suited for Parallel Enablement
J. J. Schneider **331**

Performance Tuning in the Petascale Era
F. Wolf, D. Böhme, M. Geimer, M.-A. Hermanns, B. Mohr, Z. Szebenyi, B. J. N. Wylie **339**

Hydrodynamics and Turbulence
Introduction by D. Kröner **347**

Study of Compressible and Quasi-Incompressible Flows with QUADFLOW Flow Solver
A. Bosco, B. Reinartz **349**

DNS of Finite-Size Particles in Turbulent Flows
H. Homann, J. Bec, R. Grauer **357**

Geometrical Properties of Small Scale Turbulence
N. Peters, L. Wang, J.-P. Mellado, J. H. Göbbert, M. Gauding, Ph. Schäfer, M. Gampert **365**

Shallow Moist Convection
T. Weidauer, C. Junghans, O. Pauluis, M. Pütz, J. Schumacher **373**

Computational Plasma Physics
Introduction by K.-H. Spatschek **381**

Plasma Simulation with Parallel Kinetic Particle Codes
P. Gibbon, R. Speck, B. Berberich, A. Karmakar, L. Arnold, M. Mašek **383**

Statistics, Structures and Alignment in Compressible MHD Turbulence
R. Grauer, J. Dreher, C. Beetz, C. Schwarz **391**

Entering the Petaflop-Era
–
New Developments in Supercomputing

Norbert Attig, Florian Berberich, Ulrich Detert, Norbert Eicker, Thomas Eickermann, Paul Gibbon, Wolfgang Gürich, Willi Homberg, Antonia Illich, Sebastian Rinke, Michael Stephan, Klaus Wolkersdorfer, and Thomas Lippert

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

Recent developments in the field of supercomputing on the European and the German level are presented. JUGENE, the highly scalable IBM Blue Gene/P system at the Jülich Supercomputing Centre (JSC), the first Petaflop supercomputer in Europe, started operation in July 2009. Furthermore, the JSC has installed a 300-Teraflop supercomputer, a next-generation general purpose cluster system. The JSC designed and constructed JUROPA and HPC-FF in a two-years effort, supported by its industrial partners Bull, SUN, INTEL, Mellanox, Novell and ParTec. HPC-FF is a system for fusion research, JUROPA succeeds JSC's p690 system JUMP. After achieving rank 10 in the TOP500 list and number 2 in Europe, JUROPA/HPC-FF became operational in August 2009. With these systems, the JSC has finally realized its concept of a *dual system complex*, which is adapted to the NIC application portfolio in an optimal manner. Additionally, the JSC has benchmarked and hosted a 100-Teraflop/s QPACE system developed by the SFB/TR 55 "Hadron Physics", the most energy-efficient supercomputer as of 2010. Coordinated by the Jülich Supercomputing Centre, the preparation phase project of the Partnership for Advanced Computing in Europe is about to be finalized. It will be followed by the first implementation phase project, again coordinated by the JSC, presumably starting mid-2010. In spring 2010, the statutes of the European Supercomputing Research Infrastructure are going to be signed, and operation of JUGENE as the first European supercomputer is foreseen to start in autumn 2010.

1 Introduction

With the inauguration of the IBM Blue Gene/P supercomputer "JUGENE" at the Jülich Supercomputing Centre (JSC) in May 2009, the Forschungszentrum Jülich has initiated the beginning of the era of Petaflop supercomputer performance in Europe. In June 2009, JUGENE was ranked position 1 in Europe and position 3 internationally in the TOP500 list[1]. As of today, JUGENE takes position 4 among the TOP500 and still is the most powerful and scalable machine worldwide which is solely devoted to free research. Operation started in July 2009. JUGENE is funded[a] in equal parts by the BMBF and the government of North Rhine-Westphalia, with the objective to become the first leadership-class system providing cycles to the new European supercomputing infrastructure "Partnership for Advanced Computing in Europe"[3].

[a]The underlying technical infrastructure as well as technical and scientific support is funded by the Helmholtz Association through the Programme "Supercomputing"[2].

After 5 years of successful operation, JSC's general purpose system, the IBM p690 Constellation JUMP[4] found a more than adequate replacement in the cluster system JUROPA – named after the project "Jülich Research on Petaflop Architectures"[5]. JUROPA indeed is a paradigm change in the provisioning of high-end systems at JSC. The system was designed within a two-years process by the computer scientists at JSC, supported by industrial partners: BULL as the integrator, SUN as hardware partner, INTEL providing the fastest general purpose processor, Mellanox delivering the high-bandwidth/low-latency network and Novell/SUSE contributing the node operating system. However, most important for the success of JUROPA was the long-term collaboration of JSC with the Munich cluster software expert ParTec within the ParaStation[6] consortium. The consortium has developed the open cluster software ParaStation currently providing the fastest and most scalable cluster operating system worldwide. In addition to the 200-Teraflop system JUROPA, the 100-Teraflop cluster HPC-FF[7] has been installed. HPC-FF is reserved for the researchers in the field of fusion science and technology, funded in equal parts by the EFDA[8] and FZJ. Both systems went into operation in August 2009. Thanks to ParaStation, JUROPA and HPC-FF can be operated as a joint cluster. Together, they achieve a peak performance of more than 300 Teraflop/s, with rank 10 in the TOP500 list of June 2009 and position 2 in Europe[1].

Thus, for the first time ever, a European site was in a position to operate two systems among the TOP10 in the world[1]. FZJ's dual system complex together with world-leading user support by JSC is optimally adapted to the application portfolios of the NIC and FZJ's Institute for Advanced Simulation (IAS). The complex is utilized by scientists and engineers all over Germany and Europe. As a key technology for industrial product development and production optimization, Petascale supercomputing will enable solving computational problems of hitherto unprecedented complexity.

The computer time on JUGENE and JUROPA is distributed through the John von Neumann Institute for Computing (NIC)[b][9]. In more than 25 years of its function, the NIC has developed a highly respected peer review scheme that is strictly based on the scientific quality of the proposals. With the start of the PRACE infrastructure foreseen in autumn 2010, part of the JUGENE capability will be distributed through the Europe-wide peer review process by the PRACE Scientific Steering Committee.

The JSC also took an active role in the development of the QPACE special purpose computer for lattice quantum chromodynamics simulations. The development of QPACE was directed by Prof. Tilo Wettig in the framework of the SFB/TR 55 "Hadron Physics". Partners are the Universities of Wuppertal, Regensburg and Frascati, DESY-Zeuthen and JSC. The JSC has been benchmarking and hosting the 100-Teraflop/s machine, which turned out to be evaluated as the most energy-efficient supercomputer in the Green TOP500 list[10] of autumn 2009. With the establishment of two innovative HPC research centres together with IBM and INTEL, the future will see an even more active role of IAS-JSC in the field of highly scalable HPC systems.

In 2009, the JSC started operation of its novel support structures, called "Simulation Laboratories"[11] (SimLab). Striving for optimized user support, the SimLabs will guarantee the ongoing development of simulation methodology, parallel algorithms and new programming and visualization techniques. The SimLabs are carrying out own research in

[b] For the Jülich applicants, the proposals are evaluated through the VSR.

core areas of the computational sciences. Each simulation lab is targeted at a specific disciplinary scientific area. Currently the SimLabs “Plasma Physics”, “Biology”, “Molecular Systems”, “Earth and Environment” and “Energy” are in operation or under construction. In the context of the Jülich-Aachen Research Alliance (JARA), the JSC intends to implement the SimLabs “Engineering” and “Computational Chemistry”.

One of the most important activities of JSC/FZJ, the management of the EU preparatory phase project for PRACE[3], which is coordinated by the Jülich chairman, Prof. Achim Bachem, is about to be successfully finalized. PRACE is an initiative of meanwhile 20 European countries to create a so-called tier-0 supercomputer infrastructure in the multi-Petaflop/s range that is on par with the US infrastructure. In spring 2010, the statutes of the European Supercomputing Research Infrastructure are going to be signed and the new legal entity will be formed. JSC’s project management office, headed by Dr. Thomas Eickermann, has submitted the proposal for the first implementation phase project of PRACE, again coordinated by the JSC and foreseen to start mid-2010. Germany’s representative in this project is the Gauss Centre for Supercomputing e.V. (GCS)[12]. The GCS was initiated by the Minister for Research, Prof. Schavan, in 2006, with Forschungszentrum Jülich, Höchstleistungsrechenzentrum Stuttgart and Leibniz-Rechenzentrum Garching, as well as the chairs of the scientific councils, *i.e.* the NIC for FZJ, as members.

This report reviews the status of GCS (section 2) and PRACE (section 3), and will describe the developments in supercomputing capability, support and research at IAS-JSC (section 4).

2 Gauss Centre for Supercomputing

In order to continue to provide computing resources at the highest performance level for computer-based science and engineering in Germany, in June 2006, the Federal Ministry of Education and Research (BMBF) has initiated the foundation of the Gauss Centre for Supercomputing (GCS)[12]. The Ministries of Innovation, Science, Research and Technology of the State of North Rhine-Westphalia, the Bavarian State Ministry of Science, Research and the Arts, and the Ministry of Science, Research and the Arts of Baden-Württemberg unreservedly support the GCS. Since September 2007, the GCS is a registered association (e.V.).

The three German national supercomputing centres joined to form the GCS with combined computing power, located in Stuttgart, Garching, and Jülich. The GCS is the largest national association for high-performance computing in Europe. The mission of the GCS is the provision of Petascale supercomputing power for computational science and engineering for Germany. In the next six years, GCS will be funded by the ministries with a total amount of € 400 million. The procurement of hardware will be closely coordinated, applications for computing time will be scientifically evaluated on a common basis, and software projects will be jointly developed. It is planned that each GCS-site, LRZ, HLRS and JSC, will install and operate high-performance computer systems of different architecture with processing speed of one to three Petaflop/s. Jülich is the first GCS partner which could benefit from the funding. Beginning in January 2009, JSC has installed an IBM supercomputer Blue Gene/P with a peak performance of one Petaflop/s and runs production since July 2009. On behalf of GCS, the peer review process and the allocation of compute time is organized and carried out by the extended committees of the NIC.

Another important area of activities of the GCS will be the training of users and the education of young scientists. The work of specialized researchers will be supported and promoted by harmonizing the services and organizing joint schools, workshops, and conferences on simulation techniques. Methodologically oriented user support is also a mission of the Gauss Centre.

With the established national institution of GCS, its funding and its first Petaflop supercomputer in Europe, the GCS plays a major role in the development of the European supercomputer infrastructure directed by PRACE. GCS as one of six hosting members of PRACE has committed itself to provide compute cycles to the value of € 100 million for large scale scientific projects of European communities in a five years period.

3 PRACE – A Pan-European HPC Service

To remain internationally competitive, European scientists and engineers must be provided with leadership-class supercomputer systems that are competitive on an international scale. PRACE, the Partnership for Advanced Computing in Europe, is creating a persistent pan-European high-performance computing (HPC) service and infrastructure. European scientists and technologists will be able to access leadership-class supercomputers with capabilities equal to or better than those available in the USA and Japan. The service will comprise three to six world-class leadership Tier-0 HPC centres, strengthened by regional and national supercomputing centres working in tight collaboration through interoperable grid technologies. The infrastructure created by PRACE will form the top level of the European HPC ecosystem, *cf.* Figure 1.

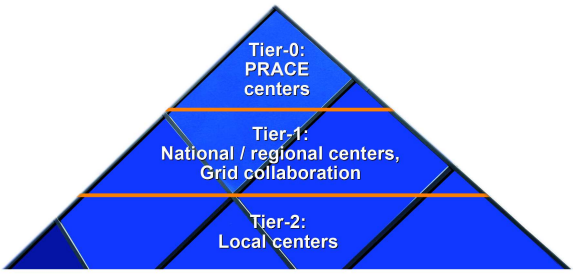


Figure 1. Performance / peer review pyramid.

During the last two years, PRACE has raised significant awareness and is well recognized as the major HPC project in Europe. PRACE has completed its work plan and has achieved all milestones in both the legal and administrative tasks as well as the technical work. The tasks of selecting a suitable legal form and its governance structure and the definition of a funding and usage model are closely interlinked and form the very centre of the project activities. After a thorough analysis of the available options, taking into account existing European Research Infrastructures, proposals for the legal form and a draft governance structure for PRACE have been derived. Negotiations involving representatives from the national ministries have led to the agreement to establish the PRACE Research Infrastructure initially as an association located in Lisbon, Portugal. Signature-ready contracts for this association have been prepared. The pan-European peer review system for PRACE is also laid out in the statutes. It is based on the principle “best systems for best science” and takes into account the results of the HET group[13] and an analysis of national HPC peer review systems.

In the initial phase, the HPC service will be based on coordinated national procurements. Subsequently, the legal entity is intended to evolve into a more tightly integrated structure. The four Principal Partner countries Germany, France, Spain and Italy have already now made binding commitments to contribute Tier-0 resources each with a value of € 100 million within the next five years.

An analysis of the HPC ecosystem has been performed that identifies key stakeholders at all levels including funding, policy makers, resource providers and potential users. PRACE is a founding member of the European e-Infrastructure Forum, which strives for interoperation of European e-Infrastructure. There is a close cooperation with DEISA[14] at the technical level, synchronizing the work and sharing results among the projects.

PRACE also has undertaken extensive dissemination and outreach activities. The web site `www.prace-project.eu` is a central point of public information about all activities within the project. It is complemented by active dissemination channels such as press releases, newsfeeds, a quarterly newsletter and contributions to the most relevant HPC conferences. A Summer and a Winter School on Petascaling as well as five code porting workshops were organized. The DEISA-PRACE symposium in May 2009 in Amsterdam was jointly organized with DEISA and attracted about 200 participants from all over the world.

One focus of the technical work in PRACE has been the gathering of information to prepare important decisions regarding the type, performance, costs, and hosting of the future Petaflop/s production systems and the prototypes of these systems. The analysis of the application requirements and the survey of available and promising future technologies matching these requirements have led to a selection of prototype systems that cover the most relevant architectures and technologies. The JSC hosted three of the 15 prototypes. This included the use of the IBM Blue Gene/P (JUGENE), the Intel Nehalem/Infiniband cluster (JUROPA) in cooperation with CEA and the IBM Cell+FPGA system (QPACE).

An application benchmark suite used for the assessment of the prototype systems has been created to be representative of the likely usage of European Petaflop/s systems. It is based on a survey of major application codes on European HPC systems and an in-depth survey of the identified applications, leading to a utilization matrix, providing information about HPC resource usage by scientific domain and algorithm.

The requirements for Tier-0 systems management and integration with the ecosystem, in particular the Tier-1 systems, have been analyzed based on experience and expectations of potential Tier-0 sites and existing Tier-1 infrastructures. A uniform user environment, using standardized authentication and data-transfer mechanisms, has been defined and deployed on the prototype systems. A close cooperation has been established with DEISA as a major provider of distributed HPC systems management solutions. This has led to a consistent environment throughout Tier-0 and Tier-1.

STRATOS, the PRACE advisory group for strategic technologies, has been created as the vehicle to foster the evaluation and active initiation of research on future multi-petascale technologies beyond 2010[15]. Within STRATOS, partners from PRACE and industrial consortia, including more than 80 organizations, co-operate on the specification and developments of such components. The STRATOS MoU was signed in December 2008 by 12 PRACE partners and the consortia PROSPECT and Ter@Tec. Its long-term work plan comprises joint activities on Exascale Software, Green IT and HPC Leadership resources.

PRACE submitted the proposal for the first implementation phase project under the European Union's 7th Framework Programme in November 2009. This new project will support the evolution of the research infrastructure by refining and extending the administrative, legal and financial framework structure the European HPC ecosystem by interacting with the stakeholders and the national and topical Tier-1 centres and focus on the specific requirements of industry. To enable world-class science on novel systems the project will support users in porting, optimizing and petascaling applications to the different architectures and deploy consistent services across the research infrastructure.

4 Supercomputing and Support

The JSC strives to work at the forefront of the technologically possible in HPC. Following its dual system philosophy JSC has replaced the general purpose constellation system JUMP by JUROPA and HPC-FF, two fully connected clusters increasing the available power for general purpose computing at JSC by a factor of more than 30. Since July 2009, the JSC also is the first site in Europe to provide Petaflop capability with its IBM Blue Gene/P, nicknamed JUGENE. With QPACE, alongside the University of Wuppertal, the JSC hosts one of the two most energy-efficient systems worldwide[10], developed by the SFB/TR 55 "Hadron Physics". Together with an online storage of 6 PetaBytes, connected to an automatized tape archive of up to 25 Petabytes and several satellite clusters for special purposes, the JSC currently operates the most advanced and flexible supercomputer complex in Europe. Its world-leading user support, featuring the novel concept of "Simulation Laboratories" is well adapted to the application portfolios of the NIC and FZJ's Institute for Advanced Simulation (IAS) and will be exemplary for the support structures to be built in the European supercomputing infrastructure PRACE.

4.1 Heterogeneous System Philosophy

Since the installation of JUMP the Parateam experts at JSC, responsible for the scaling of user applications, realized that, for the given number of users and allocation of computing time, a further growth of capability computing on JUMP would have decreased the overall efficiency. This is due to the fact that users mainly work at *their* "sweet spot" of effectivity and cannot profit from policies preferring jobs with very large processor numbers. Consequently, the JSC introduced the concept of a "Dual Supercomputer Complex" as a solution, *cf.* the white paper "Ausbau des Supercomputing in der Helmholtz-Gemeinschaft und Positionierung im europäischen Forschungsraum".

With the first IBM Blue Gene/L system JUBL complementing JUMP with 46 Teraflop/s, the JSC started realizing the dual concept. The latest pair JUGENE and JUROPA followed in 2009. Before being admitted to the machine, the scalability of a code must be proven. Following this philosophy, the number of accepted projects on JUGENE is limited. A substantial increase of the scalability up to 294,912 cores of JUGENE could be reached for several applications in quite a few workshops, like the "Big Blue Gene Weeks" and the "Blue Gene Scaling Workshops".

The dual system complex is supported by a common storage infrastructure which was further expanded in 2009. Key part of this infrastructure is the new Jülich storage cluster (JUST) now with an online disk capacity of 6 PetaByte. The maximum I/O bandwidth is 60 GB/s. JUST is connected to the supercomputers via a new switch technology based on 10-Gigabit Ethernet, serving the file systems GPFS on JUGENE and LUSTRE on JUROPA.

4.2 JUGENE – Europe's First Petaflop System

The history of leadership-class systems for capability computing applications at JSC started in mid-2005 with a one-rack Blue Gene/L, JUBL, which had been extended to eight racks in January 2006 and was replaced by the 16-rack Blue Gene/P system JUGENE in November 2007, with 65,536 processors and a peak performance of 222.8 Teraflop/s.

Based on the long experience with the IBM Blue Gene technology and after one year of successful operation, JSC decided to extend JUGENE to a 72-rack Blue Gene/P system forming Europe's first computer with a peak performance of more than one Petaflop/s[16].

After a major upgrade of the infrastructure (water cooling and power supply of 6 MW) at the beginning of 2009, the 72 Blue Gene/P racks were installed within four weeks and became available for users on July 1st, 2009, *cf.* Figure 2. This installation includes about 23 km of copper cables (Blue Gene/P data cables) and 21 km fibre cables (10-Gigabit Ethernet), plus 7.5 km Cat5 cables (1 Gigabit Ethernet). JUGENE was No. 3 worldwide in the Top500 list in June 2009 and is still the fastest computer system in Europe. The system specifications are summarized in Table 1.

Figure 2. Petaflop system JUGENE.

Type	IBM Blue Gene/P – 72 racks
Peak Performance	1.008 Petaflop/s
LinPack Performance	825.5 Teraflop/s
Cores	294,912 (ppC 450 800 MHz)
Compute Nodes	72,728 (4-way SMP, 2 GB memory)
Total memory	144 TB
Communication network	3-dimensional torus (fully DMA capable)
Network latency	160 ns
Bisectional bandwidth	188 TB/s
I/O nodes	600 (connecting to external 10GigE)
Total disk storage	5.3 PB
I/O bandwidth	66 GB/s

Table 1. JUGENE system characteristics.

JUGENE is fully integrated into the Jülich supercomputer infrastructure. Its I/O nodes are connected to the fileserver JUST by four FORCE 10 E1200i switches with a maximum of 896 ports, based on 10-Gigabit Ethernet. The fileserver itself has also been upgraded. It now consists of 28 IBM Power 6 servers and 18 storage controllers serving more than 6000 disks to reach a maximum bandwidth of 66 GB/s and a capacity of 5.3 PB.

With JUGENE's increased capacity, Jülich is excited to see its users achieving further ground-breaking scientific highlights. JSC tries to support those highlights by the advanced training of the users during high-level Blue Gene scaling workshops, as successfully practiced in the past. The last workshop in October 2009 showed several applications able to scale up reasonably to 288k cores. And the next scaling workshop will soon take place in March 2010.

The new system will help to ensure that the Forschungszentrum Jülich continues to play a leading role in the global high-performance computing research community, especially in the context of PRACE. It is the first German supercomputer system that was selected and purchased in the context of the Gauss Centre for Supercomputing (GCS) and is funded by the German Federal Ministry of Education and Research (BMBF) and the Ministry for Innovation of North Rhine-Westphalia.

4.3 JUROPA – JSC's General Purpose Supercomputer

Figure 3. Cluster Computer JUROPA.

JUROPA is a general-purpose supercomputer with more than 200 Teraflop/s peak performance, *cf.* Figure 3[5]. It is based on INTEL's most recent processor technology combined with a high-speed/low-latency Infiniband interconnect. JUROPA has been implemented as a co-development with industrial partners, focussing on scalability, interconnect topology, storage technology and cluster management. Partners in this project were FZJ/JSC, Bull, Intel, Mellanox, Novell, ParTec and Sun. Applications on JUROPA follow the JUMP portfolio and originate from various disciplines like chemistry, many-particle physics, elementary particle physics, biology/biophysics, materials science, soft matter and others. One important new source of applications has been the European fusion community (EFDA – European Fusion Development Agreement) which became interested in the JUROPA concept and signed a contract on using a 100-Teraflop/s JUROPA cluster of their own named HPC-FF (HPC for fusion). HPC-FF follows the same design principles as JUROPA and is also operated by JSC. HPC-FF and JUROPA are closely coupled with each other through a common Infiniband network with full bi-sectional bandwidth, designed by the JSC. Both systems can be operated as one big 300 Teraflop/s system.

JUROPA is built up from innovative components combined into a "best-of-breed" HPC-cluster. Intel Xeon X5570 processors (Nehalem-EP) form the computational basis

of the cluster nodes. Each node comprises two of them, providing 8 CPU cores per node attached to 24 GB of main memory. Running at 2.93 GHz clock speed, each node reaches a peak performance of 94 Gigaflops. Hence, 2208 compute nodes yield a total peak performance of 207 Teraflops. JUROPA is based on Sun Microsystem's Blade SB6048 technology, while HPC-FF is implemented with Bull NovaScale R422-E2 server technology.

All nodes are connected via Infiniband/QDR with fat-tree topology. Performance measurements show a network latency around 1.5 to 2.0 μs on the MPI-level and a sustained bandwidth of over 2800 MB/s per link and direction. The cluster management, monitoring and MPI-communication is provided by the open-source cluster operating system ParaStation[6], co-developed by JSC and ParTec Cluster Competence Centre (Munich). The MPI provided by ParaStation fully scales to the 24,000 cores of JUROPA/HPC-FF, and special environment variables actually allow to scale Infiniband communication to millions of cores. It is ParaStation that renders JUROPA/HPC-FF to a cluster of hitherto unprecedented scalability while providing an utmost communicational flexibility.

JUROPA and HPC-FF share a common Lustre storage pool providing a total of 850 TB of usable disk space. The Lustre storage pool is used for a fast scratch filesystem striped across all available disks as well as for moderately striped home filesystems. Access to data residing on Jülich's storage cluster JUST running GPFS is provided through GPFS gateway nodes. They bridge between the Lustre file server running over Infiniband and the GPFS file server running via 10-Gigabit-Ethernet. A total of 7 login nodes form the interface to the JUROPA users. They allow for interactive program development and testing as well as pre- and postprocessing.

4.4 QPACE – World Champion in Green Computing

QPACE (Quantum Chromodynamics Parallel Computing on the Cell) is a next-generation massively parallel and scalable computer architecture optimized for lattice quantum chromodynamics (LQCD), *cf.* Figure 4. It has been developed in cooperation between several academic institutions together with the IBM development laboratory in Böblingen, Germany, under the framework of the SFB/TR 55, directed by the University of Regensburg.

Figure 4. Energy Saver QPACE.

At JSC, a 4-rack QPACE system is installed. It is a sister system to a machine at Wuppertal University. The building block is a node card comprising an IBM PowerXCell 8i processor and a custom FPGA-based network processor. 32 node cards are mounted on a single backplane and eight backplanes are arranged inside one rack, hosting

a total of 256 node cards. The closed node card housing, which is connected to a liquid-cooled cold plate, acts as a heat conductor thus making QPACE a highly energy-efficient system. To remove the generated heat a cost-efficient liquid cooling system has been developed which enables high packaging densities. The maximum power consumption of one QPACE rack is about 32 kW. The 3-dimensional torus network interconnects the node cards with nearest-neighbour communication links, driven by a lean custom protocol optimized for low latencies. For the physical layer of the links q10 Gigabit Ethernet PHYs are used providing a bandwidth of 1 GB/s per link and direction.

In the context of PRACE work-package 8, promising technologies for future supercomputers have been evaluated, and QPACE was identified as one of the advanced prototypes. In order to extend the range of applications for QPACE beyond LQCD, an implementation of the high-performance LINPACK (HPL) benchmark was carried out as a first step. Here, different from QCD applications, the transfer of large messages and collective operations must be supported efficiently. This required extensions of both network communication and the software stack. A special FPGA bit-stream and a library of MPI-functions for message passing between compute nodes has been developed at JSC in cooperation with IBM. The HPL benchmark produced excellent results (43.01 Teraflop/s on 512 nodes), with the 4-rack system in Jülich giving an aggregate peak performance of more than 100 Teraflop/s.

It is a remarkable outcome of this research project that, delivering 723 MFlop/s/W, QPACE was recognized as the most energy-efficient supercomputer worldwide. It is ranked top on the Green500 list at the Supercomputing conference 2009 in Portland[10].

4.5 Simulation Laboratories

The relentless advances in supercomputing capability over the last decade have prompted a major paradigm shift in the field of computational sciences. In the past, it was possible to develop complex theoretical models without too much regard for algorithmic or numerical elegance: the raw number-crunching power of available processors would sooner or later compensate for sub-optimal code design. With the increasing prevalence of architectures based on massively parallel and multi-core processor topologies, more and more simulation scientists are now compelled to take scalability into account when developing new models or when porting long-established codes to new machines. This can pose significant problems for the small research groups making up the majority of users of the JSC computing facility, which might not have the resources or expertise for application petascaling. Many important applications are only able to exploit a small fraction of the available computing power because of their poor parallelism, and there is a real possibility that whole communities may be left behind as supercomputing hardware advances further.

To meet this challenge the Jülich Supercomputing Centre has proposed a new research and support structure, named Simulation Laboratory (SimLab), *cf.* Figure 5. This is a community-oriented research and support structure, *i.e.* each lab is tailored to a specific scientific community. In Jülich these communities mainly comprise neighbouring institutes of FZJ and local university groups. A SimLab consists of a core group located at an HPC centre and a number of associated scientists outside. It is staffed to the level of a small, self-contained research team, having high-level expertise on community-related codes and algorithms. The know-how transferred to its co-members in the research community thus far exceeds the support level of a traditional expert advisor. The well-known user-support

model is not completely substituted by SimLabs but is naturally augmented by them. To define and monitor the scope of each SimLab a Steering Committee will be established jointly by the community and the HPC centre, whose task is to evaluate incoming work packages, selecting and ranking them. As well as pursuing its own research, the core group of the Simulation Laboratory is also responsible for community-building measures such as education, training and guest programmes, and is well equipped with state-of-the-art HPC resources.

In addition to its newly established Simulation Laboratories for Biology, Plasma Physics, Molecular Systems and Earth Science (Energy will follow), JSC also pursues basic research in HPC methods and algorithms and in performance analysis tools through Cross-Sectional Teams. These teams carry out basic research in computational methods and assist simulation scientists in applying those methods. In the present context they will support Simulation Laboratories and JSC's internal Research Groups in their domain-specific tasks and at the same time also interact with individual users to improve their simulation codes. In this sense, they act both as research and as service groups. The combination of Simulation Laboratories and the Cross-Sectional Teams forms a first-class science-based partnership between disciplinary communities and the supercomputing centre.

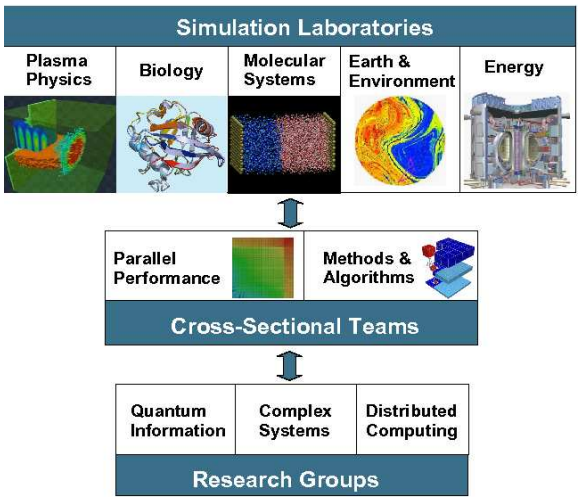


Figure 5. Simulation Labs, cross-sectional groups and research labs at IAS-JSC.

Currently there are ongoing activities to extend the SimLab concept further: JSC's partnership with KIT in the HGF programme Supercomputing will lead to the foundation of additional SimLabs with Karlsruhe. Freshly established collaborations with DLR and DWD and the intense partnership with Aachen University also call for a Fluid Engineering and a Chemistry SimLab. In the European context a joint distributed SimLab on Nuclear and Particle Physics is planned together with partners from the Eastern Mediterranean. Early engagement by JSC's SimLabs with local groups has already been successful in bringing new codes onto Petascale systems. These experiences build the basis for successfully applying the JSC blueprint on a national and European scale in the near future.

5 Concluding Remarks

The year 2009 marks several breakthroughs in supercomputing at Jülich: JUGENE, the first Petaflop system in Europe, has been brought into operation, JUROPA, a Petaflop-class general purpose cluster computer was constructed in co-development with industrial partners, the QPACE special purpose QCD system was implemented at JSC, developed in a

record time of two years only by the SFB/TR 55 “Hadron Physics” – directed by University of Regensburg – and IBM, with partners Jülich, DESY and Universities of Wuppertal and Ferrara. QPACE became the most energy-efficient supercomputer in the world. The first simulation laboratories at IAS-JSC, strongly endorsed by the Helmholtz Senate in the Evaluation of the Helmholtz Programme Supercomputing in 2009, have started operation. The PRACE preparation phase project, coordinated by JSC, is about to be completed successfully. The first PRACE implementation phase project has been submitted, again coordinated by JSC and the PRACE statutes will be signed in Spring 2010. IAS-JSC is currently finalizing the contracts for two joint development laboratories with IBM and INTEL, looking forward to multi-Petaflop/s and Exascale. JSC’s activities will guarantee that German and European scientists and engineers can continue working at the forefront of simulation sciences[17].

References

1. http://www.top500.org/list/2009/06/100.
2. http://www.helmholtz.de/forschung/schluesseltechnologien/supercomputing/.
3. http://www.prace-project.eu/.
4. http://www.fz-juelich.de/jsc/service/sco_ibmRegatta.
5. http://www.fz-juelich.de/jsc/juropa.
6. http://www.parastation.com/.
7. http://www.fz-juelich.de/jsc/juropa/configuration/.
8. http://www.efda.org/.
9. http://www.fz-juelich.de/nic/.
10. http://www.green500.org.
11. http://www.fz-juelich.de/jsc/csim.
12. http://www.gauss-centre.eu.
13. http://www.hpcineuropetaskforce.eu.
14. http://www.deisa.eu/.
15. http://www.prace-project.eu/news/prace-to-implement-stratos-2013-a-permanent-research-platform.
16. http://www.fz-juelich.de/jsc/jugene.
17. S. Durr *et al.*, “Ab-Initio Determination of Light Hadron Masses”, Science **322**, 1224 (2008).

Quark Mass Dependence of the QCD Equation of State

Frithjof Karsch[1,2,4], **Wolfgang Söldner**[3]
in collaboration with
C. Jung[4], **O. Kaczmarek**[2], **E. Laermann**[2], **C. Miao**[4], **S. Mukherjee**[4], **P. Petreczky**[4], **C. Schmidt**[2], **and W. Unger**[2]

[1] GSI Helmholtzzentrum für Schwerionenforschung, Planckstr. 1, D-64291 Darmstadt, Germany

[2] Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany

[3] ExtreMe Matter Institute EMMI, GSI Helmholtzzentrum für Schwerionenforschung Planckstr. 1, D-64291 Darmstadt, Germany

[4] Physics Department, Brookhaven National Laboratory, Upton, NY 11973, USA

We are presenting lattice calculations of quantum chromodynamics of thermodynamic quantities which are of great interest in relativistic heavy ion collisions and in the early universe. With an improved staggered fermion action (p4fat3) we use a strange quark mass that has been tuned to its physical value and two light quark masses $m_s/20$. This corresponds to an almost physical pion mass. We present results on the equation of state of QCD comparing to data obtained in earlier calculations with a light quark mass twice as heavy. This enables us to quantify the quark mass dependence. We also comment on the quark mass dependence of other thermodynamic quantities like the chiral condensate, the renormalized Polyakov loop, and the disconnected chiral susceptibility.

1 Introduction

A major topic in thermodynamics of quantum chromodynamics (QCD) and, especially, lattice QCD is the calculation of the equation of state (EoS)[1–4]. The QCD EoS is of great relevance in relativistic heavy ion collisions and in the early universe. In large experiments, like LHC, RHIC, and SPS and future experiments like FAIR, hadronic matter is studied under extreme conditions in relativistic heavy ion collisions. In these experiments hadronic matter can be studied under such conditions similar to those apparent milliseconds after the Big Bang.

In those experiments two heavy ions are colliding at very high energies (presumably) producing what one calls a quark-gluon plasma (QGP). During the expansion of the plasma the QGP cools down undergoing a chiral and confinement phase transition where free quarks and gluons become confined to hadrons. The temperature regime where this dramatic change appears is at about $200\mathrm{MeV}$. While there has been great progress in the recent years in calculating the QCD EoS in the vicinity of that phase transition within the framework of lattice QCD, still several issues remain unresolved. One of them is the observed deviation in the EoS of the lattice computation compared to what one expects from a Hadron Resonance Gas model (HRG). The HRG model is expected to give an approximation to the EoS in the low temperature regime. Indeed, there is quite good agreement with experiment[8], and one would expect that the HRG model also gives a good description for the low temperature region of the lattice QCD equation of state. However, it turns out the HRG description shows deviations from the lattice QCD data[3,5], especially in the low

light (up, down) and	$m_q = 0.1m_s$	$m_q = 0.00081 - 0.00370$
strange quark masses	$m_q = 0.05m_s$	$m_q = 0.00120 - 0.00145$
pion masses	$m_q = 0.1m_s$	$m_\pi \approx 220\ \mathrm{MeV}$
	$m_q = 0.05m_s$	$m_\pi \approx 160\ \mathrm{MeV}$
volume ($T \neq 0$)	$m_q = 0.1m_s$	$32^3 \times 8$
	$m_q = 0.05m_s$	$32^3 \times 8$
volume ($T = 0$)	$m_q = 0.1m_s$	32^4
	$m_q = 0.05m_s$	32^4
statistics ($T \neq 0$)	$m_q = 0.1m_s$	8,000-37,000 trajectories
	$m_q = 0.05m_s$	5,000-70,000 trajectories
statistics ($T = 0$)	$m_q = 0.1m_s$	2,000 - 6,000 trajectories
	$m_q = 0.05m_s$	1,500-3,500 trajectories

Table 1. Overview of the simulation details.

temperature regime. A possible source for the discrepancy, aside from lattice discretization errors, may be due to the fact that those lattice studies of the EoS had been performed with light quark mass values which are about a factor two larger than in nature, i.e. a light pseudo-scalar mass of about $220\ \mathrm{MeV}$. To quantify the effects of the too large light quark masses we have performed calculations at a smaller light quark mass $m_q = 0.05m_s$ compared to our earlier calculation with $m_q = 0.1m_s$ where m_s is the strange quark mass. The pion mass lowers to about $160\mathrm{MeV}$ which is almost at its physical value.

We also make use of the data for $m_q = 0.05m_s$ to investigate the properties of the chiral and confinement phase transition. In particular, we will look at the scaling with the light quark mass of the chiral condensate and the chiral susceptibility which is relevant for identifying the critical point of the chiral phase transition. The renormalized Polyakov loop is used to study the confinement phase transition.

The paper is organized as follows. In section 2 we give an overview of numerical details and outline the calculation of the EoS on the lattice. Results of the EoS with $m_q = 0.05m_s$ are presented in section 3 and results for the Polyakov loop, the chiral condensate and susceptibility are shown in section 4. Finally, we comment on performing a continuum limit.

2 Numerical Details and Setup

Large computational resources are required to compute the QCD equation of state within the framework of lattice QCD. The calculations we are presenting have been performed on IBM Blue Gene/L and Blue Gene/P, as well as on QCDOC architecture, where we have been using those supercomputers located at the Jülich Supercomputing Centre, the Lawrence Livermore National Laboratory, and the New York Center for Computational Sciences (NYCCS).

In order to calculate expectation values, like the energy density ε or the pressure p, we are generating an ensemble of gauge field configuration for several temperatures T by using a Rational Hybrid Monte Carlo (RHMC) algorithm[3]. For our calculations we have been using the p4fat3 action which is an improved action in the sense that cutoff

effects are reduced. We are simulating QCD with flavours $N_f = 2 + 1$, i.e. we have two degenerate light quarks and one (heavier) strange quark. We have generated two sets with different (physical) light quark masses m_q keeping the strange quark mass m_s fixed. Some computational details are summarized in Table 1 and Refs. 6, 9. Further information can be found in Refs. 10–12.

In the following we give a brief sketch of the calculation of the EoS on the lattice. The basic quantity is the trace anomaly $\Theta^{\mu\mu} \equiv \varepsilon - 3p$ with the energy-momentum tensor $\Theta^{\mu\nu}$, energy density ε, and pressure p. The trace anomaly is computed in terms of the action density, and the strange and light quark chiral condensates on the line of constant physics, i.e. the physical values of the simulation parameters are kept fixed for all temperatures. We are only interested in the thermal part of the trace anomaly and subtract the zero temperature part,

$$\Theta^{\mu\mu} \equiv \varepsilon - 3p = (\varepsilon - 3p)_T - (\varepsilon - 3p)_{T=0}. \tag{1}$$

From this expression we can determine the pressure through the thermodynamic relation $\Theta^{\mu\mu}/T^4 = T\partial(p/T^4)/\partial T$ which is integrated to yield

$$\frac{p(T)}{T^4} - \frac{p(T_0)}{T_0^4} = \int_{T_0}^{T} \mathrm{d}\bar{T} \frac{1}{\bar{T}^5} \Theta^{\mu\mu}(\bar{T}) \; . \tag{2}$$

Here T_0 is chosen to be in the deep hadronic region where $p(T_0)$ is already exponentially small. The energy density ε then is determined given $\Theta^{\mu\mu}$ and p. For further details see Ref. 3.

3 Mass Dependence of the QCD Equation of State

The results for the trace anomaly for $m_q = 0.05 m_s$ are plotted in Figure 1. We compare to the $m_q = 0.1 m_s$ case from Refs. 3,9 as well as to the HRG model result. The curves in the plot correspond to an interpolation of the lattice data. Note that for our calculation of the trace anomaly for the HRG all resonances have been chosen to be at their physical mass. All resonances up to the mass $M_{max} = 2.5\mathrm{GeV}$ are taken into account. The (physical) temperature scale has been determined by calculating the static quark potential where we used a value for $r_0 = 0.469\mathrm{fm}$[7].

From Figure 1 we observe in the transition region a shift towards lower temperature for $(e - 3p)/T^4$ for $m_q = 0.05 m_s$ of about $5\mathrm{MeV}$ compared to $m_q = 0.1 m_s$ (with temporal lattice extent $N_\tau = 8$). A small part of that shift of about $0.5\ \mathrm{MeV}$ comes from a change in the scale for the $m_q = 0.05 m_s$ data. Comparing to the HRG model, for the larger temperature regime we find that for the smaller quark mass there is to some extent agreement with the HRG model. However, for the lower temperature region we observe deviations from the HRG.

On the right panel of Figure 1 we plotted the corresponding energy density ε and the pressure p. We observe a rapid change in the transition region $T = 180 - 190\mathrm{MeV}$ in the energy density and also, a more moderate change, in the pressure. The energy density is increasing with larger temperatures towards the Stefan–Boltzmann limit ε_{SB}. A difference in the quark mass shows up only in the transition region.

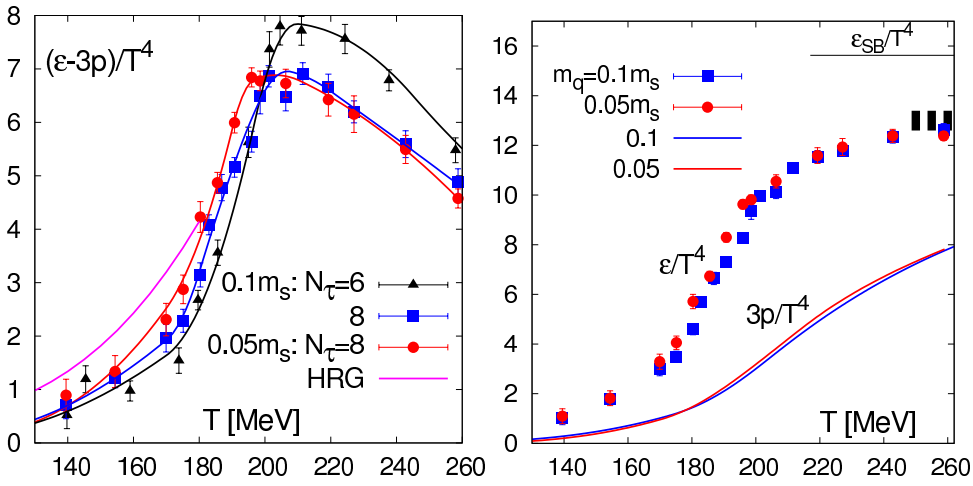


Figure 1. On the left hand side the trace anomaly from lattice QCD at temporal lattice extent $N_\tau = 6, 8$ for $m_q = 0.1m_s$ and $m_q = 0.05m_s$ is plotted. Deviations are observed between HRG and lattice QCD for the low temperature regime. The plot on the right hand side shows the corresponding energy density ε and pressure p for $N_\tau = 8$. The black boxes indicate the shift when using different values for the integration constant $p_0 = p(T_0)$[6].

4 Chiral Condensate

In this section we present results on the chiral condensate and the renormalized Polyakov Loop. The chiral condensate is an order parameter of the chiral phase transition, the Polyakov loop is an indicator of the deconfinement phase transition. We consider a combination of the light and strange quark condensate to eliminate quadratic divergencies with respect to the additive quark mass renormalization

$$\Delta_{l,s} = \frac{\langle\bar{\psi}\psi\rangle_{l,\tau} - \frac{\hat{m}_l}{\hat{m}_s}\langle\bar{\psi}\psi\rangle_{s,\tau}}{\langle\bar{\psi}\psi\rangle_{l,0} - \frac{\hat{m}_l}{\hat{m}_s}\langle\bar{\psi}\psi\rangle_{s,0}}. \tag{3}$$

We normalize this combination by its zero temperature value to cancel the multiplicative renormalization factor. In Figure 2 we show on the left hand side results for our two data sets $m_q = 0.05m_s$ and $m_q = 0.1m_s$. We observe a sharp drop of $\Delta_{l,s}$ in the region of $T = 180 - 190\mathrm{MeV}$ for both data sets. For the smaller quark mass $m_q = 0.05m_s$ we observe a shift in the transition region of about $5\mathrm{MeV}$. We note that a future goal is to probe the chiral limit $m_q \to 0$. With only two different light quark masses we cannot investigate this limit at the moment since near the critical temperature T_c terms proportional to m_q as well as $\sqrt{m_q}$ are present.

The renormalized Polyakov loop is calculated from the bare Polyakov loop by multiplying it by the renormalization constant $z(\beta)$, see Ref. 6,

$$L_{ren} = z(\beta)^{N_\tau} L_{bare} = z(\beta)^{N_\tau} \frac{1}{3}\mathrm{tr} \prod_{x_0=0}^{N_\tau-1} U_0(x_0, \mathbf{x}). \tag{4}$$

In the above formula $U_0(x_0, \mathbf{x})$ denotes the temporal link variables. Also for the Polyakov

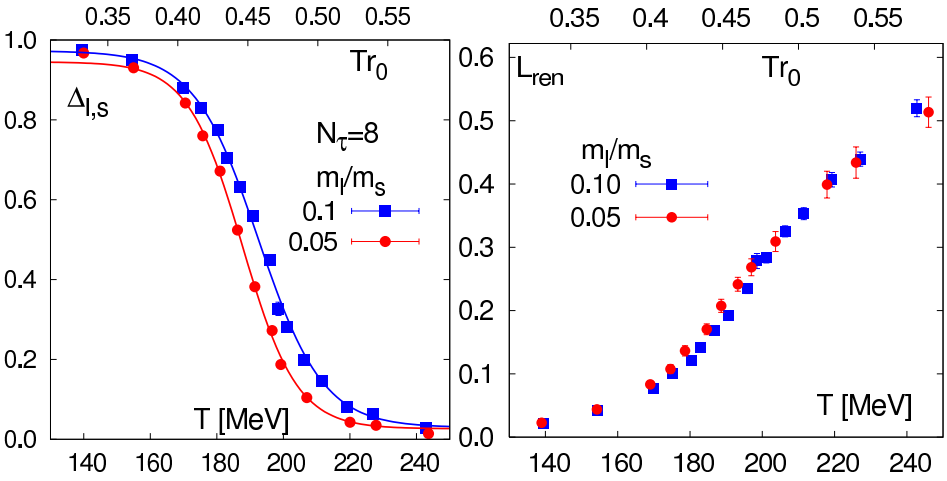


Figure 2. We plot $\Delta_{l,s}$ for the two different light quark masses (left panel). The renormalized Polyakov loop is shown on the right hand side. The transition region is in the temperature interval $T = 180 - 190\mathrm{MeV}$.

loop we observe a rapid change in the transition region $T = 180 - 190\mathrm{MeV}$. Note that this is also the region where the energy density shows its rapid change.

On the right hand side of Figure 3 we plotted the disconnected light chiral susceptibility $\chi_{l,disc}$ for different light quark masses. We find that $\chi_{l,disc}$ shows a strong quark

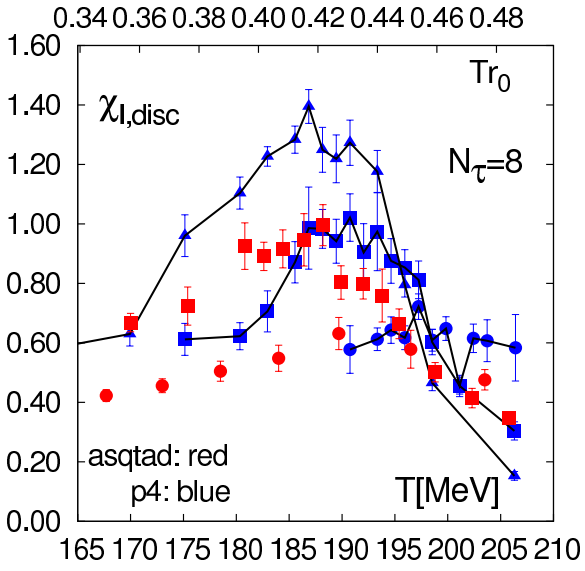


Figure 3. Disconnected chiral susceptibility for light quark masses $m_q = 0.2m_s$ (circles), $m_q = 0.1m_s$ (squares), and $m_q = 0.05m_s$ (triangles). We compare to two different actions, p4fat3 and asqtad, having different discretization errors[5].

mass dependence over a wide temperature range. This can be understood as an increase in the fluctuations of Goldstone modes below the transition temperature. The transition temperature should then be found near the right edge of the peak. Further evidence for this picture is given by the observation that the height of the peak of $\chi_{l,disc}$ scales as $1/\sqrt{m_q}$.

5 Continuum Limit

The ultimate goal of any lattice study of the QCD EoS is to perform a reliable continuum limit. For a given temperature $T = 1/(aN_\tau)$ the lattice spacing a is sent to zero keeping the temperature fixed, i.e. the temporal lattice spacing N_τ goes to infinity. To reduce cut-off effects and hence improve the reliability of a continuum limit we are using an improved lattice framework in our calculation which is $\mathcal{O}(a^2)$ improved. We emphasize that our lattice studies for $N_\tau = 4$ show sizable cut-off effects, most prominent in the transition region. It is therefore necessary in order to perform a reliable continuum limit to include $N_\tau = 12$ data. These calculations are part of an ongoing project at the Jülich Supercomputing Centre (NIC).

6 Summary

We performed lattice QCD calculations for $2 + 1$ flavours at an almost physical pion mass of about $160\mathrm{MeV}$. We compare two data sets with quark masses $m_q = 0.05m_s$ and $m_q = 0.1m_s$ at fixed m_s. For the EoS we observe a total shift towards smaller temperatures of about $5\ \mathrm{MeV}$ in the transition region in the trace anomaly compared to earlier results for $m_q = 0.1m_s$. Comparing to the Hadron Resonance Gas model (HRG) we find that at larger temperatures the results for $m_q = 0.05m_s$ are more consistent with HRG model while there are deviations at lower temperatures.

We have studied the behaviour of $\Delta_{l,s}$ and the renormalized Polyakov loop for our two light quark masses. A rapid change in the vicinity of the critical temperature is observed. Furthermore, we have presented results for the disconnected light chiral susceptibility. In the transition region $\chi_{l,disc}$ shows a broad peak for the different light quark masses which rises as $1/\sqrt{m_q}$. This gives support for a picture where the Goldstone modes are causing the fluctuations below T_c.

Acknowledgements

We are grateful to the Jülich Supercomputing Centre, LLNL, NNSA, and New York Center for Computational Sciences for providing access to the computational resources.

References

1. Y. Aoki, Z. Fodor, S. D. Katz and K. K. Szabo, JHEP **0601**, 089 (2006) [arXiv:hep-lat/0510084].
2. C. Bernard *et al.*, Phys. Rev. D **75**, 094505 (2007).

3. M. Cheng *et al.*, "The QCD Equation of State with almost Physical Quark Masses," Phys. Rev. D **77**, 014511 (2008) [arXiv:0710.0354 [hep-lat]].
4. A. Bazavov *et al.*, Phys. Rev. D **80**, 014504 (2009).
5. W. Söldner, "Quark Mass Dependence of the QCD Equation of State on $N_\tau = 8$ Lattices," arXiv:0810.2468 [hep-lat].
6. M. Cheng *et al.*, "Equation of State for physical quark masses," arXiv:0911.2215 [hep-lat].
7. A. Gray *et al.*, Phys. Rev. **D72**, 094507 (2005).
8. P. Braun-Munzinger, K. Redlich and J. Stachel, arXiv:nucl-th/0304013; A. Andronic, P. Braun-Munzinger and J. Stachel, Nucl. Phys. A **772**, 167 (2006) [arXiv:nucl-th/0511071].
9. R. Gupta, "The EOS from simulations on Blue Gene/L Supercomputer at LLNL and NYBlue", PoS **LAT2008** 170 (2008).
10. C. DeTar, "Recent Progress in Lattice QCD Thermodynamics", PoS **LAT2008** (2008).
11. F. Karsch, "Recent lattice results on finite temperature and density QCD", PoS **LAT2007** 015 (2007).
12. C. DeTar, R. Gupta [HotQCD Collaboration], "Towards a precise determination of Tc with 2+1 flavors of quarks", PoS **LAT2007** 179 (2007) [arXiv:hep-lat/0710.1655].
13. M. Cheng *et al.*, Phys. Rev. D **74**, 054507 (2006) [arXiv:hep-lat/0608013].

Numerical Studies of Structures and Phases in (Nano-) Systems in Reduced Geometry

Daniel Mutter, Christine Schieback, Jörg Neder, Florian Bürzle, Kerstin Franzrahe, Annette Geng, and Peter Nielaba

Physics Department, University of Konstanz
78457 Konstanz, Germany
E-mail: peter.nielaba@uni-konstanz.de

In the statistical mechanics of condensed matter phenomena, computer simulations[1] have become more and more important in the last years. Nano-systems in confined geometry contain about 10-10^7 particles, which marks them as nearly ideal systems for the investigation by computer simulation methods. Our group has been able to contribute many interesting results, which have been obtained to a large extend in our project HMZ07, using computer time from the NIC.

1 Simulations of Structural Phase Transitions in Fe, FeNi and NiTi

The approach of our work is to analyze, whether the shape-memory effect and the underlying martensitic phase transformation is suited for the purpose of designing and constructing functional switching devices at the nano-scale. This kind of phase transition occurs between different crystal structures at certain temperatures or external stresses, and it is observed for many alloys, for example FeNi, NiTi or TiV, but also for pure iron, which shows the "α-γ-transition" at 1184 K between a low-temperature bcc (called martensite) and a high-temperature fcc (austenite) structure.
Here we show results[2] of molecular-dynamics (MD) simulations of iron, iron-nickel-, and nickel-titanium-alloys with a number of particles between 1000 and 10000, which were carried out under different boundary conditions (bc). While the temperature was varied stepwise, the structural changes were analyzed by the use of a bond-orientational order-parameter[3] (BOP) (Fe, FeNi), or by the deformation of the simulation box and the calculation of the Gibbs free energy (NiTi). The MD calculations were carried out with a Velocity-Verlet-algorithm, and the temperature was imposed by a Nose-Hoover thermostat. Model potentials from the literature were used (Fe, FeNi, NiTi)[4,5]. In Figure 1, one can observe that there is a hysteresis between the martensite- and the austenite-transition, which is more pronounced for bigger systems, a visualization of the system at different steps during heating is shown if Figure 2. In another study FeNi-alloy systems were simulated under periodic boundary conditions. The effect of the nickel concentration and the number of particles on the austenitic transition temperature τ_a was analyzed, see Figure 3 (left). Higher τ_a than in the case of free bc are observed, but the tendency of increasing τ_a with higher N holds for periodic bc, too. One can further see that τ_a decreases with increasing nickel concentration. The error-bars result from averaging over 3 random Ni distributions. While Fe and FeNi-alloys undergo a structural transition from bcc to fcc during heating, in NiTi-alloys a change from a low-temperature B19' to a high-temperature B2 phase occurs. NiTi consists of 50% nickel and 50% titanium in an ordered structure (see inset Figure 3 (right)). With a slight modification of the potential proposed in Ref. 5, a good agreement of

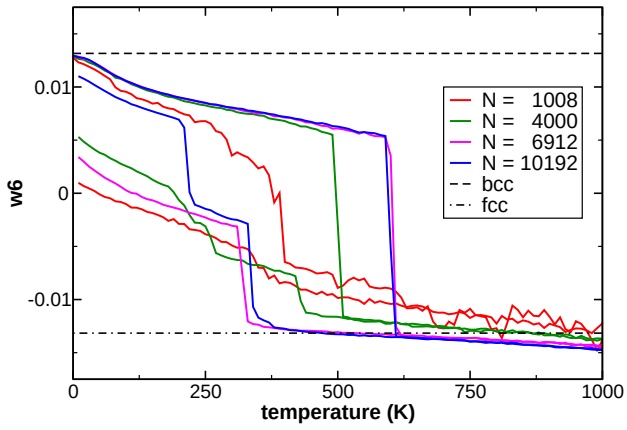


Figure 1. Fe in a initial bcc structure and free bc. Arithmetic means of the BOP w6 of all particles during a temperature cycle for different system sizes.

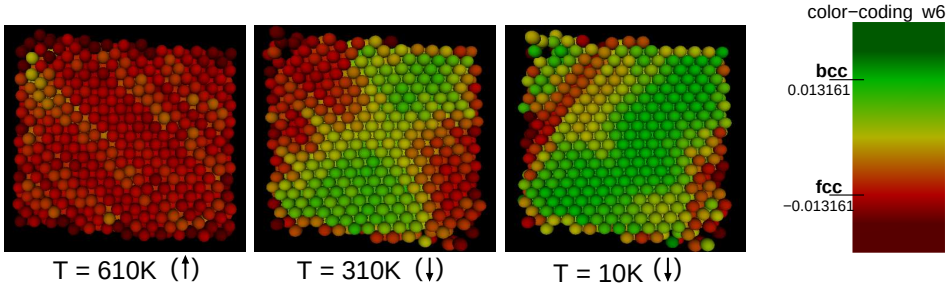


Figure 2. Visualization of the N=6912 system (Fe) at different steps during heating (↑) and cooling (↓).

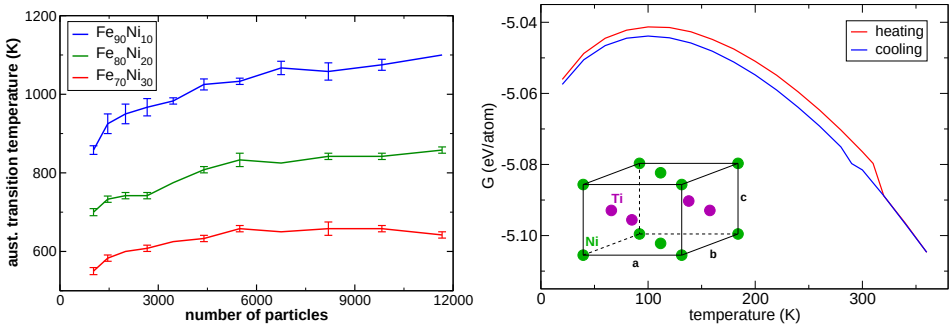


Figure 3. **Left:** Austenite transition temperature in FeNi alloys as function of the number of particles for various Fe concentrations. **Right:** Gibbs free energy during a temperature cycle for NiTi with 2048 atoms and periodic bc. Inset: B2 lattice structure.

		a [$\mathring{A}$]	b [$\mathring{A}$]	c [$\mathring{A}$]	α
B19'	this work	4.460	4.029	3.005	97.8°
	exp.[6]	4.646	4.108	2.898	97.8°
B2	this work	4.255	4.255	3.012	90.0°
	exp.[7]	4.264	4.264	3.015	90.0°

Table 1. NiTi lattice parameters for B19' and B2 structures. α denotes the angle between the a- and the c-axis.

the lattice parameters and the shear angle for both B19' and B2 structure with experiments could be achieved (see Table 1).

The B19' phase can be stabilized at low temperatures. During the heating process, the crystal undergoes the transformation to B2 at about 310 K, which can be seen by calculating the Gibbs free energy at different temperatures (Figure 3 (right)). A drawback of the model potential is the fact, that the reverse B2→B19' transition cannot be observed correctly in the simulation. Instead, a transition to another crystal structure with a slightly lower Gibbs free energy occurs, which could not be identified yet. Therefore, simulations with more sophisticated model potentials will be done in the future to reproduce experimental and *ab initio* studies in a better way.

2 Spintronics Simulations

Nanostructures in an external magnetic field as well as under the influence of spin-polarized currents have become interesting research fields in recent years. A particularly conspicuous example is the motion of domain walls (DWs). The nanoscale spin structure of head-to-head DWs in strips, wires, and rings was observed to be strongly dependent on the width and thickness of the element and plays a crucial role in determining the DW velocity in current-induced DW motion[8,9]. Furthermore, a discrepancy between room temperature measurements and 0 K calculations of the DW velocity exists[10].

To address this problem computer simulations of a classical spin model were performed[11,13–15]. Thermal activation of the system was taken into account by numerically solving the stochastic Landau-Lifshitz-Gilbert equation. Additionally spin torque effects can be modeled by further inclusion of an adiabatic and a non-adiabatic torque term[11].

At finite temperatures we compute the behaviour of DWs in a 1D chain when currents are injected along the x-direction using adiabatic and non-adiabatic spin torque terms. Very different behaviour is observed depending on the value β of the non-adiabatic term. Theoretically four different cases can be distinguished: The so-called adiabatic spin torque effect with $\beta = 0.0$, the non-adiabatic spin torque effect with $\beta < \alpha$ (damping constant α), the non-adiabatic spin torque effect with $\beta > \alpha$, and the non-adiabatic case $\alpha = \beta$ which was not considered in these studies.

For the adiabatic spin torque case ($\beta = 0$), in Figure 4 (a) simulation results of the time dependent DW position x_0 is monitored for different applied spin currents and temperatures. In Figure 4 (b) simulation results (symbols) for the long term DW velocity $\langle v \rangle$ as function of u_x (u_x is proportional to the applied spin current) are shown. The dashed line shows a fit to analytical predictions[12]. Open symbols denote that the DW rotates around the x-axis. Simulation results for finite temperatures ($k_\mathrm{B}T/J \neq 0$) and $0\,\mathrm{K}$[11] are shown. The

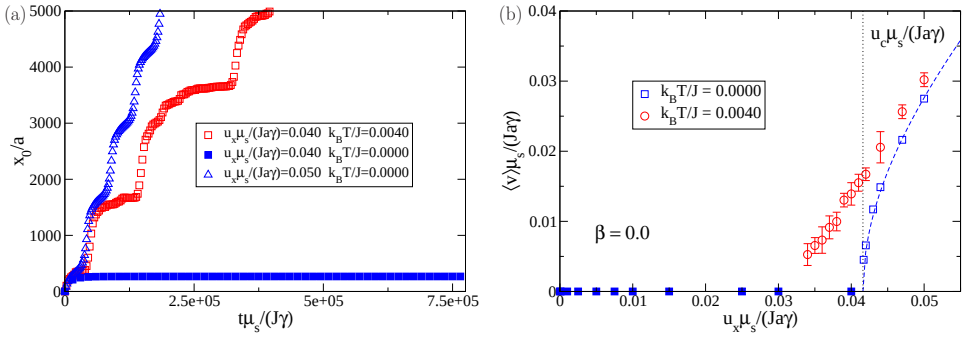


Figure 4. Simulation results (symbols) are shown for (a) the time dependent domain wall position x_0, and (b) the long term domain wall velocity $\langle v \rangle$ as function of u_x (u_x is proportional to the applied current) for the adiabatic spin torque. Open symbols denote that the domain wall precesses along the wire axis. Simulation results for finite temperatures ($k_B T/J \neq 0$) as well as 0 K are shown. The dashed lines show fits to analytical predictions[12].

simulations are performed with ferromagnetic exchange coupling and anisotropy constants $d_x/J = 0.01$ and $d_y/J = 0.005$ and Gilbert damping constant $\alpha = 0.02$.

For the 0 K adiabatic case a critical spin current u_c exists. No long term DW velocity can be observed for spin currents $u_x < u_c$. For $u_x > u_c$ the DW velocity is given by[12] $\langle v \rangle = \sqrt{u_x^2 - u_c^2}/(1+\alpha^2)$ with $u_c \mu_s/(Ja\gamma) = 0.0416$. For finite temperatures $\langle v \rangle \neq 0.0$ for $u_x < u_c$ in contrast to results at 0 K. The DW velocity $\langle v \rangle$ increases with increasing temperature. For the non-adiabatic case with $\beta < \alpha$ for all spin currents a finite DW velocity can be observed at $T = 0.0$ K. The long term DW velocity shows a linear dependence of u_x for $u_x < u_{\mathrm{Walker}}$:[12] $\langle v \rangle = u_x \beta/\alpha$. For $u_x > u_{\mathrm{Walker}}$ the DW velocity $\langle v \rangle$ increases stronger than linearly with increasing u_x. For finite temperatures the DW velocity increases with increasing temperature. For the non-adiabatic case $\beta > \alpha$ at $T = 0.0$ K the long term DW velocity shows a linear dependence of u_x for $u_x < u_{\mathrm{Walker}}$:[12] $\langle v \rangle = u_x \beta/\alpha$. For $u_x > u_{\mathrm{Walker}}$ $\langle v \rangle$ first decreases and later increase in contrast to the adiabatic case and the non-adiabatic case $\beta < \alpha$. For finite temperatures $\langle v \rangle$ decreases with increasing temperature.

However, 1D models including temperature effects are still not enough to describe the experimental findings of DW motion in Permalloy nanowires[10]. The influence of the exact spin structure has to be taken into account which is quite complicated in these systems[10]. Simulations of higher dimensional systems with temperature effects are thus necessary.

3 Lipid Bilayers under Tension

Lipid bilayers and incorporated proteins form biological membranes. Usually the microscopic surface tension of membranes is small or vanishes altogether, but there may also be situations, where membranes are under considerable stress due to osmotic pressure differences or during conformational changes, fission or fusion of cells. Epithelial cells, e.g. exposed to transmembrane osmotic gradients, can be expected to develop a state of tension under physiological conditions[16]. Therefore, the aim of our work is to study the effect of an applied tension on the physical properties of a model bilayer. Using a generic coarse-grained model[17] we analyze model bilayers systematically ranging from simulation temperatures, which correspond to the gel-phase of the bilayer, up to temperatures, where the

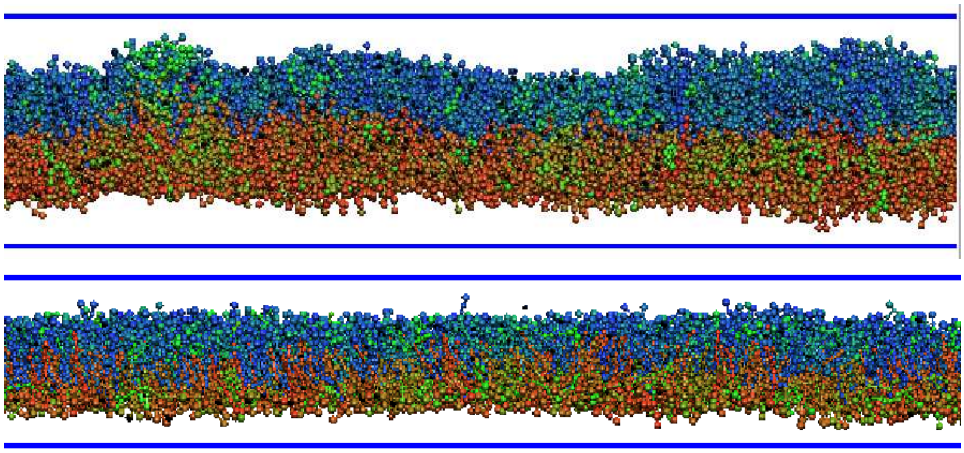


Figure 5. Snapshots of a tensionless bilayer (top) and a bilayer under stress of $\Gamma = 2.0\,\epsilon/\sigma_t^2$ (bottom) in the fluid phase. To visualize the interdigitation of the two monolayers as the tension increases they are coloured in blue and red, respectively.

system is within the fluid phase. The focus of our work lies on conformational variations and changes in mechanical properties of the bilayers when external tension is applied[18,19]. In the fluid phase, for example, an interdigitation of the lipids takes place as the tension increases. This, in turn, influences not only properties like the thickness of the bilayer and the average area per lipid but also affects the fluctuation spectra considerably. It is already apparent by looking at the snapshots in Figure 5 that both fluctuations in height and peristaltic modes of the bilayer are suppressed in systems under tension. A fact that can be quantitatively confirmed by analyzing the fluctuation spectra of the model membranes.

4 Binary Mixtures in External Fields

Minimisation trends in physics and technology have caused a lot of interest in monolayers and their interactions with a substrate lately. Colloidal suspensions have proven to be ideal model systems for studies on such systems. Elastic properties of two-dimensional colloidal crystals have been analysed recently[20], as well as non-equilibrium phenomena of colloidal systems in external gravitational fields[21]. Monodisperse 2D colloidal systems in interaction with a substrate potential have been studied extensively in experiments[22,23], computer simulations[24–30] and theory[31–33] over the last decades. Reentrant phase transition scenarios like *Laser Induced Freezing* (LIF) and *Laser Induced Melting* (LIM) have been observed. Here we address the question of how the addition of another length scale into such a system influences the intricate competition between adsorbate-adsorbate and adsorbate-substrate interaction by studying a binary 50% mixture under the influence of a 1D spatially periodic substrate potential. The colloids are modelled by hard disks. Following the approaches in the monodisperse studies, we use the following external potential: $V(\vec{r}) = V_0 \sin(\vec{K} \cdot \vec{r})$ with $\vec{K} = (4\pi/a, 0)$ and a the lattice parameter of the $S_1(AB)$ lattice. Thus the wavelength $\lambda = 2\pi/|\vec{K}|$ of the external potential is commensurate to the

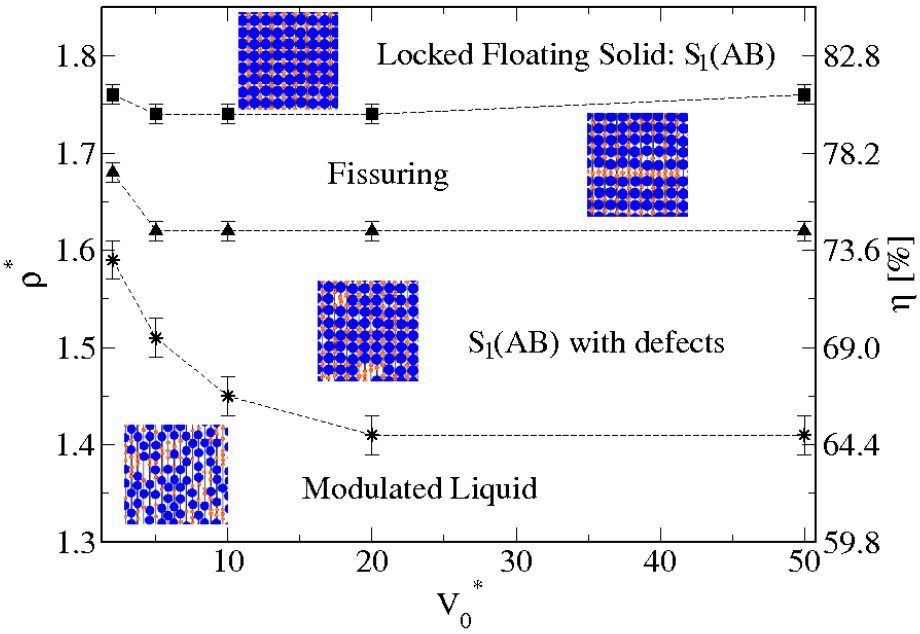


Figure 6. The phase diagram for an equimolar binary mixture with diameter ratio $\sigma_B/\sigma_A = 0.414$ exposed to an external periodic potential, commensurate to the $S_1(AB)$ square lattice for the case that both components of the mixture interact with the external potential[37]. The fissuring regime and the $S_1(AB)$ lattice with defects are stabilised down to far lower packing fractions η as compared to case (I)(see Ref. 35). Lines are a guide to the eye.

periodicity of the $S_1(AB)$ lattice. MC-Simulations in the NVT ensemble with periodic boundary conditions were carried out[34,13]. Three cases can be distinguished: (I) only the smaller component interacts with the external field[35], (II) both components interact with the external field, (III) only the larger component interacts with the external field.

For low amplitudes V_0 we found a novel phenomenon. In contrast to the monodisperse LIF scenario a Laser Induced De-mixing[35] sets in for all three cases. We observe the coexistence of a smaller component enriched fluid with a droplet of a monodisperse lattice formed by the larger component. The ordering mechanisms at work and the resulting monodisperse lattice structures differ for the three cases[34,13,36]. The ordering mechanisms stem from the attempt of the system to minimise its energy via an alignment of the components interacting with the external potential with the potential minima and the constraint of a fixed overall number density. While the coupling of only the smaller component to the external potential (case (I)) results in the coexistence of a monodisperse triangular lattice, the cases (II) and (III), where the larger component interacts directly with the external potential, show a condensation of a monodisperse rhombic lattice.

In the studied binary mixture the laser induced melting transition, as it occurs in monodisperse systems, is geometrically blocked. This leads to an enhanced stabilisation of the locked floating solid, a $S_1(AB)$ crystal. The phase diagram for the case that both components interact with the external field (case (II)) is shown in Figure 6[37].

Here an increase in the amplitude of the external potential at constant density leads to a competition between rhombic, monodisperse domains of the large component and square lattice domains. As the rhombic domains are directly stabilised by the external potential, these structures hamper the growth of the square lattice domain. The resulting lattice structure is a $S_1(AB)$ square lattice with frozen in disorder. A comparison of the phase diagrams for case(I)[35] and (II) shows that both phase boundaries, the one from the fissuring regime to the $S_1(A)$ lattice with frozen in disorder and the one to the Modulated Liquid, are shifted to considerably lower packing fractions for the case that both components of the mixture couple to the external potential[37]. This is due to the direct stabilisation by the external field of the occurring ordered phases.

5 Atomic-Sized Contacts and Si_n Clusters in External Fields

Metallic nanowires fabricated by means of scanning-tunneling microscope and break-junction techniques have turned out to be a unique playground to test basic concepts of electronic transport at the atomic scale[38]. We have studied[39–41,13] several types of nano-contacts (Au, Ag, Fe, Co, Ni) by combining Molecular Dynamics methods at non-zero temperature, tight binding parameters and electron transport methods (Keldysch formalism). We analyzed the effect of structural re-arrangements and of magnetic effects on the conductance evolution under stress and on conduction histograms. In several systems good agreement of our computations with experimental observations has been achieved.

Experimental results show that Si_4 and Si_7 and Si_{10} are possible candidates for a cluster material. The question, if Si_n clusters can be used as building blocks for cluster material was addressed by computing the binding energy of two clusters as a function of distance. For the simulations[42–45] the implementation of DFT available at Ref. 46 has been used. According to these computations[44], Si_7 seems to be a possible candidate for a cluster material building block. The effect of additional external fields on the properties of magic clusters has been investigated[45] for the case of Si_4 clusters. These results indicate that fusion barriers can be induced by external potentials, and the fields eventually can be used as structuring devices in future experiments.

6 Magnetic Fields in Star Formation and in Galaxies

This research project aims at simulating stars, stellar clusters and galaxies as well as modelling the insterstellar medium. Especially interesting is the role magnetic fields play in molecular clouds prior to as well as during the gravitational collapse. In addition, magnetic fields pervade probably all galaxies in the universe. Observational evidence showed that the magnetic field structures in most (spiral, flocculent and elliptical) galaxies offer spiral patterns. The presence of the magnetic fields are thought to obtain a great influence on the structure as well as on the dynamics of matter, gas and dust within the galaxies. We plan to study the influence of intergalactic magnetic fields on the structure of colliding and merging galaxies on time scales of few Gyr and to analyze the effect of magnetic fields on the galaxy evolution process.

Our method of choice to attack the outlined problems is the particle method SPH (Smoothed Particle Hydrodynamics)[47]. Within this project, the SPH method is utilized

using the code GADGET-2[48] which was developed for use on massively parallel computers with distributed memory. Our present version of the code includes additional, state of the art modifications, making it more suitable for star formation studies (courtesy of Ralf Klessen's star formation group, University of Heidelberg). Our implementation[13,49,50] of SPHMHD follows the version outlined in the paper by Dolag & Stasyszyn[51], although other algorithmic solutions are currently being implemented and successfully tested.

The knowledge concerning the numerical treatment of hydrodynamics and magnetism by SPHMHD at a later stage of the project shall be transferred to the investigation of hydrodynamics effects in (para-) magnetic colloidal systems.

Acknowledgements

We thank the SFB767 and SFBTR6 for support and the NIC for computer time.

References

1. *Bridging Time Scales: Molecular Simulations for the Next Decade*, edited by P. Nielaba, M. Mareschal, G. Ciccotti, Springer, Berlin (2002).
2. D. Mutter, Dissertation, U. Konstanz (in preparation).
3. P. J. Steinhardt *et al.*, Phys. Rev. B **28**, 784 (1983).
4. R. Meyer and P. Entel, Phys. Rev. B **57**, 5140 (1998).
5. W. S. Lai and B. X. Liu., J. Phys.: Cond. Mat. **12**, L53-L60 (2000).
6. X. Huang *et al.*, Nature Materials **2**, (2003).
7. E. Goo and R. Sinclair, Acta Metall. **33**, 1717 (1985).
8. M. Kläui *et al.*, APL **87** 102509 (2005); Phys. Rev. Lett. **95** 026601 (2005).
9. M. Kläui, J. Phys.: Cond. Mat. **20**, 313001 (2008).
10. M. Laufenberg, Phys. Rev. Lett. **97**, 046602 (2006).
11. C. Schieback, M. Kläui, U. Nowak, U. Rüdiger, P. Nielaba, Eur. Phys. J. B**58**, 429 (2007).
12. A. Thiaville *et al.*, EPL **69**, 990 (2005).
13. C. Schieback, F. Bürzle, K. Franzrahe, J. Neder, M. Dreher, P. Henseler, D. Mutter, N. Schwierz, P. Nielaba, in: *High Performance Computing in Science and Engineering'08*, edited by W.E. Nagel, D. Kröner, M.M. Resch, Springer Verlag (2009), 41.
14. C. Schieback, Dissertation, U. Konstanz (in preparation).
15. D. Backes, C. Schieback, et al., Appl. Phys. Lett. **91**, 112502 (2007).
16. G. Soveral, R. I. Macey, and T. F. Moura, Biology of the Cell **89**, 275 (1997).
17. O. Lenz, F. Schmid, J. Mol. Liq. **117**, 147 (2005); Phys. Rev. Lett. **98**, 058104 (2007).
18. J. Neder, Dissertation, U. Konstanz (in preparation).
19. J. Neder, B. West, P. Nielaba, F.Schmid, preprint.
20. K. Franzrahe, P. Keim, G. Maret, P. Nielaba, S. Sengupta, Phys. Rev. **E78**, 026106 (2008).
21. M. Köppl, P. Henseler, A. Erbe, P. Nielaba, P. Leiderer, Phys. Rev. Lett. **97**, 208302 (2006).
22. Chowdhury *et al.*, Phys. Rev. Lett. **55**, 833 (1985).

23. Q.-H. Wei, C. Bechinger, D. Rudhardt, P. Leiderer, Phys. Rev. Lett. **81**, 2606 (1998).
24. J. Chakrabarti, H.R. Krishnamurthy, A.K. Sood, S. Sengupta, Phys. Rev. Lett. **75** 2232 (1995).
25. W. Strepp, S. Sengupta, P. Nielaba, Phys. Rev. **E63**, 046106 (2001); Phys. Rev. **E66**, 056109(2002).
26. W. Strepp, S. Sengupta, M. Lohrer, P. Nielaba, Comp. Phys. Commun. **147**, 370 (2002).
27. P. Chaudhuri et al., Phys. Rev. **E 72**, 061404 (2005).
28. F. Bürzle, P. Nielaba, Phys. Rev. **E76**, 051112 (2007).
29. K. Franzrahe *et. al.*, Comp. Phys. Commun. **169**, pp. 197-202 (2005).
30. P. Nielaba, W. Strepp, in: *Path Integrals - New Trends and Perspectives*, edited by W. Janke and A. Pelster, World Scientific, Singapore, pp. 321 (2008).
31. J. Chakrabarti, H.K. Krishnamurthy, A.K. Sood, Phys. Rev. Lett. **73**, 2923 (1994).
32. E. Frey, D.R. Nelson, L. Radzihovsky, Phys. Rev. Lett. **83**, 2977 (1999).
33. L. Radzihovsky, E. Frey, D.R. Nelson, Phys. Rev. E **63**, 031503 (2001).
34. K. Franzrahe, P. Nielaba *et al.*, J. Phys.: Condens. Matter **20**, 404218 (2008).
35. K. Franzrahe and P. Nielaba, Phys. Rev. E **76**, 061503 (2007).
36. K. Franzrahe, Dissertation, U. Konstanz (2008).
37. K. Franzrahe, P. Nielaba, Phys. Rev. **E79**, 051505 (2009).
38. N. Agraït, A. Levy Yeyati, and J.M. van Ruitenbeek, Phys. Rep. **377**, 81 (2003).
39. M. Dreher, F. Pauly, J. Heurich, J.C. Cuevas, E. Scheer, P. Nielaba, Phys. Rev. Lett. **72**, 075435 (2005).
40. F. Pauly, M. Dreher, J.K. Viljas, M. Häfner, J.C. Cuevas, P. Nielaba, Phys. Rev. Lett. **74**, 235106 (2006).
41. M. Häfner, J.K. Viljas, D. Frustaglia, F. Pauly, M. Dreher, P. Nielaba, J.C. Cuevas, Phys. Rev. Lett. **77**, 104409 (2008).
42. M. Grass, D. Fischer, M. Mathes, G. Ganteför, P. Nielaba, Appl. Phys. Lett. **81**, 3810 (2002).
43. M. Dreher, et al., Phase Transitions **78**, 751-772 (2005).
44. F. von Gynz-Rekowski, W. Quester, R. Dietsche, Dong Chan Lim, N. Bertram, T. Fischer, G. Ganteför, M. Schach, P. Nielaba, Y.D. Kim, Eur. Phys. J. **D45**,409 (2007).
45. M. Schach, Diplomarbeit, U. Konstanz (2007).
46. CPMD. Copyright IBM Corp 1990–2001, Copyright MPI für Festkörperforschung Stuttgart 1997–2004. `http://www.cpmd.org/`.
47. J. J. Monaghan, Rep. Prog. Phys. **68**, 1703 (2005).
48. V. Springel, MNRAS **364**, 1105 (2005).
49. F. Bürzle, Dissertation, U. Konstanz (in preparation).
50. A. Geng, Dissertation, U. Konstanz (in preparation).
51. K. Dolag, F. Stasyszyn, MNRAS **398**, 1678 (2009).

Towards Precision B-Physics from Non-Perturbative Heavy Quark Effective Theory

ALPHA Collaboration
Michele Della Morte[1], Jochen Heitger[2], and Rainer Sommer[3]

[1] Johannes Gutenberg Universität Mainz, Institut für Kernphysik
Johann-Joachim-Becher Weg 45, D-55099 Mainz, Germany
E-mail: morte@kph.uni-mainz.de

[2] Westfälische Wilhelms-Universität Münster, Institut für Theoretische Physik
Wilhelm-Klemm-Straße 9, D-48149 Münster, Germany
E-mail: heitger@uni-muenster.de

[3] NIC, DESY, Platanenallee 6, D-15738 Zeuthen, Germany

We convey an idea of the significant recent progress, which opens up good perspectives for high-precision *ab initio* computations in heavy flavour physics based on lattice QCD. Rather than surveying the latest results, this contribution focuses on the concept and the challenges of fully non-perturbative computations in the B-meson sector, where the b-quark is treated within an effective theory. We outline its use to determine the b-quark mass and report on the results obtained in the quenched approximation and on the status in the two dynamical flavour theory.

1 B-Physics and Lattice QCD

The plenty of beautiful results from recent and still ongoing B-physics experiments[1,2], which require the knowledge of QCD matrix elements for their interpretation in terms of parameters of the Standard Model and its possible extensions, motivates investigations in lattice QCD. The importance of this interplay of experiment and theory is further expressed by the fact that one of its main objectives, the phenomenon of CP violation, is closely related to the symmetry breaking mechanism the 2008's Nobel Prize was dedicated to.

Lattice QCD represents our best founded theoretical formulation of QCD and allows for the computation of low-energy hadronic properties in the non-perturbative domain, where the usual power series expansion in the coupling constant fails, through the Monte Carlo evaluation of the Euclidean path integral after a discretization of space-time on a lattice with spacing a in all $3+1$ dimensions[3,4]. While such numerical computations necessarily involve approximations, one of the key features of the lattice approach is that all approximations can be systematically improved. For an overview of results from the field of heavy flavour physics, which reflect some of these improvements by the small error bars quoted for many quantities, we refer to the reviews of past Lattice Conferences[5–7].

1.1 Challenges

Among the various considerable challenges one faces in an actual lattice QCD calculation on the theoretical and technical levels, let us only highlight the multi-scale problem, which is also particularly relevant in view of B-physics applications. This is illustrated in Figure 1. There are many disparate physical scales to be covered simultaneously, ranging

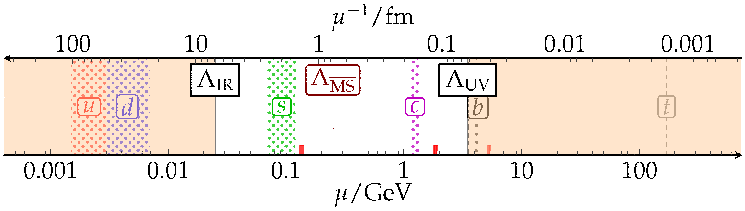


Figure 1. Large range of energy (μ) scales in lattice QCD, where shaded areas refer to quark mass values (in the $\overline{\mathrm{MS}}$ scheme) quoted by the Particle Data Group[8]. Red marks indicate the pion, the D- and the B-meson mass.

from the lightest hadron mass of $m_\pi \approx 140\,\mathrm{MeV}$ over $m_\mathrm{D} \approx 2\,\mathrm{GeV}$ to $m_\mathrm{B} \approx 5\,\mathrm{GeV}$, plus the ultraviolet cutoff of $\Lambda_\mathrm{UV} = a^{-1}$ of the lattice discretization that has to be large compared to all physical energy scales for the discretized theory to be an approximation to the continuum one. Moreover, the finiteness of the linear extent of space-time, L, in a numerical treatment entails an infrared cutoff $\Lambda_\mathrm{IR} = L^{-1}$ so that the following scale hierarchy is met:

$$\Lambda_\mathrm{IR} = L^{-1} \quad \ll \quad m_\pi\,, \ldots, m_\mathrm{D}\,, m_\mathrm{B} \quad \ll \quad a^{-1} = \Lambda_\mathrm{UV}\,. \tag{1}$$

This implies $L \gtrsim 4/m_\pi \approx 6\,\mathrm{fm}$ to suppress finite-size effects in the light quark sector and $a \lesssim 1/(2m_\mathrm{D}) \approx 0.05\,\mathrm{fm}$ to still properly resolve the propagation of a c-quark in the heavy sector. Lattices with $L/a \gtrsim 120$ sites in each direction would thus be needed to satisfy these constraints, and since the scale of hadrons with b-quarks was not even included to arrive at this figure, it is obvious that the b-quark mass scale has to be separated from the others in a theoretically sound way before simulating the theory. In section 2 we describe, how this is achieved by recoursing to an effective theory for the b-quark.

Another non-trivial task is the renormalization of QCD operators composed of quark and gluon fields, which appear in the effective weak Hamiltonian, valid at energies far below the electroweak scale. Besides perturbation theory[9], powerful non-perturbative approaches have been developed[3,4], and we will come back to the non-perturbative subtraction of power-law divergences in the context of the effective theory for the b-quark later.

1.2 Perspectives

As for the challenges with light quarks, we only mention that the condition $L \gtrsim 6\,\mathrm{fm}$ may be relaxed by simulating at unphysically large pion masses, combined with a subsequent extrapolation guided by chiral perturbation theory[10] and its lattice-specific refinements.

Regarding the algorithmic side of a lattice QCD simulation, the Hybrid Monte Carlo[11] (HMC) as the first exact and still state-of-the-art algorithm has received considerable improvements by multiple time-scale integration schemes[12,13], the Hasenbusch trick of mass-preconditioning[14,15], supplemented by a sensible tuning of the algorithm's parameters[16], and the method of domain decomposition (DD) applied to QCD[17–19], just to name a few. In addition, low-mode deflation[20] (together with chronological inverters[21]) has led to a substantial reduction of the critical slowing down with the quark mass in the DD-HMC.

Finally, in parallel to the continuous increase of computer speed (at an exponential rate) over the last 25 years and the recent investments into high performance computing at many

places of the world, the Coordinated Lattice Simulations[22] (CLS) initiative is a community effort to bring together the human and computer resources of several teams in Europe interested in lattice QCD. The present goal are large-volume simulations with $N_{\rm f} = 2$ dynamical quarks, using the rather simple $\mathrm{O}(a)$ improved Wilson action[4] to profit from the above algorithmic developments such as DD-HMC, and lattice spacings $a = 0.08, 0.06, 0.04\,\mathrm{fm}$, sizes $L = (2-4)\,\mathrm{fm}$ and pion masses down to $m_\pi = 200\,\mathrm{MeV}$, which altogether help to diminish systematic and statistical errors. Amongst others, charm physics[23] as well as our B-physics programme outlined here are being investigated.

2 Non-Perturbative Heavy Quark Effective Theory

Heavy Quark Effective Theory (HQET) at zero velocity on the lattice[24] offers a reliable solution to the problem of dealing with the two disparate intrinsic scales encountered in heavy-light systems involving the b-quark, i.e., the lattice spacing a, which has to be much smaller than $1/m_{\rm b}$ to allow for a fine enough resolution of the states in question, and the linear extent L of the lattice volume, which has to be large enough for finite-size effects to be under control (recall also Figure 1).

Since the heavy quark mass ($m_{\rm b}$) is much larger than the other scales such as its 3–momentum or $\Lambda_{\rm QCD} \sim 500\,\mathrm{MeV}$, HQET relies upon a systematic expansion of the QCD action and correlation functions in inverse powers of the heavy quark mass around the static limit ($m_{\rm b} \to \infty$). The lattice HQET action at $\mathrm{O}(1/m_{\rm b})$ reads:

$$S_{\rm HQET} = a^4 \sum_x \overline{\psi}_{\rm h} \left\{ D_0 + \delta m - \omega_{\rm kin} \mathbf{D}^2 - \omega_{\rm spin} \boldsymbol{\sigma} \mathbf{B} \right\} \psi_{\rm h} \,, \tag{2}$$

with $\psi_{\rm h}$ satisfying $P_+ \psi_{\rm h} = \psi_{\rm h}$, $P_+ = \frac{1+\gamma_0}{2}$, and the parameters $\omega_{\rm kin}$ and $\omega_{\rm spin}$ being formally $\mathrm{O}(1/m_{\rm b})$. At leading order (static limit), where the heavy quark acts only as a static colour source and the light quarks are independent of the heavy quark's flavour and spin, the theory is expected to have $\sim 10\%$ precision, while this reduces to $\sim 1\%$ at $\mathrm{O}(1/m_{\rm b})$ representing the interactions due to the motion and the spin of the heavy quark. As crucial advantage (e.g., over NRQCD), HQET treats the $1/m_{\rm b}$–corrections to the static theory as space-time insertions in correlations functions. For correlation functions of some multi-local fields $\mathcal{O}$ and up to $1/m_{\rm b}$–corrections to the operator itself (irrelevant when spectral quantities are considered), this means

$$\langle \mathcal{O} \rangle = \langle \mathcal{O} \rangle_{\rm stat} + a^4 \sum_x \left\{ \omega_{\rm kin} \langle \mathcal{O} \mathcal{O}_{\rm kin}(x) \rangle_{\rm stat} + \omega_{\rm spin} \langle \mathcal{O} \mathcal{O}_{\rm spin}(x) \rangle_{\rm stat} \right\} \,, \tag{3}$$

where $\langle \mathcal{O} \rangle_{\rm stat}$ denotes the expectation value in the static approximation and $\mathcal{O}_{\rm kin}$ and $\mathcal{O}_{\rm spin}$ are given by $\overline{\psi}_{\rm h} \mathbf{D}^2 \psi_{\rm h}$ and $\overline{\psi}_{\rm h} \boldsymbol{\sigma} \mathbf{B} \psi_{\rm h}$. In this way, HQET at a given order is (power-counting) renormalizable and its continuum limit well defined, once the mass counterterm δm and the coefficients $\omega_{\rm kin}$ and $\omega_{\rm spin}$ are fixed non-perturbatively by a matching to QCD.

Still, for lattice HQET and its numerical applications to lead to precise results with controlled systematic errors in practice, two shortcomings had to be left behind first:

1. The exponential growth of the noise-to-signal ratio in static-light correlators, which is overcome by a clever modification of the Eichten-Hill discretization of the static action[25].

2. As in HQET mixings among operators of different dimensions occur, the power-divergent additive mass renormalization $\delta m \sim g_0^2/a$ already affects its leading order.

Unless HQET is renormalized non-perturbatively[26], this divergence — and further ones $\sim g_0^2/a^2$ arising at $\mathrm{O}(1/m_\mathrm{b})$ — imply that the continuum limit does not exist owing to a remainder, which, at any finite perturbative order[27,28], diverges as $a \to 0$. A general solution to this theoretically serious problem was worked out and implemented for a determination of the b-quark's mass in the static and quenched approximations as a test case[29]. It is based on a *non-perturbative matching of HQET and QCD in finite volume*.

3 Application: The b-Quark Mass from HQET at $\mathrm{O}(1/m_\mathrm{b})$

Let us first note[4] that in order not to spoil the asymptotic convergence of the series, the matching must be done non-perturbatively — at least for the leading, static piece — as soon as the $1/m_\mathrm{b}$–corrections are included, since as $m_\mathrm{b} \to \infty$ the *perturbative* truncation error from the matching coefficient of the static term becomes much larger than the power corrections $\sim \Lambda_\mathrm{QCD}/m_\mathrm{b}$ of the HQET expansion.

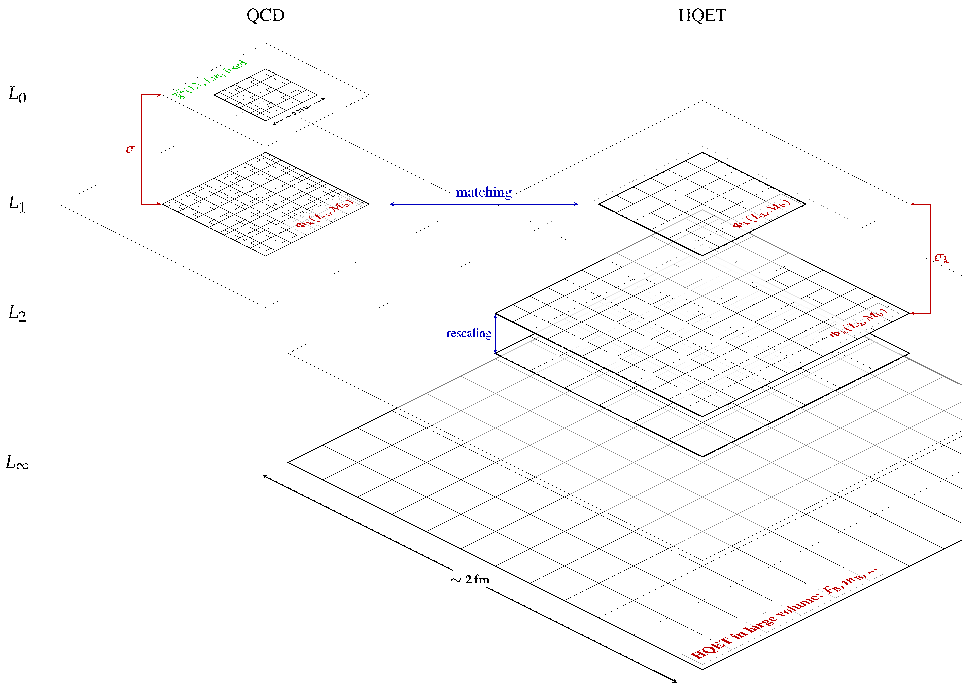


Figure 2. Idea of lattice HQET computations via a non-perturbative determination of HQET parameters from small-volume QCD simulations. For each fixed L_i, the steps are repeated at smaller a to reach a continuum limit.

In the framework introduced in Ref. 29, matching and renormalization are performed simultaneously *and* non-perturbatively. The general strategy, illustrated in Figure 2, can be

explained as follows. Starting from a finite volume with $L_1 \approx 0.5\,\mathrm{fm}$, one chooses lattice spacings a sufficiently smaller than $1/m_\mathrm{b}$ such that the b-quark propagates correctly up to controllable discretization errors of order a^2. The relation between the renormalization group invariant (RGI) and the bare mass in QCD being known, suitable finite-volume observables $\Phi_k(L_1, M_\mathrm{h})$ can be calculated as a function of the RGI heavy quark mass, M_h, and extrapolated to the continuum limit. Next, the power-divergent subtractions are performed non-perturbatively by a set of matching conditions, in which the results obtained for Φ_k are equated to their representation in HQET (r.h.s. of Figure 2). At the same physical value of L_1 but for resolutions $L_1/a = \mathrm{O}(10)$, the previously computed heavy-quark mass dependence of $\Phi_k(L_1, M_\mathrm{h})$ in finite-volume QCD may be exploited to determine the bare parameters of HQET for $a \approx (0.025 - 0.05)\,\mathrm{fm}$. To evolve the HQET observables to large volumes, where contact with some physical input from experiment can be made, one also computes them at these lattice spacings in a larger volume, $L_2 = 2L_1$. The resulting relation between $\Phi_k(L_1)$ and $\Phi_k(L_2)$ is encoded in associated step scaling functions (SSFs) σ_k, indicated in Figure 2. By using the knowledge of $\Phi_k(L_2, M_\mathrm{h})$ one fixes the bare parameters of the effective theory for $a \approx (0.05 - 0.1)\,\mathrm{fm}$ so that a connection to lattice spacings is established, where large-volume observables, such as the B-meson mass or decay constant, can be calculated (bottom of Figure 2). This sequence of steps yields an expression of m_B, the physical input, as a function of M_h via the quark mass dependence of $\Phi_k(L_1, M_\mathrm{h})$, which eventually is inverted to arrive at the desired value of the RGI b-mass within HQET. The whole construction is such that the continuum limit can be taken for all pieces.

3.1 Review of the Quenched Computation of the b-Quark Mass[30]

To apply this to M_b, the task is to fix δm and ω_kin non-perturbatively by performing a matching to QCD, after restricting to spin-averaged quantities to get rid of the contributions proportional to ω_spin. For sensible definitions of the required matching observables, Φ_1 and Φ_2, we work with the Schrödinger functional (SF), i.e., QCD with Dirichlet boundary conditions in time and periodic ones in space (up to a phase θ for the fermions): $\Phi_1^\mathrm{QCD}(L, m_\mathrm{h})$ exploits the sensitivity of SF correlation functions to θ and $\Phi_2^\mathrm{QCD}(L, m_\mathrm{h}) \equiv L\Gamma_1(L, m_\mathrm{h})$, where Γ_1 is a finite-volume effective energy. When expanded in HQET[a], $\Phi_1^\mathrm{HQET}(L)$ is given by ω_kin times a quantity defined in the effective theory (called $R_1^\mathrm{kin}(L, \theta, \theta')$), whereas $\Phi_2^\mathrm{HQET}(L)$ is a function of ω_kin and $m_\mathrm{bare} = \delta m + m_\mathrm{h}$ involving two other HQET quantities, $\Gamma_1^\mathrm{stat}(L)$ and $\Gamma_1^\mathrm{kin}(L)$. According to the strategy sketched above, by equating $\Phi_k^\mathrm{QCD}(L_1, m_\mathrm{h})$ and $\Phi_k^\mathrm{HQET}(L_1)$ one can determine the bare parameters m_bare and ω_kin as functions of m_h at the lattice spacings belonging to the volume L_1^4. To employ the spin-averaged B-meson mass, m_B^av, as phenomenological input, the Φ_k are evolved to larger volumes through proper SSFs, where the resulting $\Phi_k^\mathrm{HQET}(2L_1, m_\mathrm{h})$ still carry the dependence on m_h inherited from the matching to QCD in L_1^4. After 2 evolution steps (and taking continuum limits), linear extents of $\gtrsim 1.5\,\mathrm{fm}$ are reached, and m_bare and ω_kin, expressed in terms of SSFs, $\Phi_k^\mathrm{QCD}(L_1, m_\mathrm{h})$ as well as R_1^kin, Γ_1^stat and Γ_1^kin, are obtained — again as functions of m_h. Now, the b-quark mass is extracted by solving

$$m_\mathrm{B}^\mathrm{av} = E^\mathrm{stat} + \omega_\mathrm{kin}(m_\mathrm{h})E^\mathrm{kin} + m_\mathrm{bare}(m_\mathrm{h}) \tag{4}$$

[a]Here, $\delta m = 0$ in the action; its effect is accounted for in the overall energy shift m_bare in HQET versus QCD.

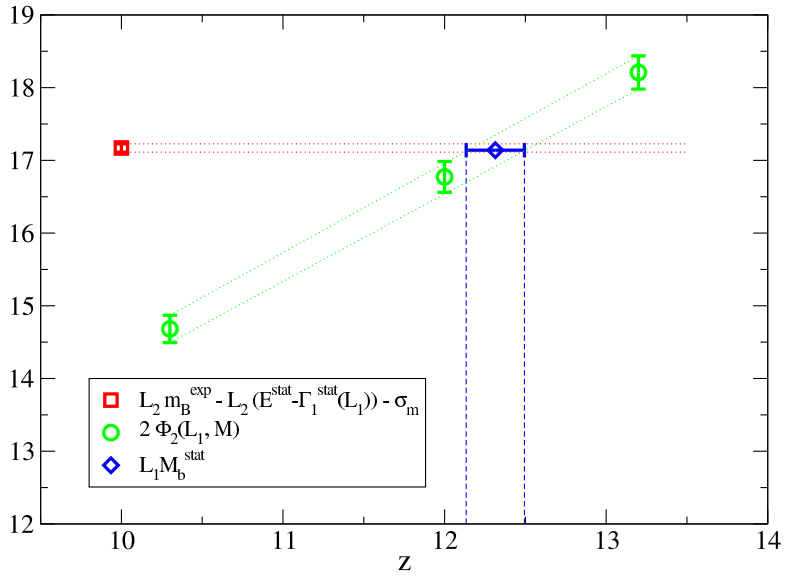


Figure 3. Graphical solution of Eq. 5 in the quenched case[30]. The quantity used in the finite-volume matching step is $\Phi_2^{\rm QCD}(L_1, M) = L_1\Gamma_1(L_1, M)$, where $z = L_1 M$ and $M \equiv M_{\rm h}$ the RGI heavy quark mass.

for $m_{\rm h}$, with $E^{\rm stat} = \lim_{L\to\infty}\Gamma_1^{\rm stat}$ and $E^{\rm kin} = -\langle\,{\rm B}\,|\,a^3\sum_{\boldsymbol{z}}\mathcal{O}_{\rm kin}(0,\boldsymbol{z})\,|\,{\rm B}\,\rangle_{\rm stat}$. All quantities entering Eq. 4 have a continuum limit either in QCD or HQET, which implies that all power divergences have been subtracted non-perturbatively.

In case of the leading-order, static approximation, where only $m_{\rm bare}$ needs to be determined, the small- and large-volume matching conditions simplify to $\Gamma_1(L_1, m_{\rm h}) = \Gamma_1^{\rm stat}(L_1) + m_{\rm bare}$ and $m_{\rm B}^{\rm av} = E^{\rm stat} + m_{\rm bare}$, respectively. To be able to solve the first equation for $m_{\rm bare}$ and replace it in the second, we bridge the volume gap in two steps by inserting a SSF $\sigma_{\rm m}(L_1) = 2L_1[\Gamma_1^{\rm stat}(2L_1) - \Gamma_1^{\rm stat}(L_1)]$ and arrive at the *master equation*

$$L_1\left[\,m_{\rm B}^{\rm av} - (E^{\rm stat} - \Gamma_1^{\rm stat})\,\right] - \tfrac{\sigma_{\rm m}(L_1)}{2} = L_1\Gamma_1(L_1, m_{\rm h})\,, \qquad (5)$$

where Γ_1 originates from QCD in L_1^4 and any reference to bare parameters has disappeared. Its graphical solution is reproduced in Figure 3 and yields $M_{\rm b}^{\rm stat} = 6.806(79)\,{\rm GeV}$.

The inclusion of the sub-leading $1/m_{\rm b}$–effects is technically more involved and exploits the freedom of choices for the angle(s) θ and an alternative set of matching observables[30]. We just quote the final value $\overline{m}_{\rm b}^{\overline{\rm MS}}(\overline{m}_{\rm b}) = 4.347(48)\,{\rm GeV}$ with the remark that, upon including the $1/m_{\rm b}$–terms, differences among the static results w.r.t. the matching condition chosen are gone, which signals practically negligible higher-order corrections.

3.2 Status in Two-Flavour QCD

The renormalization of HQET through the non-perturbative matching to $N_{\rm f} = 2$ QCD in finite volume, to do the power-divergent subtractions, is under way[31–33]. As an important prerequisite, the calculated non-perturbative relation between the RGI and subtracted bare

heavy quark mass[34] allows to fix RGI heavy quark masses in the matching volume L_1^4:

$$L_1 M = Z_{\rm M}(g_0) Z(g_0)\,(1 + b_{\rm m}(g_0) a m_{\rm q,h})\, L_1 m_{\rm q,h}\,, \quad M = M_{\rm h}\,. \tag{6}$$

The extent L_1 is defined via a constant SF coupling, $\bar{g}^2(L_1/2) = 2.989$, and the PCAC masses of the dynamical light quarks are tuned to zero.

Figure 4 shows two examples for the heavy quark mass dependence of finite-volume QCD observables in the continuum limit, which enter the non-perturbative matching.

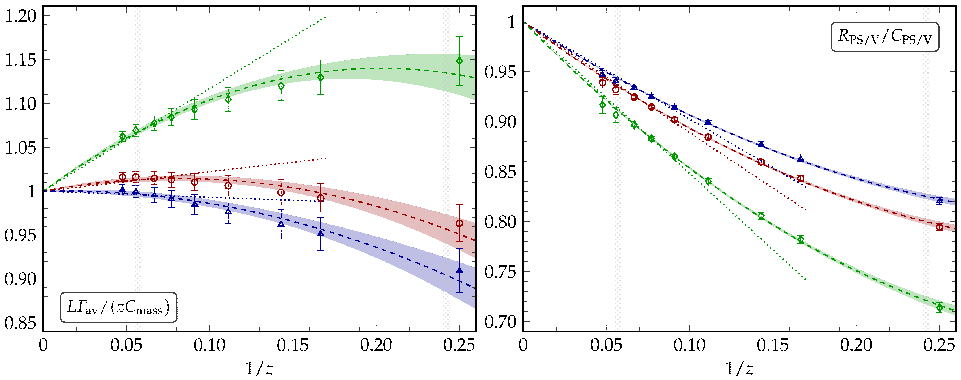


Figure 4. z–dependence of two QCD observables, which enter the matching of QCD with HQET in $N_{\rm f} = 2$ with massless dynamical quarks. $z = L_1 M$ as before. Left: Spin-averaged B-meson energy. Right: Ratio of axial-vector to vector current matrix elements. The conversion functions $C_{\rm X}$, which only introduce small perturbative uncertainties[32,33], translate the HQET predictions to the corresponding QCD quantities at finite values of z.

These results also allow to perform non-perturbative tests of HQET in the spirit of the corresponding quenched investigation[35]. The calculation of the step scaling functions in HQET is expected to be finished soon, and the large-volume part of our strategy is currently being implemented within the CLS effort[22].

4 Outlook

The non-perturbative treatment of HQET including $1/m_{\rm b}$–terms can lead to results with unprecedented precision for B-physics on the lattice. It also greatly improves our confidence in the use of the effective theory. The striking agreement, for example, between the decay constant $F_{\rm B_s}$ computed including $1/m_{\rm b}$–corrections and the value resulting from the interpolation between the static number and data around the charm[36,37], though still in the quenched approximation, provides a strong internal check of the approach. In addition, the HQET parameters at $\mathrm{O}(1/m_{\rm b})$ calculated non-perturbatively by the ALPHA Collaboration[37] can be employed for several other quantities. The programme aiming to reach the same accuracy in the $N_{\rm f} = 2$ dynamical case is well advanced and progressing fast[33].

Acknowledgements

We are indebted to our colleagues in CLS and ALPHA for a fruitful collaboration, and to P. Fritzsch for a part of the figures. We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG) in the SFB/TR 09-03, "Computational Particle Physics", and by the European Community through EU Contract No. MRTN-CT-2006-035482, "FLAVIAnet". J.H. also acknowledges a DFG grant under HE 4517/2-1. Our simulations are performed on BlueGene and APE Machines of the John von Neumann Institute for Computing at Forschungszentrum Jülich, at DESY, Zeuthen, and INFN, University of Roma "Tor Vergata", and on other platforms available to CLS. We thank all institutions for providing their computer resources and technical support.

References

1. *H*eavy *F*lavor *A*veraging *G*roup, `http://www.slac.stanford.edu/xorg/hfag`.
2. M. P. Altarelli, *LHCb status and early physics prospects*, arXiv: 0907.0926.
3. R. Sommer and H. Wittig, *Fundamental parameters of QCD*, NIC Symposium 2001 (eds. Horst Rollnik, Dietrich Wolf), **NIC Series Volume 9**, 149, 2002.
4. R. Sommer, *Non-perturbative QCD: renormalization, O(a)-improvement and matching to Heavy Quark Effective Theory*, hep-lat/0611020, and references therein.
5. M. Della Morte, *Standard Model parameters and heavy quarks on the lattice*, PoS, **LAT2007**, 008, 2007.
6. E. Gamiz, *Heavy flavour phenomenology from lattice QCD*, arXiv:0811.4146.
7. C. Aubin, *Lattice studies of hadrons with heavy flavors*, arXiv:0909.2686.
8. C. Amsler et al., *Review of particle physics*, Phys. Lett., **B667**, 1, 2008.
9. S. Capitani, *Lattice perturbation theory*, Phys. Rept., **382**, 113, 2003.
10. J. Gasser and H. Leutwyler, *Chiral perturbation theory to one loop*, Ann. Phys., **158**, 142, 1984.
11. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, Phys. Lett., **B195**, 216, 1987.
12. J. C. Sexton and D. H. Weingarten, *Hamiltonian evolution for the hybrid Monte Carlo algorithm*, Nucl. Phys., **B380**, 665, 1992.
13. C. Urbach, K. Jansen, A. Shindler, and U. Wenger, *HMC algorithm with multiple time scale integration and mass preconditioning*, Comput. Phys. Commun., **174**, 87, 2006.
14. M. Hasenbusch, *Speeding up the Hybrid-Monte-Carlo algorithm for dynamical fermions*, Phys. Lett., **B519**, 177, 2001.
15. M. Hasenbusch and K. Jansen, *Speeding up lattice QCD simulations with clover-improved Wilson fermions*, Nucl. Phys., **B659**, 299, 2003.
16. M. Della Morte, F. Knechtli, J. Rolf, R. Sommer, I. Wetzorke, and U. Wolff, *Simulating the Schrödinger functional with two pseudofermions*, Comput. Phys. Commun., **156**, 62, 2003.
17. M. Lüscher, *Lattice QCD and the Schwarz alternating procedure*, JHEP, **05**, 052, 2003.
18. M. Lüscher, *Solution of the Dirac equation in lattice QCD using a domain decomposition method*, Comput. Phys. Commun., **156**, 209, 2004.

19. M. Lüscher, *Schwarz-preconditioned HMC algorithm for two-flavour lattice QCD*, Comput. Phys. Commun., **165**, 199, 2005.
20. M. Lüscher, *Deflation acceleration of lattice QCD simulations*, JHEP, **12**, 011, 2007.
21. R. C. Brower, T. Ivanenko, A. R. Levi, and K. N. Orginos, *Chronological inversion method for the Dirac matrix in hybrid Monte Carlo*, Nucl. Phys., **B484**, 353, 1997.
22. `https://twiki.cern.ch/twiki/bin/view/CLS/WebHome`.
23. G. von Hippel, R. Sommer, J. Heitger, S. Schaefer, and N. Tantalo, $\mathrm{D_s}$ *physics from fine lattices*, PoS, **LAT2008**, 227, 2008.
24. E. Eichten and B. Hill, *An effective field theory for the calculation of matrix elements involving heavy quarks*, Phys. Lett., **B234**, 511, 1990.
25. M. Della Morte, A. Shindler, and R. Sommer, *On lattice actions for static quarks*, JHEP, **08**, 051, 2005.
26. L. Maiani, G. Martinelli, and C. T. Sachrajda, *Nonperturbative subtractions in the heavy quark effective field theory*, Nucl. Phys., **B368**, 281, 1992.
27. G. Martinelli and C. T. Sachrajda, *Computation of the b-quark mass with perturbative matching at the next-to-next-to-leading order*, Nucl. Phys., **B559**, 429, 1999.
28. F. Di Renzo and L. Scorzato, *The residual mass in lattice heavy quark effective theory to α^3 order*, JHEP, **02**, 020, 2001.
29. J. Heitger and R. Sommer, *Non-perturbative heavy quark effective theory*, JHEP, **02**, 022, 2004.
30. M. Della Morte, N. Garron, M. Papinutto, and R. Sommer, *Heavy Quark Effective Theory computation of the mass of the bottom quark*, JHEP, **01**, 007, 2007.
31. M. Della Morte, P. Fritzsch, J. Heitger, H. B. Meyer, H. Simma, and R. Sommer, *Towards a non-perturbative matching of HQET and QCD with dynamical light quarks*, PoS, **LAT2007**, 246, 2007.
32. M. Della Morte, P. Fritzsch, J. Heitger, and R. Sommer, *Non-perturbative quark mass dependence in the heavy-light sector of two-flavour QCD*, PoS, **LAT2008**, 226, 2008.
33. ALPHA Collaboration, work in progress.
34. G. M. de Divitiis, P. Fritzsch, J. Heitger, and N. Tantalo, to appear.
35. J. Heitger, A. Jüttner, R. Sommer, and J. Wennekers, *Non-perturbative tests of heavy quark effective theory*, JHEP, **11**, 048, 2004.
36. M. Della Morte, S. Dürr, D. Guazzini, J. Heitger, A. Jüttner, and R. Sommer, *Heavy-strange meson decay constants in the continuum limit of quenched QCD*, JHEP, **02**, 078, 2008.
37. ALPHA Collaboration, to appear.

Supercomputing in Astrophysics in the 21st Century

Peter L. Biermann

MPI for Radioastronomy, Bonn, Germany
E-mail: plbiermann@mpifr-bonn.mpg.de

Dept. of Physics and Astronomy, University of Bonn, Germany

Dept. of Physics and Astronomy, University of Alabama, Tuscaloosa, AL, USA

Dept. of Physics, University of Alabama at Huntsville, AL, USA

FZ Karlsruhe and Physics Dept., University of Karlsruhe, Germany

Where do we come from and where do we go? This leads directly to questions of cosmological structure and evolution, to the mechanism of the last and the next nearby stellar explosion, and to the formation of planetary disks as a precursor to the Solar system. For all three explorations super-computers are key instruments for improving and testing our understanding and insight. Comparing with data will push us to new theoretical developments, to enable us to then ask new further questions.

Numerical simulation has taken its parallel place to observations, theory, and developing instrumentation to use on ground and space based experiments. In all four lines of research there are consequences for modern society and its evolution, like a) rapid deconvolution of multiple 2D images into a 3D view of a human body for detection of disease, like b) observations of Earth to predicts tsunamis and earth quakes, like c) theory for predicting weather and space weather, involving sophisticated plasma physics with magnetic fields, and like d) technology for rapid observations of traffic and airplanes to eliminate accidents in any conditions. Here we focus on the first point, super-computer use, and narrow it down to astrophysics. Astrophysics is an example for all.

In the following we use three examples to illustrate the importance of super-computing to our physical understanding of the universe:

Clusters of Galaxies

Clusters of Galaxies in the Mare-Nostrum Universe, by Stefan Gottlöber, Gustavo Yepes, Jaime Forero-Romero, and Victor Turchaninov. The universe is dominated by dark energy, dark matter and a large fraction of solar invisible normal baryonic matter. Despite these drawbacks, of not knowing what about 98 percent of the energy content of the universe consists of, where it is, and what it does, it is remarkable, how far we have come in our modelling of the large scale structure of the universe. The team around Gottlöber is focussing on clusters of galaxies in a very large simulation[4, 6, 15]; this entails trying to understand the interaction between the gravitational structures, mostly in dark matter, and the gas flows, out of which stars form at the end. These stars in turn reionize the normal matter, dirty it up with heavy elements, fill it with magnetic fields, and also with energetic particles. And yet we cannot be sure, that larger scale phenomena do not inject more energy and more magnetic fields than these smaller scale phenomena[3, 14].

These simulations allow us to ask the questions, that will be relevant to interpret the forthcoming satellite data on the fine structures of the microwave background (the PLANCK mission); once we approach a better understanding of the physics, the universe may become the prime laboratory also for our deeper understanding of the structure of matter, perhaps far beyond what the LHC at CERN may do, on whose results we also need to build.

Supernovae

Unravelling the Mechanism how Stars Explode, by Andreas Marek, Bernhard Müller, and Hans-Thomas Janka. The seemingly most extreme explosions in the universe are those of stars, sometimes in the form of Gamma Ray Bursts. There is no lack of ideas about the driving mechanism of the explosion, with some of the oldest ideas centering on rotation and magnetic fields[1,2]. However, once it became clear that in the collapse preceding the explosion most of the available energy is turned into neutrinos, they became the popular choice for investigation; This is where Janka and his collaborators work[8,9,5]. We have learnt from this work, that instabilities rule the explosion, and data confirm this from the discovery of mixed-up unstable isotopes. And what really drives the use of super-computers, is the clear necessity, that we require a 3D simulation of the explosion, quite independent of what the mechanism is, that we wish to consider. Here Janka and his collaborators have forged ahead, and have been able to explain some of the observations. It is hoped that the hoped-for neutrino detection of more supernovae and Gamma ray Bursts will test and challenge all these models. These explosions are necessary to produce the heavy elements, which sustain life.

Protoplanetary Disks

Protoplanetary Discs in Young Dense Clusters, by Susanne Pfalzner, Christoph Olczak, Thomas Kaczmarek, and Manuel Steinhausen. What is the role of disks in star and planet formation, and in accretion to compact stars, is a really old question[7,13]. Pfalzner and her collaborators are exploring the role of disks in star formation and early evolution, closely checking with the orion nebula cluster observations[10–12]. Among the important results are that capture of stars into compact binary systems is a key ingredient, that disks are sometime large, but can lose mass. The physics explored and tested here will find further application in the early universe, when the very first massive stars formed, and with them the very first massive black holes in the universe.

The interplay between all these three explorations, large scale structure, stellar explosions, and planetary disk formation, will ultimately be used to further our understanding both of the deep structure of matter, and of the answer to the question, where do we come from, and where do we go.

Acknowledgements

Support for work with PLB has come from the AUGER membership and theory grant 05 CU 5PD 1/2 via DESY/BMBF and VIHKOS.

References

1. G. S. Bisnovatyi-Kogan, S. G. Moiseenko, N. V. Ardelyan, *Different Magneto-rotational Supernovae* , Astron. Rep. **52**, 997–1008, 2008; translat. from Russian Astron. Zh. **85**, 1109–1121, 2008.
2. G. S. Bisnovatyi-Kogan, S. G. Moiseenko, *Core-Collapse Supernovae: Magnetorotational Explosions and Jet Formation*, Phys. Part. and Fields **39**, 1128–1135, 2008.
3. S. Das, et al., *Propagation of Ultra-High-Energy Protons through the Magnetized Cosmic Web*, Astroph. J. **682**, 29–38, 2008.
4. S. Gottlöber, G. Yepes, Astroph. J. **664**, 117–122, 2007.
5. N.-J. Hammer, H.-Th. Janka, E. Müller, *Three-Dimensional Simulations of Mixing Instabilities in Supernova Explosions*, e-prints 0908.3474, Aug. 2009.
6. M. Hoeft, M. Brüggen, G. Yepes, S. Gottlöber, A. Schwope, Monthly Not. Roy. Astron. Soc. **391**, 1511–1526, 2008.
7. R. Lüst, *Die Entwicklung einer um einen Zentralkörper rotierenden Gasmasse. I. Lösungen der hydrodynamischen Gleichungen mit turbulenter Reibung*, Zeitschrift Naturf. Part A **7**, 87, 1952.
8. A. Marek, H.-T. Janka, *Delayed Neutrino-Driven Supernova Explosions Aided by the Standing Accretion-Shock Instability*, Astroph. J. **694**, 664–696, 2009.
9. A. Marek, H.-T. Janka, E. Müller, *Equation-of-state dependent features in shock-oscillation modulated neutrino and gravitational-wave signals from supernovae*, Astron. & Astroph. **496**, 475–494, 2009.
10. S. Pfalzner, C. Olczak, A. Eckart, *The fate of discs around massive stars in young clusters*, Astron. & Astroph. **454**, 811, 2006.
11. S. Pfalzner, C. Olczak, *Quasi-binarity of massive stars in young dense clusters - the case of the Orion nebula cluster*, Astron. & Astroph. **475**, 875, 2007.
12. S. Pfalzner, *Universality of young cluster sequences*, Astron. & Astroph. **498**, L37, 2009.
13. K. H. Prendergast, G. R. Burbidge, *On the Nature of Some Galactic X-Ray Sources*, Astroph. J. Letters **151**, L83, 1968.
14. D. Ryu, et al., *Turbulence and Magnetic Fields in the Large-Scale Structure of the Universe*, Science **320**, 909, 2008.
15. G. Yepes, R. Sevilla, S. Gottlöber, J. Silk, Astroph. J. Letters **666**, L61–L64, 2007.

Clusters of Galaxies in the MareNostrum Universe

Stefan Gottlöber[1], Gustavo Yepes[2], Jaime Forero-Romero[1], and Victor Turchaninov[3]

[1] Astrophysical Institute Potsdam
An der Sternwarte 16, 14482 Potsdam, Germany
E-mail: {sgottloeber, jforero}@aip.de

[2] Grupo de Astrofísica, Universidad Autónoma de Madrid
Madrid E-28049, Spain
E-mail: gustavo.yepes@uam.es

[3] Keldysh Institute for Applied Mathematics
Miusskaya Sq. 4, 125047, Moscow, Russia
E-mail: vturch@utech.ru

In 2005 we have performed a non-radiative gasdynamical simulation of structure formation in a large volume of the universe. The main goal of this simulation was to study the formation and evolution of galaxy clusters. This simulation has been analysed during the last four years at NIC Jülich. With more than 2 billion dark matter and gas particles it is one of the largest gas-dynamical cosmological simulations which needs a supercomputer with large memory for analysis and the possibility to store a huge amount of data. JUMP was very well suited for this task. Here we discuss some recent results about interacting clusters of galaxies and summarize our studies about Sunyaev-Zeldovich effect in clusters, their radio envelopes and strong lensing properties.

1 Introduction

During the last decade cosmology has entered in a new era of precision. The cosmological parameters have been determined to a typical precision of a very few percents which resulted in a standard model for cosmology. This is a flat Friedmann universe whose mass-energy content is dominated by a cosmological constant (the Λ term), a cold dark matter (CDM) component and the baryons. The convergence to a standard model of cosmology sets the framework for studying the formation of the large scale structure and of galaxies within that model.

The expansion rate of the universe and the clustering properties are described by only a few parameters: the Hubble constant, H_0, the mass density parameter, $\Omega_{\rm mat}$, the value of the cosmological constant, Ω_Λ, the primordial baryon abundance, Ω_b, and the overall normalisation of the power spectrum of initial density fluctuations, typically characterised by σ_8, the present-day rms mass fluctuations on spheres of radius $8h^{-1}$ Mpc. Structure formation is predicted to be essentially scale invariant, the same structures occur both at the scales of galaxy clusters and on scales of galaxies. The latter prediction is problematic because much more small scale structure is predicted than is observed.

Already 30 years ago the standard picture of structure formation was suggested by White and Rees[9]. Gravitational instability drives the dark matter (DM) to cluster hierarchically in a bottom-up fashion. The gaseous baryons settle into to the DM halos (namely, bound virial DM structures) and via gas-dynamical dissipative processes, cool and fragment eventually forming stars. This simple picture prevails today, although the details and

the modelling of the actual physical processes have evolved dramatically since then. The non-linear nature of the gravitational dynamics and the gas-astrophysical processes makes the problem of structure formation virtually intractable analytically, and therefore the field relies on numerical simulations. Galaxy formation depends on many physical processes, some of them are very poorly understood, and some others take place on sub-grid scales which cannot be modelled fully and consistently and therefore need to be fudged numerically by *ad hoc* recipes. This sets the stage for one of the most interesting and difficult problems of modern astrophysics, namely the formation of galaxies and other structures in the ΛCDM cosmology.

The model specifies the cosmological expansion history, the initial conditions and the material content of the universe. That together with the knowledge of the physical laws governing the dynamics of the DM, baryons and radiation provide the framework within which a successful model of galaxy formation has to be developed.

Numerical simulations have been the driving force behind much of the theoretical progress in the field. The simulations can be classified by either the type of physical processes they consider, or by the sizes of the computational volumes that are simulated. Thus, we can make two main categories of simulations: pure N-body ones, in which only DM is included, and the N-body/gas-dynamical ones where both the gravitational dynamics (DM and baryons) and the hydrodynamics of the gas are accounted for. The gas-dynamical simulations are generally divided into "non-radiative" and those that include the "full" *gastrophysics* such as radiative cooling, star formation and feedback. Moreover, cosmological simulations always pretend to reproduce a statistical representative volume of the universe in which the comparison with observations is done by statistical means, such as mass and correlation functions. On the other hand, there are some other simulations which are focused on reproducing the evolution of individual objects at the highest possible resolution with the aim of studying the structure, dynamics and kinematics of specific objects, like galaxies, groups of galaxies or clusters of galaxies. In what follows we will discuss the results from one of the largest non-radiative gas-dynamical simulation of a comoving cube of $500h^{-1}\mathrm{Mpc}$ on a side.

2 Numerical Simulations

This simulation, dubbed THE MARENOSTRUM UNIVERSE, was performed in the MareNostrum supercomputer at the Barcelona Supercomputer Center using the equivalent of about 29 years on a single CPU (250,000 CPU hours). 135 time steps, equally separated by 100 Myears, have been stored and transferred to the Jülich Multi Processor (JUMP) machine and analysed there. The simulation has been performed with the entropy conserving TreeSPH code GADGET2 developed by V. Springel[8]. It followed the non linear evolution of structures in the gas and dark matter from $z = 40$ to the present epoch ($z = 0$) within a comoving cube of $500h^{-1}\mathrm{Mpc}$ on a side. We assumed the spatially flat concordance cosmological model with the following parameters: the total matter density $\Omega_m = 0.3$, the baryon density $\Omega_b = 0.045$, the cosmological constant $\Omega_\Lambda = 0.7$, the Hubble parameter $h = 0.7$, the slope of the initial power spectrum $n = 1$ and the normalisation $\sigma_8 = 0.9$. The power spectrum used to generate the initial conditions for the simulation was kindly provided by Wayne Hu in a tabulated form. It was obtained from a direct integration of the Boltzmann code for the parameters described above. We did numerical interpolation

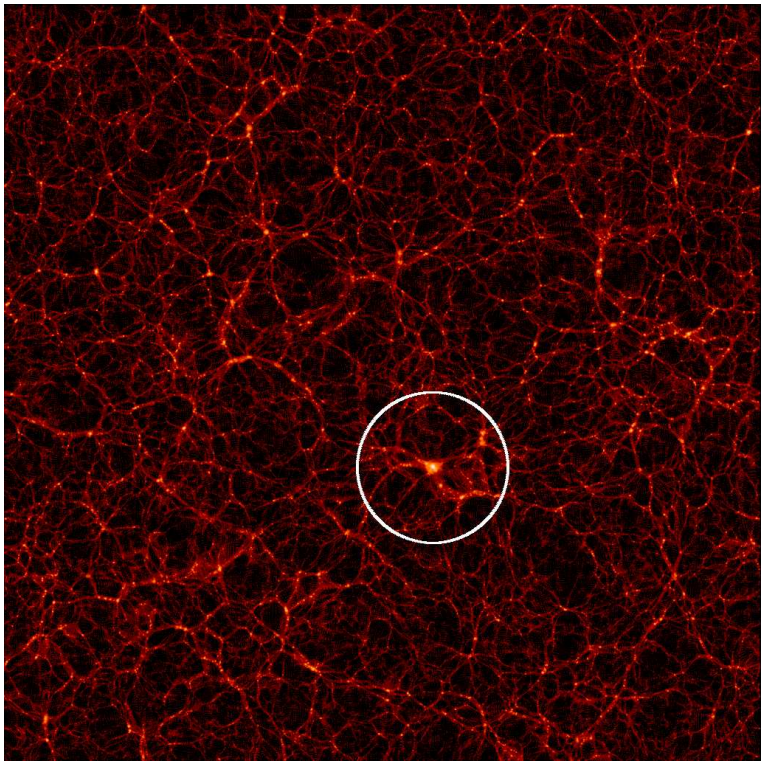


Figure 1. This slice through the simulation volume at redshift $z = 0$ contains the second most massive cluster of galaxies (marked by the circle).

to compute the contribution of the different Fourier modes. Both components, the gas and the DM, were resolved by 1024^3 particles, which resulted in a mass of $8.3 \times 10^9 h^{-1}\mathrm{M}_\odot$ for the DM particles and $1.5 \times 10^9 h^{-1}\mathrm{M}_\odot$ for the gas particles, respectively.

GADGET2 uses the TREEPM algorithm on a homogeneous Eulerian grid to compute large scale forces by the Particle-Mesh algorithm. In this simulation we employed 1024^3 mesh points to compute the density field from particle positions and FFT to derive gravitational forces. In this simulation we employed 1024^3 mesh points to compute the density field from particle positions and FFT to derive gravitational forces. Since the baryonic component is also discretized by the gas particles all hydrodynamical quantities have to be determined using interpolation from the gas particles. Within GADGET the equations of gas dynamics are solved by means of the Smoothed Particle Hydrodynamics method in its entropy conservation scheme. To follow structure formation until redshift $z = 0$ we have restricted ourselves to the gas-dynamics without including dissipative or radiative processes or star formation. The spatial force resolution was set to an equivalent Plummer gravitational softening of $15\ h^{-1}$ comoving kpc. The SPH smoothing length was set to the distance to the 40^{th} nearest neighbour of each SPH particle.

Gravitationally bound halos of different mass have been formed during the cosmological evolution. These halos consist of dark matter and gas. Due to the hydrodynamical interaction of the gas in general the ratio of gas to dark matter within the halos is different

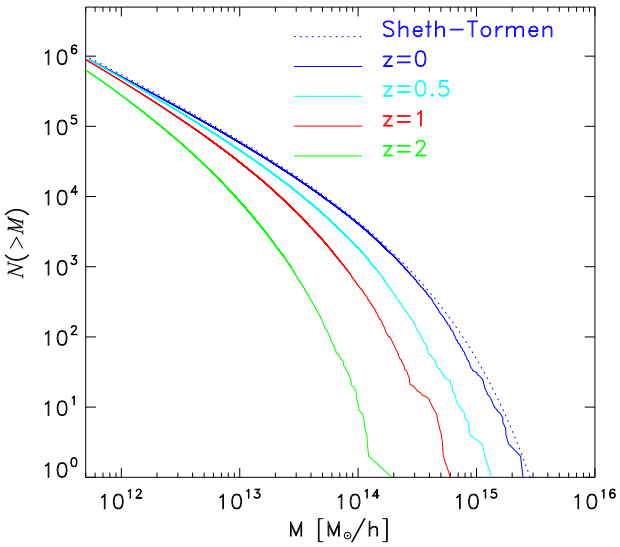


Figure 2. Mass function of halos at redshift $z = 0, 0.5, 1, 2$ and Sheth-Tormen mass function (blue dotted line) at redshift $z = 0$.

from the mean ratio of 0.15 assumed in the simulation. Moreover, the spatial distributions of gas and dark matter differ inside the halos. Even if one can see the halos by naked eye as white spots in Figure 1. it is a challenge to identify numerically all those halos within the distribution of two billion particles and to determine their properties.

To this end we have used a newly developed parallel version of the hierarchical friends-of-friends (FOF) algorithm[5]. In a first step we construct the minimum spanning tree for the distribution of gas and DM particles. After topological sorting of the set of clusters we get a cluster ordered sequence, i.e. a sequence in which each FOF cluster is represented by a segment of the sequence. After this preparation we can easily extract these clusters by simple cutting the tree at the desired linking length. We use a basic linking length of 0.17 of the mean inter-particle separation to extract the FOF clusters. We divide this linking length by 2^n ($n = 1, 3$) to find substructures and in particular the centres (density peaks) of our objects. We were running the minimum spanning tree and the FOF analysis independently over DM and gas particles to find their distribution.

3 Clusters of Galaxies

Using a linking length of 0.17 at redshift $z = 0$ we have identified more than 2 million objects with more than 20 DM particles which closely follow a Sheth-Tormen mass function. More than 4000 cluster sized objects with masses larger than $10^{14}h^{-1}\mathrm{M}_{\odot}$ have been found and about 58,000 with masses larger than $10^{13}h^{-1}\mathrm{M}_{\odot}$. At redshift $z = 1$ more than 30,000 objects with masses larger than $10^{13}h^{-1}\mathrm{M}_{\odot}$ were detected. In Figure 2 we show the mass function of the FOF-halos at different redshifts as total number of objects in the

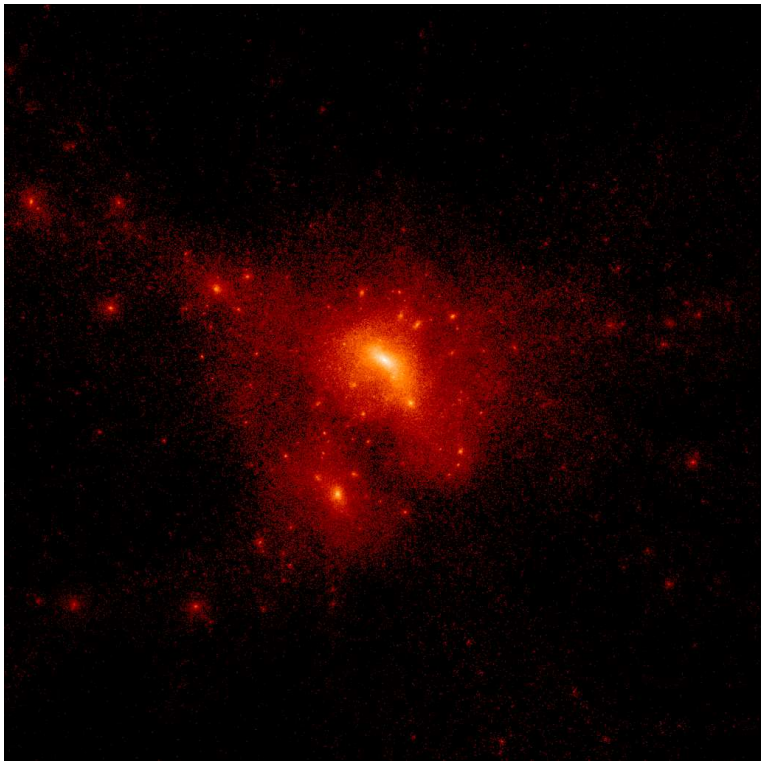


Figure 3. Zoom-in on the cluster marked in Figure 1. The size of the box is 20 $h^{-1}\mathrm{Mpc}$.

box. Already at redshift $z = 2$ first few objects with masses larger than $10^{14}h^{-1}\mathrm{M}_{\odot}$ have been formed.

In Figure 3 we show a small sub-box of $20h^{-1}\mathrm{Mpc}$ size which contains a massive galaxy cluster. Substructures of the cluster can be clearly seen. In this plot we show only the dark matter distribution. The gas distribution in the clusters differs slightly from the dark matter distribution, but both the gas and the dark matter components tend to be prolate, although the gas is much more spherically distributed[3].

The observed cluster baryon fraction is an important tool for the determination of cosmological parameters. Typically the gas fraction in clusters is measured by X-ray observations at overdensities 500 or larger, i.e. well inside the virial radius. A certain fraction of the baryons reside in stars. Using a non-radiative simulation to determine the baryon fraction is by sure a simplification, but since we are interested in the baryon fraction at virial radius we expect not to be affected very much by neglecting cooling and star formation processes. Within our simulation we have explored the baryon fraction at the virial radius in objects with virial masses larger than $5 \times 10^{13}h^{-1}\mathrm{M}_{\odot}$. The assumed virial overdensity at redshift $z = 0$ is 334, at redshift $z = 1$ it is 201. At redshifts $z = 0$ we found more than 10000 objects and about 2500 at redshift $z = 1$. The scatter in the measured baryon fraction of our cluster was found to be quite large. At redshift $z = 0$ it ranges between

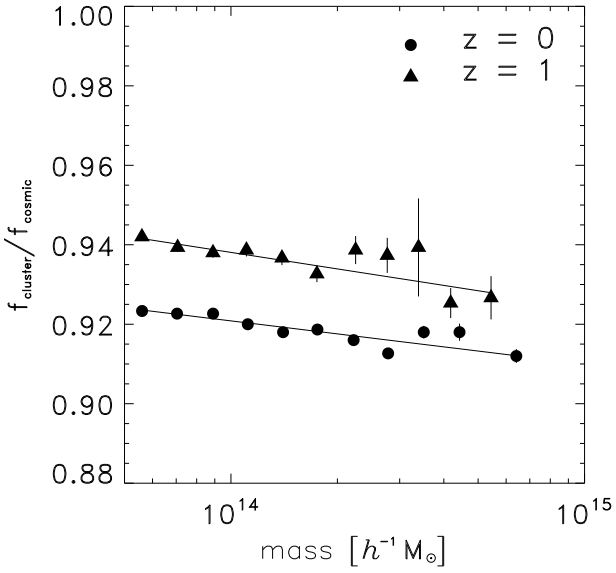


Figure 4. Baryon fraction, normalised to the cosmic baryon fraction, in clusters as a function of halo mass for two different redshifts.

0.85 and 1.0 with a mean of 0.92. In Figure 4 we show the ratio Y_b as a function of the virial mass of clusters. Due to the large number of objects the standard deviation of the mean value of Y_b for different mass bins is small and we could fit a linear relation in the semi-logarithmic plot, $Y_b = \alpha \ln(M) + \beta$, with the slope $\alpha = -0.005 \pm 0.001$ at redshift $z = 0$ and $\alpha = -0.006 \pm 0.001$ at redshift $z = 1$. There is a 2% decrease of the baryon fraction between redshifts one and zero.

Due to the halo interactions, the gas can be displaced with respect to the dark matter distribution. This has been observed in the famous bullet cluster[1]. In Figure 5 we show a similar situation detected in our simulation. This cluster is located at redshift $z = 0.3$, the coordinates are comoving and the relative speed between the two dark matter cores is 560 km/s. In this case, the physical displacement between the centres of gas and DM is of the order of the observed one in the Bullet cluster, but its mass is one order of magnitude lower.

We have derived the expected physical 2D separation distribution for clusters with a dark matter mass larger than $1.0 \times 10^{14} h^{-1}\mathrm{M}_{\odot}$. We found that around 0.6% of these clusters show separations between the dark matter and gas distributions equal or larger than the observed in the Bullet Cluster[6]. In a restricted population of objects which are as massive as the Bullet Cluster ($\sim 10^{15}\ h^{-1}\mathrm{M}_{\odot}$) we found that between 2% and 15% of those clusters should show a dark matter-gas relative separation as the observed one in the Bullet Cluster. The uncertainty in these percentages is due to redshift evolution of the distribution of displacements. The shape of this distribution can be considered as a new

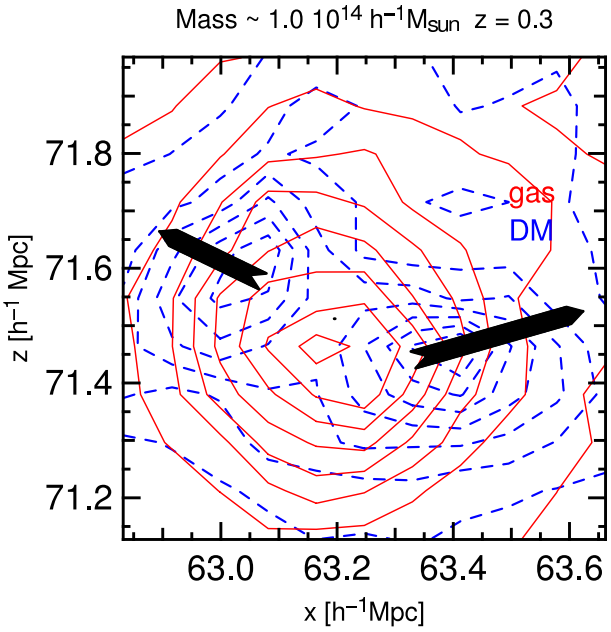


Figure 5. Linearly spaced surface density contours for the DM (dashed blue) and gas (continuous red) components of one of the clusters with a high displacement of the gas and dark matter distribution. The arrows are proportional to the velocities respect to the centre of the gas core.

prediction of the ΛCDM model, which could eventually be compared against observations. Upcoming large optical and X-rays surveys can make this feasible in the next years.

4 Summary

The simulation discussed here has been performed in 2005. This non-radiative simulation is with more than 2 billion dark matter and gas particles still one of the largest ones. It contains more than 4000 galaxy clusters at redshift $z = 0$ and the history of their formation over the 135 stored time steps. Therefore, it is an excellent data base to study the formation and evolution of clusters as well as their present day properties[3, 10]. Here we have discussed the probability of finding objects like the recently observed Bullet Cluster.

We have studied in this simulation the formation of shocks around galaxy clusters and estimated their radio emission[4]. We found several clusters with similar radio morphology as observed whereas about half of the clusters show only very little radio emission. We concluded that one can reproduce the number density and luminosity of radio relics assuming a moderate efficiency of shock acceleration and moderate magnetic fields in the region of the relics.

Motivated by the detection at 33 GHz of a strong temperature decrement in the cosmic microwave background towards the core of the Corona Borealis supercluster we have used this simulation to predict the thermal and kinetic Sunyaev–Zel'dovich effect associated with superclusters of galaxies[2]. We have shown that the Warm–Hot Intergalac-

tic Medium lying in the intercluster regions within the supercluster produces a thermal Sunyaev–Zel'dovich effect much smaller than the observed value and concluded that the observed temperature decrement most probably arises from an unknown galaxy cluster along the line of sight.

Recently, we have investigated the strong lensing properties of the galaxy cluster population in this simulation by means of a large number of ray-tracing simulations[7]. This is the largest sample of numerically simulated clusters ever studied. It allows us to accurately define the characteristics of the clusters which behave as strong lenses. In particular, we have divided the strong lenses into two sub-classes, namely the clusters which are able to develop critical lines, or equivalently to produce multiple images of the same background sources, and the clusters which are also able to produce large distortions, i.e. to map distant galaxies into extended gravitational arcs. We have quantified the minimal masses for critical and large distortion clusters as a function of their redshift assuming a fixed source distance and found that typically large distortion clusters are at least one order of magnitude more massive than critical lenses. Both critical and large distortion clusters tend to have the major axis of their inertia ellipsoid aligned with the line of sight. This orientation bias is particularly strong for large distortion lenses. A large fraction of critical lenses have elongated core structures, indicating the presence of substructures projected on to the inner regions which enhance their ability to behave as strong lenses. Finally, we found a correlation between the deviation from virial equilibrium of dark matter halos and their lensing efficiency, indicating that efficient strong lenses are preferentially dynamically active systems.

Acknowledgements

The MARENOSTRUM UNIVERSE simulation has been performed at the Barcelona Supercomputer Center and analysed at NIC Jülich. We acknowledge support by the ASTROSIM network of the European Science Foundation. VT thanks the Deutsche Forschungsgemeinschaft for supporting his visit at AIP Potsdam.

References

1. Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C., & Zaritsky, D. ApJL **648**, L109, 2006.
2. I. Flores-Cacho, J.A. Rubiño-Martín, G. Luzzi, R. Rebolo, M. De Petris, G. Yepes, L. Lamagna, S. De Gregori, E.S. Battistelli, R. Coratella, S. Gottlöber, MNRAS, in print.
3. S. Gottlöber, G. Yepes, ApJ **664**, 117-122, 2007.
4. M. Hoeft, M. Brüggen, G. Yepes, S. Gottlöber, A. Schwope MNRAS **391**, 1511-1526, 2008.
5. A. Klypin, S. Gottlöber, A. V. Kravtsov, A. M. Khokhlov, ApJ **516**, 530, 1999.
6. C. Llinares, J. E. Forero-Romero, S. Gottlöber, G. Yepes, submitted to APJL.
7. M. Meneghetti, C. Fedeli, F. Pace, S. Gottlöber, G. Yepes, submitted to A&A.
8. Springel V., MNRAS **364**, 1105, 2005.
9. S. D. M. White and M. J. Rees. MNRAS **183**, 341–358, 1978.
10. G. Yepes, R. Sevilla, S. Gottlöber, J. Silk ApJ **666**, L61-L64, 2007.

Unravelling the Mechanism how Stars Explode

Andreas Marek, Bernhard Müller, and Hans-Thomas Janka

Max Planck Institute for Astrophysics
Karl-Schwarzschildstr.1, 85748 Garching, Germany
E-mail: {amarek, bjmuellr, thj}@mpa-garching.mpg.de

Understanding how massive stars end their lives in supernova explosions is one of the most important problems in stellar astrophysics. With the most sophisticated computer simulations carried out to date, researchers at the Max Planck Institute for Astrophysics in Garching are making progress in deciphering the complex interplay of hydrodynamics and particle physics that reverses the collapse of the stellar core and initiates the violent disruption of the star.

1 Astrophysical Context

Roughly five times a second somewhere in the Universe the life of a star with more than eight times the mass of our Sun is terminated in a gigantic supernova explosion. These cosmic catastrophies are the most violent events after the Big Bang. Within only fractions of a second more energy is released than the Sun produces during all its evolution. The hot plasma of the destroyed star, expanding into interstellar space with a tenth of the speed of light, can outshine a whole galaxy for several weeks.

It is these cataclysmic blasts, which we owe our existence to. They enrich the galaxy with carbon, oxygen, silicon, and iron, the building blocks of the earth and of all the creatures on it. Assembled during millions of years of nuclear burning as the dying star has been aging, or forged in the inferno of its final explosion, these chemical elements are disseminated into the galactic medium when the star is disrupted in the supernova event.

Observations of supernova remnants tell us that not all the stellar gas is ejected in the outburst. The core of the star with the size of the Moon but more massive than the Sun, collapses to an ultra compact object, a neutron star, which has only the diameter of a big city. Such a neutron star is still visible as a point source in many of the gaseous clouds that are left behind as heralds of past explosions (see Figure 1).

2 The Supernova Puzzle

The gravitational binding energy released in the neutron star formation is hundred times more than needed for powering the supernova. But how does this happen? How is the implosion of the stellar core reversed to the explosion of the overlying layers of the star? This is still a puzzle that challenges theorists' intuition and modeling abilities. One of the most popular ideas involves neutrinos as the driving agents. These elementary particles are produced in huge numbers at the extreme conditions in the newly formed neutron star, where the matter is denser than in atomic nuclei and reaches temperatures of several hundred billion degrees. Neutrinos are the leak by which the collapsing stellar core loses its gravitational binding energy within a period of seconds. But some fraction of the emitted neutrinos is still able to deposit its energy in the matter surrounding the compact neutron

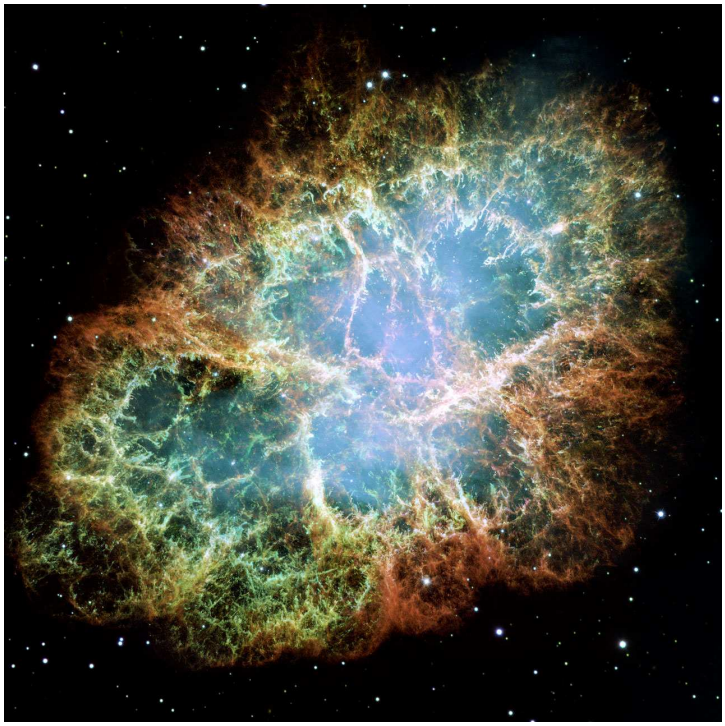


Figure 1. The Crab Nebula with the Crab pulsar, the gaseous remnant and rapidly spinning neutron star, respectively, of a supernova explosion that occurred in the year 1054 A.D. Relativistic particles, which are accelerated by the pulsar, cause the bluish glowing of the gas even 950 years after the explosion. The outer filaments consist of mostly hydrogen and helium of the disrupted star. Source: `www.spacetelescope.org`. Credit: NASA, ESA and Allison Loll/Jeff Hester (Arizona State University). Acknowledgement: Davide De Martin (`www.skyfactory.org`).

star (see Figure 2). This energy transfer could be enough powerful to accelerate the supernova shock front and to expel the overlying shells of the star. The question whether this happens or not is a central problem in supernova theory.

3 Goals of this Project

Answering this question would mean a major breakthrough in stellar astrophysics. It would not only allow us to better link the properties of supernova explosions and their remnants to the different types of progenitor stars. It would also bring us closer to an answer of the question whether supernovae are the still mysterious source of rare elements like gold, lead, thorium, and uranium. And it would allow us to more reliably predict the neutrino and gravitational wave signals, which are planned to be measured for future Galactic supernovae by a new generation of big experiments, and which are the only ways to observationally probe the processes deep inside the core of a dying star.

Until then, computer models are an indispensable tool for promoting our theoretical insight. At Garching, we have developed numerical codes that allow us to perform, with

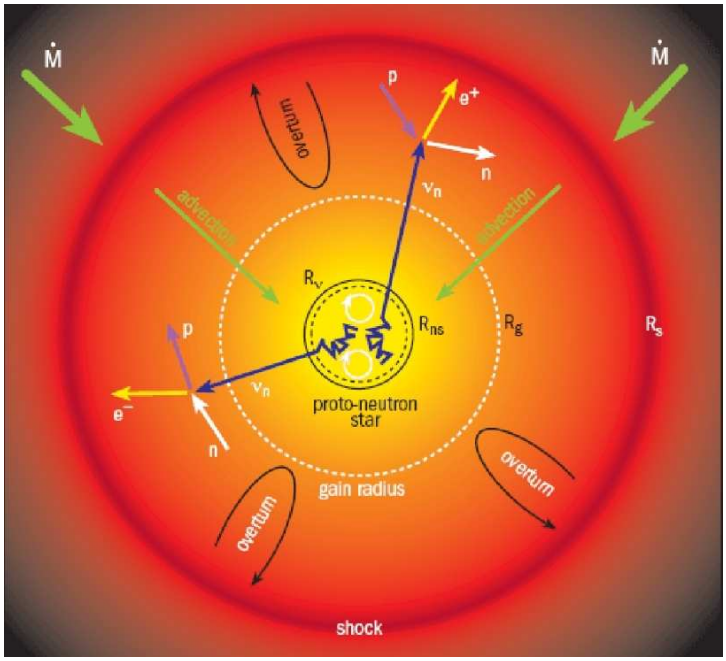


Figure 2. Neutrinos, radiated from the newly formed neutron star at the centre of a collapsing star, deposit the energy to drive the shock wave that causes the supernova explosion of the star.

unprecedented accuracy, simulations of the complex particle physics, nuclear physics, and plasma dynamics that determine the destiny of collapsing and exploding stars.[1–4]

4 Computational Challenges

The modeling of supernova explosions is in fact one of the most difficult problems in computational astrophysics. Largely different time scales, varying between microseconds and seconds, and length scales that extend from tens of meters to millions of kilometers, have to be resolved to follow neutrino interactions, nuclear reactions, turbulent convection, and sound wave propagation in different regions of the collapsing core and of the ejected outer layers of exploding stars. This is computationally extremely demanding: half a second of evolution requires 500,000 time steps and in two spatial dimensions with $500-1000$ radial zones and typically 196 lateral bins needs some 10^{18} floating point operations.

Besides integrating the Euler equations that describe the time-dependent motion of the stellar fluid, one needs to solve the transport of neutrinos in the dense stellar matter. Different from the stellar gas, where very fast electromagnetic and strong interactions ensure that equilibrium is established on dynamical timescales, neutrinos couple with matter only through weak reactions. Thus they require a transport treatment in phase space by the Boltzmann equation and its moment equations. This constitutes a high-dimensional, time-dependent problem — in spherically symmetric models the transport is three-dimensional, for axisymmetric models five-dimensional, and in full generality six-dimensional — and

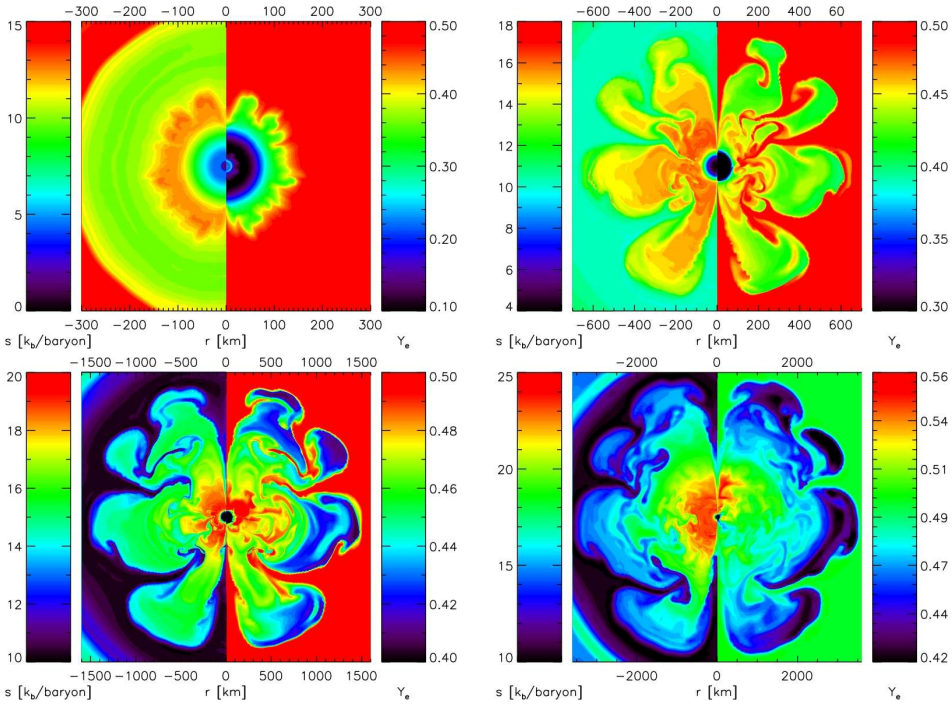


Figure 3. Gas entropy (left half of each picture) and electron-to-nucleon ratio (right half of each Figure) during the early stages (0.097, 0.144, 0.185, and 0.262 seconds) of the explosion of a star with nine solar masses. Convection causes anisotropies of the ejected gas.Important properties of this explosion model agree with characteristic observational features of the Crab Supernova (Figure 1).

poses the major computational challenge: the neutrino transport consumes the by far dominant amount of CPU time during supernova simulations.[5] The problem is also hard to be implemented efficiently on massively parallel computers. In particular the neutrino transport module has resisted such efforts so far and currently prevents us from using distributed memory architectures. Because the interaction timescale of neutrinos in neutron star matter is extremely short and the neutrino propagation happens with the speed of light after decoupling, the nonlinear transport equations of neutrinos, which as fermions are subject to phase-space blocking effects, need to be solved with fully implicit time stepping. In our current numerical implementation this leads to big, densely filled matrices that have to be inverted several times on every time level of the calculated evolution. This is computationally very expensive and defies easy parallelization. New algorithms, based on iterative multigrid solvers for hyperbolic systems of equations, are currently under construction, but not yet available for full-scale supernova calculations. Their use, however, will be unavoidable in future three-dimensional models of supernovae.

The special needs of our current two-dimensional simulations, i.e. shared-memory nodes (up to 256 CPUs) with powerful processors and the continuous availability of these nodes for many months, are satisfied on different supercomputing platforms in all three national supercomputing centres, each of which has required special code adaptation and

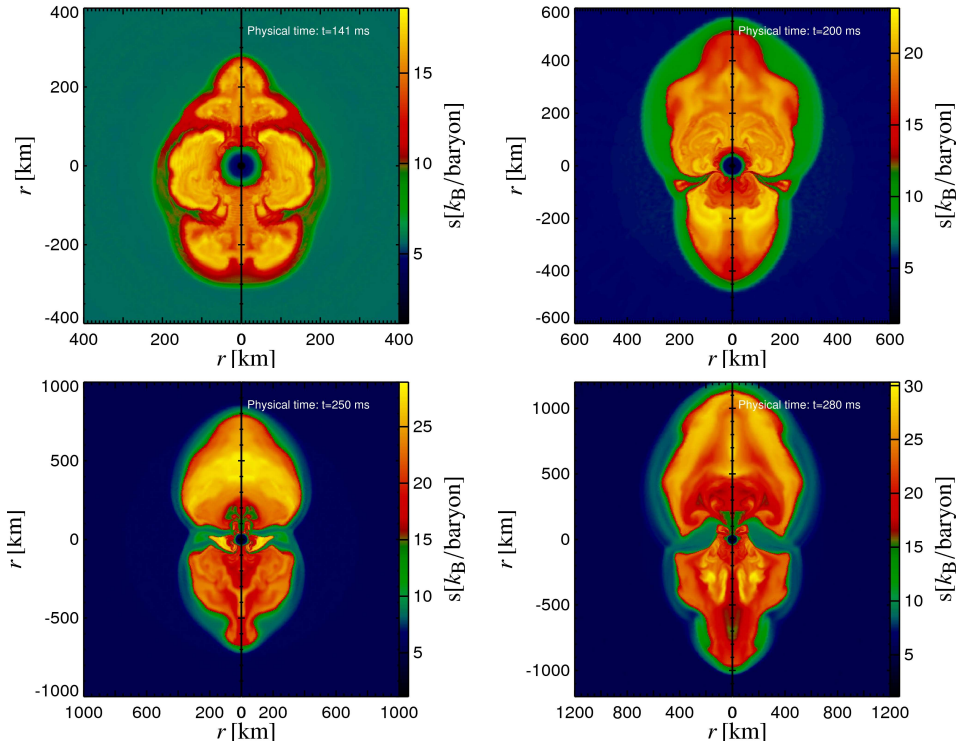


Figure 4. Gas entropy for the beginning explosion of a star with eleven solar masses. The explosion becomes globally asymmetric because of the generic nonradial instability of the developing supernova shock.

optimization: on the IBM p690 Jump and on the INTEL Juropa system of the John von Neumann Institute for Computing (NIC) in Jülich, on the NEC SX-8 and SX-9 of the High Performance Computing Center Stuttgart (HLRS), on the sgi Altix 4700 of the Leibniz-Rechenzentrum (LRZ) in Munich, and on the IBM Power6 of the Rechenzentrum der Max-Planck-Gesellschaft. Our project also has received support by CPU time within the AstroGrid-D as part of the D-Grid initiative that is funded by the German Federal Ministry of Education and Research (BMBF).

5 Recent Progress

With the most sophisticated simulations carried out to date, we have learned that the neutrino energy deposition around the newly formed neutron star (see Figure 2) is supported by different fluid instabilities that take place in the gas flow that continuously adds more matter from the collapsing stellar core to the compact remnant at the centre. The neutrino-heated gas is stirred up by vigorous convective overturn as hot matter becomes buoyant and begins to rise while cooler fluid sinks inward and is partly absorbed into the neutron star (Figure 3).

In addition to this phenomenon, which has been known already for ten years, we have

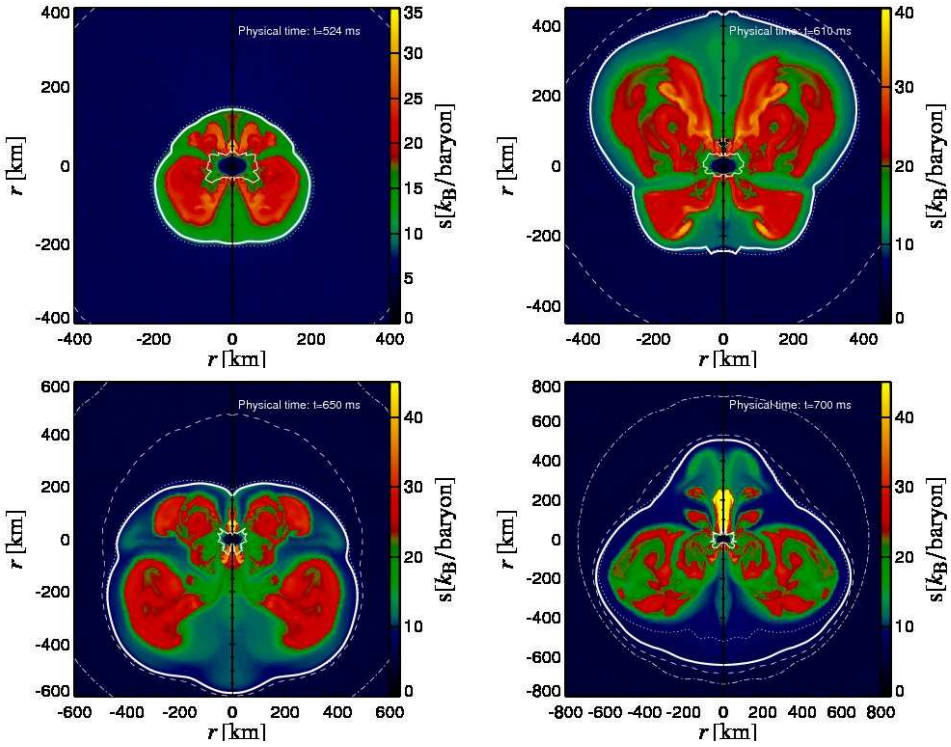


Figure 5. Gas entropy for the beginning explosion of a star with 15 solar masses. The explosion becomes globally asymmetric because of the generic nonradial instability of the developing supernova shock. The thin dotted, dashed, and dash-dotted white lines mark regions where the composition changes gradually from iron to silicon and from silicon to oxygen. The thick solid white line shows the position of the supernova shock front.

discovered a global non-radial instability of the gas flow towards the centre. The layer between neutron star and supernova shock front is unstable by a so-called advective-acoustic cycle, which constitutes an amplifying feedback cycle of inward advected vorticity perturbations and outward propagating sound waves[6,7]. This instability can grow even in conditions where convection remains weak and it can instigate violent secondary convection[8] (Figure 4). It thus improves the conditions for ongoing strong neutrino energy transfer to the supernova shock. In fact, our simulations show that for stars from about eight solar masses to at least 15 solar masses neutrino energy deposition, supported to different extent by both hydrodynamic instabilities, can initiate and drive the supernova explosion[9–11](Figures 3–5).

The onset of the explosion thus turns out to be a generically multi-dimensional phenomenon. The highly aspherical initiation of the blast (Figure 4,5), even in the absence of rapid rotation, suggests explanations for a variety of observations. The fast motions of many young neutron stars, which are measured to have average velocities of several hundred kilometers per second, some of them even of more than 1000 km/s, can be explained by the recoil imparted to the compact remnant by the anisotropically ejected supernova gas[12]. The asymmetry of the developing explosion also triggers large-scale mixing insta-

bilities in the outer layers of the disrupted star, thus accounting for the clumpiness seen in many supernovae and the asymmetric appearance of gaseous supernova remnants[13,14].

6 Outlook

Despite the significant progress of our fundamental understanding of the processes that conspire in starting the explosion, many more simulations and in particular long-time simulations are needed to establish the properties of self-consistently calculated explosions. The different structures of stars with different masses require studies of a wider range of supernova progenitors. And the incomplete knowledge of the initial conditions (e.g. of the angular momentum in the stellar core) and of various aspects of the microphysics (e.g. of the equation of state of hot neutron star matter) make it necessary to explore the full range of variability. Ultimately, three-dimensional simulations will have to be performed to confirm our findings of the present two-dimensional models. But for that to be possible, we still need to wait for the next generations of even more powerful supercomputers, and we have to advance our modeling tools to massively parallel application.

Acknowledgements

We are grateful for computer time at the John von Neumann Institute for Computing (NIC) in Jülich, the Höchstleistungsrechenzentrum of the Stuttgart University (HLRS), the Leibniz-Rechenzentrum (LRZ) in Munich, and the computer centre of the Max Planck Society (RZG). This work has been supported by the Cluster of Excellence EXC153 "Origin and Structure of the Universe", the Deutsche Forschungsgemeinschaft through the Transregional Collaborative Research Centers SFB/TR 27 "Neutrinos and Beyond", and SFB/TR 7 "Gravitational Wave Astronomy" and by AstroGrid-D, a project funded by the German Federal Ministry of Education and Research (BMBMF) as part of the D-Grid initiative.

References

1. M. Rampp and H.-Th. Janka, *Radiation hydrodynamics with neutrinos. Variable Eddington factor method for core–collapse supernova simulations*, A&A, **396**, 361–392, Dec. 2002.
2. R. Buras, M. Rampp, H.-T. Janka, and K. Kifonidis, *Two-dimensional hydrodynamic core-collapse supernova simulations with spectral neutrino transport. I. Numerical method and results for a 15 $M_\odot$ star*, A&A, **447**, 1049–1092, Mar. 2006.
3. R. Buras, H.-T. Janka, M. Rampp, and K. Kifonidis, *Two-dimensional hydrodynamic core-collapse supernova simulations with spectral neutrino transport. II. Models for different progenitor stars*, A&A, **457**, 281–308, Oct. 2006.
4. A. Marek, H. Dimmelmeier, H.-Th. Janka, E. Müller, and R. Buras, *Exploring the relativistic regime with Newtonian hydrodynamics: an improved effective gravitational potential for supernova simulations*, A&A, **445**, 273–289, Jan. 2006.
5. A. Marek, K. Kifonidis, H.-Th. Janka, and B. Müller, *The SuperN–Project: Understanding core collapse supernovae*, in: High Performance Computing in Science and Engineering 2006, Ed. Nagel, W. E. and Jäger, W. and Resch, M., pp. 19–+, Springer, Berlin. 2006.

6. J. M. Blondin, A. Mezzacappa, and C. DeMarino, *Stability of Standing Accretion Shocks, with an Eye toward Core–Collapse Supernovae*, ApJ, **584**, 971–980, Feb. 2003.
7. T. Foglizzo, *Non–radial instabilities of isothermal Bondi accretion with a shock: Vortical–acoustic cycle vs. post–shock acceleration*, A&A, **392**, 353–368, Sept. 2002.
8. L. Scheck, H.-T. Janka, T. Foglizzo, and K. Kifonidis, *Multidimensional supernova simulations with approximative neutrino transport. II. Convection and the advective-acoustic cycle in the supernova core*, A&A, **477**, 931–952, Jan. 2008.
9. A. Marek and H.-T. Janka, *Delayed Neutrino-Driven Supernova Explosions Aided by the Standing Accretion-Shock Instability*, ApJ, **694**, 664–696, Mar. 2009.
10. A. Marek, H.-T. Janka, and E. Müller, *Equation-of-state dependent features in shock-oscillation modulated neutrino and gravitational-wave signals from supernovae*, A&A, **496**, 475–494, Mar. 2009.
11. F. S. Kitaura, H.-Th. Janka, and W. Hillebrandt, *Explosions of O–Ne–Mg cores, the Crab supernova, and subluminous type II–P supernovae*, A&A, **450**, 345–350, Apr. 2006.
12. L. Scheck, K. Kifonidis, H.-Th. Janka, and E. Müller, *Multidimensional supernova simulations with approximative neutrino transport. I. Neutron star kicks and the anisotropy of neutrino–driven explosions in two spatial dimensions*, A&A, **457**, 963–986, Oct. 2006.
13. K. Kifonidis, T. Plewa, L. Scheck, H.-T. Janka, and E. Müller, *Non-spherical core collapse supernovae. II. The late-time evolution of globally anisotropic neutrino-driven explosions and their implications for SN 1987 A*, A&A, **453**, 661–678, July 2006.
14. N. J. Hammer, H.-Th. Janka, and E. Müller, *Three-Dimensional Simulations of Mixing Instabilities in Supernova Explosions*, e-prints 0908.3474, Aug. 2009.

Protoplanetary Discs in Young Dense Clusters

Susanne Pfalzner[1], Christoph Olczak[2], Thomas Kaczmarek[1], and Manuel Steinhausen[1]

[1] I. Physikalisches Institut, University of Cologne
50937 Cologne, Germany
E-mail: {pfalzner, kaczmarek, msteinh}@ph1.uni-koeln.de

[2] The Kavli Institute for Astronomy and Astrophysics, Peking University
Yi He Yuan Lu 5, Hai Dian Qu, Beijing 100871, P. R. China

In dense young clusters the gravitational interaction between the stars can influence the properties of the discs around them. The cluster regions, where observable differences in the discs are likely, are identified for a variety of cluster environments. We combine cluster simulations performed with the Nbody6++ code with particle simulations that describe the effect of a fly-by on the disc around a young star to determine the environmental imprint onto the discs.

1 Introduction

For a long time it was believed that stars form basically in isolation. However, recent observations show that the majority of stars are born as part of a cluster of stars. These star clusters can contain just a few stars but as well several 100 000 of stars. The densities in different clusters vary over many orders of magnitude. For example, the cluster NGC 6823 has a density of ~ 0.05 $M_\odot/pc^{-3}$ [1] whereas the density of the Arches cluster is $\sim 4 \times 10^5$ $M_\odot/pc^{-3}$ [2].

As the stars come into existence through a collapse inside a molecular cloud, angular momentum conservation results in the simultaneous formation of a disc around the young star. Under favourable conditions these discs can develop over the following 10 Myr into planetary systems. Naturally the question arises whether there are environments where the high concentration of stars has an influence on the development of the individual star-disc system and therefore the formation rate of planetary systems around them.

Apart from their normal development with time the circumstellar discs around the young stars might be altered by the dense star cluster environment. The present understanding is that circumstellar disks in clustered environments may be truncated due to close encounters with massive stars either through tidal stripping or photoevaporation of outer disk material[3,4]. In our work we concentrate on the investigation of gravitational interactions between the stars in such young dense clusters and their effect on the discs around the young stars.

2 Numerical Method

At present it is not possible to perform numerical particle simulations where the stars including their surrounding discs are sufficiently resolved to determine the effect of encounters on the discs quantitatively. Therefore we treated the dynamics of the stars in the cluster and star-disc encounters in separate investigations and combined the results to determine, for example, the disc destruction rate as illustrated by Figure 1.

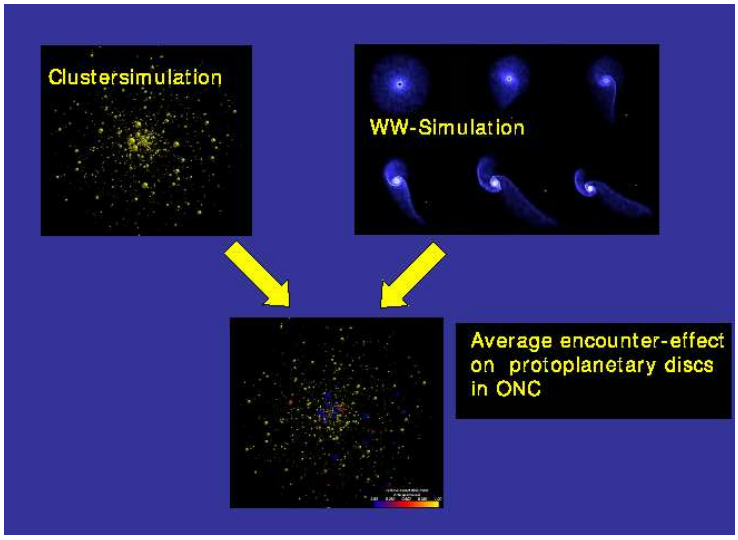


Figure 1. Schematic picture of combination of star-disc encounter simulations and cluster simulations to obtain the application domains of the effect of encounters on the discs around young stars in a star cluster environment.

2.1 Disc Evolution Inflicted by Encounters

As a other star has a close fly-by to a star-disc system several things can happen to the disc matter: (a) it can stay bound to its star, (b) it becomes unbound, (c) it can be captured by the passing star (which then might form its own disc) or (d) it can be accreted onto the star that then increases in mass (see Figure 2). The result is that the disc mass might be lowered by the encounter, its size and mass distribution altered or in extreme cases the disc can be completely destroyed.

We use N-body simulations to investigate the effect of such fly-bys. We represent the mass distribution in the disc by a distribution of particles initially circling the star on Keplerian orbits. Depending on what effect we are interested in the mass distribution is represented by between 10 000 to several million particles. If we are only interested in global properties like the total mass loss in an encounter the low resolution simulations suffice whereas for properties like disc fragmentation high resolution simulations are necessary.

In order to determine the effects of the cluster on disc properties we we performed first a parameter study of the effects of encounters on low-mass discs. This parameter study eventually had to include all possible encounter scenarios that two stars in a typical cluster environment could have. Then the dependence of the mass and angular momentum transport on the periastron distance, the relative mass of the encountering stars and eccentricity of the encounter was investigated in detail. This is done for a variety of encounter orientations as well - meaning the fly-by could be in the same direction and plane as the rotation of the disc around the star (prograde), opposit to it (retrograde) and at any possible angle (non-coplanar). It turns out that prograde, coplanar encounters are the most destructive to the disc. In the following only these were investigated further as the parameter space is otherwise not managable. This means in turn that our results are upper limits for the disc destruction.

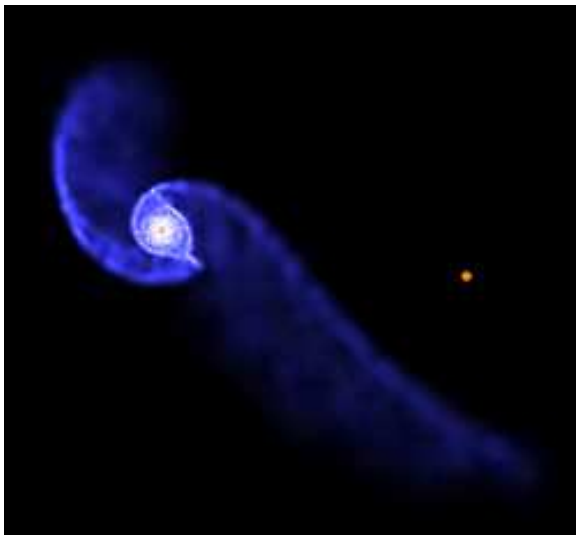


Figure 2. Gravitational interaction of a passing star onto a young disc surrounded star.

For the coplanar prograde case we were able to develop fitting formulae for the mass and angular momentum loss, which cover the whole range required in the cluster simulations. In addition, the final disc size and the mass exchange between discs are examined.

2.2 Dynamics of the Star Cluster

The second part of the investigation is the simulation of the dynamics of all stars in the cluster. The cluster simulations were performed with Nbody6++[5], which allows a high-accuracy treatment of two-body interactions and is a parallelized version of Nbody[6].

We performed simulations of different model clusters using the code Nbody6++[5]. In these simulations, for simplicity only stellar components without considering gas or any potential of the background molecular cloud are considered. Cluster models were set up with a spherical density distribution $\rho(r) \propto r^{-2}$ and a Maxwell-Boltzmann velocity distribution representing the contained stars. There is the option to either choose all stars to be initially single or to start out with a given distribution of single and binary stars. In most cases we start for simplicity with a purely single distribution because this reduces the complexity of the problem and makes the diagnostics easier. The stellar masses were generated randomly according to the mass function given by Kroupa (2002). During the simulation, information of all perturbing events on each stellar disc was recorded.

Since the aim of this work was to record the encounters in the cluster, additional routines had to be implemented: Apart from storing the closest encounter for each star a more advanced encounter list was produced: The search criterion for the next perturber of a star was modified by considering the gravitationally most dominating body instead of the closest neighbour, or in other words, the minimization of the distance was replaced by the maximization of the gravitational force, acting on disc particles.

2.3 Model Clusters

We have one cluster as standard cluster for our simulations that is a a model of the Orion Nebula Cluster (ONC). There are several reasons why we chose the ONC as our model cluster: Firstly, the ONC is a relatively nearby cluster, so there is sufficient knowledge form observations available that we can adjust our model cluster to. Secondly, the ONC is a relatively dense cluster so that interactions between the stars are likely to play a role. Thirdly, the ONC is a typical star forming region, so results obtained there should be transferable to other star forming regions[7].

In the case of the ONC-like model 4000 stars were set up in the mass range $50\mathrm{M}_\odot \geq M^* \geq 0.08\mathrm{M}_\odot$ and the most massive star was assigned a mass of 50 $\mathrm{M}_\odot$. The basic dynamical model is described in Olczak et al. (2006) with several extensions discussed in Pfalzner & Olczak (2007). The quality of the dynamical models, which we chose to be in virial equilibrium, was determined by comparison to the observational data of the ONC[9].

In addition, simulations of clusters with different densities and radial extents were performed. By contrast to the ONC-simulations there was no set mass for the most massive star but instead an upper mass limit to the initial mass function of m_{max} = 150 $\mathrm{M}_\odot$was applied. The density-scaled models were simulated with the same initial size as the ONC (R=2.5pc). The clusters are approximately characterized by densities of 1/4, 1/2, 1, 2, 4 times the density of the ONC. The size-scaled models were simulated with the same initial density as the ONC, but the extension of the system was altered to investigate the effect of the size of the stellar population. The details of the setup of these systems can be found in Olczak et al. (2009).

2.4 Combination of the Simulation Results

The temporal development of the disc properties of each individual cluster star is determined by calculating the mass and specific angular momentum loss in the succession of encounters stored in the encounter lists. The values of mass and specific angular momentum loss in the individual encounter are obtained using the before mentioned fit formulae (Eq.(1) from Pfalzner et al. (2006) and Eqs.(1) and (2) from Pfalzner (2007)). However, these fit formulae only consider parabolic, co-planar encounters, which are the most destructive ones. Therefore the losses have to be considered as upper limits. In the encounter parameter study the disc of a star of mass M_1^* is perturbed by the fly-by of a second star of mass M_2^*. In these simulations, we assumed disc parameters typical of T-Tauri stars: the discs to be of low-mass (0.01 $\mathrm{M}_\odot$), the disc surface density to have a 1/r-dependence initially and the disc size to scale with the stellar mass as $r_\mathrm{d} = 150AU\sqrt{M_1^*[\mathrm{M}_\odot]}$.

3 Results

3.1 Disc Mass Loss

Combining cluster simulations with encounter investigations, shows for the ONC cluster, that potentially up to 15-20% of the discs in the inner region of the cluster - the Trapezium region - could have been destroyed by encounters. It is plausible that the 15-20% of discless

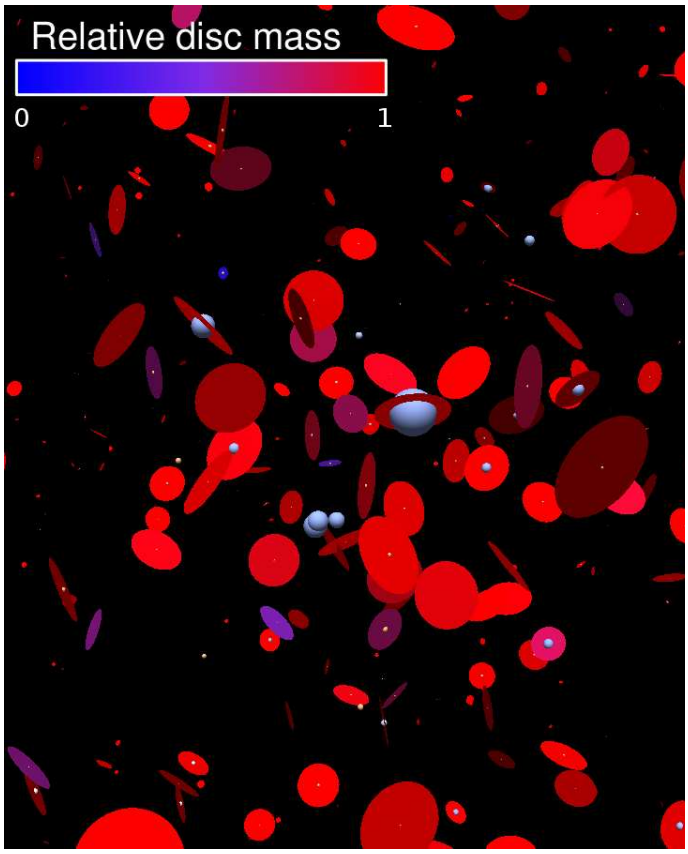


Figure 3. Snapshot of the cluster dynamics simulation. The size of the spheres represent the mass of the stars and the colours of the discs indicate the relative mass of the discs in respect to the initial disc mass. It can be seen that the most massive stars are most likely to loose their discs completely.

stars observed in the Trapezium[11] may, in a large fraction of cases, result from star-disc collisions.

One important result is that the most massive bodies dominate the disc mass loss, with significant interaction even beyond a separation of ten disc radii for a ONC-like entity. This is particulary so for the Trapezium, where some dozen massive stars are surrounded by hundreds of lighter bodies (see Figure 3). Consequently, it is the upper end of a cluster's mass distribution that to a large degree determines the fate of the circumstellar discs in its vicinity (see Figure 4) and thus there are in principle two quantities that are mainly regulating the effect of stellar encounters on the mass-loss from protoplanetary discs: namely the local stellar density which determines the encounter probability, and the upper limit of the mass range, which affects the maximum strength of the perturbing force.

The most massive stars loose their discs much faster and to a larger extent, so that it seems unlikely that planetary systems can form around such O stars in ONC-like environments[12]. This can as well be seen in the illustration of a snapshot of the dynamics

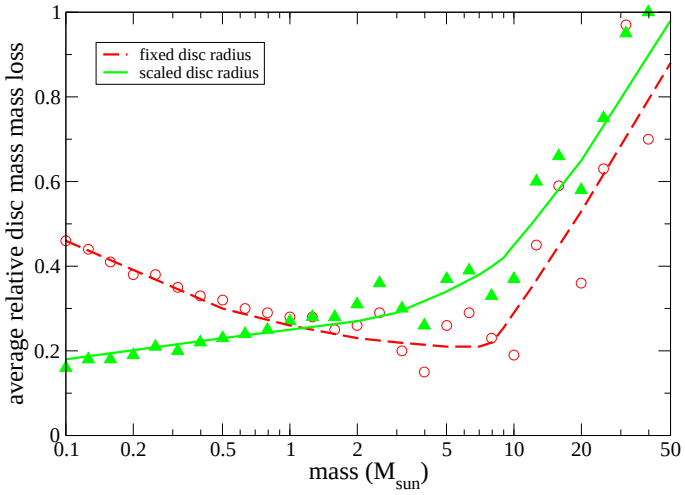


Figure 4. The relative disc mass loss as a function of the mass of the disc-surrounded star. Two cases are shown - the red line indicates the situation where all stars initially had a disc of the same size, whereas the green line indicates the situation where the disc size scales with the stellar mass.

simulation in Figure 3. The mass of the star is here represented by the size of the sphere and the relative disc mass is colour-coded as indicated. Here the most massive stars have no or only little disc material left in contrast to the lower mass stars which have discs that are less affected by the cluster environment.

3.2 Accretion of Disc Matter

The special role of the most massive stars is not restricted to the disc mass loss. Our investigations show that also the specific angular momentum is reduced most for the massive stars. This is remarkable in that way that a lower angular momentum in the disc makes accretion of disc material onto the star easier. So the massive stars become even more massive through the interaction with other stars in the cluster environment. This cluster-induced accretion is not steady but consists of relative short (100 - 10 000 yr) out burst were the accretion rate can be as high as 10^5 $M_\odot$/yr.

3.3 Formation of Binaries

The cluster simulations themselves lead to some interesting results. Observations show that for massive stars the binary frequency seems to be higher than for lower mass stars in young dense clusters. This suggests that in clusters like the ONC different mechanisms are at work in the formation of high-mass binary or multiple systems than for low-mass stars (see Figure 5). We investigate the stellar dynamics in young dense clusters to determine the role

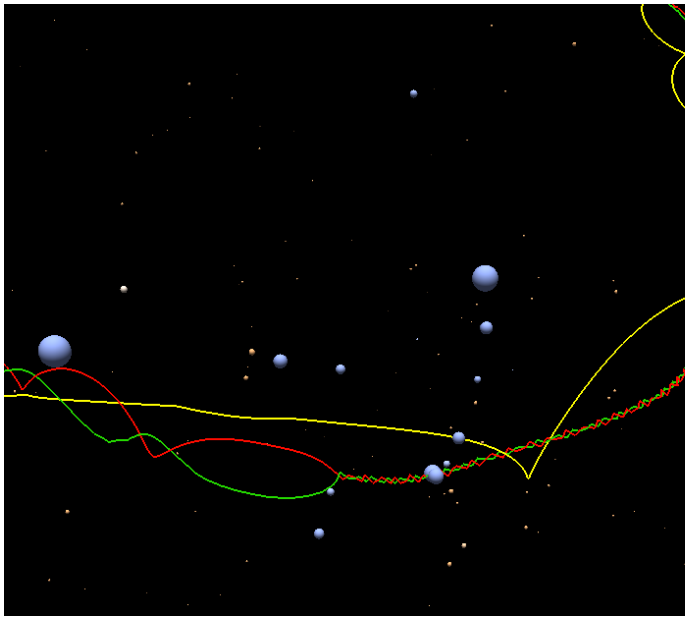


Figure 5. Illustration of the formation of a triple star system by the formation of a binary system with subsequent capture of a third star. This third star circles the binary on a relatively wide orbit. The simulation was performed using Nbody6++ for the Orion Nebula Cluster with 4000 particles representing the stars of the system.

of capture in binary formation in high-mass stars. It turns out that in contrast to lower mass stars capture is a frequent process for massive stars. However, this does not necessarily lead to long lasting binary systems but is often of transient nature. Nevertheless, capture processes could account for 15-25% of the observed binaries of the OB-stars (75%) in Orion.

Figure 4 shows an example of such a process. Here we zoomed into the cluster simulation to show the path of three specific stars. It can be seen how first a binary forms.

4 Summary

Most stars are not born in isolation, but in a cluster environment. This means that gravitational interactions might lead to changes in the disc properties. Numerically modelling the Orion Nebula Cluster, as an example, we investigate how the cluster environment influences through gravitational interactions disc properties like disc frequency, disc mass, angular momentum transport and accretion. We show that the massive stars in the cluster centre are the engines driving these processes and that their discs are most effected.

References

1. Wolff, S. C., Strom, S. E., Dror, D., Lanz, L., Venn, K., 2007, Astron. J., 132, 749.
2. Figer, D.F., 2008, “Massive Stars as Cosmic Engines”, Proc. of Intern. Astron. Union, IAU Symposium, 250, 247.

3. Olczak, C., Pfalzner, S., Spurzem, R., 2006, ApJ, 642, 1140.
4. Scally,A., Clarke, C., 2001, MNRAS 325, 449.
5. Spurzem, R., 1999, Comp. Astroph., 109, 407.
6. Aarseth, S., Gravitational N-Body Simulations 2003, CUP, Cambridge.
7. Pfalzner, S. 2009, A&A 498, L37.
8. Pfalzner, S., Olczak, C. 2007, A&A, 475, 875.
9. McCaughrean, M.; Zinnecker, H.; Andersen, M.; Meeus, G.; Lodieu, N., 2002, Messanger,109, 28.
10. Olczak et al. (2009) A&A accepted.
11. Lada, Charles J.; Muench, August A.; Haisch, Karl E., Jr.; Lada, Elizabeth A.; Alves, Joo F.; Tollestrup, Eric V.; Willner, S. P., 2000, AJ, 120, 3162L.
12. Pfalzner, S., Olczak, C., Eckart, A., The fate of discs around massive stars in young clusters, 2006, A&A, 454, 811.

Computational Biology and Biophysics

Ulrich H. E. Hansmann

Dept. of Physics, Michigan Technological University, Houghton, MI 49931, USA
E-mail: hansmann@mtu.edu

Cells are extremely efficient in controlling chemical reactions, for instance to capture, store and utilize energy. In order to exploit their abilities in medical or biotechnological applications, one needs thorough knowledge of the molecular machinery of cells. However, unveiling the fundamental processes in a cell with sufficient temporal and spatial resolution is a challenge. Not all key processes or molecules are accessible by experiments, and simulations are sometimes the only way to detect hidden processes or proteins. Hence, deciphering the life functions in a cell requires a close interplay between experiments and computing. The latter explores the fundamental mechanisms through modelling, and aims at generating a common framework that integrates all available experimental information.

Describing processes in a cell with sufficient resolution *in silico* is a computationally hard task. Whether one is interested in the time evolution of a single protein, its binding with a ligand, its association with other molecules to multi-protein complexes or fibrils, or whether one wants to probe the network of chemical reactions that make up a working cell, often such simulations are only possible on massively parallel computers as available, for instance, at the John von Neumann Institute for Computing (NIC).

Three articles in this proceedings discuss the application of supercomputing to Biochemistry and Biophysics. Bio-photolysis of water occurs in certain algae and is in principle an effective and environmental sound way of producing the hydrogen that one day could replace gasoline. However, the function of the hydrogenase, the enzyme that catalyzes the production of hydrogen from water, is inhibited by the oxygen that is released as byproduct. Using First-principle Car-Parrinello molecular dynamics and classical molecular dynamics simulations (relying on force fields) with explicit water, Dr. La Penna simulates the working of the hydrogenase enzyme and its interaction with the contaminating oxygen. His work suggests the N-terminal region of the enzyme as the most promising region for designed mutations that could enhance the enzyme's resistance to oxidation.

The collaboration by Drs. Lin, Akola and Jones, involving scientists from Taiwan, Finland and Germany, studies active transport of molecules across cell membranes. This process, requiring energy obtained from the hydrolysis of ATP, is only poorly understood, but of obvious importance in medical and pharmaceutical sciences. In the present article, the authors focus on a specific transporter protein, Sav1866. Even after reducing the number of water molecules to a minimum, they are left with a system of about 200,000 atoms which they follow with molecular dynamics over a timespan of 1μs. This impressive project clearly demonstrates the need for the supercomputing capacities provided by NIC. At the same time, the article points out some of the present bottlenecks. Much longer simulation times are necessary for their studies.

The extremely long time scales necessary for simulations of biological systems require not only supercomputers, but also sophisticated simulation techniques, and their efficient

implementation in software. Optimizing a simulation code for a few ten thousand processors as now common in todays supercomputers, is not trivial.

The article by Drs. Meinke, Mohanty and Zimmermann describes the work of the Simulation Laboratory Biology which aims at advancing the use of supercomputing in the computational biology and biophysics community. For this purpose it has to combine expertise from biology and biochemistry with such of applied computer science and computer engineering. Its mission is to develop in close collaboration with users of the NIC supercomputers the tools that help these research groups to push forward the frontiers in their field. Besides the development and implementation of novel simulation techniques, the advancement of energy functions (leading to higher accuracy), the Simulation Laboratory Biology focuses also on a connection of physics-based approaches with knowledge-based and data mining techniques. As an example the authors present a comparison of folding simulations relying on two different force fields.

The three articles demonstrate the progress and still existing bottlenecks in the use of supercomputers in biology and biophysics. They highlight the essential role of NIC in opening new research avenues. Together, these three articles give confidence that with further increasing computer power and more advanced simulation techniques supercomputing will become an ever more important tool in life sciences.

Structure and Dynamics of Large Biological Molecules: ATP-Binding Cassette (ABC) Transporters

Jung-Hsin Lin[1], **Jaakko Akola**[2,3], **and R. O. Jones**[2]

[1] Division of Mechanics, Research Center for Applied Sciences and
Institute of Biomedical Sciences, Academia Sinica, Taipei 11529, Taiwan and
School of Pharmacy, National Taiwan University, Taipei 10051, Taiwan
E-mail: jhlin@gate.sinica.edu.tw, jlin@ntu.edu.tw

[2] Institut für Festkörperforschung, Forschungszentrum Jülich
D-52425 Jülich, Germany
E-mail: {j.akola, r.jones}@fz-juelich.de

[3] Nanoscience Center, Department of Physics, FI-40014 University of Jyväskylä, Finland and
Department of Physics, Tampere University of Technology, FI-33101 Tampere, Finland

Large-scale molecular dynamics simulations (200,000 atoms for over 1 microsecond) have been performed on the JuGene supercomputer for Sav1866, an ATP-binding cassette transporter originally derived from *Staphylococcus aureus*. The overall architecture of Sav1866 is remarkably stable during the simulation, while the nucleotide binding domains appear to undergo a series of conformational transitions on a quasi-logarithmic time scale.

1 Introduction

Search engines give many thousands of websites when asked for the combination of "grand challenges" and "computational biology" or "bioinformatics". The combination of biology with information technology is one of the great growth areas in science, and the availability of massively parallel computers (such as the IBM Blue Gene series) is making an essential contribution. Well known applications include drug design, the identification of genes in DNA sequences or determining the structure of a protein given its sequence (the "protein folding problem"). Nevertheless, the use of computers in understanding the *atomistic* details of reactions in biological molecules is still at an early stage.

In principle, a means for studying such problems is provided by the density functional (DF) formalism, which is free of adjustable parameters and has the advantage of allowing chemical bonds to form and be broken. We have performed DF calculations on the hydrolysis of ATP (adenosine 5'-triphosphate)[1–3], the basic mechanism of energy production in mammalian (and other) cells. However, even the most powerful computers limit DF calculations of energy surfaces to systems with up to ca. 1000 atoms over some hundreds of picoseconds. These are severe limitations for most problems of biological interest, where we often need to study systems with tens of thousands of atoms over microseconds or even longer. Such studies still require simulations using classical force fields, and we have implemented these in the present work on ATP-binding cassette (ABC) transporters.

Active transport is one of the basic mechanisms to mediate traffic across cellular membranes. ABC transporters are membrane proteins that actively transport substrates through lipid bilayers[4]. They form the largest transporter gene family and are vital to many cellular processes in animals, plants, and prokaryotes (organisms without a cell nucleus). The

transported substrates range from ions and small organic molecules to lipids, carbohydrates, and even whole proteins. Some (including the human P-glycoprotein) contribute to multidrug resistance of cancer cells, and ABC proteins are also associated with cystic fibrosis (mutations in the CFTR protein) and other genetic disorders[5,6].

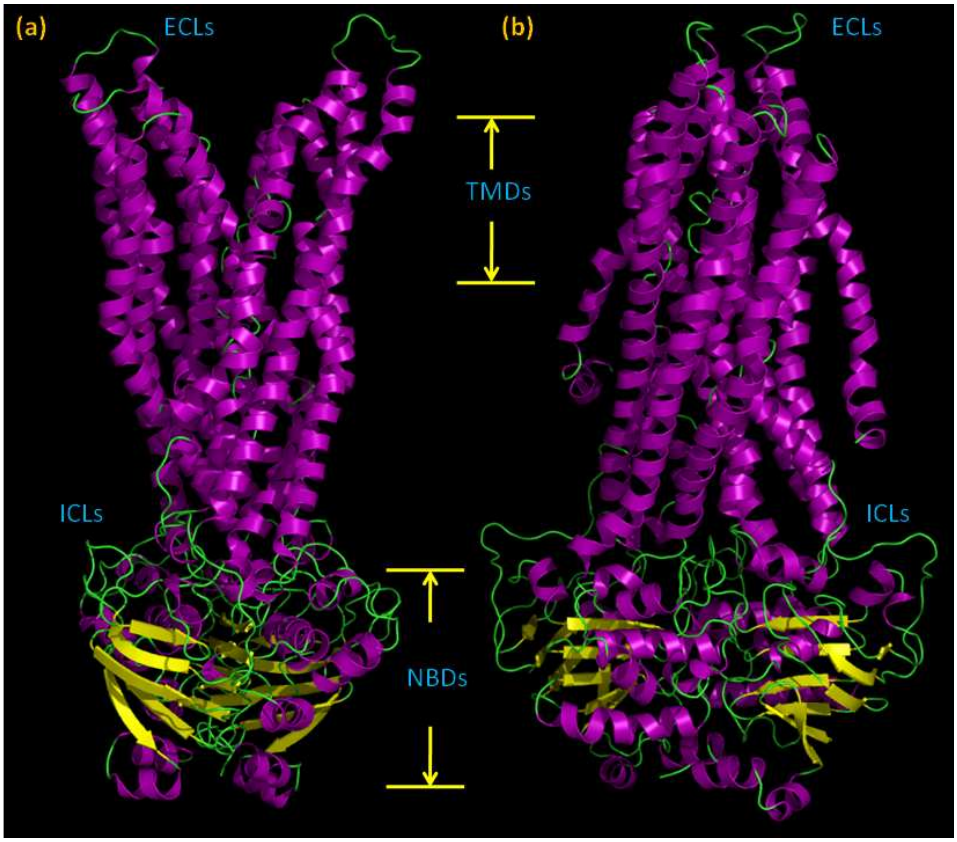


Figure 1. Sav1866 structure. (a) Backbone of the homodimeric protein, (b) View rotated by 90°. TMD denotes the transmembrane domains, NBD the nucleotide (ATP/ADP) binding domains, ICL the intracellular loops, and ECL the extracellular loops between transmembrane helices.

Selective uptake and efflux by ABC transporters is driven against concentration gradients using energy derived from the binding and hydrolysis of ATP, but the mechanism is not well understood. All ABC transporters have an ATP binding subunit whose sequence and structure is relatively well conserved. Together with the nucleotide binding domains (NBD), functional ABC proteins have two transmembrane domains (TMD) [see Figure 1] that have a more variable sequence and provide a passageway for many kinds of substrates. Structural information of crystals of these biomolecules can be obtained by using x-ray diffraction techniques, and atomistic simulations of such *in vivo* processes provide complementary information.

Questions concerning ABC-transporters include:

- How are ATP hydrolysis and the conformational changes in the protein related? Do the latter trigger the former by a "Brownian motor"[7]?
- What can we learn about the details of the atomic and electronic structures?
- What is the reaction pathway, which energies are involved, and how does the conformation change after hydrolysis?
- How does the phosphate group migrate? Does migration result in a conformational change, or is the removal of the whole nucleotide required? What is the role of the ATP-bound cation (usually Mg^{2+} or Ca^{2+})?

The ATP-binding cassette (ABC) transporter Sav1866 is a large protein with 1156 residues (578x2, a homodimer, Figure 1). It was discovered from the whole genome analysis of meticillin-resistant *Staphylococcus aureus*[8], and its structure was first determined by X-ray crystallography (resolution 3.0 Å)[5]. Apart from differences in the ATP-binding sites, no significant conformational changes were observed in complex with adenosine-5'-(β,γ-imido)triphosphate (AMP-PNP, resolution 3.4 Å)[9].

The two-nucleotide binding domains (NBDs) are in close contact in this structure, and a central cavity is formed by the two transmembrane domains (TMDs). This structure is an outward-facing conformation, i.e., the cavity opens wider towards the periplasmic side of the cell membrane than towards the cytoplasmic side. A distinctive feature of this ABC exporter is that both NBDs are in contact with both TMDs, which had not been found in the structures of three known importers[10]. The helices are strongly bent, so that the transmembrane helices lining up the cavity differ at different levels of the lipid bilayer.

2 Computational Methods and Performance on JuGene

We have performed molecular dynamics simulations of Sav1866 in the POPC lipid bilayer membrane, using the AMBER parm99SB force field[11] and the TIP3P water model. The parameters of the POPC lipid force field were adopted from a previous simulation[12]. The protonation state of the Sav1866 was assigned by the PDB2PQR web server[13], with a pH value of 7.

Version 2.6 of the highly scalable NAMD (**NA**noscale **M**olecular **D**ynamics) program[14], implemented and optimized for the Blue Gene/P architecture, was used for the 1 μs simulation. NAMD, a parallel molecular dynamics code for large biomolecular systems, has been developed and maintained at the University of Illinois at Urbana-Champaign (UIUC) and is available as freeware with source code. A recent study[15] described several techniques to scale NAMD to 8192 processors on the Blue Gene/L computer, achieving $\sim$ 1 TFlop on an ATPase benchmark. Other programs have been developed for performing massively parallel simulations on biological systems: The "Blue Matter" MD program, developed by IBM specifically for the Blue Gene architecture[16], has similar performance to NAMD in test cases, and scalable algorithms have been implemented in the "Desmond" program[17].

Extensive tests of all program and execution parameters (including number of task groups) showed that the NAMD program runs well with 4096 processors (one rack) on

the Blue Gene/P. For our first benchmark we chose P-glycoprotein (365 000 atoms), which was also run under NAMD (version 2.6) on the IBM Blade at the Academia Sinica. Benchmarks for an ATPase system with 327 000 atoms[15] had shown that it was possible to perform one time step in 21.8 ms using 2048 Blue Gene/L processors (1 rack), and we obtained somewhat better performance in our tests on the 365 000 atom system on this machine.

3 Simulations on ABC-Transporter Sav1866

The following changes were implemented after our first simulations on Sav1866:

- The system size was reduced to ca. 200 000 atoms by removing water molecules that seemed to be extraneous.
- The atomic mobility was enhanced by raising the temperature to 340 K (67°C).

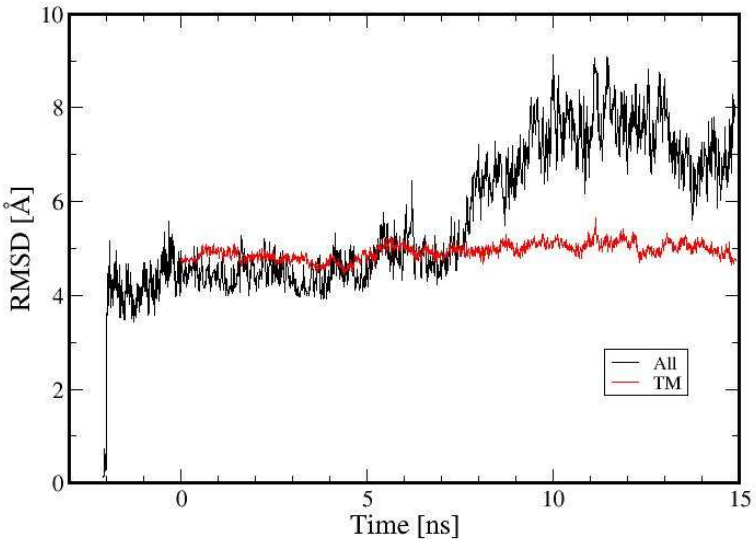


Figure 2. RMS displacement in early stages of the simulation. Black: all atoms, red: transmembrane domains, TMD).

The root-mean-square displacement of the atoms in the early stages of the simulation are shown in Figure 2. It is quite apparent that the motion of atoms in the nucleotide binding domains is already more pronounced than that of atoms in the TMD. The simulation was continued for 1 μs, making it one of the longest performed to date with an all-atom force field for a system of this size. Figure 3 shows that important structural changes occurred: the NBD and TMD have tilted and separated, causing a scissor-like movement of the whole protein. At the same time, the ICLs are exposed to the solvent. While simulations of 1 μs do not allow us to understand all details of the action of the transporter induced by ATP release, this motion is simply not visible in shorter simulations.

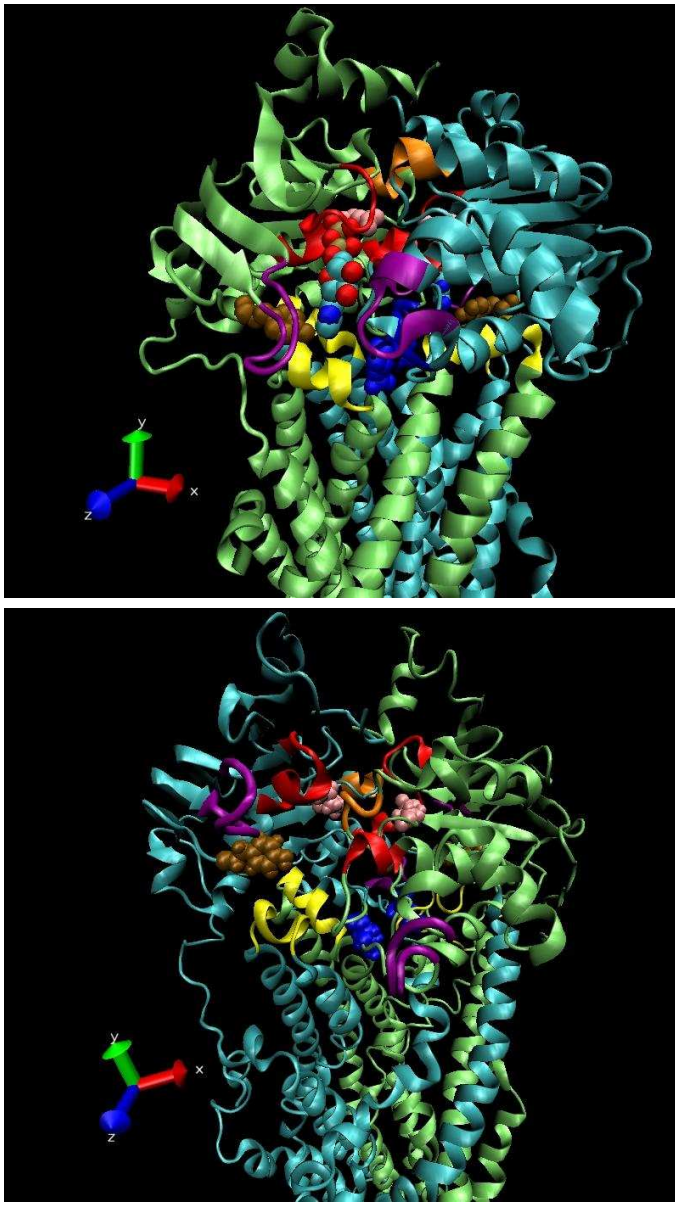


Figure 3. Upper: starting configuration, lower: structure after 870 ns simulation.

Figure 4 depicts the root-mean-square deviation of the two nucleotide binding domains (NBDs) of Sav1866 from their initial X-ray crystallographic structures. The approximately logarithmic behaviour is a sobering result for those seeking the ultimate products of the transformation, but we note that the time-dependence of conformational change of the NBDs is similar to the oscillations in "logarithmic periodic" systems[18].

In contrast to the renormalization group theory of critical phenomena, where continuous scale invariance (CSI) is observed, discrete scale invariance (DSI) is considered the physical origin of logarithmic periodicity of complex systems. In disordered systems, DSI may be associated with the spontaneous breaking of replica symmetry, which involves a hierarchical structure of the phase space[19]. It is also characterized by a discrete set of complex fractal dimension[18]. An early example of DSI was the diffusion on a one-dimensional lattice with random asymmetric transition rate[20]. This was extended to a three-dimensional lattice by Stauffer and Sornette[21], who concluded that the intermittent random walk, similar to the quasi log-periodic conformational transition of NBD in Figure 4, was punctuated by the encounter with increasingly large clusters of trapping sites.

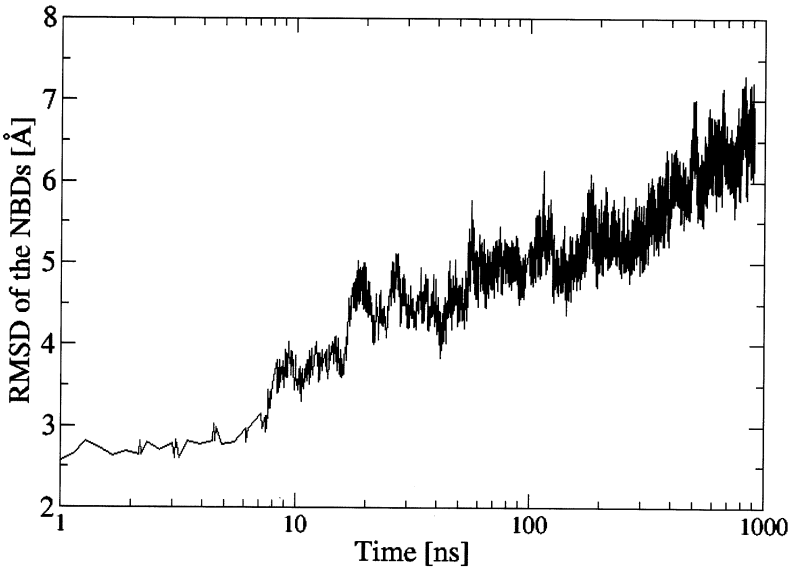


Figure 4. Root-mean-square deviation of the two nucleotide binding domains (NBDs) of Sav1866 from its initial X-ray crystallographic structure. The time axis is logarithmic.

4 Concluding Remarks and Future Challenges

Molecular dynamics simulations have been performed with a classical force field for the ABC-transporter Sav1866. A simulation of 200 000 atoms over 1 μs is large-scale by any standards, and significant atomic motion is clearly evident. Figure 4 shows, however, that longer simulations are necessary to provide a full picture of the reaction. The temperature of the simulation has been increased to 340 K (67°C) to facilitate the atomic motion, and it essential to check that this does not lead to artificial results. We shall discuss the results of a simulation on an intermediate structure at 300 K elsewhere. Future work could implement a parallel tempering strategy.

There are several interesting candidates for future studies. The combination of DF methods for the "active" part of the system with classical force field descriptions of the "inactive" surroundings ("QM/MM methods") extends the scope of DF calculations. The combination with a history-dependent algorithm (metadynamics) allows complex reactions to be studied, and our previous DF experience provides a good starting point for selecting reaction coordinates and the size of the QM region. Other systems of interest include:

- Maltose / maltodextrin ABC transporter (MalK, source: *Thermococcus litoralis*) is a subunit of the trehalosed/maltose ABC transporter of the archeon *Thermococcus litoralis*. It contains 372 residues and there are x-ray measurements of resolution 1.9-2.9 Å (PDBID: 1g29, 1q12, 1q1b, 1q1e, 2aw0)[22]. These structures can be used as starting points for studying the details of its reactions.

- QM/MM studies of ATP hydrolysis in ATPase[23,24] would be a logical extension of our earlier DF work. Another important system involving ATP hydrolysis is phosphoryl transfer and calcium ion occlusion in the calcium pump studied by Nissen and coworkers[25]. This system is larger than those mentioned above, but significantly smaller than Sav1866. It may be difficult to calculate free energy landscapes accurately, but replica exchange MD should work well for finding the most probable states, and parallel tempering should also be possible.

Acknowledgements

We thank S. Kumar (IBM, Yorktown Heights, USA) for advice on optimizing the NAMD program for Blue Gene computers. The calculations were performed on IBM Blue Gene/L, Blue Gene/P, and p690 computers in the FZ Jülich with grants from the FZJ and the John von Neumann Institute for Computing (NIC). JHL was supported by the thematic program of Academia Sinica under grant AS-95-TP-A07.

References

1. J. Akola and R. O. Jones, *ATP hydrolysis in water – a density functional study*, J. Phys. Chem. B **107**, 11774-11783, 2003.
2. J. Akola and R. O. Jones, *Density functional calculations of ATP systems I: Crystalline ATP hydrates and related molecules*, J. Phys. Chem. B **110**, 8110-8120, 2006.
3. J. Akola and R. O. Jones, *Density functional calculations of ATP systems II: ATP hydrolysis at the active site of actin*, J. Phys. Chem. B **110**, 8121-8129, 2006.
4. For a recent review of bacterial ABC systems, see A. L. Davidson *et al.*, *Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems*, Microbiol. and Molec. Biol. Rev. **72**, 317-364, 2008.
5. R. J. P. Dawson and K. P. Locher, *Structure of a bacterial multidrug ABC transporter*, Nature **443**, 180-185, 2006.
6. G. Lu, J. M. Westbrooks, A. L. Davidson, and J. Chen, *ATP hydrolysis is required to reset the ATP-binding cassette dimer into the resting-state conformation*, Proc. Natl. Acad. Sci. (USA) **102**, 17969-17974, 2005.

7. See, for example, P. Hänggi and F. Marchesoni, *Artificial Brownian Motors: Controlling Transport on the Nanoscale*, Rev. Mod. Phys. **81**, 387-442, 2009.
8. M. Kuroda *et al.*, *Whole genome sequencing of meticillin-resistant* Staphylococcus aureus, Lancet **357**, 1225-1240, 2001.
9. R. J. P. Dawson and K. P. Locher, *Structure of the multidrug ABC transporter Sav1866 from* Staphylococcus aureus *in complex with AMP-PNP*, FEBS Letters **581**, 935-938, 2007.
10. J. K. Zolnerciks, C. Wooding, and K. J. Linton, *Evidence for a Sav1866-like architecture for the human multidrug transporter P-glycoprotein*, FASEB J. **21**, 3937-3948, 2007.
11. C. Simmerling, B. Strockbine, and A. E. Roitberg, *All-atom structure prediction and folding simulations of a stable protein*, J. Amer. Chem. Soc. **124**, 11258-11259, 2002.
12. J. H. Lin, N. A. Baker, J. A. McCammon, *Bridging implicit and explicit solvent approaches for membrane electrostatics*, Biophys. J. **83**, 1374-1379, 2002.
13. T. Dolinsky *et al.*, *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations*, Nucleic Acids Res. **32**, W665-W667, 2004.
14. J. C. Phillips *et al.*, *Scalable molecular dynamics with NAMD*, J. Comput. Chem. **26**, 1781-1802, 2005. See also `http://www.ks.uiuc.edu/Research/namd`.
15. S. Kumar, C. Huang, G. Almasi, L. V. Kalé, *Achieving Strong Scaling with NAMD on Blue Gene/L*, IBM document, 2006. See `http://iffwww.iff.kfa-juelich.de/~jones/NAMDIPDPS06.pdf`.
16. B. G. Fitch *et al.*, *Blue Matter: Strong scaling of molecular dynamics on Blue Gene/L*, ICCS 2006 , 846-854, 2006. `http://iffwww.iff.kfa-juelich.de/~jones/Fetal06.pdf`
17. K. J. Bowers *et al. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*, Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, 11–17 November 2006.
18. D. Sornette, *Discrete-scale invariance and complex dimensions*, Phys. Rep. **297**, 238-270, 1998.
19. D. Sornette *et al.*, *Complex fractal dimensions describe the hierarchical structure of diffusion-limited aggregate clusters*, Phys. Rev. Lett. **76**, 251-254, 1996.
20. J. Bernasconi and W. R. Schneider, *Diffusion on a One-dimensional Lattice with Random Asymmetric Transition Rates*, J. Phys. A: Math. Gen. **15**, L729-L734, 1982.
21. D. Stauffer and D. Sornette, *Log-periodic oscillations for biased diffusion on random lattice*, Physica A **252**, 271-277, 1998.
22. See, for example, K. Diederichs *et al.*, *Crystal Structure of MalK, the ATP subunit of the trehalose/maltose ABC transporter of the archeon* Thermococcus literalis, EMBO J. **19**, 5951-5961, 2000.
23. P. D. Boyer, *A Research Journey with ATP synthase*, J. Biol. Chem. **277**, 39045-39061, 2002.
24. J. Zimmer, Y. Nam, and T. Rapoport, *Structure of a complex of the ATPase SecA and the protein-translocation channel*, Nature **455**, 936-943, 2008.
25. T. L. Sørensen, J. V. Møller, and P. Nissen, *Phosphoryl Transfer and Calcium Ion Occlusion in the Calcium Pump*, Science **304**, 1672-1675, 2004.

The Inhibition of Hydrogenase by Dioxygen

Giovanni La Penna

National research council, Institute for chemistry of organo-metallic compounds
Sesto Fiorentino, 50019 Firenze, Italy
E-mail: glapenna@iccom.cnr.it

1 Introduction

The major problem in using molecular hydrogen (H_2) as an energy vector is its production via sustainable processes. There are many of such processes, but the most sustainable is still the direct biophotolysis of water:

$$2\mathrm{H_2O} + light \rightarrow 2\mathrm{H_2} + \mathrm{O_2} \tag{1}$$

This reaction occurs in unicellular algae (*Chlamydomonas rheinardtii*, CR) and in cyanobacteria. The 4 electrons extracted from water in photosystem II (PSII), producing one O_2 molecule and 4 protons, are carried to ferredoxin. In normal metabolic conditions, the electrons are injected into the Calvin cycle where they are used to reduce oxidized carbon species (CO_2). In particular metabolic conditions electrons are transferred to [FeFe]-hydrogenase (Hyd), this latter expressed in larger amount when these environment conditions occur. The enzyme catalyzes the reduction of two protons into the molecule H_2 and this reaction is used by microorganisms to divert electrons produced by PSII from the usual Calvin cycle, when this latter is not fully sustainable: when light irradiation is low and when the conditions to keep PSII highly efficient do not occur[1].

When the normal activity of PSII recovers, electrons are again directed to the Calvin cycle because hydrogenase is demolished. This event is caused by O_2 produced in the PSII. When O_2 becomes concentrated enough, the hydrogenase protein is attacked in the most oxygen-sensitive points and the protein becomes non functioning, unfolded and suited for degradation pathways.

In the cell life the sensitivity of hydrogenase to dioxygen acts as a molecular switch for cell metabolism, but from a technological point of view O_2 is a problem, because it reduces the time during which the CR cell is able to divert electrons from PSII towards the reduction of protons. The normal alternative pathway towards CO_2 to produce polysaccharides is necessary to sustain the cell and its replication. Nevertheless, the production of H_2 together with a normal cell replication and life is possible[2] and it is prevented only by the spread of O_2 within the cell before its escape in the water environment.

The molecular engineering of hydrogenase has been proposed as a possible workaround to this problem. Hydrogenase is an iron-sulfur protein. The H-cluster, where the proton reduction occurs is schematically represented in Figure 1. The Fe_4S_4 cluster (Fe4S4 cluster, hereafter) on the lefthand side is in a resting 2Fe(II)/2Fe(III) state that drives electrons from reduced ferredoxin (FD, that forms a non-covalent complex with Hyd) towards the octahedral Fe(I) (righthand side). The two octahedral Fe(I) ions (Fe2S2 cluster, hereafter) can form transient structures with either bridging H atoms[3] or terminal

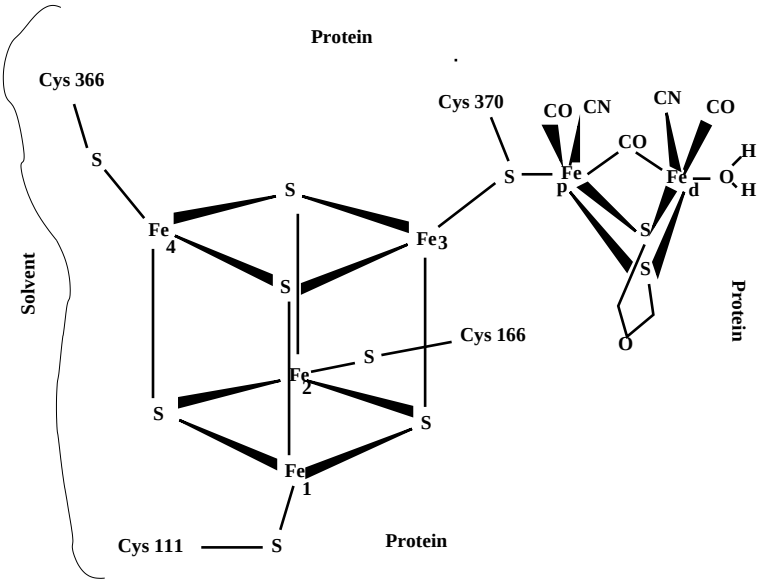


Figure 1. Schematic representation of the H-cluster of hydrogenase and of its ligand environment. The numbering of Cys residues is that of the HydA2 core of 436 residues, with M1 aligned to L140 in the sequence of CpI. On the lefthand side there is the solvent accessible space, while on the righthand side the H-cluster points towards the interior of the protein. Some bonds, like those involving bridging S, are displayed for guiding the eye.

H-bonded atoms[4]. However, it is clear that the octahedral Fe ions are forced, when the small CO and CN^- ligands are available, to adopt low oxidation states, thus attracting the electrons transiently deposited from FD to the Fe4S4 cluster.

Recent works on molecular modeling of hydrogenase dioxygen sensitivity focused on the octahedral Fe ions, assuming that dioxygen, once reaching these ions, inhibits the approaching of the proton to Fe(I) and the consequent formation of metal bound hydride. The approach of dioxygen to octahedral Fe has been investigated[5] and the mutation of residues close to several putative channels has been suggested and investigated, also in comparison with other hydrogenase sequences[6]. Nevertheless, these investigations did not give an answer on which residues can be mutated to produce a more oxygen-resistant hydrogenase.

A different perspective about the oxidation of hydrogenase is suggested by results on oxidation of Fe-S clusters. Even though the mechanism of this oxidation is not completely known, several studies suggest that oxidation occurs more favourably on the Fe4S4 cluster[7], containing tetrahedral Fe(II) ions that can easily host III oxidation states. It is important to notice that all of the hydrogenase with known structure expose one Fe4S4 cluster to the solvent, while the properly catalytic octahedral Fe ions are more buried within the protein matrix. Moreover, the role of Fe4S4 clusters in the protein folding is stronger than that of the catalytic Fe2S2 cluster, because of the many constraints imposed to the 4 bonded Cys residues. Also the general mechanism suggested for the oxidation of FeS proteins by reactive oxygen species, like superoxide, involves the η binding to tetrahedral Fe(II) sites, like those exposed by Fe4S4 clusters[8].

This perspective changes also the target for a possible FeS protection from oxidation:

the target is presumably the most solvent-exposed Fe4S4 cluster instead of the proper catalytic site. This can be important for hydrogenase of Clostridium pasteurianum (CpI), too. In this protein the Fe4S4 cluster attached to the H-cluster is buried at the interface of the two protein domains (the H- and F-domains). Nevertheless, the F-domain can unfold rapidly if one of the two Fe4S4 clusters exposed at its solvent-accessible surface is oxidized. Once the F-domain unfolds, the attack by dioxygen of the H-domain becomes easy as in the CR case, where the F-domain does not exist.

With this project I aim at providing information about the electronic structure of the resting state of the active site of hydrogenase and about the interaction of the active site with the protein matrix. The Fe4S4 cluster being exposed to the solvent and towards the disordered region of the protein, I started with a detailed study of the interactions modulating the statistical behaviour of the N-terminal region of CR hydrogenase, in the form expressed when hydrogen is produced in the largest amount (the form encoded in the HydA2 gene[9]). The disordered N-terminal region, composed of 73 residues, can in principle protect the Fe4S4 cluster of HydA2 from access by the solvent.

2 Methods

The crystal structure of the CpI H-cluster recently refined[10] was used in its resting oxidation state (the so-called H_{red} state), in first-principle Car-Parrinello molecular dynamics (CPMD) simulations for understanding the electron distribution in the resting state, for estimating the forces acting in this state because of the electronic structure, and for computing the point charges to be assigned to the atoms of the cluster and of its neighbouring Cys ligands in a relaxed configuration for further empirical simulations. The CPMD calculations were done with the Quantum-Espresso[11] code, freely available from academic sites. Two CPMD simulations were performed at T =50 K and for 2 ps: with no constraints (CPMD1) and with zero forces acting on Cα atoms of the four Cys Fe4S4 ligands (CPMD2). These latter constraints mimic part of the effects on the electronic structure due to the reduced mobility within the ordered protein core.

The point charges derived by the electronic structure of H-cluster were merged with intramolecular and Lennard-Jones parameters published in the literature. The initial structure of HydA2 core was taken from the literature[12], as it was built from homology information based on CpI. The N-terminal tail (the 62 residues not appearing in the structural model) was built in an all-trans configuration, attached to Ala 63 (the first residue appearing in the published structure) and energy minimized in the vacuum. A non-thermal random walk (Monte Carlo trajectory with random temperature in the torsional space of the chain) for the first 73 residues was performed in the vacuum, in order to generate a set of disordered configurations for the N-terminal tail, with no bad contacts with the protein core. Configurations uniformly extracted from the random walk were collapsed to the protein core by using a large non-bonding cut-off in the vacuum. This procedure produces a set of least-biased structures that, once merged into a box of water molecules, are reasonable starting points for standard molecular dynamics simulations in explicit water. All the empirical calculations were done with the Namd code[13] and with a Monte Carlo code developed in house[14]. Except for those atoms with the force-field parameters derived for FeS clusters and neighbouring Cys atoms, the force-field was a slightly modified Amber PARM94 set[15,14].

3 The Electronic Structure of the H-Cluster

The number of valence electrons evaluated in the Fe atomic basins identified through the Bader analysis[16] are 7.0 for the two Fe atoms in the octahedral coordination sites (Fe_p and Fe_d, see scheme 1), 7.1 for two of the Fe atoms in tetrahedral coordination sites (Fe 2 and 4) and 7.2 for the other two Fe atoms in the tetrahedral coordination sites (Fe 1 and 3). These values are almost constant over the CPMD simulations with no geometrical constraints (CPMD1), with the only exception of 7.1 for Fe 1 at the end (2 ps) of the simulation. As for the simulation with fixed Cys Cα positions (CPMD2), at the end of the trajectory the number of valence electrons becomes 7.1 for all the Fe 1-4 atoms in the tetrahedral coordination sites.

A change of 0.1 in the number of valence electrons is significant. As a term of comparison, when a Fe atom in a fixed coordination site is oxidized (one electron is removed from the system), the change of the number of valence eletrons identified by the Bader analysis is much less than -1, being, depending on the ligands, -0.5 at most. This observation is also consistent with atomic point charges evaluated by electrostatic potential fitting, where most of the positive charge of the metal ion is found delocalized over the atoms bonded to the metal ion. Indeed, the difference of 0.1 in the number of electrons between Fe 1/3 and Fe 2/4 in the Fe4S4 cluster is an indication of the formal distribution of oxidation states (Fe(II)/Fe(III)). On the other hand, the smaller positive charge on Fe atoms belonging to the Fe4S4 cluster (0.8-0.9) compared to that of the Fe2S2 cluster (1.0) indicates a smaller delocalization of the positive charge on the ligands in this latter: the Fe4S4 cluster is more efficient in delocalizing the positive charge than the Fe2S2 cluster.

According to the measure of electron density over metal centres reported above, the electron density is larger on the Fe atoms belonging to the tetrahedral coordination sites (Fe 1-4) than on those in the octahedral sites (Fe_p and Fe_d). According to the mechanism proposed for FeS cluster oxidation, that involves a direct η interaction between O_2 and Fe, the dioxygen molecules would choose the Fe atoms with more electrons. This observation will be integrated with the results of the empirical models below.

The performed 2 ps low-temperature CPMD simulations can be regarded as trajectories where atoms move for optimizing the electronic structure, respectively with and without the geometrical constraints imposed to Cα of Cys residues. In Figure 2 the first and last configurations obtained with no geometrical constraints (CPMD1) are displayed in left and right panels, respectively. As it is displayed in Figure 2, the Fe2S2 cluster is allowed to reach a rather stable configuration (the electron energy decreases of 263 kJ/mol) with a 60° rotation around the Fe3-Sγ(Cys370) bond, by keeping the Fe4S4 cluster almost fixed. This movement in the protein is prevented by the strong packing of the Fe2S2 cluster (see below), thus suggesting that the H-cluster in the protein structure is not able to fully relax the stress induced by its electronic structure, even in its resting state. The simulation with zero forces on Cys Cα atoms (CPMD2) does not change this time evolution: the Fe4S4 cluster moves slightly, as in CPMD1, while Fe2S2 rotates of about 60° around the same Fe-S bond, reaching about the same final configuration. This means that the constraints acting on the Cys backbone atoms do not hinder the Fe2S2 rotation, while the packing of Fe2S2 is more important.

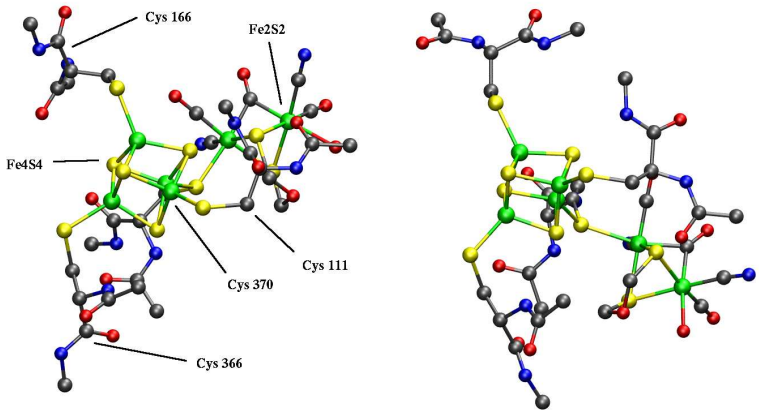


Figure 2. Starting (left) and final (2 ps, right) configurations obtained through CPMD1 (T =50 K, no geometrical constraints). Colour scheme is the following: C is in grey, N in blue, O in red, S in yellow, Fe in green, H atoms are not displayed. The program VMD[13] was used for all chemical drawings.

4 Towards a Molecular Statistics of the N-Terminal Chain of HydA2

Despite the possibility for large space extensions of the disordered region 1-73 of HydA2, the Fe4S4 cluster has large chance to be solvent exposed. If the function of this exposure is related to the accessibility of Fe4S4 by ferredoxin, the interactions responsible for this exposure are not clear. Indeed, the N-terminal tail contains a charge of +8 $|e|$, due to a large amount of Arg residues in its sequence and the electrostatic interactions between the chain and the protein core can force the former towards regions of the latter far from the accessible Fe4S4 cluster. This possibility was investigated with the computer models described below.

The collapse procedure described in the Methods section was performed with the charged residues (Glu, Asp, Arg and Lys) in the N-terminal tail (residues 1-73) in the charged and neutralized states, respectively. The neutralization was achieved by spreading a neutralization charge over the sidechain terminal atoms of the selected residues. The analysis of the set of configurations obtained after the collapse shows a significantly larger protection of the Fe4S4 cluster, measured in terms of the solvent-accessible surface area (SASA). Configurations with a SASA of Fe4S4 cluster smaller than the maximal observed value (4.8 $\AA^2$) are 11 over 50, compared to the value of 4 over 50 measured in the case of charged N-terminal residues. In Figure 3, two configurations obtained with neutralized N-terminal chain are displayed: the configuration in the right panel displays a larger covering of the Fe4S4 cluster compared to that in the left panel.

This observation was investigated in more detail by performing empirical MD simulations of some of the collapsed configurations in an explicit solvent model, in order to check, if water molecules and counterions can modify the stability of the network of interactions reached by the different collapsing procedures. Four of the 100 configurations obtained by collapsing the N-terminal chain were selected, the force-field was set as that

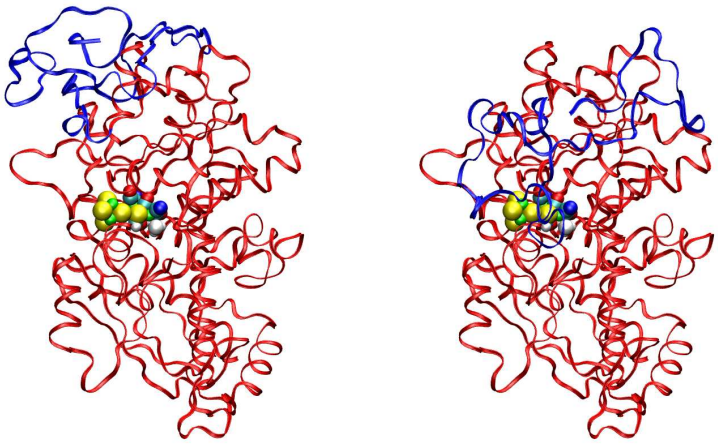


Figure 3. Two of the 50 configurations obtained after the collapse of the 73 residue N-terminal tail towards the fixed protein core: low (left panel) and high (right panel) Fe4S4 covering. The protein is represented as a ribbon interpolating backbone atoms: residues 1-73 are in blue and residues 74-498 (fixed) are in red; the H-cluster is represented with spheres. Fe ions are in green colour.

of the pH=7 state (all Glu, Asp, Arg and Lys residues are charged), the structures were merged into orthorhombic boxes of TIP3P water molecules, the resulting simulation boxes were slowly brought in 1 ns to room conditions (T =300 K and P =1 bar) and they were simulated for 1 ns at T =300 K. Table 1 summarizes the trajectories in terms of initial and final SASA of the protein and of the FeS clusters.

Trajectory	total SASA (nm^2)		SASA(Fe4S4) ($Å^2$)		SASA(Fe2S2) ($Å^2$)	
	initial	average (error)	initial	average (error)	initial	average (error)
1	228	222 (3)	4.8	8 (2)	0.0	0.0 (0)
2	232	243 (3)	0.0	2 (1)	0.0	0.0 (1)
3	222	227 (4)	4.8	8 (3)	0.0	0.0 (1)
4	213	212 (4)	0.0	1 (1)	0.0	0.0 (0)

Table 1. Initial and average SASA of the whole protein and of Fe4S4 and Fe2S2 clusters. Averages are over the last 1 ns of each MD simulation in explicit water. Errors (within parentheses) are root mean-square errors.

The table shows that the SASA of the Fe2S2 cluster is never significantly larger than zero in any of the sampled configurations. On the other hand, the SASA of the Fe4S4 cluster increases in all the investigated cases, with a more significant increase for those configurations that started from configurations where the SASA was already larger than zero. Interestingly, the simulations started from configurations built with neutralized sidechains in the 1-73 region (3 and 4), show a slow drift towards larger total SASA during the MD trajectories. This is an indication that these configurations are still searching for a region

of macromolecular stability with total SASA similar of that reached, for instance, by simulation 2.

The number of salt-bridges formed by Glu, Asp, Arg and Lys sidechains in the protein during the different simulations was also analysed. Even if simulations 1-2 started with 32 salt-bridges, while simulations 3-4 from 23, in all simulations the number of salt-bridges tends to values in the range 30-36. This observation indicates that the highly charged N-terminal region can easily find alternative negatively charged patches on the protein core surface. In order to measure the stability of these different structures more intensive calculations are necessary, especially to stabilize the trajectories started from neutralized configurations (see above discussion about SASA).

5 Conclusion & Perspectives

In this preliminary study of the electronic structure of the H-cluster and of its interactions with the HydA2 protein matrix, I have confirmed several differences between the Fe4S4 and the Fe2S2 components. The Fe4S4 cluster has three properties that make it a good candidate for interaction with O_2:

1. Fe atoms have more electrons in Fe4S4 than in Fe2S2;
2. Fe atoms are more exposed to the solvent in Fe4S4 than in Fe2S2;
3. Fe atoms in the tetrahedral coordination have the possibility to make new bonds, without breaking those already existing;
4. Fe atoms in the octahedral coordination must displace the ligands (all strong ligands, except for the water molecule on Fe_d), before binding O_2.

These observations reinforce the probability for electron donations from the Fe4S4 cluster, in the resting state, to the oxidizing O_2 molecule. Empirical models of the structure of the N-terminal region of HydA2 provide metastable conformations with low exposure of the Fe4S4 cluster to the solvent, thus indicating that it is possible to act on the N-terminal sequence for modulating the resistance of HydA2 to oxidation and further unfolding. The stability of these conformations compared to those with high solvent accessibility will be investigated in detail via further intensive numerical simulations.

Acknowledgements

The Namd computations were performed with a grant of computer time provided by NIC (2009-2010) on the Juropa HPC architecture, Jülich (64 parallel tasks). The Quantum-Espresso computations were performed on the Jump HPC architecture, Jülich (32 parallel tasks).

References

1. Anastasios Melis, *Photosynthetic H2 metabolism in Chlamydomonas reinhardtii (unicellular green algae)*, Planta, **226**, 1075–1086, 2007.

2. Giuseppe Torzillo, Alberto Scoma, Cecilia Faraloni, Alba Ena, and Udo Johanningmeier, *Increased hydrogen photoproduction by means of a sulfur-deprived Chlamydomonas reinhardtii D1 protein mutant*, Int. J. Hydrogen Energy, **34**, 4529–4536, 2009.
3. Fraser A. Armstrong, *Hydrogenases: active site puzzles and progress*, Curr. Opin. Chem. Biol., **8**, 133–140, 2004.
4. Maurizio Bruschi, Claudio Greco, Piercarlo Fantucci, and Luca De Gioia, *Structural and electronic properties of the [FeFe]hydrogenase H-cluster in different redox and protonation states. A DFT investigation*, Inorg. Chem., **47**, 6056–6071, 2008.
5. Jordi Cohen, Kwiseon Kim, Paul King, Michael Seibert, and Klaus Schulten, *Finding gas diffusion pathways in proteins: Application to O2 and H2 transport in CpI [FeFe]-hydrogenase and the role of packing defects*, Structure, **13**, 1321–1329, 2005.
6. Silvio C. E. Tosatto, Stefano Toppo, Donatella Carbonera, Giorgio Mario Giacometti, and Paola Costantini, *Comparative analysis of [FeFe]hydrogenase from Thermotogales indicates the molecular basis of resistance to oxygen inactivation*, Int. J. Hydrogen Energy, **33**, 570–578, 2008.
7. F. Wayne Outten, *Iron-sulfur clusters as oxygen-responsive molecular switches*, Nature Chem. Biol., **3**, 206–207, 2007.
8. Ivano Bertini, Harry B. Gray, Edward I. Stiefel, and Joan Selverstone Valentine, (Eds.), *Biological inorganic chemistry: Structure and reactivity*, chapter XI, pp. 328–329, University Science Books, 2007.
9. Marc Forestier, Paul King, Liping Zhang, Matthew Posewitz, Sarah Schwarzer, Thomas Happe, Maria L. Ghirardi, and Michael Siebert, *Expression of two [Fe]hydrogenases in Chlamydomonas rheinardtii under anaerobic conditions*, Eur. J. Biochem., **270**, 2750–2758, 2003.
10. A. S. Pandey, T. V. Harris, L. J. Giles, J. W. Peters, and R. K. Szilagyi, *Dithiomethylether as a ligand in the hydrogenase H-cluster*, J. Am. Chem. Soc., **130**, 4533, 2008.
11. Paolo Giannozzi et al., *QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials*, J. Phys.: Condens. Matter, **21**, 395502, 2009.
12. Christopher H. Chang, Paul W. King, Maria L. Ghirardi, and Kwiseon Kim, *Atomic resolution modeling of the Ferredoxin:[FeFe] hydrogenase complex from Chlamyodomonas rheinardtii*, Biophys. J., **93**, 3034–3045, 2007.
13. James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten, *Scalable Molecular Dynamics with NAMD*, J. Comput. Chem., **26**, 1781–1802, 2005.
14. Sara Furlan, Giovanni La Penna, and Angelo Perico, *Modeling the free energy of polypeptides in different environments*, Macromolecules, **41**, 2938–2948, 2008.
15. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Jr. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules*, J. Am. Chem. Soc., **117**, 5179–5197, 1995.
16. Richard F. W. Bader, *Atoms in molecules - A quantum theory*, Oxford University Press, Oxford, UK, 1990.

Protein Folding and Structure Prediction at the Simulation Laboratory Biology

Jan H. Meinke, Sandipan Mohanty, and Olav Zimmermann

Institute for Advanced Simulation, Jülich Supercomputing Centre (JSC), Research Centre Jülich
52425 Jülich, Germany
E-mail: {j.meinke, s.mohanty, olav.zimmermann}@fz-juelich.de

Protein folding and structure prediction are two important aspects of the scientific work at the Simulation Laboratory Biology in Jülich, an integrated research and support facility aimed at enabling highly scalable software for biology. An efficient and reliable description of the intra- and intermolecular interactions plays a central role in both these areas. Here we present a comparison of folding simulation performed with the ECEPP/3 and the LUND force fields and show how the improvements in the force fields can help within a structure prediction framework.

1 Introduction

More than a thousand genomes including the human genome have been fully sequenced and collected in the Genome OnLine Database[1]. Based on the DNA sequence, we can obtain the amino acid sequence of a myriad of proteins, but for most of them we do not know their native, folded state. The physical processes through which these molecules reach their native conformations are also not thoroughly understood.

The first problem falls under the heading of protein structure prediction. Among different approaches explored to this date, the so called knowledge-based approaches that exploit local and global similarities to the surging number of experimentally determined protein structures have been the most successful. We can often guess the structure of a protein from the known structure of a different protein with a similar sequence. We can also use the observed conformational affinities of small sequence segments and assemble the structural motifs so obtained by fitting them to representatives of different types of folds. In this way, it is sometimes possible to get excellent templates that would never be found using sequence similarity. Sometimes, there are neither proteins with high sequence similarity nor any known proteins with a similar fold. In such cases, physics-based simulations can be resorted to, which, although much slower than the knowledge-based methods, do not depend on the existence of a previously studied protein of similar sequence.

Characterization of the folding process, requires physics-based simulations. Such simulations are highly non-trivial, because of the complexity of the interactions involved and the very large number of degrees of freedom. Inevitably a large number of competing approaches have been developed over the years. A "winning strategy" has yet to emerge. Extant competing approaches differ in the resolution at which proteins and their environment are represented, the approximations used in modeling the effective interactions, as well as the manner in which the system is probed. The different approximations of the interaction potential often go by the name "force fields". Some of these force fields require a detailed knowledge of the native structure and are therefore non-transferable[2]. More transferable force fields can be obtained by a statistical analysis of many experimental structures. Yet

another group of force fields are derived from detailed theoretical calculations and experimental data on various small molecules.

Both the problems discussed above are important for an understanding of how life works at a fundamental level as well as for optimizing our chances of survival against present and future diseases. These questions are at the interface of several disciplines of science. Their many formidable challenges still demand new substantial contributions from physicists, chemists, biologists and computer scientists. The Simulation Laboratory Biology at the Jülich Supercomputing Centre was founded with the vision that supercomputing can greatly aid progress in this area.

The Simulation Laboratory Biology is a research and support facility focusing on biologically motivated problems that can be addressed using high-performance computers. While there are many problems in biology that would benefit from massively parallel processing, only few algorithms to solve these problems have been developed and implemented for large-scale machines. One of our goals is to collect the existing algorithms and make them available on our systems. Another goal is to develop and implement new solutions in collaboration with other members of the computational biology community.

Physical simulations of biological systems, such as cell membranes and proteins have a long history in high-performance computing. These simulations usually evolve the system either by solving Newton's equations of motion in a molecular dynamics (MD) simulation, or use statistical approaches such as Monte Carlo (MC) to explore the thermodynamics of the system.

For the investigation of the folding process, we maintain, at present, two Monte Carlo simulation programs: ProFASi[3] written in C++ and SMMP[4–6] written in FORTRAN. Both programs can take advantage of several thousand processors, and we are continuously working on improving their scalability[7] and harnessing new computer platforms such as CELL processors[8] and graphic processing units (GPU). They implement a variety of stochastic algorithms such as simulated annealing, simulated tempering and parallel tempering as well as several force fields, e.g. LUND[9,10] and ECEPP[11]. The use of stochastic algorithms in these programs overcomes the problem of time scales that is inherent in an MD simulation. The shortest time interval relevant for MD simulations of proteins is the period of the vibration of a single bond at about 10^{-15}s while folding of a protein chain occurs in the millisecond to second regime. Many interesting questions about complex biomolecules involve inconveniently large time scales for MD simulations, where statistical methods offer a more efficient alternative. Since the price of covering large time scales quickly is a loss of kinetic information, the stochastic methods are complementary to MD simulations.

Both ProFASi and SMMP represent protein molecules in atomistic detail. The solvent molecules are not treated explicitly for computational efficiency. Their effects are accounted for through the effective interaction potentials. Furthermore, the bond lengths and bond angles are assumed to be fixed, so that the simulations are performed with only the dihedral degrees of freedom (dof). These simplifications allow us to fold systems as large as the C-terminal fragment of Top7 (2GJH, 49 residues) and the 3-helix bundle GS-α_3W (1LQ7, 67 residues) from a stretched or random configuration to their native states and collect information about their folding behaviour[12]. A typical domain in a protein, however, consists of approximately 150 amino acids, and simulations of sufficient length for such systems remain a great challenge even with the simplifications we have made.

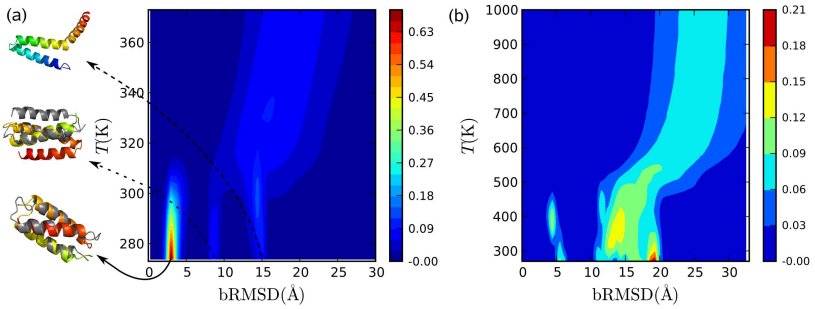


Figure 1. Comparison of the free energy landscape of 1LQ7 using (a) the LUND and (b) the ECEPP/3 force field.

Larger systems have larger number of participants in the interactions so that individual energy evaluations are more expensive. They also have many more degrees of freedom and larger conformational space for the simulations. For force fields with only short-range interactions, such as the LUND force field, the number of interactions can be kept in check by using cutoffs. Using techniques like cell lists and exploiting properties of the Monte Carlo algorithms, it is possible to achieve linear or even sub-linear scaling for the average cost of energy calculations. The number of possible configurations, however, grows exponentially. To be able to fold typical protein domains and understand their folding process, we need to further improve the force fields and the MC algorithms. Substantial computational power is needed in this endeavor. In the following, we will present some of our results for folding the protein GS-α_3W using the the LUND and ECEPP/3 force fields, as well as scenarios for the application of physics-based force fields in structure prediction.

2 Folding GS-α_3W

The three-helix bundle GS-α_3W was designed as a simple model of a tryptophanyl radical[13]. It consists of 67 amino acids and 1111 atoms. Previous computational studies of GS-α_3W used coarse-grained models[14]. We performed the simulations using two different force fields: ECEPP/3 and LUND[15,16]. The simulation with ECEPP/3 was done using SMMP and the simulation with the LUND force field used ProFASi.

For this study, we used parallel tempering Monte Carlo simulations. The free energy landscape of proteins is high dimensional, very rough, and contains deep local minima as well as large free energy barriers. The physically interesting states are normally close to the global free energy minimum at any given temperature. To find the global minimum in such a landscape and characterize it accurately, a simulation must both be able to cross large barriers and sample details around important local minima. Parallel tempering achieves this by means of a random walk through the temperature space. In parallel tempering, independent copies (replicas) of the system are simulated at different temperatures. After a sweep (1 sweep $= n_{\mathrm{dof}}$ Monte Carlo updates), we try an exchange of two replicas at neighbouring temperatures T_i and T_j. The acceptance probability is given by $P_{\mathrm{PT}}(i,j) = \min(1, \exp((\beta_i - \beta_j)(E_i - E_j))$, where $\beta_i = \frac{1}{k_B T_i}$ with k_B the Boltzmann

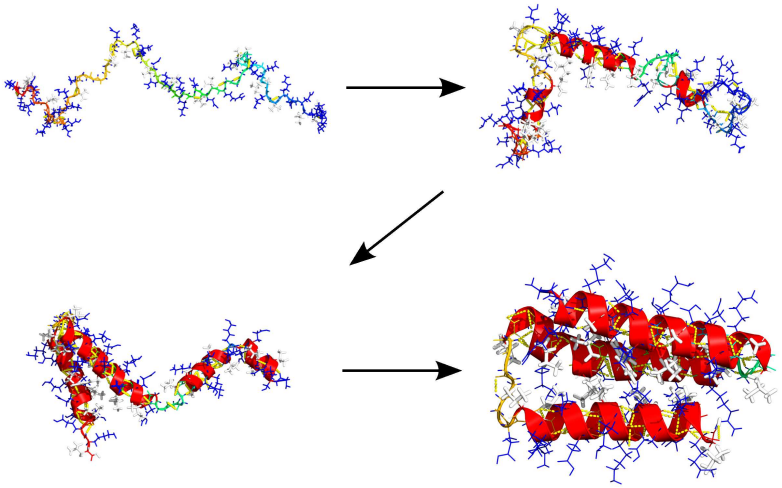


Figure 2. Schematics of the folding path of the three-helix bundle GS-α_3W. Taken from Ref. 15 with permission.

constant and E_i is the total energy at T_i. This exchange probability ensures that the replicas remain in thermal equilibrium while performing a random walk in temperature space. By moving to higher temperatures, replicas can escape local energy minima more easily, whereas at lower temperatures the simulation explores the finer details of a particular local minimum. During a simulation a replica may move back and forth between the highest and the lowest temperature many times. The temperature distribution chosen plays an important role in the effectiveness of this process.

In Figure 1, we show the free energy landscapes of GS-α_3W found with LUND and ECEPP/3 force fields. Both find free energy minima at low RMSD values. but the fraction of natively folded conformations with the LUND force field is about 80 % at the lowest temperature compared to about 15 % for ECEPP/3. The NMR analysis performed to determine the structure of GS-α_3W did not reveal any significant amounts of competing structures[13]. Thus the large population of non-native configurations at low temperatures with ECEPP/3 is likely to be an artifact. Furthermore, a single MC sweep using the ECEPP/3 force field takes 200 times the amount of CPU time compared to a single sweep using the much younger LUND force field (as implemented in ProFASi). This shows that the continued development of force-fields is important. It also shows that present day computing facilities are fast enough to collect sufficient statistics, so that genuine weaknesses of different force fields can be unambiguously identified. Such information fosters better understanding of the relevant interactions and the development of better force fields.

By looking at the changes in the observables with temperature and the configurations along a successful folding event, we can get information about the thermodynamics and the order of the folding process: The secondary structure of GS-α_3W forms very quickly starting with the central helix. Once all three helices have formed, two of the helices pair up and finally the third helix joins the other two in the native structure (see Figure 2).

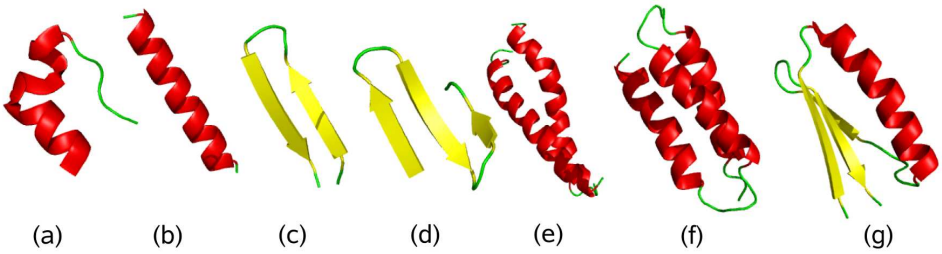


Figure 3. Schematic illustration of native geometries studied with ProFASi using the LUND force field. (a) the Trp-cage (1L2Y), (b) an α-helix, (c) a β-hairpin (1GB1,41-56), (d) a three-stranded β-sheet, (e) an α-helix dimer (1U2U), (f) a three-helix bundle (1LQ7), and (g) a mixed α/β protein (2GJH). Figure taken from Ref. 16 with permission.

3 The LUND Force Field

In contrast to ECEPP, CHARMM, and AMBER, which are derived from ground state properties, the LUND force field was developed as a folding force field. Starting from a simple ansatz for the interaction potential, the parameters and functional form of different terms in this force field were successively refined using high statistics folding simulations of well characterized small peptides. New sequences were studied at each stage of force field development. Proposed modifications to the force field were accepted if they described the properties of the new peptides while retaining their ability to describe all previously studied sequences.

The basic form of the LUND force field is given by $E = E_{\text{loc}} + E_{\text{ev}} + E_{\text{sc}} + E_{\text{hb}}$. The term E_{loc} represents local backbone electrostatics and a torsion term. These interactions are local along the sequence, and involve atoms belonging to one residue or to neighbouring residues along the sequence. The other three terms, E_{ev} representing excluded volume, E_{sc} representing side chain interactions, and E_{hb} representing hydrogen bonding are interactions non-local in sequence. For details of the force field see Ref. 16.

The current version of the LUND force field describes the folding of at least 17 peptides sized from 12 to 37 amino acids as well as a few small proteins including the C-terminal fragment of Top7 and GS-α_3W (see Figure 3 (f) and (g), respectively). For small peptides, the estimates for native population and the stability ordering of related sequences obtained from the simulations agree well with experimental data. Also, the set of peptides and proteins whose properties are well captured by the LUND force field includes chains of different secondary structures (α-helix, β-sheet), and different native folds. In addition a number of aggregation studies[17–19] have been performed using the LUND force field indicating its ability to represent protein-protein interactions. These capabilities make the LUND force field a viable candidate for applications such as structure prediction and docking.

4 Structure Prediction

The aim of structure prediction is to obtain an accurate model of the native structure of a protein from its amino acid sequence. Today's most successful structure prediction meth-

ods mainly rely on the identification of structural neighbours in databases of protein structures (PDB[20]) or protein folds (CATH[21], SCOP[22]). Often, knowledge-based potentials that have been derived from statistics of a large number of native structures are employed, as they are more robust to imperfections of the model than physics-based force fields. The Simulation Laboratory Biology investigates different ways to combine knowledge-based and physics-based approaches for structure prediction.

Depending on the degree of similarity to known structures there are three main roads for protein structure prediction from an amino acid sequence.

If the target sequence has a sequence identity $>$40%, global structural similarity can be assumed. After alignment of the sequences the structure for the backbone of matching regions is directly transferred to the target sequence. This approach is called homology modeling. Homology modeling can produce models with an RMSD of less than 2 Å, which is comparable to experimental resolution. Frequently, however, models have RMSD values between 3 and 6 Å, although the topology is correct. A critical point is the modeling of loops and the side chain placement. Physics based force fields may provide information complementary to knowledge-based approaches in both cases. Stochastic simulations with physics-based force fields could, e.g. provide information on the conformational space of a loop and thus should be of major relevance for ligand docking and drug design. The sensitivity to imperfections of the model make physics-based force fields a possible choice in the refinement process, i.e. in the exact placement of the side chain atoms. Recent attempts using an optimized version of the Amber force field look promising[23].

For many proteins there are no structures with detectable sequence similarity. A fold recognition technique called threading[24] exploits the fact that the fold space of proteins appears to be limited[25]. One can thus try in these cases to determine the best fit between the target sequence and each structure of a representative fold library. The coarse representation of the folds as C-alpha traces or polypeptide backbones makes this approach more suitable for knowledge-based potentials.

Some large groups of medically relevant proteins for which no or only few structures are known, e.g. membrane proteins and intentionally disordered proteins, are underrepresented in protein structure databases. In these and other cases where no similarity to known structures or folds are detectable one has to resort to *ab initio* approaches. Here we use machine learning based prediction of secondary and super-secondary structures[26, 27] to guide MC simulations with physics-based force fields, thereby reducing the search space considerably[28]. Due to the computational costs involved, the *ab initio* approach is particularly dependent on improvements in computational performance and scalability of the algorithms. It is thus a natural focus of the scientific work at the Simulation Laboratory Biology.

5 Conclusions

Force fields and search methods have come a long way. Using today's high-performance computers, we can investigate the thermodynamics of folding and compare the results with experimental findings for small proteins. Nevertheless, refinement and folding studies of large proteins with compact structures require further improvements of force fields as well as trial moves. The Simulation Laboratory Biology will continue its collaborative efforts to improve the quality of simulations and to produce highly scalable simulation codes.

In combination with the availability of unprecedented computational power, the size and timescales at which proteins can be simulated in atomistic detail will soon cover many biologically and medically important targets.

Acknowledgements

The simulations used in this paper were done in part using compute time grants provided by the John von Neumann Institute for Computing.

References

1. N. C. Kyrpides, *Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide*, Bioinformatics, **15**, no. 9, 773–774, 1999.
2. H. Abe and N. Go, *Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins.*, Biopolymers, **20**, no. 5, 1013 – 1031, 1981.
3. A. Irbäck and S. Mohanty, *PROFASI: A Monte Carlo simulation package for protein folding and aggregation*, J. Comput. Chem., **27**, 1548 – 1555, 2006.
4. F. Eisenmenger, U. H. E. Hansmann, S. Hayryan, and C.-K. Hu, *[SMMP] A modern package for simulation of proteins*, Comput. Phys. Commun., **138**, 192–212, Aug. 2001.
5. F. Eisenmenger, U. H. E. Hansmann, S. Hayryan, and C.-K. Hu, *An enhanced version of SMMP—An Open-Source Software Package for Simulation of Proteins*, Comput. Phys. Commun., **174**, 422–429, 2006.
6. J. H. Meinke, S. Mohanty, F. Eisenmenger, and U. H. E. Hansmann, *SMMP v. 3 — Simulating proteins and protein interactions in Python and Fortran*, Comput. Phys. Commun., **178**, no. 6, 459–470, 2007, You can download SMMP from our web page `http://smmp.berlios.de`.
7. J. H. Meinke and U. H. E. Hansmann, *Parallelization of ECEPP/3 in SMMP*, in: From Computational Biophysics to Systems Biology (CBSB07), vol. 36 of *NIC Series*, pp. 219 – 222, 2007.
8. L. Schubert, *Implementierung des Proteinkraftfeldes ECEPP/3 auf der Cell Broadband Engine*, Internal Report FZJ-JSC-IB-2009-03, Forschungszentrum Jülich, Aug. 2009.
9. A. Irbäck, B. Samuelsson, F. Sjunnesson, and S. Wallin, *Thermodynamics of α- and β- Structure formation in proteins*, Biophys. J., **85**, no. 3, 1466–1473, Sept. 2003.
10. A. Irbäck and S. Mohanty, *Folding Thermodynamics of Peptides*, Biophys. J., **88**, 1560–1569, 2005.
11. G. Nemethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, *Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides*, J. Phys. Chem., **96**, no. 15, 6472 – 6484, 1992.
12. S. Mohanty, J. H. Meinke, O. Zimmermann, and U. H. E. Hansmann, *Simulation of top7-cfr: a transient helix extension guides folding*, Proc. Natl. Acad. Sci. USA, 11 Apr. 2008.

13. Q.-H. Dai, C. Tommos, E. J. Fuentes, M. R. A. Blomberg, P. L. Dutton, and A. J. Wand, *Structure of a de Novo Designed Protein Model of Radical Enzymes*, J. Am. Chem. Soc., **124**, no. 37, 10952 – 10953, 2002.
14. A. Liwo, M. Khalili, and H. A. Scheraga, *Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains*, Proc. Natl. Acad. Sci. USA, **102**, no. 7, 2362–2367, 2005.
15. J. H. Meinke and U. H. E. Hansmann, *Free-energy-driven folding and thermodynamics of the 67-residue protein GS-α_3W - A large-scale Monte Carlo study*, J. Comp. Chem., **30**, no. 11, 1642—1648, June 2009.
16. A. Irbäck, S. Mitternacht, and S. Mohanty, *An effective all-atom potential for proteins*, PMC Biophys., **2**, no. 1, 2, 2009.
17. G. Favrin, A. Irbäck, and S. Mohanty, *Oligomerization of amyloid Aβ16–22 peptides using hydrogen bonds and hydrophobicity forces*, Biophys. J., **87**, no. 6, 3657–3664, 2004.
18. M. Cheon, I. Chang, S. Mohanty, L.M. Luheshi, C.M. Dobson, M. Vendruscolo, and G. Favrin, *Structural reorganisation and potential toxicity of oligomeric species formed during the assembly of amyloid fibrils*, PLoS Comput. Biol., **3**, 1727–1738, 2007.
19. D.W. Li, S. Mohanty, A. Irbäck, and S. Huo, *Formation and growth of oligomers: a monte carlo study of an amyloid tau fragment*, PLoS Comput. Biol., **4**, no. 12, 2008.
20. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *The Protein Data Bank*, Nucl. Acids Res., **28**, 235–242, 2000.
21. C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton, *CATH–a hierarchic classification of protein domain structures*, Structure, **5**, no. 8, 1093–1108, 1997.
22. A.G. Murzin, S.E. Brenner, T. Hubbard, and C. Chothia, *SCOP: a structural classification of proteins database for the investigation of sequences and structures*, J. Mol. Biol., **247**, no. 4, 536–540, 1995.
23. A. Jagielska, L. Wroblewska, and J. Skolnick, *Protein model refinement using an optimized physics-based all-atom force field*, Proc. Natl. Acad. Sci. USA, **105**, no. 24, 8268–8273, June 2008.
24. D.T. Jones and J.M. Thornton, *Potential energy functions for threading*, Cur, Opin. Struc. Biol., **6**, no. 2, 210–216, 1996.
25. D. Kihara and J. Skolnick, *The PDB is a covering set of small protein structures*, J. Mol. Biol., **334**, no. 4, 793–802, 2003.
26. O. Zimmermann, L. Wang, and U. H. E. Hansmann, *BETTY: prediction of beta-strand type from sequence*, In Silico Biol., **7**, 0037, 2007.
27. O. Zimmermann and U. H. E. Hansmann, *LOCUSTRA: accurate prediction of local protein structure using a Two-Layer support vector machine approach*, J. Chem. Inf. Model., **48**, no. 9, 1903–1908, 2008.
28. C. Honisch, *Protein Simulations under Constraints*, Internal Report FZJ-JSC-IB-2008-07, Forschungszentrum Jülich, Dec. 2008.

Chemistry: Atomistic Simulation from First Principles

Wim Klopper

Theoretical Chemistry Group, Department of Chemistry and Biosciences
Karlsruhe Institute of Technology, KIT Campus South
Kaiserstraße 12, 76131 Karlsruhe, Germany
E-mail: klopper@kit.edu

Ab initio quantum chemistry and first-principles atomistic simulations have developed into indispensable parts of modern chemistry and chemical biology. They provide us with concepts and simulation approaches that are central to our understanding of structural, electric, magnetic, and optical properties of (bio)chemical systems and their reaction dynamics. Powerful computer programs have been developed that, in many areas of chemistry, provide us with numerical data about chemistry, helping us to understand, confirm, or predict experimental observations and measurements.

In their contribution *Ionic Liquids*, S. Zahn and B. Kirchner report on computational investigations of ionic liquids (IL), salts with a melting temperature below 100 °C. They combine static quantum chemistry methods—for structures and energetics—with first-principles molecular dynamics (FPMD) computations—for elucidating ion dynamics. The contribution shows various aspects of modern quantum chemical computations: High-level symmetry-adapted perturbation theory (SAPT) calculations are performed to analyse the intermolecular interactions in IL systems, Møller–Plesset perturbation theory benchmark data are used to validate density functional theory augmented with corrections for dispersion (*e.g.* DFT-D), and FPMD calculations are carried out to study the ion dynamics in methylammonium nitrate.

In his contribution *Atomistic Simulations on Polymeric Nitrogen*, J. Kotakoski has investigated solid nitrogen for wide ranges of temperature and pressure. Also in this contribution, various computational approaches are combined. For example, first-principles plane-wave density functional theory calculations are performed to investigate the thermodynamical stability of different phases, and a new analytical interaction model is developed for use in classical molecular dynamics simulations.

In their contribution *Prebiotic Peptide Synthesis on Blue Gene Platforms at "Iron–Sulfur–World" conditions*, N. N. Nair, E. Schreiner, and D. Marx show how access to significant Blue Gene resources have helped to conduct a comprehensive *in silico* study of a complex network of biochemical reactions at a mineral/water interface. Car–Parrinello *ab initio* molecular dynamics simulations are performed to investigate peptide bond formation under vastly different reaction conditions.

In their contribution *Calculation of* ^{31}P *and* ^{183}W *NMR Chemical Shifts and Nuclear Spin Spin Coupling Constants in Phosphinidenoid and Other Transition Metal Complexes by the ZORA DFT Method*, G. von Frantzius and R. Streubel provide an example of how (static) quantum chemical calculations can help to understand the electronic structure and chemical bonding in novel chemical systems, here tungsten phosphinidenoid complexes.

At the *ab initio* density functional theory level, compliance constants (elements of the inverted Hessian) and NMR spectroscopic properties are computed for a series of model compounds. As a descriptor for bond strength, the tungsten–phosphorous compliances of ten model complexes correlate well with the tungsten–phosphorous bond length. Computed $\sigma(^{31}\mathrm{P})$ and $\sigma(^{183}\mathrm{W})$ absolute isotropic shieldings are analysed in terms of para-, diamagnetic, and spin–orbit contributions.

In conclusion, the four "chemistry" contributions in the present proceedings of the NIC Symposium 2010 provide beautiful examples of the variety of first-principles, quantum mechanical calculations and simulations that can be performed today to study complex chemical systems at the atomistic scale, yielding valuable structural and dynamical data.

Ionic Liquids

Stefan Zahn and Barbara Kirchner

Wilhelm-Ostwald-Institut für Physikalische und Theoretische Chemie
Universität Leipzig, 04318 Leipzig, Germany
E-mail: bkirchner@uni-leipzig.de

Our project investigates ionic liquids, salts with a melting temperature below 100° C. We combine high-level static quantum chemistry calculations, *first principles* and classical molecular dynamics simulations to point out features of ionic liquids at the molecular level which are responsible for macroscopic properties of these environmentally friendly taskspecific solvents. With the help of the three previous mentioned approaches we were able to discover one origin of the low melting point of ionic liquids. Furthermore, we revealed the role of hydrogen bonds in depressing the melting point of imidazolium based ionic liquids.

1 Introduction

Ionic liquids (ILs) are salts with a melting temperature below 100° C and have become a hot spot in a wide range of applications in recent years[1–6]. Already known since the beginning of the last century[7], fundamental questions like the origin of the low melting temperature compared to typical salts are focus of recent scientific discussions[2,8–13]. Because ILs show often unique properties as solvents and have a low vapor pressure, they are environmentally friendly taskspecific solvents. Nevertheless, the purposeful design of new task related ILs needs the ken of two important facts:

- the origin of the low melting temperature to design an IL instead of an ionic solid
- features of ILs responsible for the unique solvent properties.

Our project investigates the forces, structure and dynamics of ionic liquids employing approaches from static quantum chemistry as well as *first principles* molecular dynamics (FPMD). The combination of these approaches provide an insight into the structure of ILs and can point out the origin of the discovered features. This may improve the specific design of new task related ILs.

2 What Keeps Ionic Liquids in Flow?

In this regard it is interesting to consider the theory of Goldstein, who stated in 1969, "that portion of the potential energy surface that represents the liquid or glassy region has, unlike the portion associated with the crystalline solid, a large number of minima, of varying depths"[14]. According to Goldstein's theory, shallow energy potentials and low transition states between minima result in a gliding of the molecules and, thus, in a low melting temperature.

In order to prove the role intermolecular forces in depressing the melting temperature due to a flattening of the potential surface, we applied the symmetry-adapted perturbation theory[15] (SAPT) to 1,3-dimethyl-imidazolium chloride ([C_1mim][Cl]), monoethyl-trimethyl-ammonium chloride ([N_{2111}][Cl]), 1-methyl-pyridinium chloride ([C_1Py][Cl])

and sodium chloride ([Na][Cl])[11,13]. In all three classes of ionic liquids the equilibrium structure is not exclusively determined by the major force, i.e. the attractive electrostatic interaction. Dispersion and induction forces have a significant impact on interaction energy. Furthermore, the obtained equilibrium structure is shifted at least 40 pm towards the repulsive part of the electrostatic-exchange potential, compare lines with circles and diamonds in Figure 1a–c. This is in contrast to a typical ionic solid like [Na][Cl], where dispersion forces play no role, see Figure 1d. As a result of the intermolecular forces interplay, in ionic liquid ion pairs a shallower interaction potential is obtained compared to [Na][Cl], see Figure 1e. This may explain one origin of the low melting temperature for ionic liquids according to the energy landscape paradigm.

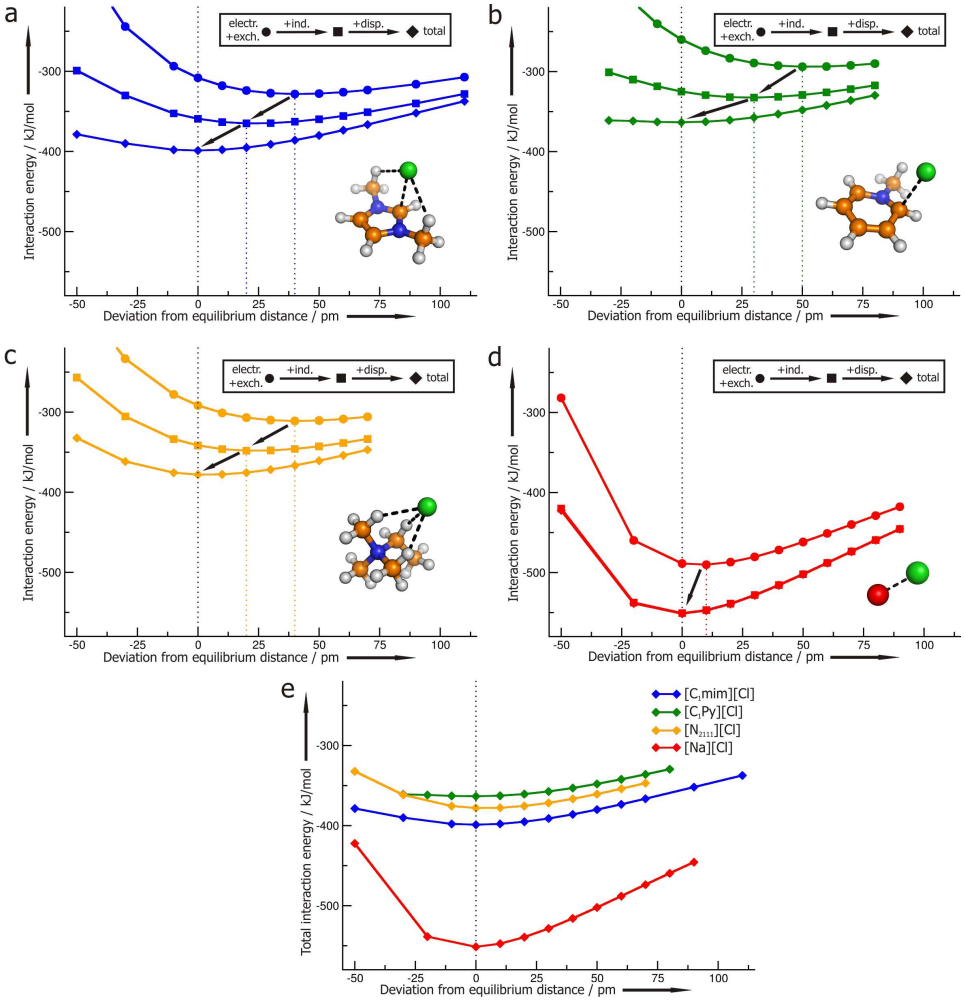


Figure 1. Interaction energies versus distance for [C_1mim][Cl] (a), [C_1Py][Cl] (b), [N_{2111}][Cl] (c) and [Na][Cl] (d). Furthermore, comparison of total interaction energy versus distance of all four investigated compounds (e).

3 Why Does a Reduction of a Hydrogen Bond Increase the Melting Temperature of an Ionic Liquid?

The substitution of the hydrogen at C2 of 1,3-dialkyl-imidazolium based ionic liquids with a methyl group replaces a hydrogen bond with a repulsive interaction and therefore a lower melting point and viscosity is expected. However, the opposite is observed[16,17]. In our investigation we combined static quantum chemistry calculations and classical molecular dynamics simulations to determine the impact of the hydrogen bond / methyl substitution on the energy potential surface[13].

At first we compared for 1,3-dimethyl-imidazolium chloride ([C_1mim][Cl]) and 1,2,3-trimethyl-imidazolium chloride ([C_1C_1mim][Cl]) the energy of the fluctuation of the anion from the structure where the anion resides above the imidazolium plane $\mathbf{D}_B/\mathbf{T}_B$ to the structure where the anion resides in front of the hydrogen bond/the methyl-substitution $\mathbf{D}_A/\mathbf{T}_A$ and finally to structure $\mathbf{D}_B/\mathbf{T}_B$ where the anion is on the opposite side of the imidazolium plane than in the starting point, see Figure 2a. The energy barrier for the side change is below 10 kJ/mol for the lower melting [C_1mim][Cl], while it is increased above 40 kJ/mol for [C_1C_1mim][Cl]. Other pathways for the side change of the anion can be excluded.

Inspired by these results we carried out classical molecular dynamics simulations of [C_2mim][Cl] and [C_2C_1mim][Cl] to determine the structural impact of the hydrogen substitution at C2 by a methyl group upon the liquid structure. In agreement with our static quantum chemistry calculations we find different probable positions if we compare [C_2mim][Cl] with [C_2C_1mim][Cl]. For [C_2mim][Cl] we observe a large sphere in front of the hydrogen donor atom H2 which slightly smears out above and below the imidazolium ring, see Figure 2b. In contrast to this no sphere is found in front of the middle methyl group in [C_2C_1mim][Cl] but a thin path around all methyl groups leading to locations above and below the imidazolium ring is apparent, see Figure 2c. The plot of the spatial distribution function at a value of 15 times of the average density (pink spheres in Figure 2b

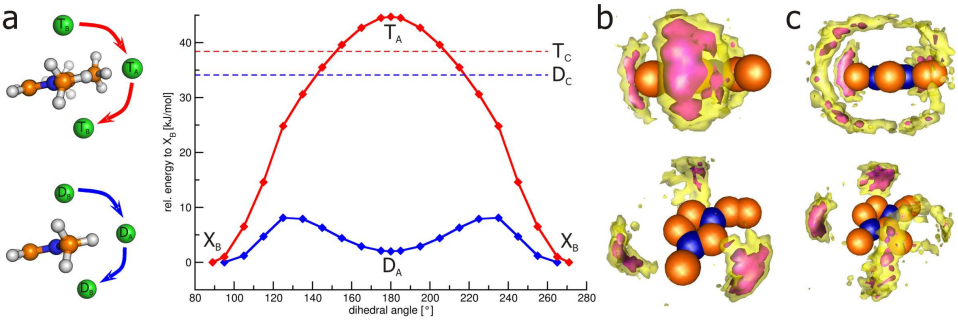


Figure 2. *a:* Potential energy surface from $\mathbf{D}_B/\mathbf{T}_B$ to $\mathbf{D}_B/\mathbf{T}_B$ (opposite side) via conformer $\mathbf{D}_A/\mathbf{T}_A$ in dependence of the dihedral angle Cl-C2-N1-C5 for [C_1mim][Cl] (*blue*) and [C_1C_1mim][Cl] (*red*). $\mathbf{D}_C/\mathbf{T}_C$ are minima structures where the anion is located between the methyl group at C1 and the hydrogen at C5.; *b+c:* Spatial distribution function of chloride for [C_2mim][Cl] (b) and [C_2C_1mim][Cl] (c) around the imidazolium plane. The isosurface is plotted at 8 (yellow outer surface) and 15 (pink interior surface) times the corresponding average density.

and 2c) reveals that the anion of [C_2C_1mim][Cl] (Figure 2c) is more localized than the one of [C_2mim][Cl] (Figure 2b). A larger sampled space corresponds to a larger part of the energy landscape with the same upper energy level. Therefore, these spatial distribution functions show similar to the static quantum chemistry approaches that the hydrogen bond enhances the mobility of the anion in imidazolium based ionic liquids. A reduction of the hydrogen bond at C2 results in a significant increase on the potential surface and therefore an increase in the melting point is observed.

4 Validation of Kohn-Sham Density Functional Theory for Ionic Liquid Systems

As discussed in section 2 weak dispersion forces have an important impact on the equilibrium structure and the interaction energy of ionic liquids. Post–Hartree–Fock methods provide proper descriptions of dispersion forces but are computationally difficult for systems larger than tens of atoms. On the other hand, Kohn-Sham density functional theory, with computational cost much lower than that of conventional correlated methods, accounts for electrostatic, exchange and induction forces very well but fails for the right description of dispersion forces[18–20]. Our investigations focus on the performance of the Kohn-Sham density functional theory for ionic liquids by common density functionals as well as two proposed approaches to improve the performance of Kohn-Sham density functional theory[21]. The first approach (DFT–D) includes a dispersion correction in the functional by a sum of damped interatomic potentials[22] while the second one includes a dispersion correction in the electronic structure by a dispersion-corrected atom-centred potential (DCACP)[23].

The investigation of several 1-butyl-3-methyl-imidazolium dicianamide ion pairs shows a mean absolute deviation from Møller-Plesset perturbation theory of 35.7 kJ/mol for Hartree–Fock and up to 33.2 kJ/mol for the density functional theory methods. The dispersion-corrected methods reduce the mean absolute deviation to less than 10 kJ/mol. Furthermore large deviations in geometries for the intermolecular distances were found for the Hartree–Fock approach (mean absolute deviation: 190 pm) and density functional theory (mean absolute deviation up to 178 pm) while for the dispersion–corrected methods the mean absolute deviation is less than 50 pm.

5 Structure and Dynamics of a Protic Ionic Liquid

The classification of ionic liquids has proven difficult. One possible subset of ILs is the family of protic ionic liquids (PILs), which are formed through the proton transfer from a Brønsted acid to a Brønsted base. As a result, PILs possess a proton-donor and -acceptor site, which can build up a hydrogen bond network. Some features like the high dielectric permittivity of ammonium based PILs are unique, but PILs and aprotic ILs share also properties and these are far from being understood. The fundamental understanding of ILs is achieved more easily with a simple compound, because numerous complications are thereby eliminated. Especially the PILs investigated by P. Walden in 1914[7] possess nearly the simplest molecular structure of all known ILs. Therefore, these ammonium based PILs provide an excellent starting point for fundamental research on ILs.

We investigated mono-methylammonium nitrate (MMAN) by means of *first principles* molecular dynamics simulations[24]. Dispersion effects were considered by a dispersion-corrected atom-centred potential meeting the requirements of our recent investigations, see section 4. No proton transfer occurred within the duration of the simulation. Hydrogen bonds between the cation and anion were established showing a large degree of directionality. Both ions hold three sites to form hydrogen bonds while only 1.8 contacts were actually used for hydrogen bonding. Only a negligible amount of two-fold (two protons bind one acceptor) or bifurcated (two acceptor bind one proton) hydrogen bonds existed in ion pairs of MMAN. Therefore, one hydrogen bond acceptor and one donor site in each ion pair of MMAN remained free. Interestingly, the same amount of free hydrogen bond donor and acceptor sites can be found in water pointing towards a similar hydrogen bond network. Furthermore, water and mono-methylammonium nitrate both exhibit a fast fluctuating hydrogen bond network. The comparable hydrogen bond network and dynamics of both liquids might explain the similar impact on reactivity and selectivity found for chemical reactions. However, the hydrogen bond network showed also differences. Considering the orientation of the hydrogen bonds in MMAN with respect to each other, we found a preference of 0° ,90° and 180° while in water the hydrogen bonds preferred an angle below 90°.

The first solvation shell of cations around themselves as well as the first solvation shell of anions around themselves divides into two subshells. Hydrogen bonded ion pairs arranged differently depending on whether they are located in the inner or in the outer subshell. While the orientation of two cations interconnected via a hydrogen bonded anion held angles between 50° and 110° in the inner subshell, they preferred an angle above 80° in the outer subshell. Two anions interconnected via a hydrogen bonded cation gave a perpendicular arrangement in the inner subshell, and a parallel arrangement is preferred in the outer subshell.

The analysis of the ion dynamics revealed long lived ion pairs that exceeded with their life time our simulation time by far. If we considered the life time of ion pairs that start from certain conformations, we found that ion pairs starting from a hydrogen bonded conformation lived longest followed by those ion pairs that start with the anions in close proximity to the ammonium group. The shortest life span was observed for those ion pairs that begin with the anions in close proximity to the methyl group. If we only considered the life time of the conformation the picture changes. The conformations decayed within 0.5 ps to 9 ps. The hydrogen bonded ion pair conformation existed shortest, followed by the conformations of those ion pairs that keep the anions in close proximity to the ammonium group. The longest life span was observed for those ion pairs that keep the anions in close proximity to the methyl group. This indicates the longest memory for conformations in which the anion and the methyl group are close. However the measured reassociation was much faster for the hydrogen bonded conformation than for the methyl group conformation. In order to clarify this behaviour we introduced an initial conformation memory function. This function showed indeed that hydrogen bonded ion pairs had the longest memory of initial conformation. If one considered the first change from one sort of conformation to the other it is apparent that ion pairs that start from the methyl conformation behave very simple in their memory loss, i.e the function just decays linearly. The process was more diverse if the initial conformation is started from either the ammonium group ion pair or from hydrogen bonded ion pairs. Both changing from an ammonium to a methyl or hydro-

gen bonded conformation results in a very fast memory loss with a non-linear behaviour. This shows that a fast cation side change of the anion is possible and in the latter case the hydrogen bond forms and breaks very fast. Interestingly the change between hydrogen bonded ion pair conformation and the ion pair conformation where the anion is closer to the ammonium group is very fast while the side change between ammonium group and methyl group is slower and depends on wether the anion moves towards (slow) or from the methyl group (slowest). This diversity of dynamics infers the following picture. The ion pairs are conserved in a cage, because ion pairs are long lived. However, there are many fluctuations and conformational changes occurring, which is reflected in the very short conformation life times. Therefore, the ion pair seems to rattle in a cage of neighbouring cat- and anions, see Figure 3 for an illustration.

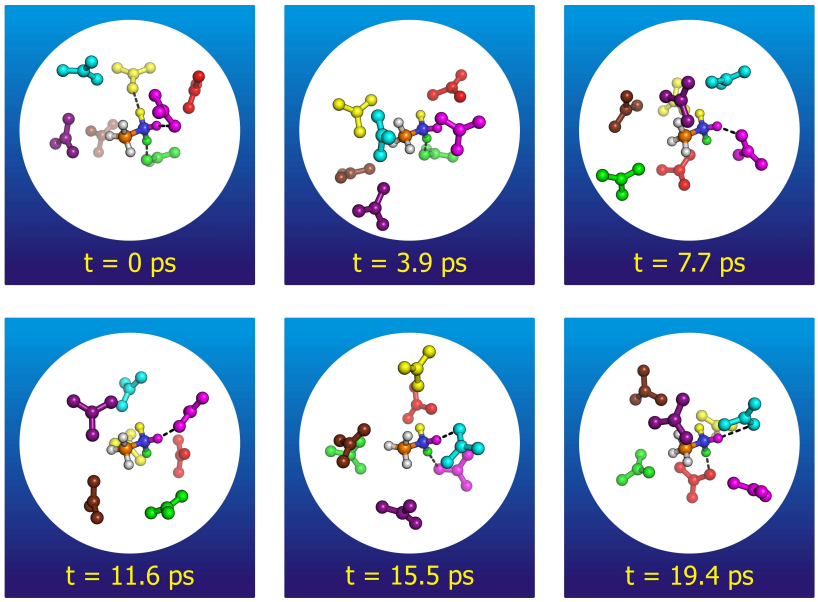


Figure 3. Illustration of the ion dynamics of mono-methylammonium nitrate. While all anions are kept in the first solvation shell of the cation, the cation rattles within the cage formed by the anions.

6 Why Are Ionic Liquid Ions Mainly Associated in Water?

Interestingly, ion pairs of a typical salt have a longer life time in water than ion pairs of ionic liquids. This is unexpected, since typical salts like sodium chloride have a higher interaction energy than ionic liquids. We investigated the ion pair association of an ionic liquid solvated in water by *first principles* molecular dynamics simulations[25]. Our investigation revealed whereas the interface of the cationic and anionic solvation shells in case of spherical ionic compounds (e.g., Na^+ and Cl^-) shows rather strong structural differences from the hydrogen bond network in pure water, the tangential orientation of water

around the hydrophobic parts of the ionic liquid cation fits better to the orientation around the anion in terms of hydrogen bonding, see Figure 4. This yields a favouring contribution towards the ion pair association of the ionic liquid. Therefore, the structure of the solvation shell interface determines whether ion pair dissociation occurs or not.

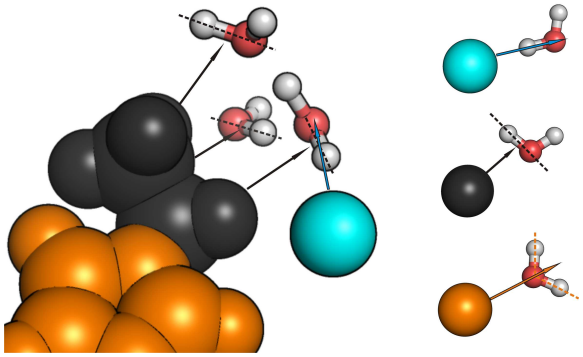


Figure 4. Illustration of the orientation of water molecules in solvation shells around different functional parts of the [C_2mim][Cl] ion pair. *cyan*: anionic part; *orange*: cationic part; *black*: hydrophobic part.

Acknowledgements

This work was supported by the DFG, in particular by the projects KI-768/4-1 from the ERA-chemistry and KI-768/5-2 SPP-IL program.

References

1. T. Welton; *Room-Temperature Ionic Liquids. Solvents for Synthesis and Catalysis.*, Chem. Rev. **99**, 2071–2084, 1999.
2. P. Wasserscheid and T. Welton; *Ionic Liquids in Synthesis*, (Wiley-VCH, Weinheim, 2003).
3. R. D. Rogers and K. R. Seddon; *Ionic liquids–solvents of the future?*, Science **302**, 792–793, 2003.
4. K. Binnemans; *Ionic liquid crystals.*, Chem. Rev. **105**, 4148–4204, 2005.
5. P. Wasserscheid; *Volatile times for ionic liquids.*, Nature **439**, 797, 2006.
6. H. Weingärtner; *Understanding ionic liquids at the molecular level: facts, problems, and controversies.*, Angew. Chem. Int. Ed. **47**, 654–670, 2008.
7. P. Walden; *Ueber die Molekulargrösse und elektrische Leitfähigkeit einiger geschmolzenen Salze.*, Bull. Acad. Sci. St Petersburg , 405–422, 1914.
8. E. A.Turner, C. C. Pye and R. D. Singer; *Use of ab Initio Calculations toward the Rational Design of Room Temperature Ionic Liquids.*, J. Phys. Chem. A **107**, 2277–2288, 2003.

9. S. Tsuzuki, H. Tokuda, K. Hayamizu and M. Watanabe; *Magnitude and directionality of interaction in ion pairs of ionic liquids: relationship with ionic conductivity.*, J. Phys. Chem. B. **109**, 16474–16481, 2005.
10. I. Krossing, J. M. Slattery, C. Daguenet, P. J. Dyson, A. Oleinikova and H. Weingärtner; *Why are ionic liquids liquid? A simple explanation based on lattice and solvation energies.*, J. Am. Chem. Soc. **128**, 13427–13434, 2006.
11. S. Zahn, F. Uhlig, J. Thar, C. Spickermann and B. Kirchner; *Intermolecular Forces in an Ionic Liquid ([Mmim][Cl]) versus Those in a Typical Salt (NaCl).*, Angew. Chem. Int. Ed. **47**, 3639–3641, 2008.
12. K. Fumino, A. Wulf and R. Ludwig; *Strong, localized, and directional hydrogen bonds fluidize ionic liquids.*, Angew. Chem. Int. Ed. **47**, 8731–8734, 2008.
13. S. Zahn, G. Bruns, J. Thar and B. Kirchner; *What keeps ionic liquids in flow?*, Phys. Chem. Chem. Phys. **10**, 6921–6924, 2008.
14. M. Goldstein; *Viscous Liquids and the Glass Transition: A Potential Energy Barrier Picture.*, J. Chem. Phys. **51**, 3728–3739, 1969.
15. B. Jeziorski, R. Moszynski and K. Szalewicz; *Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes.*, Chem. Rev. **94**, 1887–1930, 1994.
16. P. Bonhôte, A.-P. Dias, N. Papageorgiou, K. Kalyanasundaram and M. Grätzel; *Hydrophobic, Highly Conductive Ambient–Temperature Molten Salts.*, Inorg. Chem. **35**, 1168–1178, 1996.
17. L. Cammarata, S. G. Kazarian, P. A. Salter and T. Welton; *Molecular states of water in room temperature ionic liquids.*, Phys. Chem. Chem. Phys. **3**, 5192–5200, 2001.
18. S. Kristyán and P. Pulay; *Can (semi)local density functional theory account for the London dispersion forces?*, Chem. Phys. Lett. **229**, 175–180, 1994.
19. J. M. Pérez-Jordá and A. D. Becke; *A density-functional study of van der Waals forces: rare gas diatomics.*, Chem. Phys. Lett. **233**, 134–137, 1995.
20. E. J. Meijer and M. Sprik; *A density-functional study of the intermolecular interactions of benzene.*, J. Chem. Phys. **105**, 8684–8689, 1996.
21. S. Zahn and B. Kirchner; *Validation of Dispersion–Corrected Density Functional Theory Approaches for Ionic Liquid Systems.*, J. Phys. Chem. A **112**, 8430–8435, 2008.
22. S. Grimme; *Semiempirical GGA-type density functional constructed with a long-range dispersion correction.*, J. Comput. Chem. **27**, 1787–1799, 2006.
23. I–C. Lin, M. D. Coutinho–Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli and U. Rothlisberger; *Library of dispersion–corrected atom–entered potentials for generalized gradient approximation functionals: Elements H, C, N, O, He, Ne, Ar, and Kr*, Phys. Rev. B **75**, 205131, 2007.
24. S. Zahn, J. Thar and B. Kirchner; *Structure and dynamics of the protic ionic liquid mono-methylammonium nitrate ($[CH_3NH_3][NO_3]$)*, submitted to J. Chem. Phys.
25. C. Spickermann, J. Thar, S. B. C. Lehmann, S. Zahn, J. Hunger, R. Buchner, P. A. Hunt, T. Welton and B. Kirchner; *Why are ionic liquid ions mainly associated in water? A Car–Parrinello study of 1-ethyl,3-methyl-imidazolium chloride water mixture.*, J. Chem. Phys. **129**, 104505, 2008.

Atomistic Simulations on Polymeric Nitrogen

Jani Kotakoski

Institut für Materialwissenschaft, Technische Universität Darmstadt
Petersenstr. 32, D-64283 Darmstadt, Germany
E-mail: jani.kotakoski@iki.fi

Single-bonded polymeric form of nitrogen, formed under high pressures, has been considered as a promising candidate for environmentally safe high energy density material. However, synthesis of this material has proven to be exteremely difficult due to the complex enthalpy landscape. We have conducted density functional theory calculations on the way towards drawing the full phase diagram to assist the experimentalists on this quest. First, we compared the most promising candidates for metastable or stable structures, and then set out to find new stable high pressure phases after we noticed that the previously considered structure was dynamically unstable. We also developed an analytical interaction model for classical molecular dynamics which can be used to study the synthesis process at atomistic precision and to explain the pitfalls while releasing the pressure from the synthesis conditions.

1 Introduction

Molecular nitrogen can be transformed under high compression into a variety of chemically bonded solid phases. This transformation to non-molecular phases is of fundamental interest for understanding the physics of both molecular solids and the chemistry of nitrogen. It may also have applications in energy technology, since non-molecular nitrogen can potentially be used as high energy density material (more than 140 kJ/mol or 2.0 MJ/g *i.e.* roughly 500 times larger value than for TNT). During the recent years, after the experimental realization[1] of non-molecular crystalline nitrogen structure, *cubic gauche*, *cg* (see Figure 1) – proposed in 1992 based on creative thinking and density functional theory (DFT) calculations[2] – several possible new non-molecular phases have been suggested by theorists[3–13]. The multitude of proposed phases calls for a consistent comparison of their thermodynamical stability allowing to draw the phase diagram of materials under a wide pressure and temperature range. This would be of vital importance for the experimentalists working in the field. Moreover, as the high pressure experiments are extremely difficult, simulations can also assist in designing experiments which would maximize the probability for a successful synthesis.

2 Method

In order to fully understand the behaviour of nitrogen under various conditions, the free energies of the various phases must be calculated under different pressures and temperatures. From these energies, we can then deduce the phase diagram of the material. The first step in such calculations is to compare the zero temperature enthalpies of the phases for the complete pressure range. While the proposed structures have been compared using a wide variety of different methods, we carried out an extensive comparison[14] regarding their thermodynamical stability using plane-wave basis set DFT code VASP[15] with the general

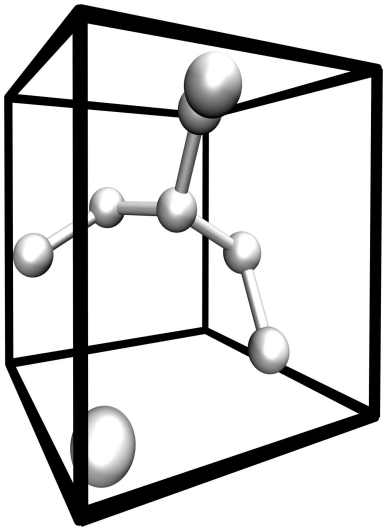


Figure 1. Cubic unit cell (twice the primitive cell) of the *cg* structure.

gradient approximation (GGA) exchange-correlation functional by Perdew *et al.*[16] . The core ($1s^2$) electrons were described with a projector augmented wave pseudopotential[17,18]. A high kinetic energy cutoff (1200 eV) was used due to the lack of p or d states in the core. Brillouin zone sampling was performed using the scheme of Monhorst-Pack[19] (including the Γ-point). For consistency, a $13 \times 13 \times 13$ k-point mesh was used for all the relaxation simulations.

To estimate the dynamical stability of the different nitrogen phases, we also calculated the phonon dispersion relation following the scheme proposed by Kunc and Martin[20]. Within this method, a supercell of sufficient size to limit the forces inside the cell is first created, and all irreducible directions within the cell are searched for. After this, each plane of atoms perpendicular to each direction is displaced along the 3 degrees of freedom to calculate the forces exerted on all atoms by the displacement. The force constant matrix is then constructed and used to calculate the dynamical matrix Ψ_k for the primitive cell for different $\mathbf{k}$ vectors along the high symmetry axes. The square roots of the eigenvalues, $\omega_i(\mathbf{k})$, of $\Psi_{\mathbf{k}}$ give the frequencies of the phonon modes for each k. Imaginary frequencies correspond to directions of barrierless phase transitions indicating dynamical instability of the phase.

3 Phase Comparison

In this study, we included two molecular structures (α-N_2 and ϵ-N_2), one chain-like structure (coordination $N_c = 2$) and seven of the most likely structures proposed before the time of the study ($N_c = 3$). Our results agree well with the previous calculations[2], although for the layered phases we obtained qualitative agreement only with a fixed simulation cell. When the lattice vectors were relaxed, we observed that – similar to graphite – it is energetically favoured to separate the layers rather than to expand the structure in all three

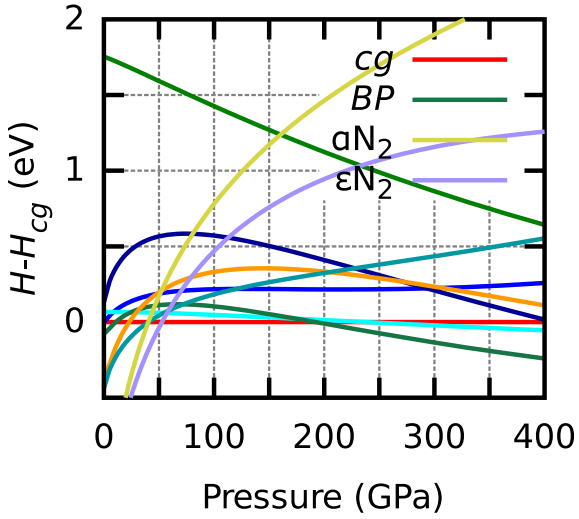


Figure 2. Enthalpy comparison of the relevant phases. Only molecular phases with *cg* and *bp* are labeled.

dimensions, leading to a flat tail in the equation of state curve (energy as a function of the atomic volume). Also the properties calculated for *cg* agree well with the results published by others[2,1,21]. The enthalpies of the various structures as compared to that of *cg* is plotted in Figure 2 as a function of the pressure.

The next step towards calculating the complete phase diagram would be to include the temperature into the equation via phonon density of states calculations for all phases under the complete range of pressures. This work has partly been already carried out[14]. However, while many new phases have been proposed[7,10–13] since the last comparison was made, the previously made calculations must be extended.

3.1 High Pressure Phases

One of the curious features we noticed while calculating the phonon spectra for some of the structures, was that the phase with the lowest enthalpy at pressures above 200 GPa (*black phosphorus, BP*), is dynamically unstable at pressures close to 200 GPa[14]. The obvious conclusion from this fact is that there must exist another structure with lower enthalpy, perhaps not previously observed because of a large number of atoms in the primitive unit cell. Keeping this in mind, we started to look for new high-pressure structures for nitrogen using an evolutionary search method with DFT as the method for total energy calculation[13]. We were able to identify four new structures. Still, after including these in the enthalpy comparison, the *cg* remains most stable up to pressure of 188 GPa. One new structure – an orthorhombic $Pba2$ is stable at the pressure range of $188 - 320$ GPa, and above 320 GPa the $P2_12_12_1$ is the ground state. All four described structures displayed real phonon modes (no dynamical instabilites) throughout the interesting pressure ranges.

The optimized lattice parameters at 250 GPa for the $Pba2$ structure are $a = 4.17$ Å, $b = 4.15$ Å and $c = 3.95$ Å with N atoms occypying four inequivalent Wyckoff $4c$ positions: (0.2933, 0.2172, 0.6608), (0.7852, 0.2167, 0.3403), (0.0209, 0.3363, 0.177) and

(0.8381, 0.0171, 0.8219). This is a layered structure, which consists of N_4 units (one central N atom being covalently bonded to three others, among which two of such N-N single bonds are equivalent). In contrast to the layered structure, the orthorhombic $P2_12_12_1$ adopt fully 3D packing consisting of helical tunnels connected to each other by covalent N-N bonds. At 350 GPa the optimized lattice parameters are $a = 4.01$ Å, $b = 3.31$ Å, $c = 2.39$ Å with atomic positions: N1 at $4a$ (0.8298, 0.261, 0.2384) and N2 $4a$ (0.5704, 0.3963, 0.5025). Again, each atom has three covalent bonds, but they are all nonequivalent. As the pressure increases, nitrogen tends to decrease local symmetry of nitrogen atoms *cg* $\rightarrow Pba2 \rightarrow P2_12_12_1$. Both of the stable high-pressure structures are insulators with band gaps of 5.2 eV and 2.64 eV, respectively, under above-mentioned pressures. The actual band gap can be expected to be twice as large, due to the typical underestimation of band gaps within DFT.

Due to the complex enthalpy landscape with very small differences between many structures under different pressures, experimental realization of also other non-molecular phases than *cg*, $Pba2$ and $P2_12_12_1$ is possible, at least in high temperatures (say, 2000 K). However, we suggest the experimentalists to especially look for these structures while synthesizing polymeric nitrogen.

3.2 Analytical Potential

Although DFT is most likely the best method for searching for new atomic structures, due to its computational requirements, it remains impossible to carry out dynamical simulations with large system sizes for modeling phase transitions or amorphous structures with DFT. To this end, we have developed a new analytical interaction model[22], similar to an earlier model[23], to be used in classical molecular dynamics simulations. Our preliminary studies indicate that the model can be used to study the transition from a molecular structure to the ordered polymeric phase. Since our model is constructed using *cg* phase as the main structure for fitting, our results can be expected to improve on the earlier synthesis calculations[24].

Further, it is somewhat surprising from the theoretical point of view that bringing the *cg* phase to ambient conditions has proven to be, if not impossible, at least extremely difficult. We use the recently developed interaction model also to attack this problem. So far, we have been able to show that both intrinsic vacancies – introduced into the structure during synthesis – and grain boundaries (see Figure 3) have a significant effect on serving as seeds for the collapse of the structure during pressure release.

4 Concluding Remarks

The enthalpy landscape of nitrogen under high pressures has proven out to be extremely complicated, well highlighted by the number of predicted structures which can exist under extreme conditions[2–13]. First determining the relevant structures from this pool, and then calculating their free energies for wide pressure and temperature ranges, will allow drawing the phase diagram for this material to assist experimental work on realizing its potential as an high energy density material. We have conducted the first two steps on this path by first comparing the different known structures (at the time of the study)[14] and then finding stable high pressure structures[13] after we realized that the previously considered

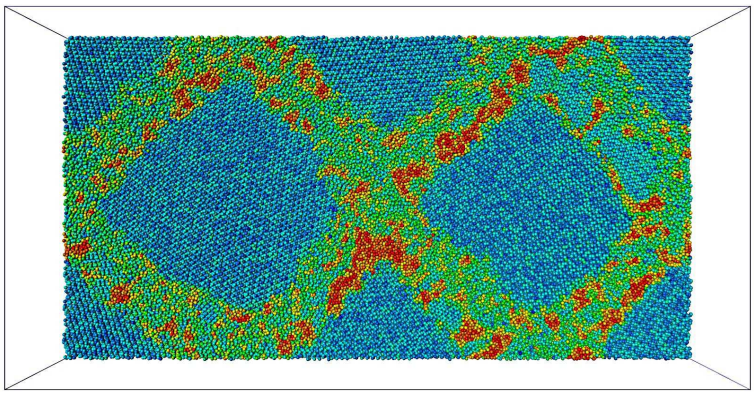


Figure 3. Polycrystalline *cg* as described by our analytical interaction model. The colouring is based on potential energy of each atom. The grain boundaries have clearly higher energy than the crystalline parts of the system.

high pressure structure is actually unstable. Also, we have developed a new interatomic interaction model for nitrogen[22] which can be used to study the synthesis process and to describe the problems and their solutions on the path towards ambient conditions after successful synthesis of the material.

Acknowledgements

Grants of computer time from the Jülich Supercomputing Centre in Jülich, Germany are gratefully acknowledged. This work has been supported by Deutsche Forschungsgemeinschaft through Project No. AL 578/3-1 and -2. We also acknowledge financial support through a program funded by the German foreign exchange server (DAAD) and the Academy of Finland.

References

1. M. I. Eremets and V. V. Gavriliuk, *Single-bonded cubic form of nitrogen*, Nature Mater., **3**, 558–563, 2004.
2. C. Mailhiot, L. H. Yang, and A. K. Mcmahan, *Polymeric Nitrogen*, Physical Review B, **46**, no. 22, 14419–14435, Dec 1 1992.
3. Eugene Gregoryanz, Alexander Goncharov, Russell J. Hemley, and H. K. Mao, *High-pressure amorphous nitrogen*, Phys. Rev. B, **64**, 052103, 2001.
4. M. M. G. Alemany and J. L. Martins, *Density-funtional study of nonmolecular phases of nitrogen: Metastable phase at low pressure*, Phys. Rev. B, **68**, 024110, 2003.
5. M. J. Lipp, J. Park Klepeis, B.J. Baer, H. Cynn, W.J. Evans, V. Iota, and C.-S. Yoo, *Transformation of molecular nitrogen to nonmolecular phases at megabar pressures by direct laser heating*, Phys. Rev. B, **76**, 014113, 2007.
6. W. D. Mattson, D. Sanchez-Portal, S. Chiesa, and R. M. Martin, *Prediction of new phases of nitrogen at high pressure from first-principles simulations*, Physical Review Letters, **93**, no. 12, 125501, Sep 17 2004.

7. F. Zahariev, J. Hooper, S. Alavi, F. Zhang, and T. K. Woo, *Low-pressure metastable phase of single-bonded polymeric nitrogen from a helical structure motif and first-principles calculations*, Physical Review B, **75**, no. 14, 140101, Apr 2007.
8. A. R. Oganov and C. W. Glass, *Crystal structure prediction using ab initio evolutionary techniques: Principles and applications*, Journal of Chemical Physics, **124**, no. 24, 244704, Jun 28 2006.
9. F. Zahariev, S. V. Dudiy, J. Hooper, F. Zhang, and T. K. Woo, *Systematic method to new phases of polymeric nitrogen under high pressure*, Physical Review Letters, **97**, no. 15, 155503, Oct 13 2006.
10. R. Caracas, *Raman spectra and lattice dynamics of cubic* gauche *nitrogen*, J. Chem. Phys., **127**, 144510, 2007.
11. X. L. Wang, Z. He, Y. M. Ma, T. Cui, Z. M. Liu, B B. Liu, J. F. Li, and G. T. Zou, *Prediction of a new layered phase of nitrogen from first-principles simulations*, J. Phys.: Condens. Matter, **19**, 425226, 2007.
12. Y. Yao, J. S. Tse, and K. Tanaka, *Metastable high-pressure single-bonded phases of nitrogen predicted via genetic algorithm*, Physical Review B, **77**, no. 5, 052103, Feb 2008.
13. Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, *Transparent Dense Sodium*, Nature, p. in press, 2009.
14. J. Kotakoski and K. Albe, *First-principles calculations on solid nitrogen: A comparative study of high-pressure phases*, Physical Review B, **77**, no. 14, 144109, Apr 2008.
15. G. Kresse and J. Furthmüller, *Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set*, Comput. Mat. Sci., **6**, 15, 1996.
16. J. P. Perdew, K. Burke, and M. Ernzerhof, *Erratum: Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett., **78**, 1396, 1997.
17. P. E. Blöchl, *Projector augmented wave method*, Phys. Rev. B, **50**, 17953–17979, 1994.
18. G. Kresse and D. Joubert, *From ultrasoft pseudopotentials to the projector augmented wave method*, Phys. Rev. B, **59**, 1758, 1999.
19. H. J. Monkhorst and J. D. Pack, *Special points for Brillouin-zone integrations*, Phys. Rev. B, **13**, 5188, 1976.
20. K. Kunc and Richard M. Martin, Ab Initio *Force Contants of GaAs: A New Approach to Calculation of Phonons and Dielectric Properties*, Phys. Rev. Lett., **48**, 406–409, 1982.
21. T. Zhang, S. Zhang, Q. Chen, and L.-M. Peng, *Metastability of single-bonded cubic-gauche structure of N under ambient pressure*, Phys. Rev. B, **73**, 094105, 2006.
22. J. Kotakoski and K. Albe, to be published elsewhere.
23. J. Nord, K. Albe, P. Erhart, and K. Nordlund, *Modelling of compound semiconductors: analytical bond-order potential for gallium, nitrogen and gallium nitride*, Journal of Physics-Condensed Matter, **15**, no. 32, 5649–5662, Aug 20 2003.
24. K. Nordlund, A. Krasheninnikov, N. Juslin, J. Nord, and K. Albe, *Structure and stability of non-molecular nitrogen at ambient pressure*, Europhysics Letters, **65**, no. 3, 400–406, Feb 2004.

Prebiotic Peptide Synthesis on Blue Gene Platforms at "Iron-Sulfur-World" Conditions

Nisanth N. Nair[1,2], **Eduard Schreiner**[1,3], **and Dominik Marx**[1]

[1] Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum
44780 Bochum, Germany
E-mail: dominik.marx@theochem.ruhr-uni-bochum.de

[2] Current address:
Department of Chemistry, Indian Institute of Technology,
208016 Kanpur, India
E-mail: nnair@iitk.ac.in

[3] Current address:
Theoretical and Computational Biophysics Group, Beckman Institute, University of Illinois
Urbana IL 61801, USA
E-mail: eschrein@ks.uiuc.edu

The exploration of possible scenarios for prebiotic molecular synthesis, including biopolymers such as peptides, is paramount to understanding how primitive life emerged on the young Earth. Amongst a vast amount of different proposals, evidence accumulated that mineral surfaces and water at extreme thermodynamic conditions might offer favourable reaction environments. Using large-scale *ab initio* computer simulations a full synthesis cycle for peptide formation involving hot-pressurized water (at about 500 K and 20 MPa) and a pyrite mineral surface has been generated and compared to ambient aqueous conditions. An indirect route to activated amino acids (Leuchs anhydrides) via an isocyanate intermediate is found to be preferred. Compared to the ambient environment, extreme conditions and the pyrite surface both favour the constructive branch, i.e. peptidization and chain growth, over the degrading reaction, i.e. hydrolysis, and hence yield overall a productive peptide synthesis cycle. This comprehensive *in silico* study of a complex network of biochemical reactions at a mineral/water interface has only been possible due to significant access to capability resources provided by the Blue Gene platforms at JSC and NIC in conjunction with the multi-level massively-parallel `CPMD` simulation package.

1 Setting the Stage: Chemical Reactions at Extreme Conditions

The exploration of possible scenarios for prebiotic molecular synthesis, including biopolymers such as peptides, is paramount to understanding how primitive life emerged on the young Earth. Amongst a vast amount of different hypotheses, evidence accumulated that mineral surfaces in conjunction with water at extreme thermodynamic conditions might offer favourable reaction environments. In particular, an unusually precise proposition for biomolecular synthesis on iron/sulfur minerals in hot-pressurized water as found close to deep-sea hydrothermal vents has been detailed in the framework of the so-called "Iron-Sulfur-World" (ISW) scenario[1]. Certainly, understanding the interplaying fundamental issues of unusual chemical reactivity at extreme conditions, liquid state theory of solvation, and physical chemistry of mineral/water interfaces turn out to be of overriding importance here. In the intricate reaction chain from small molecules to functional proteins the formation of the peptide bond as such is, without any doubt, a key step. Although significant

experimental support of the major ingredients has been accumulated by several groups, pertinent experiments lack detailed molecular insight into how small "inorganic" reactants transform into biomacromolecular products. Still, preliminary *ab initio* molecular dynamics (AIMD) simulations at ISW conditions were carried out only a few years back[2–4].

In the long-term project "Full *In Silico* Exploration of Possible Routes to Prebiotic Peptide Synthesis by *Ab Initio* Metadynamics" devoted to fundamental research in Chemistry the importance of high temperature and pressure ISW reaction conditions including mineral surfaces is being assessed by AIMD techniques. The primary goal here is to provide vital molecular level understanding about the pertinent reactions in the "virtual lab"[5] which is otherwise difficult or even impossible to obtain in real laboratory experiments. Greatly expanding our initial work[2–4], our computations on the IBM Blue Gene systems JUBL and JUGENE at the John von Neumann Institute for Computing (NIC) at Forschungszentrum Jülich (FZJ) during the last two years have unveiled very important mechanistic and energetic details of peptide synthesis at ISW conditions. It is stressed that the unprecedented computational complexity of our *in silico* prebiotic peptide synthesis demanded an investment of up until now about *five nanoseconds* of AIMD simulation time in total, which was only possible due to generous access to these efficient resources at NIC. In 2008 the above-mentioned project has been elected to be the first "NIC Excellence Project of the Year".

2 The "Virtual Lab" Approach to Chemistry

Recent advances in both computer technologies and simulation methods, in particular Car-Parrinello AIMD[6,7] in conjunction with efficient sampling methods like the powerful "metadynamics" technique developed by Laio and Parrinello (see Ref. 10 for a review), make it possible to study truely complex chemical reaction networks in the "virtual lab"[5]. All calculations presented here were performed using Hohenberg-Kohn-Sham density functional theory in its efficient plane wave pseudopotential implementation[7,9] within the `CPMD` software package[8,9].

Since the beginning of the 1990ies, the `CPMD` code[8] has been designed by Jürg Hutter from the onset to run efficiently on all kinds of parallel platforms as explained in detail in NIC Lecture Notes[9]. However, low-latency interconnects are required to run this parallel AIMD code efficiently. To get around load balancing problems on platforms of ever growing processor numbers a second level of parallelization named the "task-grouping" of processors has been implemented into `CPMD` some time ago[9]. Furthermore, the hierarchical multi-level strategies[9] that combine distributed-memory and shared-memory parallelization are highly suited for ultra-dense massively parallel HPC machines such as the Blue Gene architecture[11] in particular.

However, electronic structure calculations and thus AIMD require non-trivial parallelization strategies as the character of the underlying off-lattice quantum problem is not easily suited for partitioning without making use of further approximations such as done in linear scaling algorithms. This is due to the spatially non-local character of quantum-mechanical wavefunctions. The multiple walker metadynamics algorithm[10], which we have successfully implemented, is a linear scaling algorithm in itself and thus improves the net scaling behaviour of `CPMD` for our given system sizes at hand on a large number of processors.

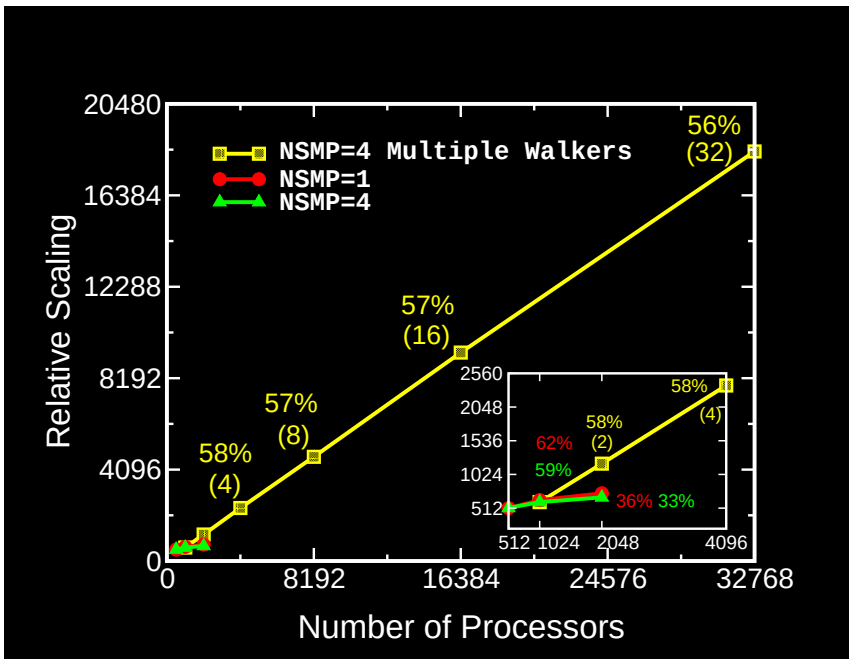


Figure 1. Relative scaling performance when using N processors with respect to 512 processors (i.e. the ratio of the computer time for one AIMD step per processor using N processors to that for 512 processors multiplied by N) for the CPMD code for one of the systems studied on JUGENE at NIC. Red and green lines are using Open MP threads 1 and 4, respectively. The yellow line is the performance using the multiple walker technique where the number of walkers, n, is reported in parenthesis. In the inset the scaling behaviour going from 512 processors to 4098 processors is magnified for clarity.

Figure 1 shows the performance of CPMD on the JUGENE Blue Gene/P installation at NIC employing up to one half of the whole machine with and without multiple walkers. Still, due to the inherent scaling limitations in any quantum simulation code for typical system sizes like that in the test case shown, it is much less efficient to go beyond a Blue Gene mid-plane for practical simulations; see for instance in Figure 1 that the scaling dropped down to about 33 % when using a full mid-plane with respect to using only a quarter of a mid-plane. However, the parallelization among walker replicas is extremely efficient due to the loose coupling of the walkers such that the communication characteristics of Blue Gene systems can be exploited using many racks for a single AIMD simulation. As usage of n walkers will decrease the total length of the simulation by a factor of about n the effective scaling of a multiple walker algorithm can be estimated based on CPU time per AIMD step divided by n. It is evident based on the scaling behaviour shown in Figure 1 that use of all processors of JUGENE is now possible for our system of interest without the need to increase the size of the system even when certain processor groups must communicate *via* high-latency interconnects.

3 The Key Result: Peptide Synthesis Cycle

The simulations of peptide bond formation[12–14] between two glycine molecules were carried out using three vastly different reaction conditions: ambient bulk water at about 300 K and 0.1 MPa (ABW), hot-pressurized bulk water at about 500 K and 20 MPa (HPW), and

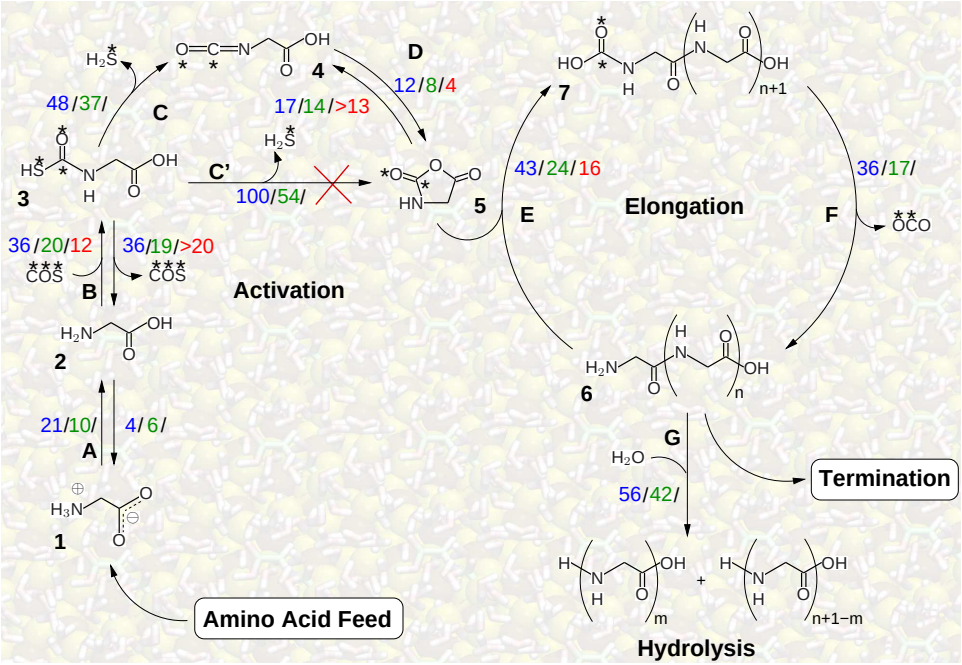


Figure 2. Full peptide synthesis cycle comprising input of an amino acid (or peptide), here glycine, and its activation followed by elongation using another amino acid (or peptide), here another glycine, as well as termination and hydrolysis as a major reverse reaction. The calculated free energy barriers (given in $k_{\mathrm{B}}T$ energy units) for individual steps of the mechanism leading to diglycine formation are colour coded. Blue: ambient bulk water (ABW), green: hot-pressurized bulk water (HPW), red: hot-pressurized water at the pyrite interface (PIW). The crossed direct formation path **C′** is very unlikely in view of its high activation free energy compared to the indirect path *via* isocyanate **4**. See Ref. 12 for details.

hot-pressurized water at the pyrite interface (PIW). The effective free energy barriers estimated along the peptide synthesis cycle leading to diglycine are reported in Figure 2.

It is evident from the mechanism depicted in Figure 2 that it is the neutral form of the amino acid glycine **2** that is required and not the zwitterionic form **1** for its reaction with the COS molecule to form thiocarbamate (see step **B**). The thiocarbamate **3**, in turn, leads to an activated amino acid in form of its so-called Leuchs anhydride **5** that easily adds to another amino acid (or peptide) to form a peptide bond (in step **E**) which finally yields an elongated peptide. As apparent by the computed free energy surface Figure 3 HPW extreme conditions stabilize the neutral form **2**, consistent with the lowering of the dielectric constant of HPW, whereas in ABW neutral glycine converts easily to the zwitterion **1** on an ultrafast timescale of $\approx$ 1 ps. In other words, such extreme thermodynamic conditions are found to increase the concentration of the neutral amino acid by shifting the equilibrium between the neutral and the charged zwitterionic forms of amino acids toward neutral form, thus favouring the formation of peptides. This is an interesting result and immediately reveals the importance of hot-pressurized conditions for this route to peptides.

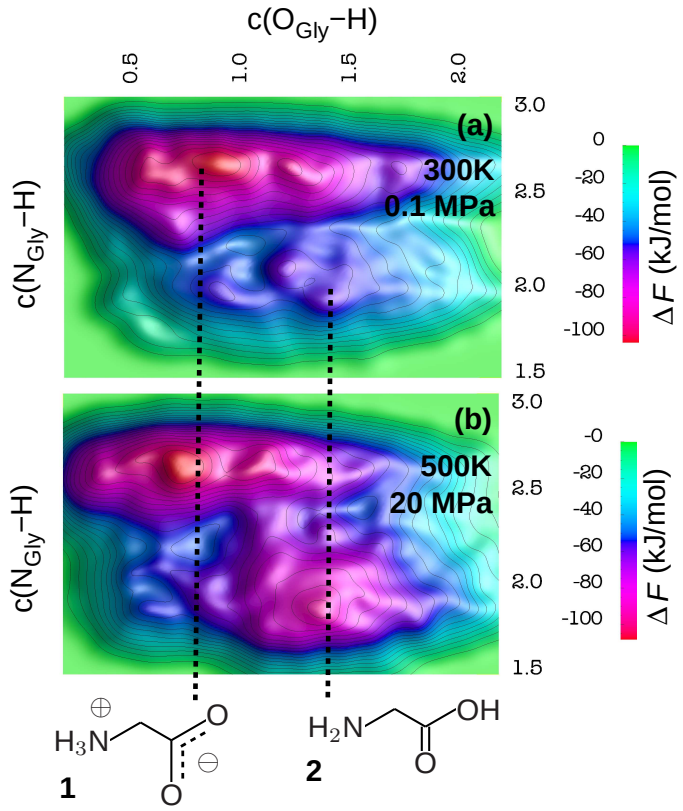


Figure 3. Free energy surface for the conversion between zwitterionic **1** and neutral form **2** of glycine (a) in ambient bulk water (ABW) and (b) in hot-pressurized bulk water (HPW); colour bar shows the relative free energy ΔF in kJ/mol energy units. Metadynamics AIMD simulations were performed using two collective variables: coordination numbers of nitrogen and carboxy oxygen to all hydrogen atoms in the system, $c(N_{Gly} - H)$ and $c(O_{Gly} - H)$, respectively. See Ref. 13 for details.

Moreover, it was found that these extreme HPW conditions speed up the production of peptides by accelerating individual steps of the whole peptide synthesis cycle according to the free energy barriers reported in Figure 2.

Another discovery from our simulations is the so-called isocyanate pathway leading to the formation of the activated form of amino acid, Leuchs anhydride **5**, from thiocarbamate **3**. Compared to a direct cyclization of the thiocarbamate **3** to form Leuchs anhydride **5**, the indirect isocyanate pathway, i. e. first forming an isocyanate **4** which rapidly cyclizes to Leuchs anhydride, is very much lower in terms of free energy barriers. This result confirms earlier experimental speculations about such a route including an equilibrium between **5** and **4**. Our calculations also shed light on the productivity of the cycle including the formation of isocyanate being the rate-determining step of the whole peptide cycle. Based on a simple estimate, the time scale for forming peptide bonds along this route is in the order of a few minutes at hot-pressurized conditions whereas it would be several years at ambient conditions! It is also shown by the simulations that hydrolysis of the

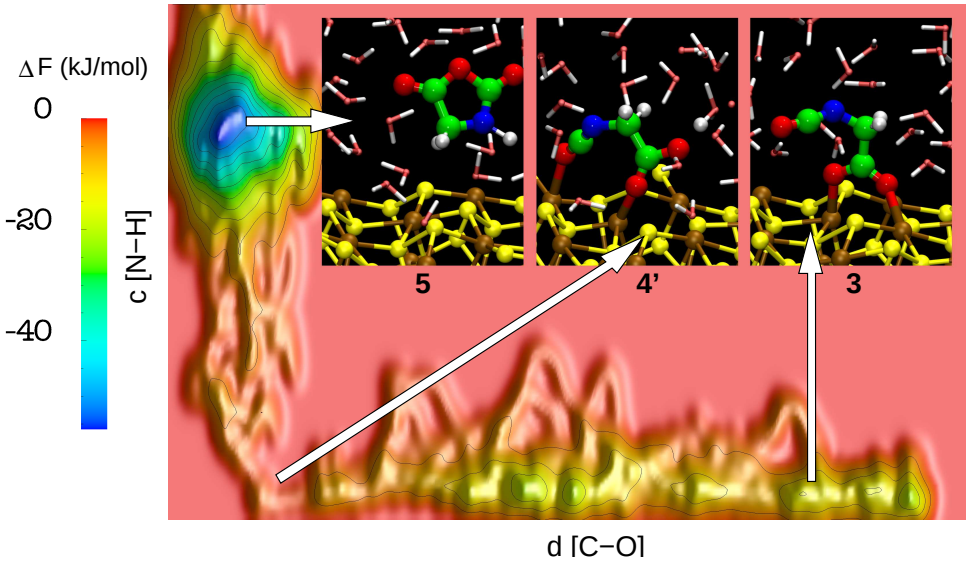


Figure 4. Mechanism of the formation of Leuchs anhydride **5** from thiocarbamate **3** in hot-pressurized water at the pyrite interface (PIW) based on the free energy surface; colour bar shows the relative free energy ΔF in kJ/mol energy units. Metadynamics AIMD simulations were performed using two collective variables: distance between the carbon atom of the COS entity to one of the carboxylate oxygen atoms d[C-O] and coordination number of nitrogen to all hydrogen atoms in the system c[N-H]. Three representative real space configurations sampled from these simulations at the pyrite-water interface demonstrate scaffolding due to FeS_2(001) by preformation of the cyclic topology of Leuchs anhydride upon bidentate chemisorption. Colour code: hydrogen (white), oxygen (red), carbon (green), nitrogen (blue), sulfur (yellow), iron (ocher); labeling is according to Figure 2. See Ref. 12 and upcoming publications for details.

synthesized peptide is slower than the rate-determining formation step and, therefore, a net accumulation of peptide can be expected in agreement with experimental conclusions.

Possible roles of Fe/S mineral surfaces were also investigated *via* simulations using an ideal pyrite surface, FeS_2(001), as the simplest model. By decreasing the entropic contribution to the free energy barriers this surface is found to accelerate several reaction steps in the peptide synthesis cycle by lowering the corresponding free energy barriers up to a factor of two, which increases the respective reaction rates exponentially. In addition, more interesting effects like scaffolding of reactant molecules favouring the formation of transition states and thereby speeding up the reaction by several orders of magnitude have been identified. In particular, the pyrite surface is found to favour the preformation of the five-membered ring that is characteristic of Leuchs anhydride **5** as demonstrated in Figure 4.

Based on free energy calculations and careful examination of various reaction mechanisms, our studies underpin the importance of extreme conditions and mineral surfaces for peptide synthesis along a putative route[12–14]. These comprehensive simulations delineate pathways connecting the crucial activation and elongation steps through which peptides can be produced out of amino acids and COS *via* an indirect isocyanate/Leuchs anhydride route. Importantly, the data provide convincing evidence that all steps along the proposed

synthesis cycle are clearly favoured in hot-pressurized water when compared to ambient conditions providing in total a productive synthesis cycle[12]. Beyond the specific case, these findings imply that "different chemistry" must be considered when discussing putative prebiotic synthesis scenarios at extreme aqueous conditions.

4 What Next?

Although a significant step forward, the cycle in Figure 2 is based on a set of disconnected free energy calculations whereas a single, global free energy landscape is necessary to fully explore this rather complex reaction network including the roles of reverse and side reactions. This, again, will be a challenge to both algorithms to sample free energies beyond three or four dimensions and platforms to carry out such a unified, ultramassive simulation. Another key issue that still remains largely unexplored is the role of mineral surfaces since defective surfaces are known to be most active in heterogeneous catalysis whereas an ideal, non-defective pyrite has been used up to now in our "virtual lab"[5]. Including defects at the mineral/water interface, most desirably as "dynamical degrees of freedom" in an AIMD simulation, certainly adds a lot more complexity to the problem but would be another necessary step forward. However, looking back at recents developments it appears to optimists that such dreams might become true much sooner than currently expected.

Acknowledgements

We are most grateful to Alessandro Curioni and Jürg Hutter for help in porting `CPMD` onto the Blue Gene/L and P platforms JUBL and JUGENE at NIC. This research is supported by Deutsche Forschungsgemeinschaft *via* Normalverfahren MA 1547/7.

References

1. G. Wächtershäuser *Groundworks for an Evolutionary Biochemistry: The Iron-Sulfur World*, Prog. Biophys. Molec. Biol. **58**, 85-201, 1992.
2. C. Boehme, D. Marx, *Glycine on a Wet Pyrite Surface at Extreme Conditions*, J. Am. Chem. Soc. **125**, 13362-13363, 2003.
3. R. Pollet, C. Boehme, D. Marx, *Ab Initio Simulations of Desorption and Reactivity of Glycine at a Water-Pyrite Interface at "Iron-Sulfur World" Prebiotic Conditions*, Origins Life Evol. Biospheres **36**, 363-379, 2006.
4. N. N. Nair and, E. Schreiner, D. Marx, *Glycine at the Pyrite-Water Interface: The Role of Surface Defects*, J. Am. Chem. Soc. **128**, 13815-13826, 2006.
5. D. Marx, *Theoretical Chemistry in the 21st Century: The "Virtual Lab"*, In: Proceedings of the "Idea-Finding Symposium: Frankfurt Institute for Advanced Studies", (Greiner, W.; Reinhardt, J., Eds.), pp 139-153 (EP Systema, Debrecen 2004).
6. R. Car, M. Parrinello, *Unified Approach for Molecular Dynamics and Density-Functional Theory*, Phys. Rev. Let. **55**, 2471-2474, 1985.
7. D. Marx, J. Hutter, *Ab initio Molecular Dynamics: Basic Theory and Advanced Methods* (Cambridge University Press, Cambridge 2009).

8. CPMD, J. Hutter *et al.*, Copyright: IBM Corp 1990-2009 and MPI für Festkörperforschung Stuttgart 1997-2001; http://www.cpmd.org.
9. D. Marx, J. Hutter, *Ab Initio Molecular Dynamics: Theory and Implementation*, In: Modern Methods and Algorithms of Quantum Chemistry, (Grotendorst, J., Ed.), Volume 3, pp 301-449. (John von Neumann Institute for Computing NIC, Forschungszentrum Jülich, 2000).
10. A. Laio, M. Parrinello, *Computing Free Energies and Accelerating Rare Events with Metadynamics*, In: Computer Simulations in Condensed Matter: From Materials to Chemical Biology, (Ferrario, M.; Ciccotti, G.; Binder, K., Eds.), Volume 1, pp 315-347 (Springer Verlag, Berlin 2006).
11. J. Hutter, A. Curioni, *Car-Parrinello Molecular Dynamics on Massively Parallel Computers*, ChemPhysChem **6**, 1788-1793, 2005.
12. E. Schreiner, N. N. Nair, D. Marx, *Influence of Extreme Thermodynamic Conditions and Pyrite Surfaces on Peptide Synthesis in Aqueous Media*, J. Am. Chem. Soc. **130**, 2768-2770, 2008.
13. N. N. Nair, E. Schreiner, D. Marx, *Peptide Synthesis in Aqueous Environments: The Role of Extreme Conditions on Amino Acid Activation*, J. Am. Chem. Soc. **130**, 14148-114160, 2008.
14. E. Schreiner, N. N. Nair, and D. Marx, *Peptide Synthesis in Aqueous Environments: The Role of Extreme Conditions on Peptide Bond Formation and Peptide Hydrolysis* J. Am. Chem. Soc. **in press**, xx-xx, 2009.

Calculation of ^{31}P and ^{183}W NMR Chemical Shifts and Nuclear Spin Spin Coupling Constants in Phosphinidenoid and Other Transition Metal Complexes by the ZORA DFT Method

Gerd von Frantzius and Rainer Streubel

Institute of Inorganic Chemistry, University of Bonn
53121 Bonn, Germany
E-mail: r.streubel@uni-bonn.de

Phosphinidenoid complexes [Li(12-cr-4)][$(CO)_5$M{P(R)X}] (M = W, R = $CH(SiMe_3)_2$), X = F, Cl, Br, I) are a new class of compounds, which in analogy with carbenoids or silylenoids, can release a shortlived intermediate phosphinidene complex [$(CO)_5$M(PR)] by elimination of an alkaline halide 12-crown-4 complex [Li(12-cr-4)X]. At the current stage of the project isomeric phosphinidenoid model complexes [Li(12-cr-4)][$(CO)_5$W{P(Me)F}] are to be characterized both by compliance constants (elements of the inverted hessian in internal valence coordinates, the diagonal elements of which are inversely proportional to bond strengths) and their NMR spectroscopic properties ($\delta(^{31}P)$, $\delta(^{183}W)$, ^1J(W,P)). In contrast to the corresponding phosphine complexes [$(CO)_5$W{PH(Me)X}] the phosphorus NMR chemical shifts of phosphinidenoid complexes are dominated by the paramagnetic contribution. This does not hold for the tungsten NMR chemical shift, which for both classes, phosphine- and phosphinidenoid complexes is dominated by the ground state part of the absolute isotropic shielding.

1 Introduction: Computation of NMR Chemical Shifts and Nuclear Spin-Spin Coupling Constants

Physical application of a magnetic field to a molecule produces a shielding field at each nucleus which is due to the diamagnetic current of the electrons and which is proportional to the external field. The shielding field of the bound electrons can be either in the same (negative shielding) or in the opposite direction of the external field (positive shielding). The shielding constant σ_A of a nucleus A is the (negative) component of the shielding field in direction of an applied external field of unit strength and, according to Ramsey[1], consists of a diamagnetic and a second-order paramagnetic part:

$$\sigma_A = \sigma^d + \sigma^p. \tag{1}$$

Since in an actual calculation the magnetic field is represented in the electronic Hamiltonian by its vector potential, the separation into a diamagnetic and a paramagnetic part is not unique – every gauge transformation will yield a new vector potential for the same magnetic field – but their sum is, if the basis set is complete. In a finite basis the quality of the calculation of the diamagnetic part in terms of the ground state electronic wavefunction only is not balanced by the quality of the calculation of the paramagnetic shielding, which in principle involves all excited states wave functions[2]. The problem of incomplete basis sets can be overcome by the use of gauge invariant/including atomic orbitals (London orbitals, gauge dependent orbitals), which make all matrix elements made up of differences in the vector potential and thus all properties independent of the gauge origin. The wave function, however, depends on the origin by means of a complex phase factor[3].

In the usual derivative formulation of molecular properties the energy is expanded in a Taylor series in the perturbation strength x

$$E\left(x\right)=E\left(0\right)+\frac{\partial E}{\partial x}x+\frac{1}{2}\frac{\partial^{2}y}{\partial x^{2}}x^{2}\ldots\,, \tag{2}$$

such that the n'th order property is the n'th order derivative of the energy, $\partial^{n}E/\partial x^{n}$, with or without the 1/n! factor[3]. Now the shielding constant σ_A is obtained as

$$\sigma_{A}=\frac{1}{3}tr\left(\sigma\right),\quad\sigma\propto\left(\frac{\partial^{2}E}{\partial B\partial I}\right), \tag{3}$$

where the shielding tensor σ is proportional to the mixed second derivative of the energy with the external magnetic field B and the internal magnetic field I_A due to the nuclear spin. Similarily the indirect nuclear spin-spin coupling constant *J*(A,A′) is

$$J(A,A')=\frac{1}{3}tr\left(J_{AA'}\right),\quad J_{AA'}=h\frac{\Upsilon_{A}\Upsilon_{A'}}{4\pi^{2}}K_{AA'},\quad K_{AA'},\propto\left(\frac{\partial^{2}E}{\partial I_{A}\partial I_{A'}}\right), \tag{4}$$

(the Υs are the nuclear gyro-magnetic ratios) where the reduced indirect spin-spin coupling tensor $K_{AA'}$ is proportional to the mixed second derivative of the energy with the internal magnetic fields at nucleus A and A′. Both shielding and spin-spin coupling tensor consist of two parts (not shown): a diamagnetic part, corresponding to the expectation value of the unperturbed state, and a paramagnetic part, which represents the relaxation of the wave-function in response to the external field. For the shieldings there are no spin contributions whereas for the couplings the paramagnetic spin contributions are the most important ones. Shielding and coupling constants are of the order

$$\left|\sigma_{A}\right|\approx\alpha^{2},\quad\left|K\left(A,B\right)\right|\approx\alpha^{4},\quad\alpha\approx1/137 \tag{5}$$

where α ist the fine-structure constant[3,4].

1.1 Computational Methods

All optimizations and frequency calculations were performed using GAUSSIAN03[5] and ADF-2007.01[6]. The standard DFT method with GAUSSIAN03 in this work is B3LYP on the standard basis set 6-311G(d,p) eventually combined with an effective core potential description using LanL2DZ by Hay and Wadt[7]. The standard DFT method with ADF-2007.1 is VWN[8]B[9]P[10]86/TZ2P SO ZORA. Stationary points have been characterized by analytical second derivatives (the Hessian) with respect to redundant cartesian coordinates. Transformation into nonredundant internal force constants was done either by using algorithms of Pulay and Fogarasi (INTC/FCTINT)[11], where inversion of the Hessian into the compliance matrix is accomplished by standard methods[12], or by inversion of the hessian within a G03 output.

The general theoretical methods for the DFT calculation of chemical shifts and NSSCCs as implemented in the ADF program code was put forward by Ziegler[13–16].

^{31}P and ^{183}W nmr chemical shifts and nuclear spin-spin couplings werde computed using ADFs SO ZORA GIAO-DFT included in the CPL[13,14] module (see standard method above) where absolute isotropic shieldings were converted to common chemical shifts via

$$\begin{aligned}\delta(^{31}\text{P compound, calc.})&=\sigma(\text{PH}_3)-\sigma(\text{compound, calc.})-266.1,\\ \delta(^{183}\text{W compound, calc.})&=\sigma(\text{W(CO)}_6)-\sigma(\text{compound, calc.})-3486.\end{aligned} \tag{6}$$

2 Influence of the Electronic Structure at Phosphorus

To investigate the influence of the electronic structure at phosphorus on the W-P compliance and associated NMR spectroscopic properties ($\delta(^{31}P)$, $\delta(^{183}W)$, $^1J(W,P)$) a set of reference complexes $[(CO)_5W(L)]$ (Table 1) is calculated. The compounds of the second row are the protonated compounds of the first row etc. Phosphido complexes of the type $[L_nW{\equiv}P]$ were reviewed by Scheer[19], who also obtained the structure of dinuclear tungsten phosphido complex $[PPh_4][\{(CO)_5W\}_2PH_2]$[20]. Lithiated phosphine complexes $[(CO)_5W(R_2PLi)]$, $[(CO)_5W(RPLi_2)]$ and $[(CO)_5W(PLi_3)]$ were studied both by Karaghiosoff[21] and Mathey[22], (similarily Hoefler et al. proposed polylithiation of phosphine complex $[CpMn(CO)_2PH_3]$ to yield $[CpMn(CO)_2PLi_3]$[23], while phosphine complexes $[(CO)_5M(PR_3)]$ became the subject of recent calculations and a solid state investigation by the Frenking group (X-ray: M = Mo, W; R = H, Cl; calculation: M = Cr, Mo, W; R = H, Me, F, Cl)[24].

P^+	P^-	P^{3-}	
PH^{2+}	PH	PH^{2-}	
	PH_2^+	PH_2^-	$[(CO)_5W(PH_2)]^-$
		PH_3	

Table 1. Scheme of ligands L of reference complexes $[(CO)_5W(L)]$.

So far $\delta(^{31}P)$ and $\delta(^{183}W)$ values of the unusual (and experimentally unknown) phosphido complex $[(CO)_5W{\equiv}P]^-$ (**2**) have to be regarded unreliable (the extraordinary phosphorus chemical shift is caused by a spin orbit coupling contribution of -9608 ppm, which is unphysical). NMR data of the complex $[(CO)_5W(PH)]^{2+}$ (**4**) were yet not obtained.

W-P and W-C^{tr} compliances of complexes **1-10** are compared in Figure 1, left. A strong tungsten-phosphorus bond (small compliance) is accompanied by a weak ligand bond of the carbonyl *trans* to the phosphorus ligand - a strong *trans*-effect, quantified by compliance constants. The correlation of W-P compliance and associated bond length is

Compound	Nr	W-P [Å/mdyn]	$^1J(W,P)$ [Hz]	W-P [Å]	$\delta(^{31}P)$ [ppm]	$\delta(^{183}W)$ [ppm]
$[(CO)_5WP]^+$	**1**	0.264	-19	2.189	-1758	-261
$[(CO)_5W{-}P]^-$	**2**	0.730	-21	2.495	”12589”	”-5722”
$[(CO)_5WP]^{3+}$	**3**	1.715	8	2.837	-435	-2916
$[(CO)_5W(PH)]^{2+}$	**4**	0.248	-	2.168	-	-
$[(CO)_5W(PH)]$	**5**	0.517	-50	2.417	2044	-1050
$[(CO)_5W(PH)]^{2-}$	**6**	1.432	13	2.775	19	-3278
$[(CO)_5W(PH_2)]^+$	**7**	0.398	316	2.312	373	-1260
$[(CO)_5W(PH_2)]^-$	**8**	1.104	116 (73)[22]	2.708	-276 (-276)[22]	-3221
$[\{(CO)_5W\}_2(PH_2)^-$	**9**	1.015	84	2.638	-250	-3226
			(146)[20]	(2.579(2))[20]	(-268)[20]	(-3350)[20]
$[(CO)_5W(PH_3)]$	**10**	0.858	160 (216)[22]	2.521 (2.493)[24]	-198 (-183)[22]	-3370

(i) B3LYP/6-311g(d,p), (ii) VWNBP86/TZ2P SO ZORA; experimental values in parentheses

Table 2. W-P compliance,[(i)] $^1J(W,P)$,[(ii)] bond lengths,[(ii)] $\delta(^{31}P)$[(ii)] and $\delta(^{183}W)$[(ii)] of complexes $[(CO)_5W(L)]$, where L = P^+, P^-, P^{3-}, PH^{2+}, PH, PH^{2-}, PH_2^+, PH_2^-, $[(CO)_5W(PH_2)]^-$, PH_3.

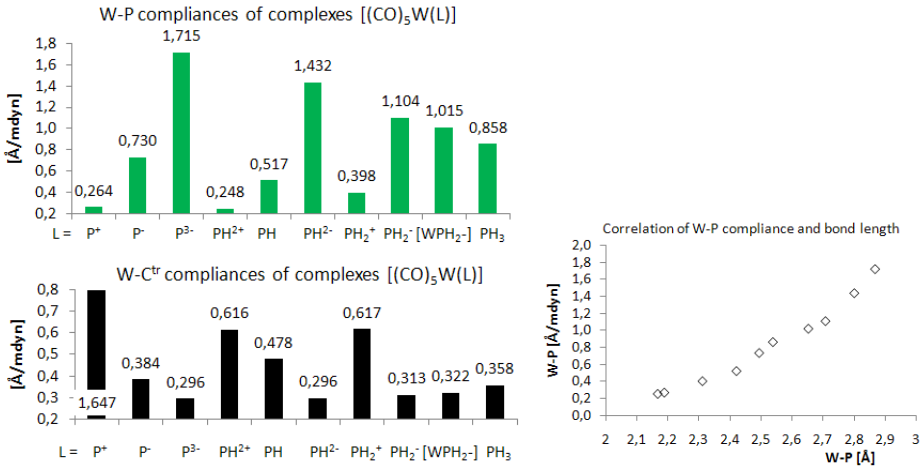


Figure 1. Left: W-P and W-C^{tr} compliances of complexes $[(CO)_5W(L)]$ (**1-10**). Right: Correlation of W-P compliance and bond length of complexes **1-10**.

shown in Figure 1, right. For this class of compounds the usual rule holds: the stronger the W-P blond (small compliance) the shorter the bond.

For comparison the W-P compliances of cyclopentadienyl carbonyl complexes $[Cp(CO)_nW(PR^1R^1)]$ (Table 3, Figure 2) were calculated. From the coordination mode, W-P compliances and bond lengths of $[Cp(CO)_2W{=}PH_2]$ (**11**, compare the DFT calculation at VWNBP86 by Berces)[25] complexes **11** and **12**[26] are tungsten-phosphorus doubly bonded compounds, corresponding with phosphenium complex $[(CO)_5W{=}PH_2]^+$ (**7**), while $[Cp(CO)_3W(PH_2)]$ is W-P singly bonded, analogously to phosphido complex $[(CO)_5W(PH_2)]^-$ (**8**). Complexes with W-P double bond are further characterized by large W-P couplings (>300 Hz) and phosphorus chemical shifts at lower field >300 ppm.

Compound	Nr	W-P [Å/mdyn]	^1J(W,P) [Hz]	W-P [Å]	$\delta(^{31}P)$ [ppm]
$[Cp(CO)_2W{=}PH_2]$	**11**	0.378	-	2.277	-
$[Cp(CO)_2W{=}P(tBu)_2]$	**12**	0.408	483 (552)[26]	2.305 (2.284(4))[26]	453 (373)[26]
$[Cp(CO)_3W(PH_2)]$	**13**	1.002	-	2.664	-

(i) B3LYP/6-311g(d,p), (ii) VWNBP86/TZ2P SO ZORA; experimental values in parentheses

Table 3. W-P compliance,[(i)] ^1J(W,P),[(ii)] bond lengths,[(ii)] $\delta(^{31}P)$[(ii)] and $\delta(^{183}W)$[(ii)] of complexes $[Cp(CO)_nW(L)]$, where $L = PR^1R^2$.

The influence of a fluorine substituent at phosphorus on $\delta(^{31}P)$ and ^1J(P,M) in mono- and bis-borane and mono- and dinuclear pentacarbonyl tungsten complexes was studied with the ligands PR^1R^2 (R^1/R^2 = H, F, Me; Table 4). By successive borane complexation of PH_2^- the lone pair is shown to have a shielding effect, which is stepwise released in going from $[(BH_3)PH_2]^-$ to $[(BH_3)_2PH_2]^-$. Simililarly tungsten complexation effects a deshielding compared with free PH_2^-. The situation is different however in case of fluo-

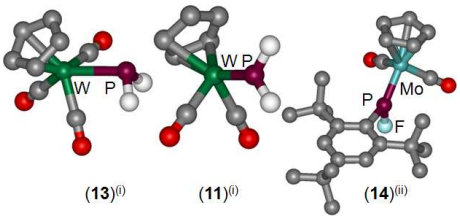


Figure 2. Different coordination modes of PR^1R^2 at a $CpM(CO)_n$ metal fragment for n = 2,3; (i) B3LYP/6-311g(d,p), (ii) VWNBP86/TZ2P.

rine substitution, as the lone pairs at phosphorus and at fluorine effect the chemical shift. Bis-borane complexation both at PF_2^- and at $PMeF^-$ has a shielding effect while contributing negatively to the phosphorus fluorine coupling. Thus for tungsten complex anion $[(CO)_5W\{P(Me)F\}]^-$ (**25**) a phosphorus downfield shift of 359 ppm and a P-F coupling constant of -596 Hz results, which is small compared with common P^{III}-F derivatives. Fluorine and the lone pair at phosphorus give a large negative contribution to the tungsten phosphorus coupling constant, which almost vanishes (7 Hz).

[M]	PH_2^- $\delta(^{31}P)$ [ppm]	^1J(P,M) [Hz]	PHF^- $\delta(^{31}P)$ [ppm]	^1J(P,F) [Hz]	PF_2^- $\delta(^{31}P)$ [ppm]	^1J(P,F) [Hz]	^1J(P,M) [Hz]	PF_2^- $\delta(^{31}P)$ [ppm]	^1J(P,F) [Hz]	^1J(P,M) [Hz]
-	-371	-	525	-31	1182	-603	-	652	-105	-
[B]	-203	-	-	-	-	-	-	318	-511	-23
$[B]_2$	-80	-	-	-	360	-1086	8	251	-750	13
[W]	-276	116	-	-		-	-	359	-596	7
$[W]_2$	-250	84	-	-	-	-	-	-	-	-

(i) VWNBP86/TZ2P SO ZORA; [B] = BH_3, [W] = $W(CO)_5$

Table 4. $\delta(^{31}P)^{(i)}$ and $^1J(P,M)$,$^{(i)}$ (M = B, W) of reference phosphides = PR^1R^2 (R = H, F, Me) and borane / pentacarbonyltungsten complexes $[M]PR^1R^2$ thereof.

Experimentally the influence of a bonded fluorine atom at phosphorus can be seen by comparison of $\delta(^{31}P)$ in proton-bridged dinuclear phosphido complexes $[M_2Cp_2(\mu\text{-}H)(\mu\text{-}PR^1R^2)(CO)_4]$ (Table 5, M = W, Mo): while in the parent compound (R^1, R^2 = H) the phosphorus is shielded (M = W, $\delta(^{31}P)$ = -40 ppm, $^1J(P,W)$ = 200 Hz) or mildly deshielded (M = Mo, $\delta(^{31}P)$ = 34 ppm), P-bound fluorine (R^1 = F, R^2 = Mes, Cy, Ph) exerts a deshielding of more than 330 ppm downfield. For molybdenum phosphido complex **14**[28] the relativistic calculation of $\delta(^{31}P)$ including spin-orbit coupling improves the (gas phase) chemical shift by 28 ppm relative to the experimental value (CH_2Cl_2 solution).

Phosphinidenoid complexes[30–32] $[Li(12\text{-cr-}4)][(CO)_5M\{P(R)X\}]$ (M = W, R = $CH(SiMe_3)_2$, X = F, Cl, Br, I) are a new class of compounds, which in analogy with carbenoids[33] or silylenoids[34], can release a shortlived intermediate phosphinidene complex $[(CO)_5M(PR)]$ by elimination of an alkaline halide 12-crown-4 complex $[Li(12\text{-cr-}4)X]$[35]. Experimentally they can be obtained by lithiation of the corresponding phosphine complexes $[(CO)_5W\{PH(R)X\}]$ (Table 6), the W-P compliance, $^1J(W,P)$ and $\delta(^{31}P)$ / $\delta(^{183}W)$ of which were calculated for model (R = Me) and fully substituted complexes (R = $CH(SiMe_3)_2$).

Compound	Nr.	Exp.	Non-rel.	SO ZORA
$[W_2Cp_2(\mu{-}H)(\mu{-}PH_2)(CO)_4]$		-40[27]		
$[Mo_2Cp_2(\mu{-}H)(\mu{-}PH_2)(CO)_4]$		34[27]		
$[Cp(CO)_2Mo\{PF(2,4,6\text{-}^tBu_3{-}C_6H_2)\}]$	**14**	324[28]	395(i)	367(ii)
$[Mo_2Cp_2(\mu{-}H)\{\mu{-}PF(Mes)\}(CO)_4]$		368[29]		
$[Mo_2Cp_2(\mu{-}H)\{\mu{-}PF(Cy)\}(CO)_4]$		417[29]		
$[Mo_2Cp_2(\mu{-}H)\{\mu{-}PF(Ph)\}(CO)_4]$		375[29]		

(i) VWNBP86/TZ2P, (ii) VWNBP86/TZ2P SO ZORA

Table 5. $\delta(^{31}P)$ of reference W/Mo phosphido complexes.

		W-P	^1J(W,P)		W-P		$\delta(^{31}P)$		$\delta(^{183}W)$	
R/X	Nr	[Å/mdyn]	[Hz]	Exp.	[Å]	Exp.	[ppm]	Exp.	[ppm]	Exp.
a/F	**15**	0.708	200	-	2.470	-	158	-	-3259	-
b/F	**16**	-	-	286[31]	2.499	2.4463(9)[30]	176	157	-	-3300[30]
a/Cl	**17**	0.774	187	-	2.486	-	51	-	-3143	-
b/Cl	**18**	-	-	269[36]	2.513	-	77	57	-	-
a/Br	**19**	0.794	182	-	2.493	-	23	-	-3094	-
b/Br	**20**	-	-	265[30]	2.520	-	54	23	-	-
a/I	**21**	0.818	173	-	2.499	-	-47	-	-3039	-
b/I	**22**	-	-	254[30]	2.523	-	-14	-49	-	-
a/CN	**23**	0.826	168	-	2.504		-80	-	-3250	-
b/CN	**24**	-	-	243[37]	2.537	2.4753(5)	-68	-70	-	-3280[30]

(i) B3LYP/6-311g(d,p), (ii) VWNBP86/TZ2P SO ZORA, a = Me, b = $CH(SiMe_3)_2$, exp. values in parentheses

Table 6. W-P compliance,(i) ^1J(W,P),(ii) bond lengths,(ii) $\delta(^{31}P)$(ii) and $\delta(^{183}W)$(ii) of phosphine complexes $[(CO)_5W\{PH(R)X\}]$.

A comparison of experimental and calculated phosphorus chemical shifts is plotted in Figure 3. The systematic increase in shielding of the phosphorus (chemical shifts directed upfield) in going from X = F^- to CN^- is reproduced by the model calculations by $\pm$4 ppm. Tungsten shielding leads to chemical shifts upfield from -3259 to -3034 ppm, while agreement to the experiment is $\pm$40 ppm.

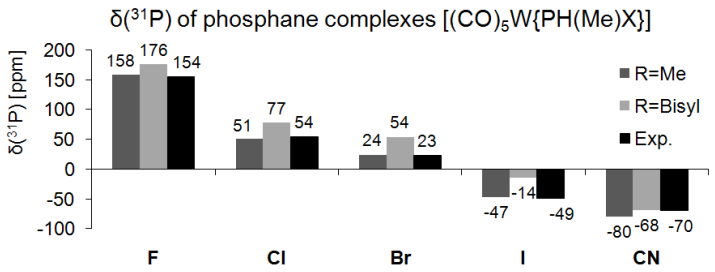


Figure 3. Comparison of experimental and calculated $\delta(^{31}P)$ of complexes $[(CO)_5W\{PH(R)X\}]$, R = Me, $CH(SiMe_3)_2$, X = F, Cl, Br, I, CN at VWNBB86/TZ2P SO ZORA (ADF).

Para- (σ^{p}), diamagnetic (σ^{d})- and spin-orbit contributions (σ^{SO}) to the absolute isotropic shielding $\sigma(^{31}P)$ at phosphorus and $\sigma(^{183}W)$ at tungsten in phosphine complexes $[(CO)_5W\{PH(R)X\}]$ are plotted in Figures 4. (Note that the paramagnetic shielding is plotted as $|\sigma^{p}|$).

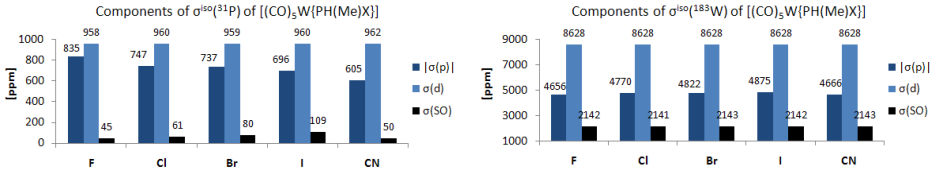


Figure 4. Para- ($|\sigma^{p}|$), diamagnetic (σ^{d})- and spin-orbit contributions (σ^{SO}) to the absolute isotropic shielding $\sigma(^{31}P)$ at phosphorus and $\sigma(^{183}W)$ at tungsten in phosphine model complexes $[(CO)_5W\{PH(Me)X\}]$ (**15**,**17**,**19**,**21**,**23**), X = F, Cl, Br, I, CN.

The reference calculations above set the expected values of $\delta(^{31}P)$ and $^{1}J(W,P)$ for the phosphinidenoid class of compounds (structure types **A**-**D**, Figure 5): phosphorus chemical shifts should be downfield >300 ppm, while tungsten-phosphorus coupling constants are expected to be small. In fact calculated $\delta(^{31}P)$ varies from downfield 322 to 522 and $^{1}J(W,P)$ <<100 Hz throughout (Table 7). Although this general trend reflects the experiment, agreement between experiment and theory is yet unsatisfactory.

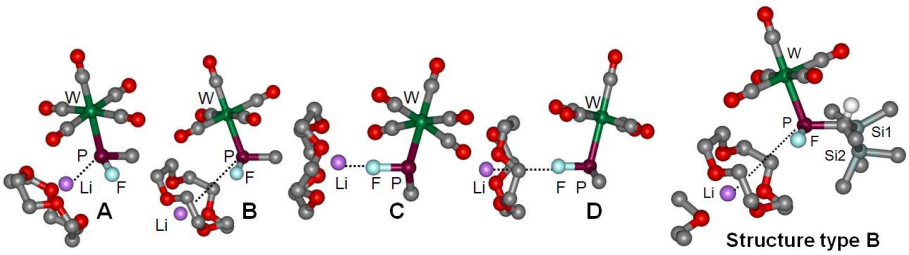


Figure 5. Isomers **A**-**D** of model phosphinidenoid complexes [Li(12-cr-4)][$(CO)_5W\{P(Me)F\}$] (**26**,**27**,**29**,**30**) and the fully substituted complex [Li(12-cr-4)·Et_2O][$(CO)_5W\{P(CH(SiMe_3)_2)F\}$] (**28**) of structure type **B** including a coordinated solvent molecule (Et_2O) at Li (VWNBP86/TZ2P SO ZORA, ADF 2007.1).

All structure types are tight ion-pairs with the $[Li(12\text{-cr-}4)]^{+}$ cation either in contact to phosphorus (**A**,**B**) or to fluorine (**C**,**D**), while this contact can be either direct (**A**,**C**) or mediated by the 12-crown-4 ligand (**B**,**D**). In the latter cases (**B**,**D**) further complexation of the Li^{+} cation by solvent molecules is to be expected, which is included in the calculation of the fully substituted compound **28** (Figure 5).

In contrast to the corresponding phosphine complexes $[(CO)_5W\{PH(Me)X\}]$ the phosphorus NMR chemical shifts of phosphinidenoid complexes are dominated by the paramagnetic contribution, which means that excited states (simulated by virtual orbitals) have a significant influence on the shielding property in these compounds. This does not hold for

R	Nr		W-P [Å/mdyn]	^1J(W,P) [Hz]	Exp.	W-P [Å]	$\delta(^{31}P)$ [ppm]	Exp.	$\delta(^{183}W)$ [ppm]	Exp.
a	**25**		1.022	7	-	2.658	359	-	-3137	-
a	**26**	**A**	0.897	43	-	2.588	322		-3164	-
a	**27**	**B**	-	28	-	2.615	364		-3130	-
b	**28**	**B**	-	-	71	2.637	358	306	-3109	-3230
a	**29**	**C**	0.971	8	-	2.608	522	-	-3108	-
a	**30**	**D**	0.979	11	-	2.608	456	-	-3137	-

(i) B3LYP/6-311g(d,p), (ii) VWNBP86/TZ2P SO ZORA, a = Me, b = $CH(SiMe_3)_2$

Table 7. W-P compliance,[(i)] ^1J(W,P),[(ii)] bond lengths,[(ii)] $\delta(^{31}P)$[(ii)] and $\delta(^{183}W)$[(ii)] of complexes $[(CO)_5W\{P(Me)F\}]^-$ (**25**) and [Li(12-cr-4)][$(CO)_5W\{P(R)F\}$] of structure types **A**-**D**.

the tungsten NMR chemical shift, which for both classes, phosphine- and phosphinidenoid complexes is dominated by the ground state part of the absolute isotropic shielding.

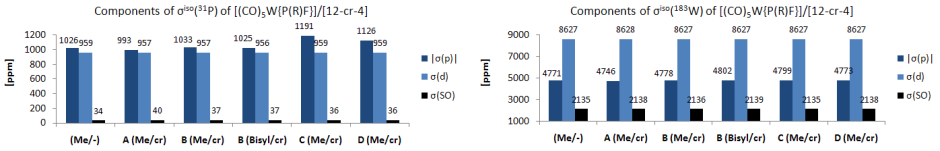


Figure 6. Para- ($|\sigma^p|$), diamagnetic (σ^d)- and spin-orbit contributions (σ^{SO}) to the absolute isotropic shielding $\sigma(^{31}P)$ at phosphorus and $\sigma(^{183}W)$ at tungsten in free phosphido $[(CO)_5W\{P(Me)F\}]^-$ (Me/-) and phosphinidenoid complexes [Li(12-cr-4)][$(CO)_5W\{P(R)F\}$] **A**-**D**, R = Me, $CH(SiMe_3)_2$ (Bisyl); cr = 12-crown-4.

So far the compliance constants of structure type **B** could not be obtained, since the B3LYP calculations failed to converge for this starting geometry and ran into structure type **D** instead. The VWNBP86 SO ZORA calculations converged without difficulties.

From the ion-pair character of phosphinidenoid complexes an explanation for the yet unsatisfactory agreement between calculated and experimental NMR parameter can be inferred: given the sensitivity of spectroscopical properties to slight changes in the molecular geometry in solution as well as in solid state the variation of the ion-pairs as compared to the gas phase might be large enough to account for the differences $\Delta(\delta(^{31}P)) \approx 50$ ppm, $\Delta(\delta(^{183}W)) \approx 100$ ppm, $\Delta(^1J(W,P)) \approx 50$ Hz.

Acknowledgements

All computations on the JUMP were performed with a grant of computer time provided by the Research Centre Jülich (Project HBN12).

References

1. N. S. Ramsey, *Phys. Rev.* **1950**, *78*, 699-703.
2. C. J. Jameson, *Ann. Rev. Phys. Chem.* **1996**, *47*, 135–169.

3. F. Jensen, *Introduction to Computational Chemistry* (Wiley, Chichester, 1999).
4. T. Helgaker, M. Jaszunski, K. Ruud, *Chem. Rev.* **1999**, *99*, 293–352.
5. Gaussian 03, Revision B.02/B.03, Wallingford CT, 2004.
6. ADF 2007.1, Theoretical Chemistry,Vrije Universiteit, Amsterdam, a) G. te Velde, E. J. Baerends, *J. Comput. Phys.* **1992**, *99*, 84. b) E. J. Baerends, D. E. Ellis, and P. Ros, *Chem. Phys.* **1973**, *2*, 41. c) E. J. Baerends, P. Ros, *Chem. Phys.* **1973**, *2*, 52.
7. a) Becke, A. D., *J. Chem. Phys.* **1993**, *98*, 5648. b) R. Krishnan, J. S. Blinkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650. c) A. D. McLean, G. S. Chandler, *J. Chem. Phys.* **1988**, *53* , 1995. d) P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1988**, *82*, 270.
8. S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.
9. A.D. Becke, *Phys. Rev.* **1988**, *38A*, 3098.
10. J. P. Perdew, *Phys. Rev.* **1986**, *33B*, 8822; Erratum: J.P. Perdew, *Phys. Rev.* **1986**, *34B*, 7406.
11. G. Fogarasi, X. Zhou, P. W. Taylor, P. Pulay, *J. Am. Chem. Soc.* **1992**, *114*, 8191-8201.
12. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1994).
13. GIAO: G. Schreckenbach, T. Ziegler, *J. Phys. Chem.* **1995**, *99*, 606-611.
14. ZORA GIAO: G. Schreckenbach, T. Ziegler, *Int. J. Quant. Chem.* **1997**, *61*, 899-918.
15. ZORA NSSCC: J. Autschbach, T. Ziegler, *J. Chem. Phys.* **2000**, *113*, 936-947.
16. SO ZORA NSSCC: J. Autschbach, T. Ziegler, *J. Chem. Phys.* **2000**, *113*, 9410-9418.
17. C. van Wuellen, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2137-2144.
18. M. Minelli, J. H. Enemark, R. T. C. Brownlee, M. J. O'Connor, A. G. Weed, *Coord. Chem. Rev.* **1985**, *68*, 169-278.
19. B. P. Johnson, G. Balazs, M. Scheer, *Coord. Chem. Rev.* **2004**, *232*, 1-23.
20. U. Vogel, K.-C. Schwan, M. Scheer, *Eur. J. Inorg. Chem.* **2004**, -, 2062-2065.
21. T. Rotter, K. Karaghiosoff, P. Mayer, M. Westerhausen, *Het. At. Chem.* **2005**, *16*, 420-425.
22. F. Nief, F. Mercier, F. Mathey, *J. Organomet. Chem.* **1987**, *328*, 349-345.
23. M. Hoefler, H. Hausman, A. Schneider, *Z. Naturforsch. B* **1978**, *33*, 1559-1560.
24. G. Frenking, K. Wichmann, N. Froehlich, J. Grobe, W. Golla, D. Le Van, B. Krebs, M. Laege, *Organometallics* **2002**, *21*, 2921-2930.
25. A. Berces, *J. Computat. Chem.* **1997**, *18*, 45-55.
26. K. Joerg, W. Malisch, W. Reich, A. Meyer, U. Schubert, *Angew. Chem.* **1986**, *98*, 103-104; *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 92-93.
27. E. A. V. Ebsworth, A. P. McIntosh, M. Schroeder, *J. Organomet. Chem.* **1986**, *312*, C41-C43.
28. M. Alonso, M. E. Garcia, M. A. Ruiz, H. Hamidov, J. C. Jeffrey, *J. Am. Chem. Soc.* **2004**, *126*, 13610-13611.
29. C. M. Alvarez, M. A. Alvarez, D. Garcia-Vivo, M. E. Garcia, M. A. Ruiz, D. Saez, L. R. Falvello, T. Soler, P. Herson, *Dalton Trans.* **2004**, -, 4168-4179.
30. A. Oezbolat-Schoen, University of Bonn, planned PhD Thesis.
31. A. Oezbolat, G. von Frantzius, R. Streubel, *Angew. Chem.* **2007**, *119*, 9488-9491; *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 9327-9330.

32. A. Oezbolat, G. von Frantzius, W. Hoffbauer, R. Streubel, *Dalton Trans.* **2008**, -, 2674-2676.
33. A. Mueller, M. Marsch, K. Harms, J. C. W. Lohrenz, G. Boche, *Angew. Chem.* **1996**, *108* , 1639-1640; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1518-1520.
34. G. Molev, D. Bravo-Zhivotovskii, M. Karni, B. Tumanski, M. Botoshansky, Y. Apeloig, *J. Am. Chem. Soc.* **2006**, *128*, 2784-2785.
35. Review: A. V. Bajaj, N. S. Poonia, *Coord. Chem. Rev.* **1988**, *87*, 55-213.
36. R. Streubel, S. Priemer, F. Ruthe, P. G. Jones, *Eur. J. Inorg. Chem.* **2000**, -, 1253-1259.
37. E. Ionescu, G. von Frantzius, P. G. Jones, R. Streubel, *Organometallics* **2005**, *24*, 2237-2240.

Elementary Particle Physics: Lattice QCD at the Crossroads

Richard Kenway

School of Physics & Astronomy, The University of Edinburgh, Edinburgh EH9 3JZ, UK
E-mail: r.d.kenway@ed.ac.uk

After more than 35 years of effort, lattice QCD is on the verge of realising its original objectives[1]. Through the use of computer simulation, in which a finite space-time volume is approximated by a regular hypercubic lattice, we are close to proving that Quantum Chromodynamics correctly describes the strong interactions of quarks and gluons at low energies where they are permanently bound into hadrons, as well as in high-energy collisions where perturbation theory works. This has extended the range of applicability of quantum field theory in Nature from weak coupling, analytically tractable phenomena to a regime in which the fundamental quanta interact strongly and all analytical methods fail. The specific goal of computing the spectrum of bound states in QCD from first principles and showing that this precisely reproduces the experimentally observed hadron spectrum has now been achieved for a non-trivial subset of the states, as reported in Ref. 2.

Huge computational effort and theoretical breakthroughs were needed to reach this point. The rigorous construction of QCD as the continuum limit of a lattice theory required a deep understanding of how symmetries and renormalisation are realised in Nature. Some of this is explained in Ref. 3. Demonstrating that we have done this correctly has required the entire development to be repeated in different lattice formulations, carrying the calculations of physical quantities through to completion to demonstrate that they really give the same results. One of these alternative formulations is described in Ref. 4. The hard work has been to control all sources of numerical uncertainty in the limiting process, achieved through a series of simulations, in which the lattice spacing is reduced and the results are extrapolated to the continuum.

Of course, the physics of the strong interaction is enormously rich and diverse. Now that confidence in lattice QCD is well-founded, the potential applications are vast. For example, our inability to compute precisely enough some matrix elements, which describe the strong-interaction effects that mask the underlying decays of one flavour of quark into another, are restricting experimental efforts to put constraints on possible physics beyond the Standard Model. Lattice QCD is starting to reduce the theoretical uncertainties in these processes to below that of experiment, creating the potential for discoveries. The proof that lattice QCD accurately describes Nature will be essential if we are ever to use these results to conclude that we are seeing new physics. Together, the following three papers[2–4] describe the platform that has been created for this future work. While much remains to be done, even to tie up the loose ends in current calculations of the hadron spectrum, there is little doubt that Wilson's original vision will be realised. Where do we go next?

The tool we have built is expensive and will remain so for some years. Yet it has established the feasibility of applying quantum field theory from first principles and banished

the need to rely on *ad hoc* models of strong-interaction physics. One critical research path will lead to exascale and relatively cheap petascale machines. These will be needed to broaden the application of lattice quantum field theory and turn lattice QCD into a phenomenological tool accessible by non-experts. A major challenge will be to exploit strong scaling, because the computational cost of QCD grows much faster than the number of lattice sites and the parallelism in exascale machines, with hundreds of millions of cores, will outstrip that available through domain decomposition, i.e. weak scaling. A highly optimised QCD code is shown to exhibit near-perfect strong scaling up to 300,000 cores and sustain one-third of a petaflop/s in Ref. 5. Fault tolerance, validation and verification of simulation results already feature in lattice QCD methodology. They will be mandatory at exascale and so lattice QCD has a leadership role to play.

Technology push alone will not be enough in the LHC era. The field needs to articulate goals that the wider scientific community recognises justify the cost of the exascale project, when interest is shifting towards experimental discoveries and major societal challenges are demanding attention. Precise theoretical underpinning of flavour physics at LHCb must be a priority for most of the next decade. Solving the "sign problem" by finding an algorithm which enables simulations at any baryon chemical potential would have ramifications beyond the heavy-ion physics at ALICE, e.g. for understanding superconductivity, but this goal seems elusive. Nuclear physics could be reborn from first principles, but is its community ready for change? Perhaps, the most exciting opportunity is for lattice QCD to return to its roots and explore different realisations of strong coupling in quantum field theories. QCD has demonstrated that Nature uses strong coupling. Why only once? In struggling to understand the physics of LHC, such as electroweak symmetry breaking, we no are no longer constrained to use perturbation theory as our guide.

References

1. K.G. Wilson, *Confinement of Quarks*, Phys. Rev. D10, 2445-2459, 1974.
2. Z. Fodor, *Lattice QCD with Fat Clover Fermions: Ab Initio Determination of Flavour Non-Singlet Mesons and Baryons*, this volume, pp. 139–146.
3. N. Cundy, I. Koutsou, T. Lippert and A. Schäfer, *Dynamical Overlap Fermions with Small Pion Masses and Topology*, this volume, pp. 131–138.
4. K. Jansen, *QCD Simulations with Light, Strange and Charm Dynamical Quark Flavours*, this volume, pp. 147–154.
5. S. Krieg and T. Lippert, *Tuning Lattice QCD to Petascale on Blue Gene/P*, this volume, pp. 155–164.

Dynamical Overlap Fermions with Small Pion Masses and Topology

Nigel Cundy[1], **Ioannis Koutsou**[2], **Thomas Lippert**[2], **and Andreas Schäfer**[1]

[1] Institut für Theoretische Physik, Universität Regensburg, 93040 Regensburg, Germany
E-mail: {*nigel.cundy, andreas.schaefer*}*@physik.uni-regensburg.de*

[2] Jülich Supercomputing Centre, Research Centre Jülich, 52425 Jülich, Germany
E-mail: {*i.koutsou, th.lippert*}*@fz-juelich.de*

Lattice Quantum Chromodynamics is a numerical tool used to calculate observables from the theory of the strong nuclear force. Although current lattice simulations now have errors of only a few percent, they do so using formulations which break the important chiral symmetry. Using a formulation which respects chiral symmetry reduces the statistical noise, allows cleaner extrapolations to the target continuum theory, and removes the need to accurately tune additive residual masses; and will allow even more accurate lattice calculations. However, although there are known formulations of Lattice Quantum Chromodynamics which respect chiral symmetry, their simulation remains a technical, algorithmic and computational challenge. In these proceedings we outline the various problems encountered and our solutions to them.

1 Introduction

The theory of the strong nuclear force, Quantum Chromodynamics (QCD), is one component of the phenomenologically successful standard model of particle physics. In contrast to the situation for the electromagnetic and weak interactions, accurate QCD calculations for many observables have not been possible so far. The usual approach, perturbation theory, which involves an expansion in powers of the coupling constant often breaks down. In a few cases, these problems can be overcome by quite involved re-summation techniques, but in general the remaining uncertainties are very large. In fact, these uncertainties substantially reduce the possibilities to identify "New Physics" at accelerator centres like CERN or FAIR.

Fortunately it is possible to calculate many of the problematic observables purely numerically. QCD itself cannot be simulated numerically, because it has an infinite number of degrees of freedom, but a variant of QCD, lattice QCD (LQCD) can. In LQCD one discretizes space-time so that the the theory has very many, but not infinitely many degrees of freedom and can thus be simulated with Monte-Carlo techniques. As a matter of fact, LQCD is not a model, but a well defined field theory in its own right. For small enough lattice constant (the distance between two nearest lattice points) it is not possible to decide empirically whether LQCD or continuum QCD is the correct description, as there exists a formalism of LQCD, based on the overlap fermions used here, which is strictly equivalent to the continuum theory[1,2]. More precisely, LQCD approaches QCD as three limits are taken in the correct sequence. First the lattice spacing has to be extrapolated to zero, then the volume to infinity, and then the masses to their physical values. Each of these limits can be controlled systematically. The scale parameter of LQCD and quark masses are each set by pinning one observable to its continuum value; everything else can be calculated from first principles. After many years of work characterized by large systematic errors

and slow progress, now lattice QCD is capable of producing results with a precision of only a few percent[3].

The recent advances in lattice QCD were driven by two developments: firstly the availability of greater computing resources, such as the Blue Gene/P at the Jülich Supercomputing Centre, and secondly the rapid improvement in algorithms (for example, compare the predictions for the computer time needed of the Berlin 2001 lattice conference[4] with what has been achieved in the past few years[5]). Without this algorithmic and theoretical work, the increase in computing power would not have been enough to address the problems we are addressing today.

In this ongoing project, we are attempting to simulate lattice QCD with an exact lattice chiral symmetry. As discussed in section 2, this approach is very demanding both computationally and technically. The aim of our work is partly to develop and establish methods for future research, and partly to extract physics observables which can be determined on small lattice volumes. Section 2 gives a detailed motivation for the project, section 3 describes the lattice formulation used, section 4 discusses algorithmic challenges and their solution, section 5 describes the status of our current simulations, and section 6 outlines our current and future plans.

2 Motivation

Lattice QCD is successful because the symmetries of the continuum theory are either maintained or smoothly restored in the continuum limit. There is one crucial exception in most current simulations: chiral symmetry. Chiral symmetry is an exact symmetry of massless QCD, and a very good approximate symmetry for light hadrons. Much of our understanding of the light hadron spectrum is related, e.g. to the spontaneous breaking of chiral symmetry by the QCD vacuum. The most naive formulation of LQCD respects chiral symmetry, but the Dirac equation has 16 solutions instead of just one. This is the fermion doubling problem. It is possible to remove these doublers by adding a chiral symmetry breaking term to the Dirac operator, but, as shown by Nielson and Ninoyima[6], it is impossible to do so without breaking the continuum formulation of chiral symmetry. Of the currently most popular formulations of LQCD, Wilson-clover fermions explicitly break chiral symmetry, while improved staggered fermions have additional doublers. For Wilson-clover fermions, although chiral symmetry is restored in the continuum limit the restoration is not smooth. For staggered fermions the consequences are unclear. While both approaches have been successfully employed, there are thus still lingering theoretical concerns. Domain wall fermions reduce the chiral symmetry violation by a considerable amount, but at a substantial cost in computer time. Thus there may well be systematic errors affecting observables particularly sensitive to chiral symmetry, and there are severe difficulties for many observables caused by mixing between operators which would normally be forbidden by chiral symmetry. All of this adds to the total error as the continuum limit is taken.

However, even though it is not possible to construct continuum chiral symmetry on the lattice, it is none the less possible to construct a symmetry which has a smooth limit to the continuum. This solves the various problems associated with the breaking of chiral symmetry on the lattice. The way to achieve this is through the Ginsparg-Wilson formulation[7] which was derived by considering how the symmetry changes under re-normalisation

group blocking. The Ginsparg-Wilson relation for a Dirac operator D can be written as[1]

$$\gamma_L D + D\gamma_R = 0, \tag{1}$$

and any Dirac operator satisfying this relation for local γ_L and γ_R which smoothly reduce to γ_5 in the continuum limit will satisfy a lattice variant of chiral symmetry[8]. It was found that the overlap operator[9–11] satisfied the Ginsparg-Wilson relation, and possesses this key symmetry. If both the Yang-Mills action and the fermion action are expressed in terms of the overlap operator[12], one does, therefore, expect an exceptionally good scaling behaviour when the continuum limit is taken[1]. This allows, e.g. to use coarser lattices than for other formulations[13] and generally speaking overlap fermions are the theoretically cleanest approach to lattice QCD. There is no additive mass re-normalisation, no operator mixing, no need for tuning of the action, a clearly defined topological index, and the overlap fermion formulation can be used to construct Weyl and Majorana fermions on the lattice[1,14]. They are also, however, by far the most challenging formulation to implement efficiently.

The overlap operator for a mass μ is defined as

$$D = 1 + \mu + (1 - \mu)\gamma_5 \text{sign}(K), \tag{2}$$

where the kernel operator, K, can in principle be any local lattice Dirac operator with no doublers, although in practice the Wilson Dirac operator with several levels of stout smearing[15] is usually used. The difficulties are caused by the matrix sign function. Firstly, it is expensive to calculate, and it requires the eigenvectors of the kernel operator (and of the overlap operator, for preconditioning), which gives the overlap operator poor scaling properties with increasing volume compared to other formulations. Therefore, it is difficult to simulate overlap fermions at the required physical lattice volume needed for the measurement of physical observables, although once that volume is reached the advantages of overlap fermions will start to become manifest.

Lattice QCD calculations use Monte-Carlo importance sampling. The method of choice is the Hybrid Monte-Carlo algorithm[16] (HMC), which has better scaling properties with respect to the lattice volume than other approaches and generally has short autocorrelation times, but requires the differential of the action with respect to the gauge field. For the overlap function, this involves the differentiation of the matrix sign function, causing a Dirac δ-function in the force whenever one of the eigenvalues of the kernel operator crosses zero. This is invisible to standard HMC, so the HMC algorithm needs to be modified to make the δ-function force visible. Eigenvalue crossings correspond to a changes in the topological index, and are necessary to maintain ergodicity.

Therefore, overlap fermions represent both a challenge and an opportunity for lattice field theory. The challenge concerns the computing resources required, the difficulty of generating efficient code, and the development of novel algorithmic techniques. The opportunity is that once all tasks are accomplished, overlap fermions will remove a number of systematic uncertainties affecting other formulations of QCD.

3 The LQCD Overlap Action

The gauge field U in LQCD is given by an SU(3) matrix on each link connecting two lattice sites; each link can be written in terms of eight coordinates u_a. So if there are V lattice sites and $d = 4$ dimensions, the size of the system is $8 \times 4V$. The differential of a

function of the gauge field with respect to the coordinate u for a lattice site x and direction μ is defined as

$$\partial^a_{x\mu} f(U) = \lim_{t\to 0} \frac{d}{dt^a} f(U_t(x,\mu;y,\nu));$$

$$U_t(x,\mu;y,\nu) = \begin{cases} e^{iT^a t^a} U(y,\nu) & [y,\nu] = [x,\mu] \\ U(y,\nu) & \text{otherwise} \end{cases} \tag{3}$$

The aim is to weight the ensembles according to the action, which in Euclidean space-time is $e^{S[U]} = e^{-S_g[U] + \mathrm{Tr}\log D[U]}$. Observables are then measured either in terms of the inverse of the Dirac operator D or closed loops of gauge fields. The gauge action, $S_g[U]$, is a lattice representation of the Yang-Mills action $\frac{\beta}{4} F^2_{\mu\nu}$ (β is a tunable parameter which controls the lattice spacing), which can be expressed in two ways, either as the sum of traces of various closed loops of the gauge field (for example the simplest four link loop, the Wilson plaquette action, or more sophisticated combinations of four and six loops) or as a function of the Dirac operator[12,13]. For the overlap Dirac operator, the second method is theoretically superior but more expensive. For that reason we are for the moment using a standard Lüscher-Weisz gauge action.

The Dirac operator is given by Eq. (2) with the kernel

$$K = \gamma_5 \left[1 - \kappa \sum_\mu \left[(1-\gamma_\mu)\delta_{x+\hat{\mu},x'} \tilde{U}_\mu(x) + (1+\gamma_\mu)\delta_{x-\hat{\mu},x'} \tilde{U}^\dagger_\mu(x-\hat{\mu}) \right] \right]. \tag{4}$$

The determinant of the Dirac operator cannot be calculated exactly, so it must be estimated stochastically. A momentum π conjugate to the gauge fields is introduced according to a Gaussian distribution, and an energy functional is constructed $H_\tau = \pi^2/2 + S[U]$. The system is then integrated over the classical energy conserving trajectory with respect to a fictitious computer time τ using a reversible and measure conserving integrator. The symplectic Omelyan integrator is the current method of choice for lattice simulations[17]. Finally the unavoidable numerical violation of energy conservation is corrected for using a Metropolis step. The difficultly with overlap fermions, compared to other types of actions, is that the overlap operator is discontinuous because the matrix sign function needs to be differentiated. This is discussed in the next section.

4 Algorithmic Challenges

4.1 Differential of the Matrix Sign Function

The matrix sign function of a Hermitian operator is defined in terms of a spectral decomposition

$$\mathrm{sign}(K)\phi = \sum_i \psi_i(\psi_i,\phi)\mathrm{sign}(\lambda_i), \tag{5}$$

where ψ_i represent the eigenvectors of K and λ_i the eigenvalues. For the bulk of the eigenvalue spectrum, a Zolotarev rational approximation can be used[18], but for the smallest kernel eigenvalues this is impossible because this approximation cannot cope with the discontinuity to the required accuracy. The discontinuity in the sign function is precisely

what gives the overlap operator its good chiral properties[1] and continuous approximations to the sign function, which do exist and are used, e.g. for Domain Wall fermions, violate chiral symmetry and thus the property which motivated overlap fermions in the first place. Therefore, the eigenvalues closest to zero should be deflated and treated exactly. One can write the mixing between two eigenvectors when there is a small change in the gauge field in terms of mixing angles θ and ϕ, so that (if the mixing angle θ is sufficiently small)

$$\delta\psi_i = (1 - \cos\theta_{ij})\psi_i + e^{i\phi_{ij}} \sin\theta_{ij}\psi_j, \tag{6}$$

where

$$\tan 2\theta_{ij} = \frac{2\sqrt{(\psi_j, \delta K \psi_i)(\psi_i, \delta K, \psi_j)}}{\lambda_i - \lambda_j + (\psi_i, \delta K \psi_i) - (\psi_j, \delta K \psi_j)} \tag{7}$$

$$e^{i\phi_{ij}} = \sqrt{\frac{(\psi_j, \delta K \psi_i)}{(\psi_i, \delta K \psi_j)}}. \tag{8}$$

If the eigenvalues are well separated from zero, this can be expanded in $\delta K/(\lambda_i - \lambda_j)$, giving

$$\delta(\text{sign}(K))\phi = \sum_{i \neq j} \psi_i (\psi_i, \delta K \psi_j)(\psi_j, \phi) \frac{\text{sign}(\lambda_i) - \text{sign}(\lambda_j)}{\lambda_i - \lambda_j}$$

$$+ \sum_i \psi_i (\psi_i, \phi)(\psi_i, \delta K \psi_i)\delta(\lambda_i). \tag{9}$$

If there are two (or more) eigenvalues close to zero then the more exact form given in Eq. (6) must be used. This creates some difficulties with regard to conservation of measure, since Eq. (6) is a function of the momentum as well as the gauge field, and thus if it is used naively within a force, the update of the momentum fields will neither conserve the measure nor be reversible. However, these problems can be overcome[19].

4.2 The Transmission/Reflection Routine

The second issue is how to contend with the Dirac-δ function in Eq. (9). The simplest answer is to treat it as a potential wall of height ΔS in classical mechanics. One integrates up to the wall, calculates the momentum normal to the wall (in the direction η within the space of gauge field variables), and if this momentum is large enough, one transmits through the wall with a reduced momentum, $(\pi, \eta)^2/2 \to (\pi, \eta)^2/2 - \Delta S$, and if the momentum is small enough, one reflects off the wall, $(\pi, \eta) \to -(\pi, \eta)$. This original algorithm[20] is not optimal because it only conserves energy up to $\delta\tau$ rather than $(\delta\tau)^2$ as in the rest of the molecular dynamics algorithm, and because it does not maximize the probability of transmission given a particular ΔS and a Gaussian momentum distribution. These problems can be overcome by firstly modifying the momentum in directions normal to η to overcome the $\delta\tau$ errors[21], and secondly by using a more sophisticated change in the momentum $e^{-(\eta,\pi)^2/2} \to e^{-(\eta,\pi)^2/2+\delta S} + (1 - e^{\Delta S})$[22]. This naively no longer conserves area or energy, but the Jacobian of this update cancels out the discrepancy in energy conservation, so that detailed balance is still satisfied. Using this method increases the probability of transmission by about a factor of three, and reduces the auto-correlation by a similar amount.

4.3 Reducing the Potential Wall

The next problem is that using the pseudo-fermion estimate of the determinant, ΔS is still too large to allow tunnelling between different topological sectors, particularly at low quark mass. The reason for this is that the pseudo-fermion approach requires the inverse of the squared Hermitian Dirac operator, which means that the noise generated by the pseudo-fermion estimate of ΔS is inversely proportional to the quark mass squared. This leads to an autocorrelation of order $e^{-\Delta S} \sim e^{-\alpha/\mu^2}$. This large autocorrelation is now becoming a concern for other approaches to lattice QCD[23]. The change in the actual determinant, however, contains no strong mass dependence. We avoid this large autocorrelation by writing a new Dirac operator in terms of a randomly chosen parameter λ_0, $D[\lambda_0] = 1 + \mu + (1 - \mu)\text{sign}(K + \lambda_0)$, factorising the fermion determinant as $\det(D[0]) = \det(D[0]/D[\lambda_0])\det(D[\lambda_0])$. $\det(D[0]/D[\lambda_0])$ can be treated exactly without pseudo-fermions, and since the topological index change is only visible in this term and not the term $\det(D[\lambda_0])$, which is estimated by pseudo-fermions, ΔS is reduced by several orders of magnitude and loses its poor scaling with the quark mass.

4.4 Algorithmic Advantages of Overlap Fermions

All of the above, together with the additional cost needed for the rational approximation of the matrix sign function, implies that a hybrid Monte-Carlo code for overlap fermions with topological charge changes (a Japanese collaboration avoids these problems by artificially suppressing the small eigenvalues of the Dirac operator[24]) is considerably more complex and computationally slower than other lattice QCD formulations. However, we can take advantage of the chiral symmetry of the overlap operator in several ways. The first method, well developed and in use in our current production code, can be used if additional quark flavours are introduced. It is customary to use degenerate up and down quarks, so here the square of the Dirac operator is used. For additional flavours of fermions, it is necessary, in every formulation except for overlap fermions, to use a rational approximation of the square root of the squared matrix to ensure that the pseudo-fermion estimate is based on a positive definite matrix. In particular it turned out that preconditioning of the multi-shift inversion is difficult. For overlap fermions, one can instead use a chiral decomposition[22], and the usual preconditioning tricks. This saves approximately a factor of five in the additional cost when going from $N_f = 2$ to $N_f = 2 + 1$ fermions. The same will hold for any other additional fermion flavour.

The second approach, which we are currently developing and which is not yet ready to be implemented into our main production code, is to use the favourable properties of overlap fermions under re-normalisation group blocking[1] to construct a coarse lattice overlap operator from a fine lattice gauge field. Basically the action on the coarse lattice is equivalent to a lattice action on the fine lattice, which can be constructed explicitly if necessary. Using the re-normalisation group to extrapolate between the coarse and fine lattices, all physical observables can be calculated on the coarse lattice. This allows us to simulate the fermions on a lattice with a much reduced number of lattice points. Once perfected, we hope that this method might lead to gains of up to two orders of magnitude in computer time[13].

5 Simulation Progress

We intend to perform $N_f = 2+1$ flavour simulations at the Jüelich Supercomputing Centre on $16^3 \times 48$ lattices. We are currently in the process of tuning the three quark masses and the lattice spacing for a smaller volume, $12^3 \times 24$, where the simulations run considerably faster. At the same time, we are tuning the parameters of the HMC algorithm to maximise its efficiency, and we work on further improving the quality of our code.

We have developed optimised Wilson matrix and inversion routines, which dominate most lattice simulations, for the Blue Gene/P and JUROPA, to the extent that our production runs are now dominated by eigenvalue calculations and smearing routines. We are currently in the process of improving these parts of our code too.

6 Physical Observables

We intend to use our lattices to study aspects of QCD which are especially sensitive to chiral symmetry, for example, the relationship between confinement and spontaneous chiral symmetry breaking, certain aspects of hadronic structure, especially those strongly affected by the pion cloud, non-perturbative re-normalisation, the neutral kaon mixing parameter B_K, the meson η', the CKM unitarity triangle and the $K \to \pi\pi$ decay. We will start by investigating the vacuum structure of QCD, building on methods used in quenched simulations. This has the advantage that it can be done on relatively small lattices and with low statistics.

7 Conclusions

Lattice QCD with exact chiral symmetry is an enormous algorithmic and computational challenge. The algorithmic work has largely been completed. However, the required computational resources are still very large, and physics production runs are barely possible with current machines. However, once the method is established and large scale simulations become possible, it promises to produce considerably superior results to other methods, with far easier continuum and chiral extrapolations.

Acknowledgements

Computer simulations have been carried out on the IBM supercomputers JUGENE, Jump and JUROPA and a PC cluster at the Jülich Supercomputing Centre.

References

1. Nigel Cundy, *A renormalisation group derivation of the overlap formulation*, Nuclear Physics, **B824**, 42, 2010.
2. N. Cundy, in preparation.
3. S. Durr et al., *Ab-Initio Determination of Light Hadron Masses*, Science, **322**, 1224–1227, 2008.

4. A. Ukawa, *Computational cost of full QCD simulations experienced by CP-PACS and JLQCD Collaborations*, Nuclear Physics B - Proceedings Supplements, **106-107**, 195–196, 2002.
5. M. A. Clark, *The rational hybrid Monte Carlo algorithm*, PoS, **LAT2006**, 004, 2006.
6. H. B. Nielsen and M. Ninomiya, *Absence of Neutrinos on a Lattice. 1. Proof by Homotopy Theory* , Nucl. Phys, **B185**, 20, 1981.
7. P. H. Ginsparg and K. G. Wilson, *A remnant of chiral symetry on the lattice*, Phys. Rev., **D25**, 2649, 1982.
8. Martin Lüscher, *Exact chiral symmetry on the lattice and the Ginsparg-Wilson relation*, Phys. Lett., **B428**, 342–345, 1998.
9. Herbert Neuberger, *A practical implementation of the overlap-Dirac operator*, Phys. Rev. Lett., **81**, 4060–4062, 1998.
10. Rajamani Narayanan and Herbert Neuberger, *Chiral determinant as an overlap of two vacua*, Nucl. Phys., **B412**, 574–606, 1994.
11. R. Narayanan and H. Neuberger, *Chiral fermions on the lattice*, Phys. Rev. Lett., **71**, 3251–3254, 1993.
12. Ivan Horvath, *A framework for systematic study of QCD vacuum structure. II: Coherent lattice QCD*, 2006.
13. Nigel Cundy, *Multi-grid HMC for Ginsparg-Wilson fermions*, 2009.
14. Yuji Igarashi and Jan M. Pawlowski, *Majorana fermions and CP-invariance of chiral gauge theories on the lattice*, Nucl. Phys., **B821**, 228–240, 2009.
15. Colin Morningstar and Mike J. Peardon, *Analytic smearing of SU(3) link variables in lattice QCD*, Phys. Rev., **D69**, 054501, 2004.
16. S. Duane, A. Kennedy, B. Pendleton, and D. Roweth, *Hybrid Monte-Carlo*, Phys. Lett., **B195**, 216, 1987.
17. I. P. Omelyan, I. M. Mryglod, and R. Folk, *Symplectic analytically integrable decomposition algorithms: classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations*, Computer Physics Communications, **151**, no. 3, 272 – 314, 2003.
18. J. van den Eshof, A. Frommer, T. Lippert, K. Schilling, and H. A. van der Vorst, *Numerical methods for the QCD overlap operator. I: Sign- function and error bounds*, Comput. Phys. Commun., **146**, 203–224, 2002.
19. Nigel Cundy, *Small Wilson Dirac operator eigenvector mixing in dynamical overlap hybrid Monte-Carlo*, Comput. Phys. Commun., **180**, 180–191, 2009.
20. Z. Fodor, S. D. Katz, and K. K. Szabo, *Dynamical overlap fermions, results with hybrid Monte-Carlo algorithm*, JHEP, **08**, 003, 2004.
21. N. Cundy et al., *Numerical methods for the QCD overlap operator. IV: Hybrid Monte Carlo*, Comput. Phys. Commun., **180**, 26–54, 2009.
22. Nigel Cundy, *Current status of dynamical overlap project*, Nucl. Phys. Proc. Suppl., **153**, 54–61, 2006.
23. Stefan Schaefer, Rainer Sommer, and Francesco Virotta, *Investigating the critical slowing down of QCD simulations*, 2009.
24. T. Kaneko et al., *JLQCD's dynamical overlap project*, PoS, **LAT2006**, 054, 2006.

Lattice QCD with Fat Clover Fermions: *Ab Initio* Determination of Flavour Non-Singlet Mesons and Baryons

Zoltán Fodor[1,2,3]

[1] Bergische Universität Wuppertal, Gaussstr. 20, D-42119 Wuppertal, Germany

[2] Institute for Theoretical Physics, Eötvös University, H-1117 Budapest, Hungary

[3] Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany

E-mail: fodor@bodri.elte.hu

We summarize results from a series of large-scale lattice QCD computations, which were predominantly carried out on the Jülich Blue Gene installation. Emphasis is put on our choice of the lattice action and algorithm, and the procedures by which we extrapolate to the physical mass point, to the continuum and to infinite volume. In this way the spectrum of the low-lying flavour non-singlet meson and baryon states is obtained *ab initio*, with fully controlled systematics.

1 Introduction

The bulk part of the visible matter in the universe is made from protons and neutrons. Quantum Chromodynamics (QCD), the modern theory of strong interactions, describes how quarks and gluons amalgamate to form nucleons and other strongly interacting states, called hadrons. Due to the highly non-linear structure of these interactions, any form of *ab initio* determination of the properties of these hadrons from the underlying interaction of quarks and gluons has remained a challenge for decades. The goal of this contribution is to describe how lattice QCD tackles the problem and allows us to obtain results with fully controlled systematics.

At first sight QCD is rather analogous to QED, the quantum theory of relativistic electrodynamics. This is because the Lagrangian[1] (in Euclidean space)

$$L_{\rm QCD} = \frac{1}{2}{\rm Tr}(F_{\mu\nu}F_{\mu\nu}) + \sum_{i=1}^{N_f} \bar{\psi}(D{+}m_i)\psi \qquad (1)$$

closely resembles the Lagrangian of QED. The difference is that the field-strength tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + {\rm i}g[A_\mu, A_\nu]$ is (for a given μ, ν combination) a 3×3 matrix in "colour" space with a commutator term (which in the case of QED is zero). The symbol ${\rm Tr}(.)$ refers to the trace over this internal degree of freedom. A similar sum is implied in the second term of Eq. (1) which, in addition, contains an explicit sum over the "flavours" one wants to deal with, usually "up" (u), "down" (d) and "strange" (s). This commutator term and the pertinent non-linearities render QCD a fascinating theory, but also a demanding one. One key property is called "asymptotic freedom"[2,3]. It states that the effective (renormalized) version of the coupling g showing up in $F_{\mu\nu}$ becomes weak at short distances. In other words, at high energies (or high relative momentum) quarks and gluons

interact weakly, allowing for standard perturbative methods to be used. By contrast, at low relative momentum the interaction becomes strong, to the point where asymptotic states do not carry a net "colour" charge any more. This phenomenon is called "confinement".

In order to derive these properties from the QCD Lagrangian (1), the theory needs to be regularized, both in the ultraviolet and in the infrared. A particularly convenient choice is the lattice, as introduced by Wilson[4,5]. The finite lattice spacing a and the finite four-dimensional box volume $V = L^4$ render the path-integral for the generating functional

$$\begin{aligned} Z[\bar{\eta},\eta] &= Z_0 \int DU\, D\bar{\psi}D\psi\; e^{-\int S_g\, dx - \sum \int [\bar{\psi}(D+m)\psi + \bar{\eta}\psi + \bar{\psi}\eta]\, dx} \\ &= Z_0 \int DU\; \prod \det(D+m)\, e^{+\sum \int \bar{\eta}(D+m)^{-1}\eta\, dx}\, e^{-\int S_g\, dx} \end{aligned} \tag{2}$$

(with sums and products running over the flavours) well-defined and the important property of gauge symmetry is maintained exactly. The task is then to compute a given dimensionless ratio, e.g. M_N/M_Ξ, for a series of lattice spacings a, box sizes L and quark masses m_u, m_d, m_s, and to perform a well monitored extrapolation of this ratio to zero lattice spacing, infinite volume and physical values of the quark masses. If one can control all sources of error in each step, one can claim to have an *ab initio* result for the strong interactions which provide the bulk part of the mass of the visible matter in the universe.

2 Choice of Lattice Action and Algorithm

Creutz was the first author to notice[6] that well established importance sampling techniques from statistical mechanics can be used to calculate specific ratios of QCD functional integrals in Euclidean space. With this technique and the gluon and fermion discretizations given in Ref. 4, 5 one can in principle carry out Wilson's program. In practice things look somewhat different. The reason is that the cost, in terms of CPU time, grow rather dramatically along each one of the directions $a \to 0$, $V \to \infty$, $m_{u,d} \to m_{u,d}^{\rm phys}$ we aim for.

To mitigate this cost growth some theory investment is required. Here we discuss how the continuum limit is accelerated, postponing the remaining two steps for a moment. The conceptual background for controlling the costs of the continuum limit $a \to 0$ has been laid out by Symanzik[7,8]. The starting point is that on the lattice any dimensionless ratio differs from its value in the continuum by terms which scale with a certain power of a. For the nucleon-to-xi mass ratio considered before, this means that

$$(M_N/M_\Xi)(a) = (M_N/M_\Xi)(0) + \mathrm{const}(a/\Lambda)^n + O((a/\Lambda)^{n+1}) \tag{3}$$

where n is a property of the action used to sample the phase space and to construct the interpolating fields of the states considered, and Λ is a QCD intrinsic scale (e.g. $\Lambda_{\rm QCD}$ or the nucleon mass in the chiral limit).

The crucial observation by Symanzik was that one may augment a given lattice action by "irrelevant" (mass-dimension 5 and higher) terms with coefficients tuned such that "const" in the first sub-leading term of Eq. (3) is zero. This has an enormous impact on the cost, in terms of CPU time, of a physical result in the continuum, because it enlarges the smallest (and hence most expensive) lattice spacing at which simulations need to be carried out. For illustration, let us assume the scaling regime [the region where the $O((a/\Lambda)^{n+1})$ term in Eq. (3) can be neglected] starts at $0.1\,\mathrm{fm}$, and we wish to have data over $3/4$ of the

range in which we extrapolate linearly. In a theory with $O(a)$ cut-off effects, we would sample at $a/\mathrm{fm} \simeq 0.025, 0.05, 0.075, 0.1$, while in a theory with $O(a^2)$ effects we would use $a/\mathrm{fm} \simeq 0.05, 0.071, 0.087, 0.1$, which obviously amounts to great savings.

Following these ideas we use the Symanzik[9] and Sheikholeslami-Wohlert[10] actions

$$S_g = \frac{\beta}{3} \sum_x \Big\{ c_0 \sum_{1 \le \mu < \nu \le 4} [1 - \mathrm{Re}\,\mathrm{Tr}(U^{1\times 1}_{x,\mu\nu})] + c_1 \sum_{1 \le \mu \ne \nu \le 4} [1 - \mathrm{Re}\,\mathrm{Tr}(U^{1\times 2}_{x,\mu\nu})] \Big\} \qquad (4)$$

$$S_f = a^4 \sum_x \Big\{ \frac{1}{2a} \sum_{\mu=1}^{4} \bar{\psi}_x [(\gamma_\mu - I) V_{x,\mu} \psi_{x+\hat{\mu}} - (\gamma_\mu + I) V_{x-\hat{\mu},\mu} \dagger \psi_{x-\hat{\mu}} + (\frac{4}{a} + m) \psi_x] \Big\}$$
$$- a^5 \sum_x \Big\{ \frac{c_{\mathrm{SW}}}{2} \sum_{\mu<\nu} \bar{\psi}_x \sigma_{\mu\nu} F_{x,\mu\nu}[V] \psi_x \Big\} \qquad (5)$$

for the quarks and gluons, respectively, with $\sigma_{\mu\nu} = (\mathrm{i}/2)[\gamma_\mu, \gamma_\nu]$ and hermitean $F_{\mu\nu}$. The improvement coefficients are set to their tree-level values, i.e. $c_1 = -1/12$, $c_0 = 1 - 8c_1 = 5/3$ and $c_{\mathrm{SW}} = 1$. Note that in Eq. (4) the original ("thin") gauge links $U_{x,\mu}$ are used to define the 1×1 and 1×2 Wilson loops $U^{1\times 1}_{x,\mu\nu}$ and $U^{1\times 2}_{x,\mu\nu}$, whereas in Eq. (5) modified ("fat") gauge links $V_{x,\mu}$ are used which relate to the $U_{x,\mu}$ through 6 levels of stout smearing[11]. This is a legal modification with an enormous impact on both the stability of the simulation and the speed (see below). Note that the usual ultra-locality of Wilson fermions $D(x,y) = 0$ for $||x-y|| > a$ is maintained, but the ultra-locality with respect to changes in the gauge field is replaced by a locality in the modern sense. By this we mean that $\delta(D(x,y))/\delta(A(z))$ falls off exponentially as z moves away from $x = y$, see Ref. 12 for details.

The cost of a QCD simulation is dominated by frequently solving sparse systems of the form $D\dagger D x = b$ to keep track of the change of the functional determinant $\det(D+m)$ in Eq. (2). There is a battery of tricks to ease this task. We use even-odd preconditioning[15], multiple time-scale integration[16] ("Sexton-Weingarten scheme"), multiple mass preconditioning[17] ("Hasenbusch trick"), Omelyan integrator[18], and mixed-precision solver[19]. As noted in the literature[20,21] combining several of these ingredients yields a dramatic reduction of the critical slowing down that has traditionally been observed for light quark masses.

Having specified our action and algorithm, it is important to verify that it entails a nice scaling of physical observables towards the continuum and – a point particularly relevant for (improved) Wilson fermions – algorithmic stability. We have performed several stability analysis[19], including a hysteresis test (to detect a nearby bulk phase transition) and a determination of the smallest (in magnitude) eigenvalue of the hermitean Wilson operator $H_m = \gamma_5 (D+m)$. The latter test shows how the distribution of $\lambda_{\min}$ gets narrower, at a fixed physical quark mass, as the lattice spacing diminishes. In practice, it is far more convenient to monitor the inverse iteration count $1/n_{\mathrm{CG}}$ of the solver[19], since this quantity is closely related to $\lambda_{\min}$ and comes for free. An explicit example from a $N_f = 2+1$ run (with a degenerate quark mass m_{ud} for the first two flavours and a heavier m_s for the third one) is shown in Figure 1. It indicates a clear gap, providing strong evidence for the stability of the algorithm. We monitor $1/n_{\mathrm{CG}}$ in all our production runs. In Ref. 19 various scaling tests have been carried out, mostly for $N_f = 3$ at a common quark mass which roughly corresponds to the physical strange quark mass. This allows us to reach considerable statistics and hence small error-bars. An example is shown in Figure 2. Both the nucleon (N) and

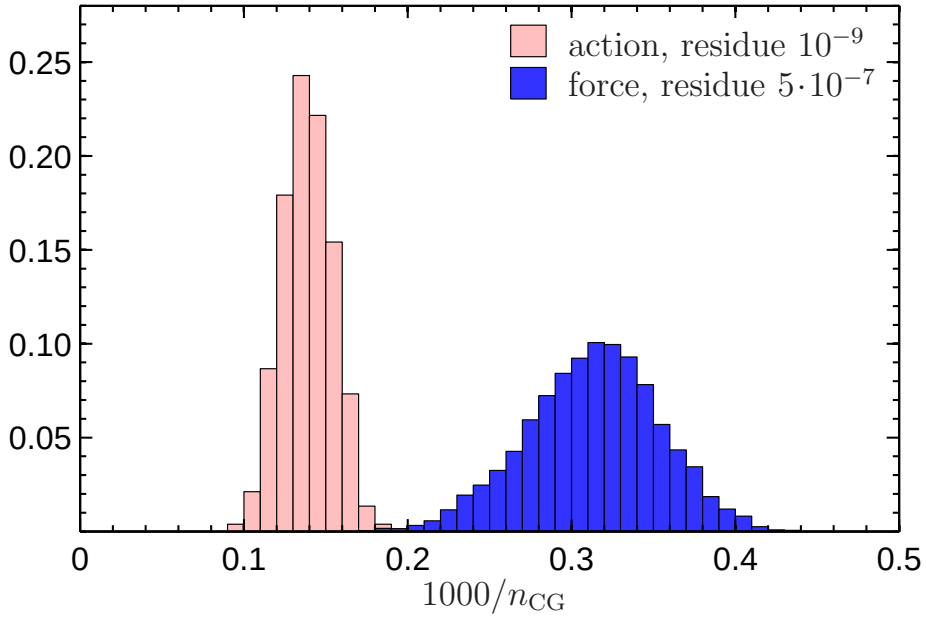


Figure 1. Histogram of the inverse iteration count of our solver for the lightest pseudofermion in the action. Results are from an $N_f = 2{+}1$ run on a $48^3{\times}64$ lattice at $\beta = 3.57$ with $am_{ud}^{\mathrm{PCAC}} \sim 0.0056$ and $am_s^{\mathrm{PCAC}} \sim 0.044$. This amounts to $a \sim 0.085\,\mathrm{fm}$ and $M_\pi \sim 190\mathrm{MeV}$. See Ref. 19 for details.

the delta (Δ) baryon seem to scale rather well, that is the term proportional to a/Λ in Eq. (3) seems to be absent and the slope [i.e. the prefactor of the $(a/\Lambda)^2$ term] seems to be small. We find it remarkable that these properties hold true out to lattice spacings around $a \sim 0.16\,\mathrm{fm}$.

3 *Ab initio* Determination of Light Hadron Masses

As mentioned in the introduction, one of the classic goals of lattice QCD is to get the spectrum of the light hadrons correct. To provide an *ab initio* calculation, such a study should (simultaneously) satisfy the following requirements:

1. Inclusion of the u, d, and s quarks in the fermion determinant with an exact algorithm and an action whose universality class is QCD.
2. A complete determination of the masses of the light flavour-nonsinglet pseudoscalar and vector mesons, as well as the baryon octet and decuplet.
3. Large box volumes to guarantee small finite-size effects on one-particle states and at least one data point at a larger volume to confirm the smallness of these effects.
4. Controlled interpolations and extrapolations to the physical values of the quark masses m_s and $m_{ud} = (m_u + m_d)/2$, respectively.
5. Controlled extrapolations to the continuum ($a \to 0$), requiring that the data are generated at no less than three values of the lattice spacing a.

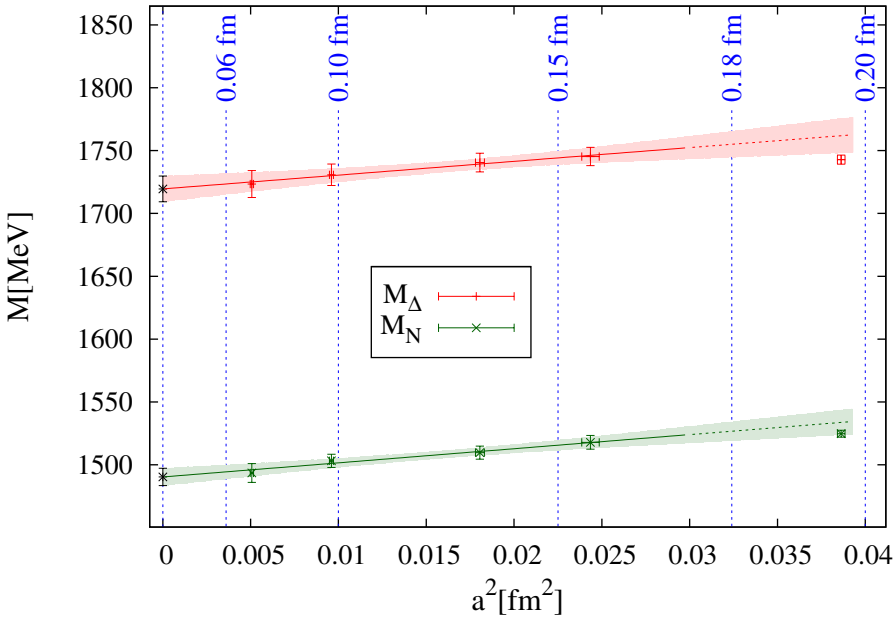


Figure 2. Scaling of the nucleon and delta for an artificial choice of quark masses, defined by $M_\pi/M_\rho = M_K/M_\rho = 0.67$, using the scale setting procedure defined around Eq. (7) of Ref. 19. The error-bars are statistical.

We stress that important progress on each individual point has been made by a number of groups, but a simultaneous fulfillment of all these requirement has been elusive prior to Ref. 12. Let us comment on those points which have not been discussed in the previous section.

Regarding point 2 we just mention that particles with a significant admixture of flavour-singlet states have been omitted, since they require a dedicated all-to-all propagator technique and massive I/O capabilities which (at the time when these simulations were performed) were not available on BG/L and BG/P.

Regarding point 3 we emphasize that the effect of the finite spatial volume has been corrected for, even though these effects were tiny. In particular for two stable states (the pion and the nucleon) the $M(L)/M$ predictions of Chiral Perturbation Theory[22,23] have been tested (by augmenting our dataset by dedicated simulations with a smaller and larger L). We found complete agreement between our data and the prediction.

Regarding point 4, two comments are in order. The first one concerns the fact that the difference $m_d^{\rm phys} - m_u^{\rm phys}$ is small compared to the typical scale in hadron physics. The effects due to this splitting are tiny compared to our error-bars, and a similar statement holds true for the effects due to the electromagnetic interactions of the quarks (which we do not attempt to simulate on the lattice). This implies that we need to correct the input data, i.e. the physical values of M_π, M_K, M_Ξ to which we want to tune our simulation, for the influence of isospin breaking and electromagnetic effects. On this point we follow

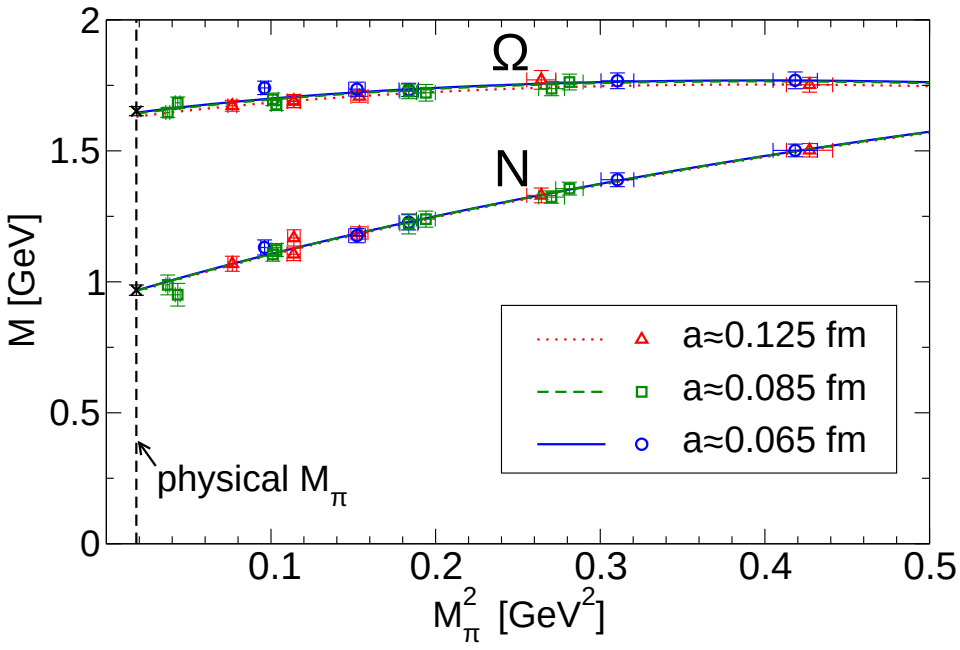


Figure 3. Synopsis of the quark mass extrapolation of our data to the physical value of m_{ud} (equivalently to the physical value of M_π). The scale is set through aM_Ξ at the physical mass point. Our data at three lattice spacings seem to lie on a universal line, confirming the good scaling properties seen in Figure 2.

Ref. 24 and chose $M_\pi = 135\mathrm{MeV}$, $M_K = 495\mathrm{MeV}$ and $M_\Xi = 1318\mathrm{MeV}$ as input values. The fact that each one of these values come with an error of a few MeV is accounted for and contributes to the final error which we quote. The second comment concerns the details of the extrapolation in the quark mass and the scale setting. As one can see in Figure 3, our data, when plotted against M_π^2, show a rather small amount of curvature. Moreover, that our pion mass reaches down to $190\mathrm{MeV}$ implies that we need to extrapolate very little to reach the target value of $135\mathrm{MeV}$. We do this via a simple Taylor fit or, alternatively, via formulae from Chiral Perturbation Theory[25]. In a similar sense the scale is set either through the value of aM_Ξ at the (extrapolated) physical mass point (the method used in Figure 2) or through the value of aM_Ξ in the actual simulation (which amounts to an analogous plot with $(M_\pi/M_\Xi)^2$ on the x-axis and M/M_Ξ on the y axis).

The last two options are, in fact, just two examples of our overall analysis strategy. Whenever there are several reasonable ways of how to set a given variable to its physical value, we do not just use one of them, but we use them all in a systematic form. Each time we perform a full analysis, including an assessment of the statistical error via a complete bootstrap procedure, leading to an overall distribution of the central values. From this distribution we draw, in the final phase of the analysis, our central value and the best estimate of the overall theoretical or systematic uncertainty. This uncertainty is then added (in quadrature) to the statistical uncertainty to give the error-bars depicted in Figure 4.

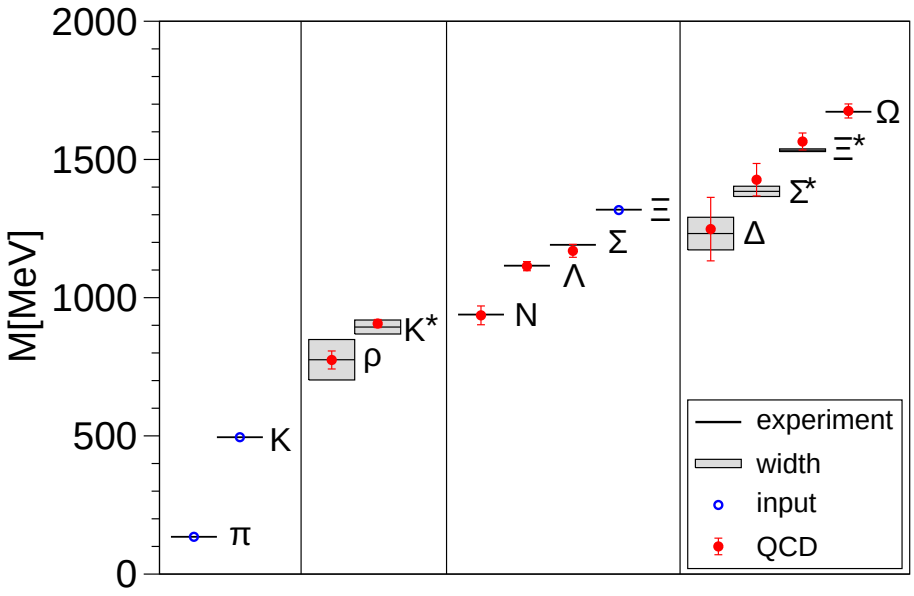


Figure 4. Our *ab initio* calculation[12] of the light hadron spectrum with three light flavours (pseudoscalar and vector meson octet, baryon octet and decuplet, excluding states with significant flavour singlet admixture), compared to experiment. In the latter case states which decay under strong interactions are shown together with the pertinent width. The error-bars on our data include statistical and systematic uncertainties. The symbols for M_π, M_K, M_Ξ have no error-bar, since these have been used as input to adjust the quark masses and to set the lattice spacing.

4 Summary and Outlook

We have performed the first *ab initio* calculation of the spectrum of the light pseudoscalar and vector mesons and octet and decuplet baryons in QCD with fully controlled systematics[12]. The outcome is in perfect agreement with experiment. As a result we can claim to understand the origin of the bulk of the visible mass in the universe; it is due to the dynamical mechanism which binds almost-massless quarks and exactly-massless gluons into protons and neutrons, thereby giving them a mass of order $1\mathrm{GeV}$ each.

With these results in hand, it is natural to extend our calculations to give the mass of the u, d and s quarks from first principles. An important step towards this goal is to compute the renormalization factor Z_S^{RI} for each of our lattice spacings. We opt for a non-perturbative determination, following the ideas of Ref. 26. We stress that to date the light quark masses m_u, m_d, m_s are among the least precisely known parameters of the Standard Model of elementary particle physics; they are essentially known[28] up to a factor 2 only. We feel confident that we can achieve a significantly more precise result.

Acknowledgements

The author thanks all members of the Budapest-Marseille-Wuppertal Collaboration for fruitful collaboration and, in particular Stephan Dürr for his help in preparing the present summary.

References

1. H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys. Lett. B **47**, 365 (1973).
2. D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).
3. H. D. Politzer, Phys. Rev. Lett. **30**, 1346 (1973).
4. K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
5. K. G. Wilson, New Phenomena In Subnuclear Physics. Part A. Proceedings of the First Half of the 1975 International School of Subnuclear Physics, Erice, Sicily, July 11 - August 1, 1975, ed. A. Zichichi, Plenum Press, New York, 1977.
6. M. Creutz, Phys. Rev. D **21**, 2308 (1980).
7. K. Symanzik, Nucl. Phys. B **226**, 187 (1983).
8. K. Symanzik, Nucl. Phys. B **226**, 205 (1983).
9. M. Lüscher and P. Weisz, Phys. Lett. B **158**, 250 (1985).
10. B. Sheikholeslami and R. Wohlert, Nucl. Phys. B **259**, 572 (1985).
11. C. Morningstar and M. J. Peardon, Phys. Rev. D **69**, 054501 (2004) [hep-lat/0311018].
12. S. Dürr *et al.* [Budapest-Marseille-Wuppertal Collab.], Science **322**, 1224 (2008), including "supplementary online material" [0906.3599, hep-lat].
13. S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, Phys. Lett. B **195**, 216 (1987).
14. M.A. Clark and A.D. Kennedy, Phys. Rev. Lett. **98** (2007) 051601 [hep-lat/0608015].
15. T. A. DeGrand and P. Rossi, Comput. Phys. Commun. **60**, 211 (1990).
16. J. C. Sexton and D. H. Weingarten, Nucl. Phys. B **380**, 665 (1992).
17. M. Hasenbusch, Phys. Lett. B **519**, 177 (2001) [hep-lat/0107019].
18. T. Takaishi and P. de Forcrand, Phys. Rev. E **73**, 036706 (2006) [hep-lat/0505020].
19. S. Dürr *et al.* [Budapest-Marseille-Wuppertal Collab.], Phys. Rev. D **79**, 014501 (2009) [0802.2706, hep-lat].
20. C. Urbach, K. Jansen, A. Shindler and U. Wenger, Comput. Phys. Commun. **174**, 87 (2006) [hep-lat/0506011].
21. S. Aoki *et al.* [PACS-CS Collaboration], 0807.1661 [hep-lat].
22. G. Colangelo, S. Dürr and C. Haefeli, Nucl. Phys. B **721**, 136 (2005) [hep-lat/0503014].
23. G. Colangelo, A. Fuhrer and C. Haefeli, Nucl. Phys. Proc. Suppl. **153**, 41 (2006) [hep-lat/0512002].
24. C. Aubin *et al.* [MILC Collaboration], Phys. Rev. D **70**, 114501 (2004) [hep-lat/0407028].
25. P. Langacker and H. Pagels, Phys. Rev. D **10**, 2904 (1974).
26. G. Martinelli, C. Pittori, C. T. Sachrajda, M. Testa and A. Vladikas, Nucl. Phys. B **445**, 81 (1995) [hep-lat/9411010].
27. K. G. Chetyrkin and A. Retey, Nucl. Phys. B **583**, 3 (2000) [hep-ph/9910332].
28. C. Amsler *et al.* [Particle Data Group], Phys. Lett. B **667**, 1 (2008).

QCD Simulations with Light, Strange and Charm Dynamical Quark Flavours

Karl Jansen

NIC, DESY, Platanenallee 6, 15738 Zeuthen, Germany
E-mail: Karl.Jansen@desy.de

for the European Twisted Mass Collaboration

We give an account on our first experiences of the NIC project simulating mass degenerate light up and down and mass-split strange and charm quarks. These simulations are the first of their kind in lattice QCD. Being able to reach rather light pion masses and small values of the lattice spacing, we enter a regime of lattice QCD simulations corresponding to an almost realistic situation. We describe our preliminary results for the light mesons, the K- and D-mesons as well as for the nucleon mass.

1 Introduction

In nature, we observe three generations of neutrinos as demonstrated by LEP. Correspondingly, there should exist 3 generations of quarks, ordered in quark pairs as (up,down), (strange, charm), (bottom,top). The masses of these quarks show a particular hierarchy with essentially mass-degenerate up and down quarks and basically exponentially increasing mass values from the strange to the top quark.

We strongly believe nowadays that the quarks are the constituent particles of all hadrons with the gluons being the interaction particles that –via the strong interaction– "glue" the quarks together to form the bound hadron states which are the ones observed in experiments. The mechanism of forming the bound states is theoretically described by quantum chromodynamics. The postulation of QCD is that at very short distances, the quarks behave as almost free particles that interact only very weakly, a phenomenon we call asymptotic freedom. At large distances, at the order of 1fm, the quarks interact extremely strong and in fact so strong that they will never be seen as final, observable states but rather form the observed hadron bound spectrum. The latter phenomenon is called confinement of quarks.

Since the interaction between quarks become so strong at large distances, analytical methods such as perturbation theory fail to analyze QCD. A method to nevertheless tackle the problem is to formulate QCD on a 4-dimensional, euclidean space-time grid. This setup first of all allows for a rigorous definition of QCD and leads to fundamental theoretical and conceptual investigations. On the other hand, the lattice approach enables theorists to perform large scale numerical simulations. In this contribution, we will describe one approach to "lattice QCD", the twisted mass formulation.

In the past, lattice physicists had to work with a number of limitations when performing numerical simulations. These simulations are extremely expensive, reaching the need for Petaflop computing and even beyond, a regime of computing power we just reach today. Therefore, for a long time the quarks were treated as infinitely heavy, indeed a crude approximation given that the up and down quarks have masses of only O(MeV). In a next

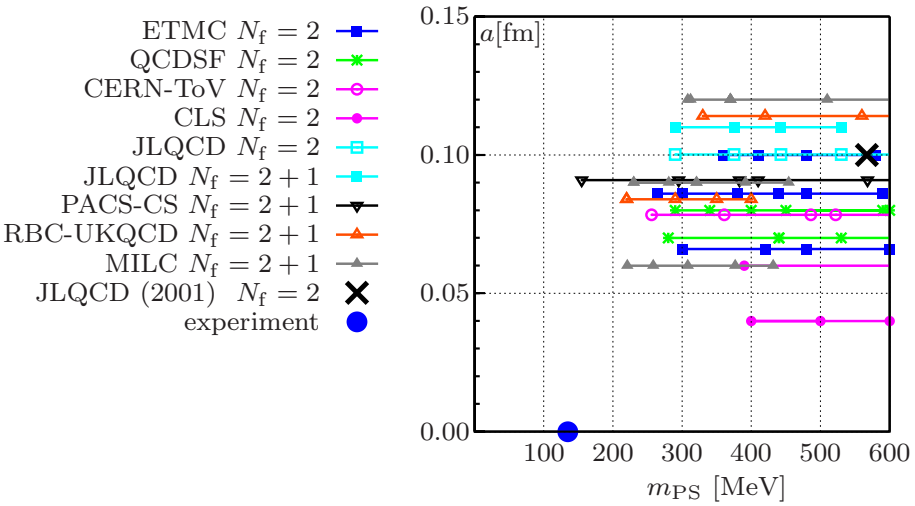


Figure 1. The values of the lattice spacing a and pseudo scalar masses $m_{\rm PS}$ as employed presently in typical QCD simulations by various collaborations as (incompletely) listed in the legend. The blue dot indicates the physical point where in the continuum the pseudo scalar meson assumes its experimentally measured value. The black cross represents a state of the art simulation by the JLQCD collaboration in 2001.

step, only the lightest quark doublet, the up and down quarks, were taken into consideration, although their mass values as used in the simulation have been unphysically large.

Nowadays, besides the up and down quarks, also the strange quark is included in the simulations. In addition, these simulations are performed in almost physical conditions, having the quarks masses close to their physical values, large lattices with about 3fm linear extent and small values of the lattice spacing such that a controlled continuum limit can be performed. The situation of the change of the simulation landscape is illustrated in Figure 1. In the figure, the blue dot indicates the physical point. The black cross represents a state of the art simulation in the year 2001. As can be seen in the graph, most of the simulations now go well beyond what could be reached in 2001 demonstrating clearly the progress in performing realistic simulations.

The drastic change in the situation is due to three main developments: i) algorithmic breakthroughs; ii) machine development; the computing power of the present BG/P systems is even outperforming Moore's law, as we see at the example of the installation at the Jülich Supercomputer Centre, iii) conceptual developments, such as the use of improved actions which reduce lattice artefacts and the development of non-perturbative renormalization.

In this article, which is to a large extent based on the proceedings contribution of S. Reker to the 2009 lattice symposium in Beijing, China, we will report our first results when also the charm degree of freedom is included in the simulations. "We" stands her for a European collaboration, the European Twisted Mass Collaboration (ETMC) which comprises 16 institutions in Europe, i.e.

- Cyprus: Univ. of Cyprus,
- France: Univ. of Paris Sud and LPSC Grenoble,

- Germany: Humboldt Univ. zu Berlin, Univ. of Münster, DESY in Hamburg and Zeuthen,
- Great Britain: Univ. of Glasgow and Univ. of Liverpool,
- Italy: Univ. of Rome I, II and III, ECT* Trento,
- Netherlands: Univ. of Groningen,
- Poland: Univ. of Poznan,
- Spain: Univ. of Valencia,
- Switzerland: Univ. of Bern.

It has to be stressed that such computations including the charm quark are the first of their kind world-wide and hence our collaboration and this works enters new territory.

In the twisted mass approach it turns out that when only one parameter of the action, namely the bare Wilson mass, is tuned to a critical value, the lattice artefacts in physical observables appear only quadratic in the lattice spacing. This is in contrast to a standard Wilson action, where these lattice spacing effects are linear. The main advantage of the twisted mass formulation of lattice QCD is then that this tuning of the bare Wilson quark mass, the so-called maximal twist situation, provides an improved, i.e. $\mathcal{O}(a^2)$, continuum limit scaling of physical observables *automatically*, i.e. without the need to supply further, operator specific improvement coefficients. The twisted mass formulation of Lattice QCD[1,2] is being studied extensively with $N_f = 2$ dynamical flavours, i.e. including only the lightest up and down quark doublet, by the European Twisted Mass collaboration[3–7]. The major goal of our present project is to include a dynamical strange and charm quark in the simulation by introducing a mass-split doublet as discussed in Ref. 8–10.

2 Lattice Setup

In the gauge sector we use the Iwasaki gauge action[11]. The fermionic action for the light doublet is given by:

$$S_l = a^4 \sum_x \left\{ \bar{\chi}_l(x) \left[D_W[U] + m_{0,l} + i\mu_l \gamma_5 \tau_3 \right] \chi_l(x) \right\}, \tag{1}$$

using the same notation as used in Ref. 10. In the heavy sector, the action becomes:

$$S_h = a^4 \sum_x \left\{ \bar{\chi}_h(x) \left[D_W[U] + m_{0,h} + i\mu_\sigma \gamma_5 \tau_1 + \mu_\delta \tau_3 \right] \chi_h(x) \right\}. \tag{2}$$

The gauge configurations are generated with a Polynomial Hybrid Monte Carlo (PHMC) updating algorithm[12–14].

2.1 Tuning Action Parameters

Tuning to maximal twist requires to set $m_{0,l}$ and $m_{0,h}$ to a sufficiently accurate value of the critical mass $m_{\rm crit} = m_{\rm crit}(\beta)$[8]. Here we set $m_{0,l} = m_{0,h} \equiv 1/(2\kappa) - 4$. As has been shown in Ref. 8, 9, this is consistent with $\mathcal{O}(a)$ improvement defined by the maximal twist condition $am_{{\rm PCAC},l} = 0$ (see also Ref. 10). The numerical precision at which the condition $m_{{\rm PCAC},l} = 0$ is fulfilled in order to avoid residual large $\mathcal{O}(a^2)$ effects when the

pion mass is decreased is, for the present range of lattice spacings, $|\epsilon/\mu_l| \lesssim 0.1$, where ϵ is the deviation of $m_{\rm PCAC,l}$ from zero[4,15]. As explained in Ref. 10, tuning to $\kappa_{\rm crit}$ was performed independently for each μ_l value. From Table 1 we observe that the estimate of $\kappa_{\rm crit}$ depends weakly on μ_l. The heavy doublet mass parameters μ_σ and μ_δ should be adjusted in order to reproduce the values of the renormalized strange (s) and charm (c) quark masses. The latter are related to μ_σ and μ_δ via[8]:

$$(m_{s,c})_{\rm R} = \frac{1}{Z_P}(\mu_\sigma \mp \frac{Z_P}{Z_S}\mu_\delta), \tag{3}$$

where the minus sign corresponds to the strange and the plus sign to the charm quark. In practice we fix the values μ_σ and μ_δ by requiring the resulting K- and D-meson masses to match experimental results.

Label	β	κ	$a\mu_l$	$a\mu_\sigma$	$a\mu_\delta$	L/a	T/a	$m_\pi L$	$\lvert\epsilon/\mu_l\rvert$
C_1	1.90	0.1632700	0.0040	0.150	0.190	20	48	3.0	0.14(14)
C_2	1.90	0.1632700	0.0040	0.150	0.190	24	48	3.5	0.07(14)
A_1	1.90	0.1632650	0.0060	0.150	0.190	24	48	4.1	0.03(3)
A_2	1.90	0.1632600	0.0080	0.150	0.190	24	48	4.8	0.02(2)
A_3D_1	1.90	0.1632550	0.0100	0.150	0.190	24	48	5.3	0.02(2)
A_4	1.90	0.1632720	0.0030	0.150	0.190	32	64	4.0	0.08(7)
A_5C_3	1.90	0.1632700	0.0040	0.150	0.190	32	64	4.5	0.04(5)
A_6	1.90	0.1632670	0.0050	0.150	0.190	32	64	5.0	0.05(2)
D_2	1.90	0.1632550	0.0100	0.150	0.197	24	48	5.3	0.35(1)
B_1	1.95	0.1612400	0.0025	0.135	0.170	32	64	3.4	0.06(6)
B_2	1.95	0.1612400	0.0035	0.135	0.170	32	64	4.0	0.02(2)
B_3	1.95	0.1612360	0.0055	0.135	0.170	32	64	5.0	0.08(1)
B_4	1.95	0.1612320	0.0075	0.135	0.170	32	64	5.8	0.05(1)
B_5	1.95	0.1612312	0.0085	0.135	0.170	24	48	4.6	0.01(2)

Table 1. Input parameters, $m_\pi L$ and $|\epsilon/\mu_l|$ for all ensembles used in this paper. Every ensemble has ~ 5000 thermalized trajectories of length $\tau = 1$. We have two main ensemble sets: A and B, at $\beta = 1.90$ and $\beta = 1.95$ respectively. Ensembles labeled C are used to check finite size effects. Ensembles labeled D are used to check/tune the strange and charm quark masses.

2.2 Determination of Heavy-Light Meson Masses

Since the twisted mass lattice Dirac operator of the non-degenerate heavy quark doublet (cf. (2)) contains a parity odd and flavour non-diagonal Wilson term, parity as well as flavour are not anymore quantum numbers of the theory. In contrast to parity and flavour conserving lattice formulations, it is not possible to compute correlation functions restricted to a single parity and flavour sector in this setup. We have developed several techniques to deal with this problem which eventually reveal completely consistent values of the D-meson mass. Hence we think that at least for the meson masses in the heavy-light sector, we have the effects of flavour breaking under control.

The left panel of Figure 2 shows the tuning of the strange quark mass by showing the difference, scaled with the chirally extrapolated value of r_0/a between twice the K-meson mass squared and the pion mass squared. Set A at $\beta = 1.90$ $a\mu_\delta = 0.190$ (green points) appears to overshoot the physical point (the black cross on the left), while set B (red points) extrapolates better. To improve the tuning of the strange quark mass for set A, we are currently applying a reweighting procedure as described in Ref. 10 in the parameters $a\mu_\delta$ and κ. The blue point with a different heavy sector splitting $a\mu_\delta = 0.197$ is a run to check this procedure. The right panel of Figure 2 shows the mass of the D-meson as a function of the pion mass squared for various simulation points as well as the experimental value from the Particle Data Group[17]. The plot demonstrates that we have tuned the charm (sea) quark mass in our simulations to a physically realistic value. As a final check, we also use an estimate of Z_P/Z_S to verify that $m_c \sim 10 m_s$.

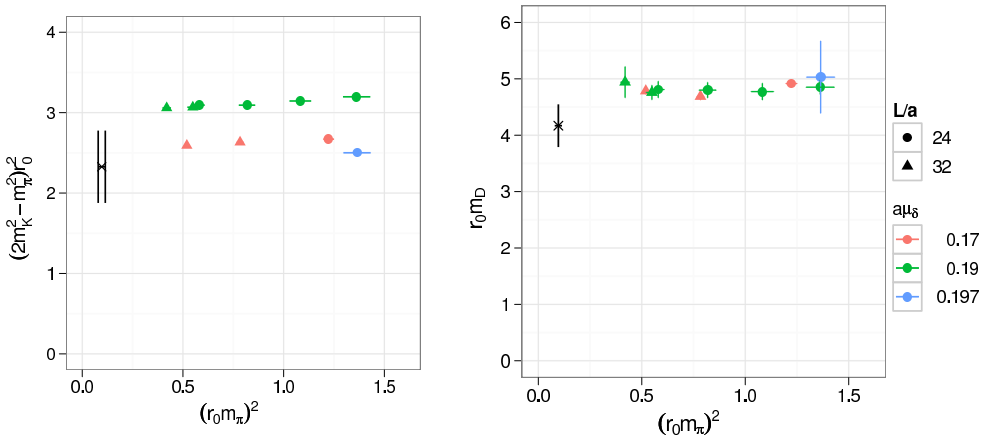


Figure 2. $r_0^2(2m_K^2 - m_\pi^2)$ and $r_0 m_D$ as functions of $(r_0 m_\pi)^2$, showing the status of the tuning of the strange and charm quark mass respectively. The experimental value from PDG is added as the black cross ($r_0 = 0.44(4)$fm was used). Red points label the $\beta = 1.95$ runs, green points label the $\beta = 1.90$ runs, where the single blue point corresponds to $\beta = 1.90$ with a different heavy sector splitting $a\mu_\delta$. Circles denote runs with $L/a = 24$, triangles indicate a volume with $L/a = 32$.

3 Results

As a first check of our data, we have compared it to the extensively analysed data set that exists for our $N_f = 2$ data. To compare the two sets, we plot dimensionless physical ratios in Figure 3. The figure shows a nice agreement between all our results, suggesting small discretisation effects and small effects of dynamical s- and c-quarks for these observables.

3.1 r_0/a Extrapolation

The static force parameter r_0 is a quantity that is often used to study the continuum limit behaviour of physical quantities. We therefore have computed r_0/a in our simulations on

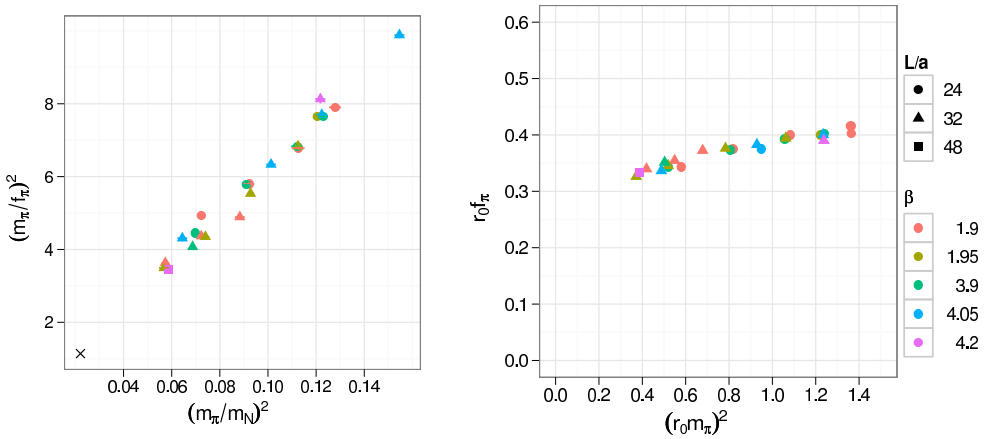


Figure 3. $(m_\pi/f_\pi)^2$ vs $(m_\pi/m_N)^2$ (left) and $(r_0 f_\pi)^2$ vs $(r_0 m_\pi)^2$ (right) for both $N_f = 2+1+1$ data (with $\beta = 1.90, 1.95$) and $N_f = 2$ data (with $\beta = 3.9, 4.05, 4.2$ and using a different gauge action). In the left plot, m_N is the nucleon mass, and the physical point is included as the black cross. In both plots finite size corrections are not applied, and in the right plot the chirally extrapolated values for r_0/a were used.

the set of configurations given in Table 1. In order to obtain r_0/a at the physical point, we extrapolate r_0/a using a polynomial fit $r_0/a = c_1 + c_2 a\mu_l + c_3 a^2\mu_l^2$, and perform a simple quadratic fit $r_0/a = c_1 + c_2 a^2\mu_l^2$ to help estimate systematic errors. c_1 is the value of r_0/a in the chiral limit. We find that based on the χ^2/d.o.f. the quadratic fit is in both cases the best fit and follows closely the polynomial fit. We therefore use the value of c_1 in the chiral limit from the polynomial fit, and take the difference with the quadratic fit as a first estimate of the systematic error in the extrapolation. Using $r_0/a = 0.44(4)$ fm, this yields as lattice spacings $0.084(7)$ fm and $0.077(7)$ fm for set A at $\beta = 1.90$ and set B at $\beta = 1.95$ respectively.

β	c_1(pol)	χ^2/dof(pol)	c_1(qua)	χ^2/dof(qua)
1.90	5.26(13)	1.8	5.24(2)	1.4
1.95	5.69(21)	3.1	5.71(4)	2.1

Table 2. Comparison of two different r_0/a extrapolation functions, for both data-sets the quadratic method is better from a $\chi^2/d.o.f.$ point of view. Statistical errors in the fits are included between brackets.

3.2 Light Meson Chiral Perturbation Theory Fits

In order to extract the lattice spacing and light quark mass from our data-sets, we perform a next to leading order $SU(2)$ chiral perturbation theory fit of the m_π and f_π data. We use continuum formulae and correct for finite size effects either without any new low energy constants à la Gasser and Leutwyler[18], or with $\bar{l}_1$ and $\bar{l}_2$ added in, as described in Ref. 19. The results are listed in Table 3. We have performed these fits for ensemble sets A and B separately, and also combined them in a single fit. In Table 3, we include a systematic error,

estimated at $2-5\%$, coming from the dispersion of the values of the fitted parameters between NLO and NNLO. Note that since the quark mass enters the χPT expression, in order to combine the two sets at different lattice spacings, we need to know the renormalization factor of the quark mass $Z_\mu = 1/Z_P$, a computation which is not yet complete. Assuming that Z_P is effectively a function of β in the range of parameters we are considering, we can fit the ratio of those Z_P-values and lattice spacings and extract lattice spacings from the combined fit. In every fit we use as inputs the physical f_π and m_π, and extract f_0, $\bar{l}_3$ and $\bar{l}_4$. A complete analysis (analogous to Ref. 20) of the systematic effects is in progress.

set	pts	f_0(MeV)	l_3	l_4	$a_{\beta=1.90}$(fm)	$a_{\beta=1.95}$(fm)
A & B	11	121(4)	3.5(2)	4.7(2)	0.086(6)	0.078(6)
A	6	121(4)	3.4(2)	4.8(2)	0.086(7)	
B	5	121(4)	3.7(2)	4.7(2)		0.078(7)

Table 3. Results from the NLO $SU(2)$ χPT fits for combined, only set A and only set B respectively. Errors are dominated by a systematic error of $2-5\%$ due to performing an NLO fit. The column "pts" refers to the number of ensembles used in that fit.

3.3 Chiral Extrapolation of the Nucleon Mass

In this section, we present preliminary results for the light quark mass dependence of the nucleon mass. We consider the one-loop result from heavy baryon chiral perturbation theory (HBχPT)

$$m_N = m_N^0 - 4c_1 m_\pi^2 - \frac{3g_A^2}{16\pi f_\pi^2} m_\pi^3 \tag{4}$$

and fix the scale and light quark mass to the point where the ratio m_N/m_π attains its physical value. We fix f_π and g_A to their physical values (130.7 MeV and 1.27 respectively) as has also been performed in Ref. 7. Using this procedure, we find a lattice spacing of $0.089(2)$ fm and $0.077(3)$ fm for $\beta = 1.90$ and 1.95 respectively. The $\chi^2/(d.o.f.)$ of these fits is not very good, and fitting a linear extrapolation appears to be consistent with the data. This is not unique to our data, and has been observed by various collaborations. We therefore perform the linear fit here as well, and absorb the difference between the two extrapolations in the systematic error. A more detailed analysis of the chiral extrapolation of the nucleon mass will be presented in an upcoming study.

4 Conclusions

We have presented first results from runs performed with mass degenerate light up and down as well as mass-split strange and charm quarks (so called $N_f = 2+1+1$ simulations). We find a nice agreement between these results and those with $N_f = 2$ twisted mass fermions as shown through dimensionless ratio plots (of m_π, f_π, m_N and r_0/a), suggesting small discretisation effects and small effects of dynamical s- and c-quarks for these

observables. Our results show first of all that simulations with $N_f = 2 + 1 + 1$ flavours are perfectly feasible. Second, it is possible to tune the action parameters such that maximal twist and hence automatic $\mathcal{O}(a)$ improvement is realized and that the physical K- and D-meson masses are matched. Clearly, these first results need careful checks of possible systematic effects. To this end, we are presently performing simulations at a lattice spacing of $a \approx 0.06$fm and pion masses down to $m_\pi \approx 160$MeV which will bring us very close to the physical point and allows to test the chiral behaviour of physical quantities.

Acknowledgements

We thank all other members of the ETM Collaboration for a very fruitful collaboration and for many valuable discussions. The HPC resources for this project have been made available by the computer centres of Barcelona, Groningen, Jülich, Lyon, Munich, Paris and Rome (apeNEXT), which we thank for enabling us to perform this work. This work has also been supported in part by the DFG Sonderforschungsbereich/Transregio SFB/TR9-03, and by GENCI (IDRIS - CINES), Grant 2009-052271.

References

1. **ALPHA** Collaboration, R. Frezzotti, P. A. Grassi, S. Sint and P. Weisz, JHEP **0108** (2001) 058.
2. R. Frezzotti and G. C. Rossi, JHEP **0410** (2004) 070.
3. **ETM** Collaboration, Ph. Boucaud *et al.*, Phys. Lett. B **650** (2007) 304.
4. **ETM** Collaboration, Ph. Boucaud *et al.*, Comput. Phys. Commun. **179** (2008) 695.
5. **ETM** Collaboration, C. Urbach, PoS **LAT2007** (2007) 022.
6. **ETM** Collaboration, B. Blossier *et al.*, JHEP **0907** (2009) 043.
7. **ETM** Collaboration, C. Alexandrou *et al.*, Phys. Rev. D **78** (2008) 014509.
8. R. Frezzotti and G. C. Rossi, Nucl. Phys. Proc. Suppl. **128** (2004) 193.
9. T. Chiarappa *et al.*, Eur. Phys. J. C **50** (2007) 373.
10. **ETM** Collaboration, R. Baron *et al.*, PoS **LAT2008** (2008) 094.
11. Y. Iwasaki, Nucl. Phys. B **258** (1985) 141.
12. R. Frezzotti and K. Jansen, Phys. Lett. B **402** (1997) 328.
13. T. Chiarappa, R. Frezzotti and C. Urbach, PoS **LAT2005** (2006) 103.
14. C. Urbach, K. Jansen, A. Shindler and U. Wenger, Comput. Phys. Commun. **174** (2006) 87.
15. **ETM** Collaboration, P. Dimopoulos, R. Frezzotti, G. Herdoiza, C. Urbach and U. Wenger, PoS **LAT2007** (2007) 102.
16. **ETM** Collaboration, K. Jansen, C. Michael, A. Shindler and M. Wagner, JHEP **0812** (2008) 058.
17. C. Amsler *et al.* [Particle Data Group], Phys. Lett. B **667** (2008) 1.
18. J. Gasser and H. Leutwyler, Nucl. Phys. B **307** (1988) 763.
19. G. Colangelo, S. Durr and C. Haefeli, Nucl. Phys. B **721** (2005) 136.
20. **ETM** Collaboration, P. Dimopoulos *et al.* PoS **LAT2009** (2009) 117.

Tuning Lattice QCD to Petascale on Blue Gene/P

Stefan Krieg[1] **and Thomas Lippert**[2]

[1] Center for Theoretical Physics, MIT, Cambridge, MA 02139, USA
E-mail: krieg@mit.edu

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

We present a highly scalable implementation of the Hybrid Monte Carlo algorithm on the Jülich IBM Blue Gene/P Petaflop supercomputer JUGENE. The program simulates Lattice Quantum Chromodynamics (LQCD) including the effects of dynamical Wilson fermions. Its core routines benefit from GCC inline assembly. All other computationally intensive parts have been optimized by means of a set of macros that make use of "intrinsics" of the IBM XLC compiler. A specific low-level communication layer has been designed for optimized communication between nodes. Direct node-to-node access through the Blue Gene/P SPIs leads to a reduction of latencies. This short report gives details how petascaling of LQCD on a leadership-class supercomputer with $O(10^5)$ cores can be achieved.

1 Introduction

Quantum Chromodynamics (QCD)[1] describes the "strong force" which binds protons and neutrons forming the atomic nucleus. Much theoretical evidence validating the theory is given in the high energy sector. Here, owing to the large momenta involved, a comparison of the theory to results from accelerator experiments is feasible by perturbative methods. In the low energy sector, however, the favorite method is based on a regularized approximation of the continuum theory on a Euclidean lattice, a method named Lattice QCD (LQCD). Many important quantities such as the hadron spectrum are only accessible through such calculations[2]. Markov chains of the gauge fields are computed by means of stochastic importance sampling. Lattice simulations need huge amounts of computational resources that require the most advanced supercomputers. To make best use of these systems the implementation needs to be highly optimized.

This paper describes the implementation of a state-of-the-art simulation software for LQCD simulations. It is based on the classical Hybrid Monte Carlo algorithm (HMC)[3] which is improved by mass preconditioning[4], multiple time step[5] integration with optimized integrators[6], and mixed precision inverters[7]. Here we focus on the secrets of the implementation of the code on the IBM Blue Gene/P (BGP) supercomputer.

In the next section, BGP's hardware elements relevant for the implementation are introduced. In section 3, the fast communication layer is presented, in section 4, the "kernel", the even/odd preconditioned[8] Wilson[9] matrix-vector multiplication is given. Section 5 is devoted to a set of macros to optimize other critical parts of the code. Finally, performance results are presented in section 6.

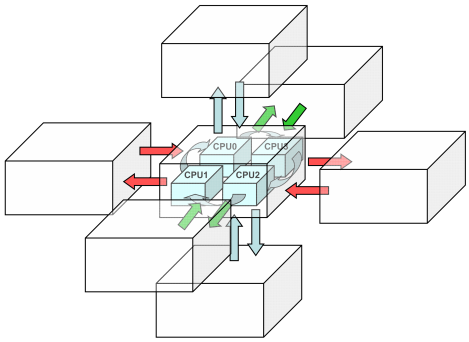


Figure 1. Parallelization scheme: Three dimensions are mapped to the torus HW, here depicted by arrows pointing from and to the neighbouring nodes (boxes). The 4^{th} dimension uses the four cores of a BGP node.

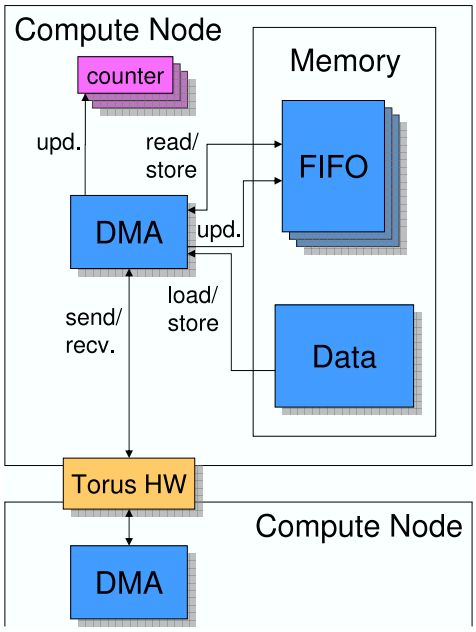


Figure 2. The DMA controller sends data independent from the CPU. In a specific mode the DMA fetches data from local memory and communicates via the torus to the DMA of the receiving node.

2 Blue Gene/P Hardware Features

As its predecessor, the Blue Gene/L (BGL), the IBM Blue Gene/P is well suited for LQCD simulations. A major advantage is the 3-dimensional torus network with very low latency (<1 μs) and high throughput (5.1 GB/s). Three freely selectable dimensions of the 4-dimensional space-time lattice can be directly mapped onto the torus hardware (HW), the 4^{th} dimension can be parallelized over the four cores of each node[a] (see Figure 1).

2.1 Communication Hardware

One important difference between BGP and BGL concerns the method of utilization of the torus network: The BGP features a DMA communication controller whereas on the BGL the CPU is feeding torus FIFOs. This new DMA communication controller can be used to move data over the torus network while the CPU is busy with computations (*cf.* Figure 2). The CPU is just required to set up and/or start the communication process. It is not involved in moving the data.

[a]This implementation is based on the VN-mode of BGP, where only a small window of shared memory is used for global operations (see section 3). This allows for maximal memory bandwidth as memory operations will not require cache coherency between the cores (see section 2.2).

In order to set up the communication process, a so called "DMA memory injection FIFO" is required (*cf.* Figure 3). The injection FIFO basically is a piece of ordinary memory, where "message descriptors" can be stored. Each of these descriptors specifies one message, *i.e.* one continuous chunk of data to be transferred. It contains a reference to an injection and a reception counter (*cf.* Figure 4). The injection counter counts down the number of bytes to be sent, the reception counter, located on the receiving node, will count down the bytes that have arrived. These counters also have a memory address associated with them. This address serves as an offset – the relative send and receive addresses of a descriptor, are added to this offset when data is fetched on the send side and stored on the receive side. To send a data packet via "direct put"[b] from node A to node B, the procedure might look as follows:

1. (send side) Allocate an injection FIFO (if necessary);
2. (send side) allocate an injection counter (if necessary);
3. (rec. side) allocate a reception counter (if necessary);
4. (send side) set the injection counter offset to the (virtual) start address of the data packet;
5. (send side) set the injection counter value to the number of bytes contained in the packet;
6. (rec. side) set the reception counter offset to the (virtual) address where the packet should be stored;
7. (rec. side) set the reception counter value to the number of bytes contained in the packet;
8. (send side) prepare the descriptor. Set the relative send and receive offset to 0, because the offset values are already the memory addresses from where the data is to be fetched and where it is to be stored. Set the size value to the number of bytes contained in the packet and give the coordinates of the target node. Finally, give the send counter and the reception counter IDs;
9. (send side) inject the descriptor into the FIFO. The DMA starts working immediately;
10. (rec. side) poll the reception counter until it has reached zero.

When the descriptor is injected into the FIFO at step 9, it will be stored at the so-called FIFO tail and the tail will be incremented. In addition to the tail, there is a FIFO head, the location of the next descriptor to be serviced by the DMA, and a FIFO starting and ending address (*cf.* Figure 3). The DMA will stop servicing the FIFO when the FIFO head has reached the FIFO tail. Both FIFO head and tail wrap around to the FIFO starting address when they reach the FIFO ending address. Thus the FIFO size is only an issue, when there is not enough time for the DMA to service the descriptors before new descriptors are injected.

The FIFO head can be controlled and the FIFO can be activated and deactivated by the user via the SPI[10]. Persistent[c] send and receive operations can be set up by

[b] There are more ways to transfer data. For LQCD this is the fastest because of minimized latencies.
[c] By persistent send/receive those communication operations are denoted that communicate the same amount of data between the same nodes and from/to the same addresses in memory.

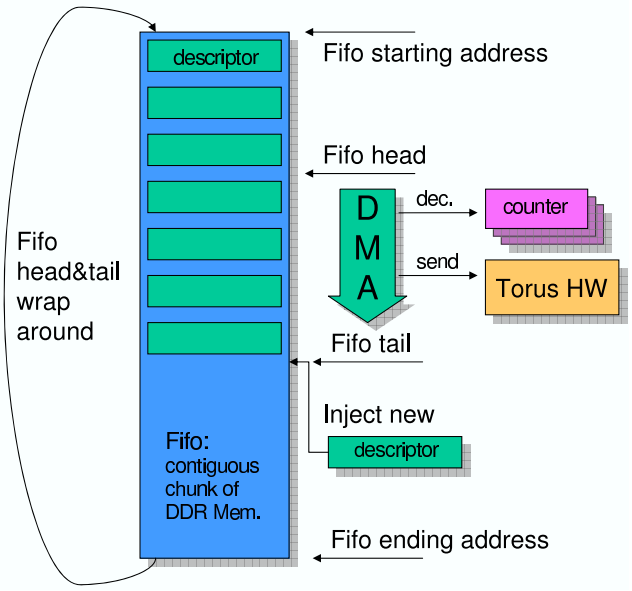


Figure 3. The BGP DMA FIFO contains message descriptors (a message equals a continuous piece of data that is to be transferred) that are read and then executed by the DMA. After executing a descriptor, the DMA moves the FIFO head towards the FIFO tail. When the tail is reached, the DMA stops servicing the FIFO.

Figure 4. "Direct Put" descriptor. The descriptor contains the XYZ coordinates of the target node, "hint bits" telling in which direction the packet will travel, send and receive offsets, message length and the reception and injection counter IDs/Group IDs.

1. Deactivating a FIFO;
2. injecting descriptors into the deactivated FIFO;
3. then moving the FIFO head to the tail manually;
4. activating the FIFO afterwards.

The communication process can then be started by setting up the reception and injection counters and by moving the FIFO head back to its original position.

It is also possible to use the DMA to perform communications that are local on a node. Actually, these are simple memory copies, but only when the DMA is used, they may be hidden behind the computations. Rearranging the data within these copies, the memory access of the computations that access the transferred data later on can be optimized.

In addition to the 3-dimensional torus network, the Blue Gene architecture features a global tree network, which is used for operations such as a global allreduce or a global barrier. It has a low latency of 5.0 μs (worst case) and a bandwidth of 1.7 GB/s. The functionality of the tree network is fully accessible by the SPI. It is actually possible to do non-blocking communication here as well. Global operations such as a global unsigned integer sum or unsigned integer max are supported by the hardware, global operations with floating point numbers have to be implemented through software.

2.2 The ppc450 Core

In addition to the new DMA communication controller, the number of cores per node has been doubled upgrading from BGL to BGP. The core itself has also been updated: while the cache coherency between the two ppc440 cores of the BGL had to be managed by software, the cache coherency of four ppc450 cores in the BGP is managed by hardware. To be able to sustain all four cores, for each core the amount of cache and the memory and cache bandwidth per cycle is the same as on BGL. However, since the hardware cache coherency requires the caching policy to be fixed to "Write-Through", the typical LQCD code performance on one core decreased a bit even though the clock frequency was increased from 700 MHz to 850 MHz.

Cache coherency related synchronization between the cores reduces the effective memory bandwidth further. In VN mode[d] however, because each node will be working in its own space, the synchronization overhead is reduced as bandwidth measurements show. Therefore, our implementation uses the VN mode. Communication between the cores is possible via a shared memory window or DMA via driven memory copies.

As it was the case for BGL, each core has a double floating point unit (FPU) where the secondary FPU is only accessible via special so-called "oedipus" instructions. *I.e.* , in order to achieve beyond 50% of theoretical peak performance, one has to exploit this special instruction set. These instructions are, however, of SIMD-type which makes automated code optimization in most cases impossible for the compiler.

There are two ways to solve this problem: The first and most time consuming is certainly to write the most computationally intensive parts of the code in raw assembly. Secondly, one can write optimized code by using the special compiler macros of the IBM XL C compiler ("intrinsics"). Each of these macros corresponds directly to a specific assembly instruction and often has the same name. So even in this case the user has to select an assembly instruction for each logical operation, however the compiler will take over scheduling and might even choose to use other instructions, if alternatives benefit performance.

The "oedipus" instruction set is used for both BGL and BGP, thus optimized routines written for BGL can also be run on the BGP. Since on BGL most codes used the "Write Back" caching policy to improve the performance, the BGL code will still have to be adapted for the BGP.

3 Optimized Communication Layer

A QCD HMC code requires a variety of communication operations. Most communication takes place between neighbouring nodes and is implemented using the torus network or DMA-driven local memory copies. Moreover, for the linear system solvers global sums are required that most effectively make use of the low latency tree network of the BGP.

[d] There are three modes of operation for the BGP: in SMP mode, all cores share the same memory and one node runs one MPI task which can have 4 threads (on BGL one core can only run a singe thread). In DUAL mode, the available memory will be split between two times two cores. In this case, one hardware node runs two MPI tasks which can have 2 threads each. Finally, in VN mode, each core has its own quarter of the memory and the hardware node runs four MPI tasks.

Additionally, global barriers will employ the tree network, whereas local, *i.e.* node-wide barriers use the so-called "lockbox" to synchronize.

The computationally costly part of the HMC-code is the Wilson (sparse) matrix vector multiplication. Fortunately, it is possible to implement the routine in such a way that "persistent communication" can be used: all pointers involved as well as the amount of data transferred is the same for each call to the routine. Since all directions should be parallelized, a total of 8 FIFOs (one for each direction), 2 reception and 2 injection counters (for positive and for negative parity) are required. Local communication is implemented in form of a DMA-driven memory copy.

Additional generic communication routines are required. For this purpose 8 additional FIFOs are set aside, again one for each direction. The number of injection (and also reception) counters used in this case is typically 8 but is not fixed and can be chosen at runtime. With 8 pairs of reception and injection counters 8 independent communication operations can proceed at any time (so that the counter value gives the progress of the communication and because the counters have the offsets for memory access associated). Any further communication operation waits until a set of counters becomes available. The descriptors will be injected into the appropriate FIFO depending on the direction in which the data are sent. Again, communication local to a node is implemented as DMA memory copy. For global operations like a global allreduce a shared memory window is used.

The global allreduce (tree) network is capable of performing unsigned integer operations only. As the code requires a global double precision floating point sum and a global double precision floating point maximum operation, it is necessary to split the floating point numbers into mantissa and exponent and to perform separate operations on these numbers. If the sign of the original floating point number was negative, the mantissa is replaced by the complement of two. To calculate the global sum, for example, the first step is to compute the global maximum of the exponent. Since this is an unsigned integer already, the maximum can be computed right away. Afterwards all mantissa have to be shifted according to the global maximum exponent. Typically for a double precision floating point number, the mantissa will be converted to a 128 Bit unsigned integer, to make sure the precision of the global unsigned sum on the shifted mantissa is precise enough. Finally mantissa and exponent are reassembled into a floating point number.

4 The Wilson Fermion Matrix

Most of the cycles are spent within a small fraction of the program. This "kernel" implements the fermion "Dslash" matrix vector multiplication, approximating the fermionic determinant. The details of the kernel follow the specific discretization scheme used.

Our HMC is designed for a simulation based on Wilson's discretization[9]. Here, the Wilson kernel is defined as the non-diagonal part of the matrix-vector multiplication:

$$\Psi_n = \sum_{m=1}^{V} D_{W\,n,m}\Phi_m = \sum_{\mu=1}^{4}(1-\gamma_\mu)\otimes U_\mu(n)\Phi_{n+\hat{\mu}} + (1+\gamma_\mu)\otimes U_\mu^\dagger(n-\hat{\mu})\Phi_{n-\hat{\mu}}.$$

Here n is the index of a lattice site of the (4-dimensional) lattice, V is the number of lattice sites and $n+\hat{\mu}$ is the neighbouring lattice site of n in positive μ direction. γ_μ are complex 4×4 and U_μ are SU(3) (3×3) matrices in a tensor product. $U^\dagger(n)$ is the hermitian conjugate of $U(n)$. The diagonal term is ignored as it is implemented outside the kernel.

The Wilson matrix connects nearest neighbouring sites only. As a consequence all communication is restricted to nearest neighbouring nodes when arranged in a 4-dimensional space-time lattice, which is the parallelization scheme for the entire code (*cf.* Figure 1).

4.1 Memory Footprint and Number of Flops

For every lattice point one has to store four 3×3 matrices $U_\mu(n)$, one input "spinor" Φ_n and one output spinor Ψ_n. Thus, the memory required for the kernel scales linearly with V. The same is true for the number of floating point operations, *i.e.* , the kernel realizes a sparse matrix-vector multiplication.

γ_μ is a 4×4 matrix. It is tensor-multiplied with an SU(3) matrix U. Φ_m and Ψ_n contain four SU(3) vectors each. One of these SU(3) vectors consists of 3 complex numbers. γ_μ itself is fixed for all lattice points, and for the γ representation used here it only contains the (signed) complex or real unit. Therefore, it is computed rather than stored in memory. A 3×3 matrix requires 9 complex numbers leading to a total memory usage of $4 \times 9 + 2 \times 12 = 60$ complex numbers per lattice site. The total memory-footprint includes additional communication buffers beyond the $60 \times V$ complex $= 120 \times V$ 4 or 8 Byte floating point numbers.

One multiplication with a γ matrix requires 12 floating point operations. One SU(3) multiplication amounts to 66 flops, the result accumulation to 24 flops. Due to the structure of the γ matrices, only two independent SU(3) multiplications are required for each of the products $(1-\gamma) \otimes U\Phi_m$ giving a total flop count of $8 \times 2 \times 66 + 8 \times 12 + 7 \times 24 = 1320$ flops per lattice site. All kernel performance numbers given in this paper are measured relative to 1320 flops/site.

4.2 Communication Setup

As mentioned above, it is desirable to use persistent communication in order to reduce latencies. In the case of the Wilson kernel this can be straightforwardly carried out, because the vectors $\lambda_{\mu,n} = (1-\gamma_\mu)\Phi_{n+\hat{\mu}}$ and $\chi_{\mu,n} = (1+\gamma_\mu) \otimes U^\dagger_\mu(n-\hat{\mu})\Phi_{n-\hat{\mu}}$ can be stored in a temporary buffer which is the same for each call to the kernel. Since this buffer has a fixed address and all relative addresses within the buffer are identical as well, the same set of descriptors can be reused in every call to the kernel. This holds as well for even/odd preconditioning, if the local lattice volume is restricted to an even number.

The complete setup therefore requires two reception and two rejection counters, one for communication in positive (forward), one for communication in negative (backward) direction. Additionally one FIFO for each direction is required, that is 8 FIFOs for the double precision and 8 FIFOS for the single precision kernel. The descriptors are injected into each FIFO at the beginning of the simulation. Scattering of the buffers can then be started by simply setting the counter values to the number of bytes that are to be communicated (different for double and single precision) and by having each FIFO head pointing to the first descriptor.

4.3 Implementation Details

Implementing the Wilson kernel, one can make full use of the BGP's DMA capabilities to overlap communication and computation. Our implementation strategy:

1. (scalar) Compute $\lambda_{\mu,n} = (1 - \gamma_\mu)\Phi_{n+\hat{\mu}}$;
2. (comm) start communicating "backward" the $\lambda_{\mu,n}$ on the boundaries to the nodes in negative directions;
3. (scalar) compute $\chi_{\mu,n} = (1 + \gamma_\mu) \otimes U^\dagger_\mu(n - \hat{\mu})\Phi_{n-\hat{\mu}}$;
4. (comm) start communicating "forward" the $\chi_{\mu,n}$ on the boundaries to the nodes in positive directions;
5. (comm) wait "backward";
6. (scalar) multiply the $\lambda_{\mu,n}$ with $U_\mu(n)$, accumulate the results;
7. (comm) wait "forward";
8. (scalar) accumulate the $\chi_{\mu,n}$ to the result of the previous step.

For typical lattice sizes the waits in step 5 and 7 return directly, indicating that all communication is already finished when the wait is called. As a matter of fact, there is no bandwidth problem caused by the DMA and the cores simultaneously accessing memory: when the steps 1 and 4 are omitted, the time elapsed after a call to the kernel is practically unchanged. This means that the overhead due to the communication start and wait routines is negligible.

4.4 Single Precision Arithmetics

The Blue Gene/P FPU is capable of double precision arithmetics but cannot perform single precision floating point calculations. Still it is possible to load single precision numbers, which are converted "on the fly" to double precision and stored in a double precision register. Results stored in these registers can be rounded to single precision and then stored as single precision numbers. Because all arithmetic operations are done in double precision, the theoretical peak performance does not increase.

However, twice as many single precision floating point numbers can be loaded from memory (or can be communicated) per cycle compared to double precision computations. Since the kernel is bandwidth bound, this will increase the kernel performance, because less time is spent waiting for data to arrive in registers. Still, latencies may actually increase due to the required extra rounding step. Therefore, the performance of the single precision kernel is significantly larger, but is constrained to be less than twice the performance of the double precision variant.

In any case, the HMC requires double precision accuracy. The linear solvers, which call the kernel, have to calculate solution vectors which are more precise than the full single precision accuracy. This problem is solved by using "mixed precision" solvers[7]. The idea is, that the solver mainly calls the single precision kernel and uses the double precision variant only to correct for potential differences due to rounding.

5 Using Macros with Intrinsics

Optimizing large fractions of the code by rewriting it in GCC inline assembly is not worth the effort on BGP, because similar performance numbers can be achieved by using the compiler macros ("intrinsics") of the IBM XL C compiler. As mentioned in section 2.2,

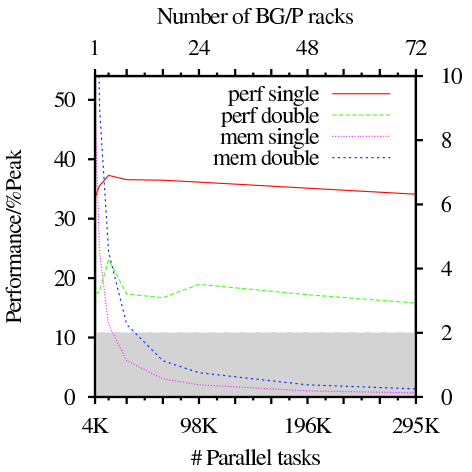


Figure 5. Performance of the Wilson kernel in a *strong* scaling analysis ($64^3 \times 144$ lattice). Single precision: (red) solid line; memory footprint per core: (purple) pointed line; double precision: (green) dashed; memory footprint: (blue) dashed curve. The shaded area is the L3-cache size per core.

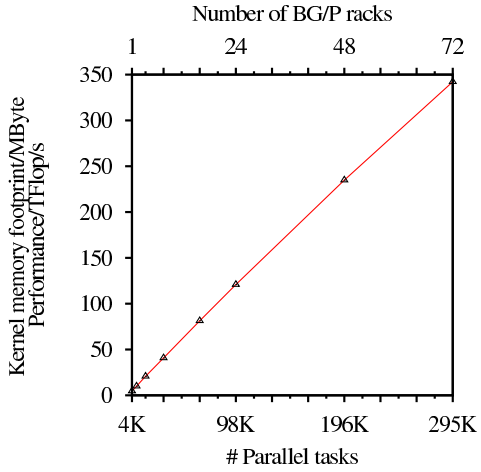


Figure 6. Shown is the same strong scaling analysis of a global $64^3 \times 144$ lattice, but this time the total performance of the single precision kernel is given. Over 335 Teraflop/s are achieved when running on 295k cores.

the intrinsics still require the user to select the exact floating point instruction to be used, but the compiler will take over scheduling the instructions.

One specific approach is to use the intrinsics inside preprocessor macros in order to implement the basic mathematical operations of the routines that are to be tuned. By using such macros tuning becomes more efficient, because the programmer can think in terms of mathematical operations that are to be optimized. Programming becomes less error-prone, since the building blocks of the tuned code can be tested and debugged separately.

Applying macros it is also possible to remove superfluous load/store operations. This is the main reason for not using inline functions. If inline functions were used, each function would have to be stand-alone, *i.e.* load input data from memory and, if the result exists, store it to memory. It would be left to the compiler to figure out which load/store operations could be avoided. When using the macros, the data is left in "virtual registers" that are used by the intrinsics. Loading and storing data then proceeds in a separate macro call.

6 Performance Figures

Most of our tuning efforts were put into the single precision kernel, because, with mixed precision inverters, it is the single precision kernel that dominates. In Figure 5 the performance of both the single and double precision kernels is shown under a strong scaling analysis based on a $64^3 \times 144$ lattice. The curve for the single precision kernel is flat, indicating that strong scaling is close to being perfect. The total performance of the single precision kernel is given in Figure 6 reaching a performance of 37.5% of machine peak.

The performance of those routines of the remaining code, optimized using macros as described in section 5, can exceed the performance of the kernel. A simple SU(3) matrix multiply add for instance reaches 50% of machine peak. Since these routines scale rather

well, the scaling of the full HMC is excellent over a large number of CPUs[e].

7 Conclusion

A highly optimized implementation of the performance critical parts of the Hybrid Monte Carlo Algorithm as used by the Budapest-Marseille-Wuppertal Collaboration on the IBM Blue Gene/P has been described. In-depth details of the optimization procedure of the Wilson kernel and other parts of the simulation code were given. The kernel achieves up to 37.5% of machine peak.

Acknowledgements

We would like to thank P. Vranas for collaboration in the early stages of the project. We grateful acknowledge the help of our friends from the BMW-collaboration headed by Zoltán Fodor and of the BGP team at Jülich Supercomputing Centre. We thank D. Chen, P. Heidelberger, B. Walkup and M. Giampapa for assistance in understanding the Blue Gene/P hardware. This work was supported by DFG grants Li 701/5-1,2 and Fo 502/1-2.

References

1. H. Fritzsch, M. Gell-Mann and H. Leutwyler, "Advantages Of The Color Octet Gluon Picture," Phys. Lett. B **47**, 365 (1973).
2. S. Durr *et al.*, "Ab-Initio Determination of Light Hadron Masses," Science **322**, 1224 (2008). [arXiv:0906.3599 [hep-lat]].
3. S. Duane, A. Kennedy, B. Pendelton and D. Roweth, Phys. Lett. B **195**, 216 (1987).
4. M. Hasenbusch, "Speeding up the Hybrid-Monte-Carlo algorithm for dynamical fermions," Phys. Lett. B **519**, 177 (2001) [hep-lat/0107019].
5. J. C. Sexton and D. H. Weingarten, "Hamiltonian evolution for the hybrid Monte Carlo algorithm," Nucl. Phys. B **380**, 665 (1992).
6. T. Takaishi and P. de Forcrand, "Testing and tuning new symplectic integrators for hybrid Monte Carlo algorithm in lattice QCD," Phys. Rev. E **73**, 036706 (2006) [hep-lat/0505020].
7. S. Durr *et al.* [Budapest-Marseille-Wuppertal Coll.], Phys. Rev. D **79**, 014501 (2009) [arXiv:0802.2706].
8. T. A. DeGrand and P. Rossi, Comput. Phys. Commun. **60**, 211 (1990).
9. K.G. Wilson Phys. Rev. D **10**, 2445 (1974); K.G. Wilson In: A. Zichichi, Editor, New phenomena in subnuclear physics, Plenum Press, N.Y. (1977) (Erice, 75).
10. The SPI is a software interface available on Blue Gene /P. Documentation is available through the files.
11. P. Vranas, J. Phys. Conf. Ser. **78**, 012080 (2007); P. Vranas et al., SC 2006: 50; G. Bhanot *et al.*, Nucl. Phys. Proc. Suppl. **140**, 823-825 (2005).
12. S. Krieg, Advances in Parallel Computing, Vol. 15, Feb. 2008.
13. S. Aoki *et al.* [PACS-CS Collab.], Phys. Rev. D **79** 034503 (2009) [arXiv:0807.1661].

[e]For comparison, a version of the single precision Wilson kernel based on the XL C intrinsics and MPI communications can reach 20% of machine peak only.

Materials Science

R. O. Jones

Institut für Festkörperforschung, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: r.jones@fz-juelich.de

The changes in scientific high performance computing (HPC) over the past 30 years have astonished all who were active during that period. The first supercomputer of NIC's predecessor HLRZ was Europe's first Cray Y-MP (1987), and I remember the excitement when this eight-processor machine performed a quantum chromodynamics (QCD) calculation with 1 gigaflop! The performance of JuGene, the newest supercomputer of NIC, is approximately five orders of magnitude greater. This is close to the Moore's Law prediction of doubling every 18 months, although the massively-parallel aspect is more relevant than the number of "transistors" on a chip. I shall reflect here on the changes brought to materials science and other areas of condensed matter physics by the dramatic increase in HPC power over the last 20 years.

All contributions to the NIC Symposium 2008 in the area of materials science were based on the density functional (DF) formalism, which was already well established in materials research in 1987. This approach allows us to calculate the total energy of an arrangement of atoms in magnetic or electric fields, so that we can study–for example–stable structures and cohesive properties without using input from experiment. The combination of DF methods with molecular dynamics (MD) [the "Car-Parrinello method"] was also available from 1986 and is now used by many groups worldwide. It is but one example of many improvements in numerical algorithms in materials science and related areas, including statistical physics. The first published CP application was to amorphous silicon. In such disordered systems, the structure and the motion of individual atoms can be calculated with a precision that is seldom attainable in measurements, and computer simulations have obvious advantages for studying materials that are radioactive or poisonous.

The theoretical groundwork necessary to use HPC for DF calculations was then in place when our first supercomputer was installed, and research in materials science has benefited greatly. Without requiring experimental input, we can study systems with hundreds of atoms without real restrictions on their atomic numbers. Furthermore, we can perform calculations simultaneously on whole families of materials in the time required for a single system only a few years ago. The improvements in other methods, such as simulations using classical force fields, have been no less dramatic.

Two years ago I wrote that "materials science is one of the great beneficiaries of changes in the landscape of scientific computing"[1], but this is not really true today. Problems in materials science (and other areas of condensed matter physics) play minor roles in JuGene allocations compared with elementary particle physics, in particular, where the resources requested routinely exceed by far the total available for distribution. Such calculations are crucial, of course, to our understanding of the basic structure of matter, but I do not think that this "fundamental" aspect alone justifies the overwhelming allocations to this area. One of the most eminent elementary particle physicists, Steven Weinberg, noted

that this field "has no direct applications that anyone can foresee"[2]. By contrast, materials scientists are challenged on a daily basis by problems with direct applications. The following contributions to this volume provide examples of materials-oriented research carried out on JuGene, JUMP, and JUROPA.

The work on magnetic layers that led to the 2007 Nobel Prize in Physics (Peter Grünberg, Jülich and Albert Fert, Paris) has been discussed on countless occasions, and significant developments in "spintronics" are inevitable. Bihlmayer *et al.* consider low-dimensional magnetic systems (films, wires) and identify the main sources of magnetic anisotropy in these systems. Sakong and Kratzer calculate the lifetime of adsorbates on semiconductor and insulator surfaces, with special reference to CO vibrations on silicon, still an important surface in semiconductor applications. Walter and Moseler focus on Na and Pd clusters supported on thin MgO films, with the long-term goal of understanding catalytic processes. Stärk and coworkers study structural and magnetic properties in multilayer systems, particularly cobalt on diamond. These four applications use DF theory and show just how much detailed information can be obtained on complex systems.

The remaining applications focus on much larger systems that remain far from accessible to DF calculations: Albe *et al.* describe large-scale MD simulations on plastic deformation in metals, particularly Pd, and Brener *et al.* focus on pattern formation across a range of length scales. The examples discussed include crack propagation and crystal growth. The range of materials science applications that can be attacked with the newest generation of HPC machines was unthinkable in 1987, and there will be challenging problems with important applications for years to come.

References

1. R. O. Jones, Proceedings, NIC Symposium 2008, NIC Series Vol. **39**, 151-152, 2008.
2. *"Although surprises are always possible, my own main research area, elementary particle physics, has no direct applications that one can foresee."* S. Weinberg, New York Review of Books, **56**, Number 16, 19-22, 2009.

Deformation Processes in Nanostructured Metals and Alloys

Karsten Albe[1], Alexander Stukowski[1], Jonathan Schäfer[1], Yvonne Ritter[1], and Diana Farkas[2]

[1] Technische Universität Darmstadt, Institut für Materialwissenschaft
Fachgebiet Materialmodellierung
Petersenstr. 32, 64287 Darmstadt, Germany
E-mail: albe@mm.tu-darmstadt.de

[2] Materials Science and Engineering, College of Engineering
Virginia Polytechnical Institute and State University
Holden Hall, Blacksburg, VA 24060, USA

The influence of solutes and twin planes on the plastic behaviour of nanocrystalline (nc) metals is studied by means of large-scale molecular dynamics simulations. First, the role of Au-solute atoms on the deformation behaviour of nc-Pd is investigated. The results reveal the importance of segregation effects in grain boundary areas even for this fully miscible material. Secondly, we study the role of as-prepared twin planes on the deformation behaviour. While Cu samples show the expected strengthening, the presence of twins in nc-Pd samples leads to a softening effect, which is explained by the occurrence of twinning dislocations.

1 Introduction

One of the key issues in materials engineering is the characterization, control and optimization of mechanical properties. Depending on the device application parameters like the materials strength and ductility need to be adjusted, which requires a detailed understanding of the deformation processes on the atomic scale.

In nanocrystalline materials the mechanical properties can become very different compared to coarse-grained polycrystalline materials. Hence, a lot of research has focused on this material class since Gleiter published his landmark paper in 1989[1] and various possible applications of nanocrystalline materials were proposed in the past[2]. Plastic deformation in general is considered to change from intra-granular to inter-granular or from bulk to interface mechanisms, when the grain size is decreased below a certain limit d_c[3]. This makes the identification of the active mechanisms of plastic deformation on the nanoscale one of the key issues to ensure that structural requirements are trustworthy fulfilled also by nc-materials. When grain size is decreased, hardness and yield stress of the material typically increase, which is known as the Hall-Petch effect[4,5]. This effect is, for example, not observed in the nanometer regime[6]. It turns out that for the smallest grain sizes, the flow stress even levels off and decreases[7] which is known as reverse Hall-Petch effect[8,9].

Since experimental approaches for sample generation and the according *in situ* or post-experimental analysis to study plasticity in various loading conditions are complicated by different factors (e.g. porosity)[10], computer simulations such as molecular dynamics are the method of choice for obtaining a detailed understanding of mechanisms operating on the atomic scale. In the past computer simulations revealed unexpected mechanisms, such as grain boundary sliding and intra-granular slip involving dislocation emission and

absorption at grain boundaries[11–13]. These deformation modes are somehow similar to thermally activated super-plasticity, but grain boundary sliding in nanomaterials takes place also at low temperatures. The precise nature of these slip mechanisms, however, is the subject of considerable debate[12] and it is assumed, that the amount of slip not only depends on grain size, but also on the grain boundary type[14].

While in the past there has been a strong focus on modelling nanocrystalline fcc-metals, like Ni, Al and Pd, there are few studies on the deformation behaviour of nanocrystalline alloys[15]. Introducing miscible solutes, however, offers an additional degree of freedom to the system. This leads to modified mechanical properties as compared to the pure metals, since alloying of nanocrystalline metals can stabilize the micostructure if the solute segregates to the boundary, modify the elastic properties and the stacking-fault energies.

In the first part of this contribution, we present results on a completely soluble model system (PdAu) where fully three dimensional samples are constructed and alloyed with a hybrid MD-MC method in the semi-grand canonical ensemble[16] before deformation behaviour is studied.

Another route to further modify the properties of nanocrystalline metals is by introducing additional twin planes in the as-prepared samples. Recent experimental studies for Cu have revealed that twins with spacings in the nanometer scale significantly increase hardness and yield strength without loss of ductility or electrical conductivity[17–19].

The widely accepted hypothesis for the strengthening effect of the twins found in both experimental and simulation work on copper is the fact that twins are effective barriers to slip transfer and therefore hinder plastic deformation that is dislocation mediated. However, it is also possible that twins act as nucleation sites for additional dislocations[20,21]. Furthermore it is also possible that twins migrate during the deformation process contributing to plasticity[22]. The interplay among these various effects of the twin boundaries is studied in the second part of this paper by comparing results of Cu and Pd .

2 Methods

Samples consisting of 54 grains with an average grain size of 15nm (alloyed nc) and 20nm (twinned nc) were generated using a Voronoi tessellation scheme. The grain orientations were chosen to be completely random with a unimodal grain size distribution. We first relaxed the as-created nanostructure at T=300K and p=0bar for 30ps. Uniaxial deformation was performed under standard strain controlled tensile test conditions with the tensile direction along the z-axis of our sample. Periodic boundary conditions were used in all directions, with the periodicity along z controlled to give a strain rate of $10^8 \mathrm{s}^{-1}$. The simulations were carried out using the open source MD-code LAMMPS[23]. For data analysis and visualization the software tool OVITO is used, which was developed in our group[24]. In order to distinguish the topology of atomic sites the common next neighbour analysis[25] scheme as implemented in OVITO was applied. For calculating atomic forces, we use interatomic potentials of the embedded-atom type (EAM) for Cu[26], Pd[27] and Au[28]. The Pd-Au potential is a cross-parameterization following the concentration-dependent EAM scheme[29].

For the alloyed samples, initial relaxation and alloying was done for 1 ns at 600 K by a parallel-efficient MC-MD hybrid simulation[16] that generates an equilibrium distribution of the solutes within grain interiors and grain boundaries.

3 Pd-Au Nanocrystals

Pd-Au is a solid solution that forms intermetallic phases at low temperatures. By equilibrating the composition of the material for a given chemical potential difference between the constituents using the hybrid MD-MC scheme the occupation of lattice sites with Pd and Au was adjusted at 600 K and averaged over 25000 MC steps to determine the local concentration. Figure 1 shows the variation of local concentration for a representative slice of a sample containing 30% gold. While the concentration is homogeneous within the grains, preferential segregation of either Pd or Au can be found in grain boundary areas. Averaging the local concentration over all atoms which have no perfect fcc surrounding shows, that the Au concentration in the GBs is lower than the total average. Grain boundary enrichment in segregating systems was observed in computer simulation[31] and experiments before. But grain boundary depletion in nanocrystalline miscible systems has to our knowledge not been reported. By comparing microstructures of different grain size, we additionally find an increasing Au depletion of the grain boundary areas with increasing grain size.

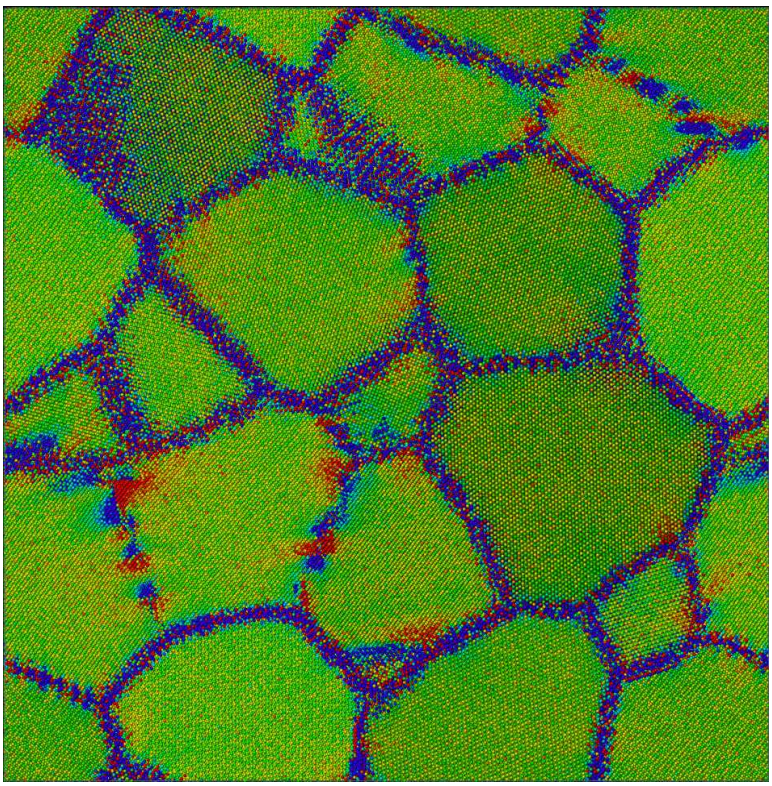


Figure 1. Concentration distribution in a Pd-Au nanocrystal containing in average 30% gold. The colour scheme shows the deviation from the average concentration (blue $>$ 10% , red $<$ -10 %).

.

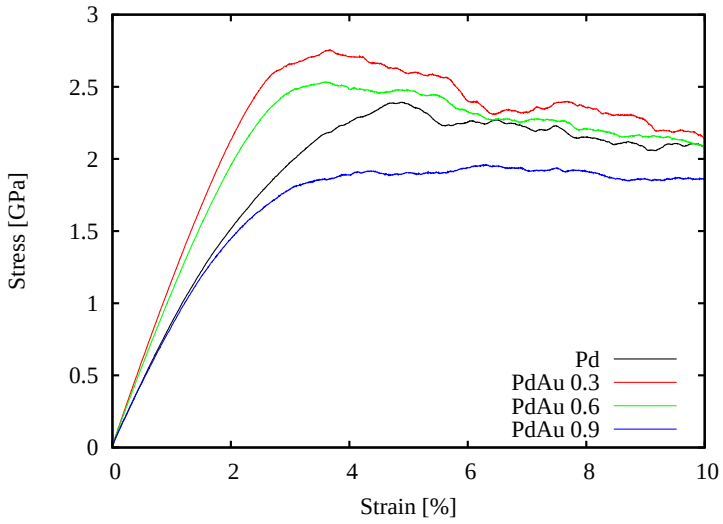


Figure 2. Stress-strain behaviour of Pd-Au nanocrystals of 15 nm with different Au concentrations calculated with a constant strain rate of $10^8 s^{-1}$.

Comparing the stress-strain behaviour under normal stress for samples with different Au content reveals another interesting feature. The gold concentration leads to an elastic hardening of the material which is reflected by the different slopes of the stress-strain curves in the elastic regime of Figure 2. While elastic moduli of the homogeneous bulk crystal have maxima at 50% Au concentration, the stress-strain curve of the nanocrystalline sample with 30% Au yields the maximum flow stress. This can be attributed to the fact that for 50% Au concentration partial dislocations have a larger dissociation width because of a lower stacking-fault energy, which in turn requires lower stresses for dislocation nucleation.

Our results reveal that the effect of alloying on the deformation behaviour of nanocrystals can (even in the case of a fully miscible system) not simply be described by concentration dependent material properties, but requires a detailed analysis of segregation effects in the grain boundary area, where local constraints lead to depletion or enrichment of the solute.

4 Twinned Nanocrystals

The extremely high strength and hardness of nanocrystalline materials relative to their coarse-grained counterparts may give rise to unique applications for these materials. In particular, twins with spacings in the nanometer scale have been observed to significantly increase hardness and yield strength in Cu. The nano-scale grains and twins also produce particular strain hardening and strain rate sensitivity behaviour. Twins are ubiquitous to many nanocrystalline fcc metals, suggesting that twinning is a major deformation mechanism when the grain sizes reach nanometer range. Twins may appear as a result of annealing and may include multi-fold twins. In this section, we present results that help to

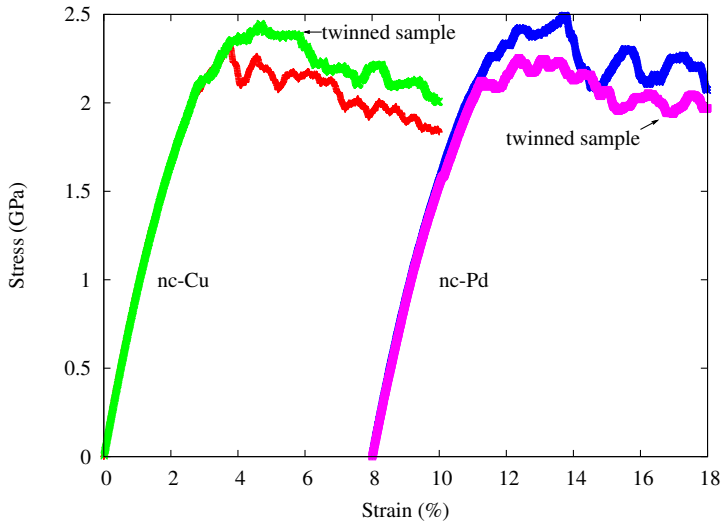


Figure 3. Stress-strain curves of nanocrystalline Cu and Pd with 20nm average grain size. Twinned samples contain a pair of twin planes in 5nm distance.

understand the precise mechanism of the strengthening observed in Cu and address the important question, if strengthening by twins is a general feature for fcc metals. We compare nanocrystals consisting of Cu and Pd. While Cu has a low stacking fault energy that results in a dissociation width of partial dislocations corresponding to the grain diameter, Pd can be understood as a material where only full dislocations are present, which suggests that cross-slip of dislocations through twin planes might be facilitated.

For our simulations, we have generated nanocrystals of 20 nm grain size with a pair of twin planes in every grain (twin distance: 5nm). The samples are prepared in a way that maintains the same relative orientation between all grains as in the non-twinned sample. For comparing Cu and Pd identical but rescaled atomic configurations are used, which makes all calculations fully comparable.

In Figure 3 the calculated stress–strain curves are presented. In the case of copper the twinned sample clearly shows a higher flow stress than the non-twinned counterpart. This is in line with the experimental finding and supports the picture of twins acting as dislocation barriers. Most interestingly, however, the opposite effect is found for Pd. Here, the twins soften the material leading to a lowering of the flow stress.

The reason for this can be revealed by close inspection of the atomic configurations. In Cu dislocation propagation through twin-planes is widely prevented, although we observe rare cases of transmission of partials through the twin plane. Figure 4 a,b shows that the sample containing twin pairs has a smaller total area of stacking faults, which suggests that partial dislocations emitted from grain boundaries cannot contribute to slip transfer. If partials are blocked by twins the flow stress increases, since additional dislocations need to be nucleated, which explains the observed stress-strain behaviour.

In Pd we observe that twin planes also act as barriers for dislocations emitted from

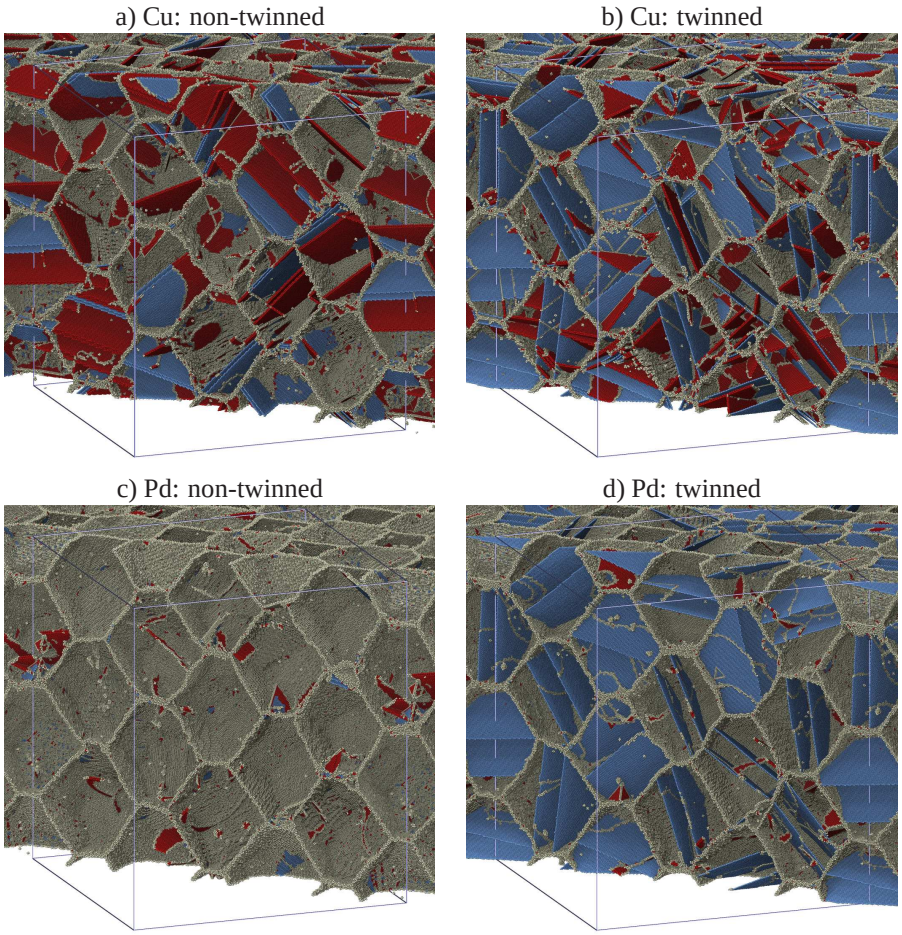


Figure 4. Structure of nanocrystalline Pd at 5% strain. Samples consisted of single-crystalline grains before deformation (non-twinned) or loaded with a twin pair before deformation (twinned). Shown are all atoms that are not in a perfect fcc-environment. Grey atoms are located in grain boundaries, blue atoms belong to twin planes (including deformation twins), red atoms are part of stacking faults.

grain boundaries, which despite of the small splitting distance cannot cross slip. This, however, would suggest that a strengthening effect similar to Cu is operational. Indeed, what finally causes the observed softening effect, is the occurrence of a large number of twinning dislocations in Pd that move on top of the twin planes and result in twin plane motion which provides the necessary slip transfer. This can be explained by the fact that the nucleation barrier for the formation of twinning dislocations is smaller than that for lattice dislocations in Pd, while the opposite is true for Cu.

Therefore, we can conclude that strengthening due to the presence of nanotwins is not a universal feature of all fcc metals, but depends on the generalized twinning and stacking-fault energy of the respective material.

5 Summary

In this contribution, we have presented molecular dynamics simulations of uniaxial plastic deformation of nanocrystalline metals and alloys. Our results reveal insights into deformation mechanisms on time and lengths scales that are not accessible by experimental methods. Specifically, we have shown how the presence of solutes or microstructural defects, like twin planes, affects the deformation behaviour of nanocrystalline metals.

Acknowledgements

This work was supported by Deutsche Forschungsgemeinschaft through project FOR714. The computations were performed with a grant of computer time provided by the John von Neumann Institute for Computing in Jülich.

References

1. H. Gleiter, *Nanocrystalline materials*, Progress in Materials Science, **33**, 223–315, 1989.
2. C. Suryanarayana and C. C. Koch, *Nanocrystalline materials - Current research and future directions*, Hyperfine Interactions, **130**, 5–44, 2000.
3. S. Yip, *Nanocrystals - The strongest size*, Nature, **391**, 532–533, 1998.
4. E. O. Hall, *The Deformation and ageing of mild steel. 3. Discussion of results*, Proceedings of the Physical Society of London Section B, **64**, 747–753, 1951.
5. N. J. Petch, *The cleavage strength of polycrystals*, Journal of the Iron And Steel Institute, **174**, 25–28, 1953.
6. G. W. Nieman, J. R. Weertman, and R. W. Siegel, *Mechanical-behavior of nanocrystalline Cu and Pd*, Journal of Materials Research, **6**, 1012–1027, 1991.
7. H. Conrad and J. Narayan, *Mechanism for grain size softening in nanocrystalline Zn*, Applied Physics Letters, **81**, 2241–2243, 2002.
8. A. H. Chokshi, A. Rosen, J. Karch, and H. Gleiter, *On the validity of the Hall-Petch relationship in nanocrystalline materials*, Scripta Metallurgica, **23**, 1679–1683, 1989.
9. R. W. Siegel and G. E. Fougere, *Mechanical properties of nanophase metals*, Nanostructured Materials, **6**, 205–216, 1995.
10. M. A. Meyers, A. Mishra, and D. J. Benson, *Mechanical properties of nanocrystalline materials*, Progress in Materials Science, **51**, 427–556, 2006.
11. K. S. Kumar, H. Van Swygenhoven, and S. Suresh, *Mechanical behavior of nanocrystalline metals and alloys*, Acta Materialia, **51**, 5743–5774, 2003.
12. H. Van Swygenhoven, P. M. Derlet, and A. G. Froseth, *Stacking fault energies and slip in nanocrystalline metals*, Nature Materials, **3**, 399–403, 2004.
13. D. Wolf, V. Yamakov, S. R. Phillpot, A. Mukherjee, and H. Gleiter, *Deformation of nanocrystalline materials by molecular-dynamics simulation: relationship to experiments?*, Acta Materialia, **53**, 1–40, 2005.
14. H. Van Swygenhoven, M. Spaczer, A. Caro, and D. Farkas, *Competing plastic deformation mechanisms in nanophase metals*, Physical Review B, **60**, 22–25, 1999.

15. S. Jang, Y. Purohit, D. L. Irving, C. Padgett, D. Brenner, and R. O. Scattergood, *Influence of Pb segregation on the deformation of nanocrystalline Al: Insights from molecular simulations*, Acta Materialia, **56**, 4750–4761, 2008.
16. S. Babak, P. Erhart, A. Stukowski, E. Martinez, and A. Caro, *A massively-parallel algorithm for atomic scale modelling of multi-component systems*, Modelling and Simulation in Materials Science and Engineering, *submitted*, 2009.
17. S. Suresh and J. Li, *Materials Science: Deformation of the ultra-strong*, Nature, **456**, 716–717, 2008.
18. Lei Lu, *Current progress of mechanical properties of metals with nano-scale twins*, Journal of Materials Science & Technology, **24**, no. 4, 473–482, JUL 2008.
19. L. Lu, X. Chen, X. Huang, and K. Lu, *Revealing the maximum strength in nano-twinned copper*, Science, **323**, 607–610, 2009.
20. Li Lan and M. Ghoniem Nasr, *Twin-size effects on the deformation of nanotwinned copper*, Physical Review B , **79**, 075444, 2009.
21. F. Sansoz, H. C. Huang, and D. H. Warner, *An atomistic perspective on twinning phenomena in nano-enhanced fcc metals*, Journal of Materials, **60**, 79–84, 2008.
22. F. Wang, P. Huang, and K. W. Xu, *Time dependent plasticity at real nanoscale deformation*, Applied Physics Letters, **90**, 16, 2007.
23. S. Plimpton, *Fast parallel algorithm for short-ranged molecular dynamics*, Journal of Computational Physics, **117**, 1–19, 1995.
24. A. Stukowski, *Visualization and analysis of atomistic simulation data with OVITO – The Open Visualization Tool*, Submitted to Model. Simul. Mater. Sci. Eng., 2009, Software available at `http://ovito.sourceforge.net/`.
25. Daniel Faken and Hannes Jansson, *Systematic analysis of local atomic structure combined with 3D computer graphics*, Computational Materials Science, **2**, 279–286, 1994.
26. Y. Mishin, M. J. Mehl, D. A. Papaconstantopoulos, A. F. Voter, J. D. Kress, *Structural stability and lattice defects in copper: Ab initio, tight-binding, and embedded-atom calculations*, Physical Review B, **63**, 22 2001.
27. S. M Foiles and J. J. Hoyt, *Computer Simulation of Bubble Growth in Metals Due to He*, Tech. Rep., Sandia National Laboratories, 2001.
28. G. J. Ackland, G. Tichy, V. Vitek, and M. W. Finnis, *Simple N-body potentials for the noble metals and nickel*, Philosophical Magazine A, **56**, no. 6, 735–756, 1987.
29. A. Stukowski, B. Sadigh, P. Erhart, A. Caro, *Efficient implementation of the concentration-dependent embedded atom method for molecular-dynamics and Monte-Carlo simulations*, Model. Simul. Mater. Sci. Eng., **17**, 075005, 2009.
30. A. J. Detor and C. A. Schuh, *Grain boundary segregation, chemical ordering and stability of nanocrystalline alloys: Atomistic computer simulations in the Ni-W system*, Acta Materialia, **55**, 4221–4232, 2007.
31. A. J. Detor and C. A. Schuh, *Tailoring and patterning the grain size of nanocrystalline alloys*, Acta Materialia, **55**, 371–379, 2007.

Anisotropic Magnetic Interactions in Low Dimensions

Gustav Bihlmayer[1], **Marcus Heide**[1], **Yuriy Mokrousov**[1,2], **Alexander Thiess**[1,2], **Stefan Heinze**[2], **and Stefan Blügel**[1]

[1] Institute for Solid State Research (IFF-1) and Institute for Advanced Simulation (IAS-1)
Research Centre Jülich, 52425 Jülich, Germany
E-mail: {g.bihlmayer, m.heide, y.mokrousov, a.thiess, s.bluegel}@fz-juelich.de

[2] Institute of Applied Physics
University of Hamburg, 20355 Hamburg, Germany
E-mail: {ymokrous, sheinze}@physnet.uni-hamburg.de

We present two examples of low-dimensional magnetic systems where effects of the magnetic anisotropy lead to important consequences for the stability of the magnetic structure. In ultra-thin films and atomic wires we can identify spin-orbit coupling effects to be the major source of the magnetic anisotropy and it further gives rise to the Dzyaloshinskii-Moriya interaction with important consequences for the magnetic structure. We think that the systematic investigation of relativistic effects in magnetic nanostructures by first-principles methods is crucial for the understanding of these materials and its use in spintronic applications.

1 Introduction

The miniaturization of magnetic structures that can be used for spintronic applications brings new challenges for the stability of these structures with respect to external perturbations, e.g. caused by temperature or external magnetic fields. The fact, that it is possible to stabilize (at least locally) magnetic order can only be explained by the fact that the material properties change fundamentally when the dimensionality of the system is reduced. As compared to a macroscopic bulk sample, the magnetic interactions in a nanostructure can be considerably enhanced or, even more fundamental, new interactions can get important that have been forbidden by symmetry in an extended, three-dimensional system.

A quantitative description of the interactions of a spin with its environment can only be obtained on the basis of relativistic quantum-mechanical calculations. In our institute IFF-1 and IAS-1 we have developed a computational method to apply this theory to materials that are actually used in experimental setups, which enables us to explain and predict properties that can be verified by highly advanced measurement tools as the spin-polarized scanning tunneling microscope or spin- and angle resolved photoemission. To this end, also on our, i.e. the computational side the most advanced equipment is necessary to make these collaborations a success. In this respect, only the close collaboration between the Jülich Supercomputing Centre (JSC) and the IFF, joined in the IAS of the Research Centre Jülich, made it possible to perform material science on the highest level.

The present contribution will briefly outline the employed computational methods and parallelization strategies. Then, as an example, we show the effects of magnetic anisotropy in iron nanostructures on an iridium and a tungsten substrate and end with a discussion of magnetic domain walls in this system. We demonstrate, how density functional theory (DFT) calculations in conjunction with other models (here a classical Heisenberg model

and a micromagnetic model) can be used to extrapolate the results into regimes that are difficult to access with in DFT, e.g. to finite temperatures or to very extended length scales.

2 Method & Computational Scheme

Density functional theory is an *effective* one-particle theory, i.e. the output quantities (charge density, total energy) can be described in terms of auxiliary single-particle wavefunctions that are solutions of a Schrödinger-type equation, the Kohn-Sham (KS) equation. It contains – in addition to the usual kinetic and potential energy terms – a potential in which the many-particle interactions are embedded. For magnetic systems, this framework can be extended to use explicitly spinor wavefunctions, $\Psi = (\psi^{\uparrow}, \psi^{\downarrow})$, to extract also spin-dependent quantities like the magnetization density from the KS equation. Still, this Hamiltonian is rotationally invariant in spin-space, i.e. we can transform the solutions by a rotation of all spins without changing the density or total energy of the system. This means, in particular, that the magnetization *direction* of an e.g. ferromagnetic film cannot be predicted. But for any magnetic application the knowledge of this quantity is, besides the magnetic moments and the magnetic order, of crucial importance.

In this theory the spin-space and real-space directions are decoupled, i.e. the orientation of the magnetic moments is independent of the spatial orientation of the crystal lattice. To introduce a coupling between real space and spin space, we have to consider the relativistic motion of the electron, that carries the spin, in the potential of the (screened) atomic nuclei. In the moving rest-frame of the electron, the potential gradient, i.e. an electric field $\mathbf{E}$, is Lorentz-transformed into a magnetic field, $\mathbf{B}$, that can couple to the electron's spin ($\boldsymbol{\sigma}$). This leads in the Dirac theory of the electron to a term in the Hamiltonian that has the form

$$\frac{\mu_{\mathrm{B}}}{2c}\boldsymbol{\sigma} \cdot (\mathbf{E}(\mathbf{r}) \times \mathbf{p}) \tag{1}$$

where $\mathbf{p}$ is the momentum of the electron and $\mu_{\mathrm{B}} = \frac{e\hbar}{2mc}$ is the Bohr magneton. Already from this formula it is obvious, that the effects we can expect are small (note the velocity of light, c, in the denominator), and calculations of the relativistic effects resulting from this term are challenging. Despite its smallness, this so called spin-orbit coupling (SOC) term (Eq. 1) is responsible for a large class of important properties, like the magnetic anisotropy that determines the magnetization orientation, or the antisymmetric exchange interaction we will encounter later. (One has to add here, that SOC is just *one* contribution to the magnetic anisotropy, other sources, like the dipole-dipole interaction, can get important, too. However, in the low-dimensional systems studied here spin-orbit coupling effects are found to dominate).

In an atomic system $\mathbf{E}$ in Eq. 1 is radially symmetric and the SOC operator can be written as $\xi\boldsymbol{\sigma} \cdot \mathbf{L}$, where $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is the orbital momentum operator. ξ includes the gradient of the atomic potential and scales approximately with the square of the nuclear number, Z, of the atom. For heavy elements like Bismuth spin-orbit effects can induce level-splittings in the order of 1 eV for the free atom. But in a solid the crystal field can overcome spin-orbit effects and quenches the orbital momentum. Related quantities like the magnetic anisotropy are reduced to a few μeV in the bulk. This is mainly due to the fact that changes due to a perturbation given by Eq. 1 contribute to the energy only in second order perturbation theory when we compare two magnetization directions as e.g.

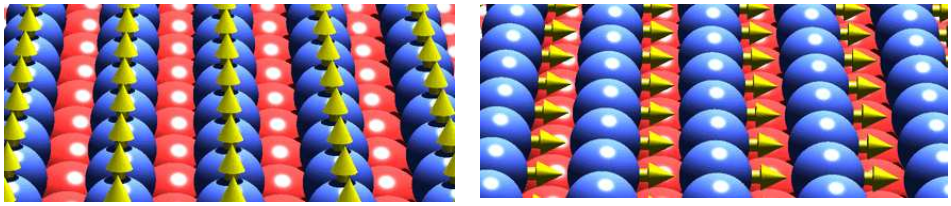


Figure 1. Two different magnetization directions of an ferromagnetically ordered thin film on a substrate. The energy difference between these configurations is typically in the order of 1 meV, requiring very well converged calculations.

shown in Figure 1. This means, that only states near the Fermi level can contribute via hybridization to a (direction dependent) de-quenching of the orbital moment that gives rise to a magnetization direction dependence of the total energy. To capture this effect even qualitatively a very fine sampling of the reciprocal space is necessary which makes this kind of calculations ideal candidates for massive parallelization. How this is actually implemented is described in another volume of this series[1].

While the magnetic anisotropy described above can be seen as a local, atomic quantity, spin-orbit coupling can also modify the interaction of spins on neighbouring atoms. While usually the coupling between two spins, $\mathbf{S}_i$ and $\mathbf{S}_j$, can be described by a Heisenberg-type term, $J_{ij}(\mathbf{S}_i \cdot \mathbf{S}_j)$, in the presence of SOC an interaction of the spins of the form $\mathbf{D}_{ij} \cdot (\mathbf{S}_i \times \mathbf{S}_j)$ can come into play. The appearance of this so-called Dzyaloshinskii-Moriya interaction (DMI) is coupled to certain symmetry conditions which are seldom met in bulk metals but are always present at surfaces. Therefore, in magnetic nanostructures DMI is not an exotic phenomenon but it has been mainly overlooked in the last decades. Our recent findings on ultrathin magnetic films clearly support this view[2,3].

A hallmark of the Dzyaloshinskii-Moriya interaction is the occurrence of homochiral spiral spin-structures. Spin-spirals are structures where the local magnetization turns from atom to atom along a certain direction, $\mathbf{q}$, by a fixed angle (characterized by $1/\|\mathbf{q}\|$). While such long-ranged spin-structures can be also a consequence of *scalar* interactions of the form $J_{ij}(\mathbf{S}_i \cdot \mathbf{S}_j)$, these interactions will not fix the sense of rotation (left- or right-handed). But the DMI, that contains the *vector* product of two spins, $\mathbf{S}_i \times \mathbf{S}_j$, can be responsible for the occurrence of only one sense of rotation, a so-called homochiral structure.

The calculation of spin-spirals, that can sometimes extend over many chemical unit cells, within DFT is a challenging task. In absence of spin-orbit coupling this can be achieved exploiting the symmetry of the problem, a feature built into our computer code FLEUR[4], which is tuned to the calculation of complex magnetic structures in low dimensions[5]. Only recently we succeeded to extend this formalism to spin-spirals in the presence of SOC which allow a calculation of the $\mathbf{D}_{ij}$ vectors from first principles[6]. For a reliable calculation of these tiny – but crucial – relativistic effects in huge magnetic structures the use of massively parallel supercomputers is indispensable.

3 Applications

In the following, we will illustrate the effects of spin-orbit coupling on the magnetic properties of iron double-chains and ultrathin Fe films on substrates with a large nuclear number,

Z. These materials are often chosen to enhance SOC effects in these systems to stabilize magnetic order at finite temperatures by an increase of magnetic anisotropy. But, as will be demonstrated below, SOC can lead not only to larger anisotropies, it can even influence the magnetic order on the atomic scale.

3.1 Magnetic Chains on Noble Metal Substrates: Fe on Ir(001)

As a first example to demonstrate the interplay of exchange interactions and relativistic effects in a low-dimensional magnetic system we chose Fe chains on Ir(001). Regular growth of monoatomic or diatomic chains can be accomplished by using either vicinal surfaces where the chains grow on step-edges[7] or reconstructed surfaces where the atoms arrange in pre-formed trenches of the substrate. One example for the latter is the (5×1) reconstructed Ir(001) surface where the growth of Fe chains could be accomplished experimentally[8]. From the application point of view, such well-ordered chains are promising since they provide well separated magnetic entities which are densely packed on a surface. The magnetic structure of these double-chains, however, is still not completely clear and spin-polarized scanning tunneling microscopy (SP-STM) studies hint at a dependence of the magnetic order on the structural details of the chains[9].

These experiments suggest the formation of ladder-like structures as shown in Figure 2. DFT calculations can single out two energetically favourable structures, the so-called C1-type, shown on the left of Figure 2 with a small distance (2.35 Å) between the two strands, and the C4-type with a larger chain separation (4.23 Å, right of Figure 2). These structures differ by less than 10 meV/Fe atom, i.e. depending on the kinetics of growth it is possible that both types are observed. From these calculations, which were performed assuming both ferromagnetic and antiferromagnetic order along the chains, it can also be deduced that the magnetic order is ferromagnetic for the C1-type and antiferromagnetic for the C4 chain[10]. But not only the magnetic order is different in the two cases, also the magnetization direction depends sensitively on the intrachain separation: while in the C1 chain the magnetization prefers to align perpendicular to the substrate surface, in the C4 case an in-plane easy axis is found, oriented along the chain direction.

Comparing the energies for different magnetization directions, we see that the two in-plane orientations are energetically very unfavourable in the C1-type (Figure 2 a), leading to an uniaxial anisotropy in this case. The C4-type, on the other hand, shows very little anisotropy in the (x, z)-plane, the energy needed to rotate the magnetization from in-plane to out-of-plane in this case is just 0.2 meV/Fe atom (Figure 2 b). This has not only significant impacts on the thermal stability of the magnetization in both cases (temperature can easily excite fluctuations of the magnetization in the (x, z)-plane in the C4-type, while the spins are rather confined to the z-direction in the C1 chain), it can also allow for more complex spin-structures in the C4 chains. That more involved spin-structures can occur in these systems is also supported by a comparison to thin Fe films on Ir(111): in monolayer Fe films on the hexagonal Ir(111) surface a complex spin-structure including at least 15 atoms in the magnetic unit cell was observed[11]. It still has to be explored, whether higher-order spin interactions or the above mentioned Dzyaloshinskii-Moriya interaction are strong enough to overcome the collinear arrangement of spins that is favoured by Heisenberg-type exchange and the magnetic anisotropy. This involves relativistic, non-collinear calculations of large supercells which are still challenging on state-of-the-art supercomputers. But in any case it is clear that spin-spirals are more likely to occur in

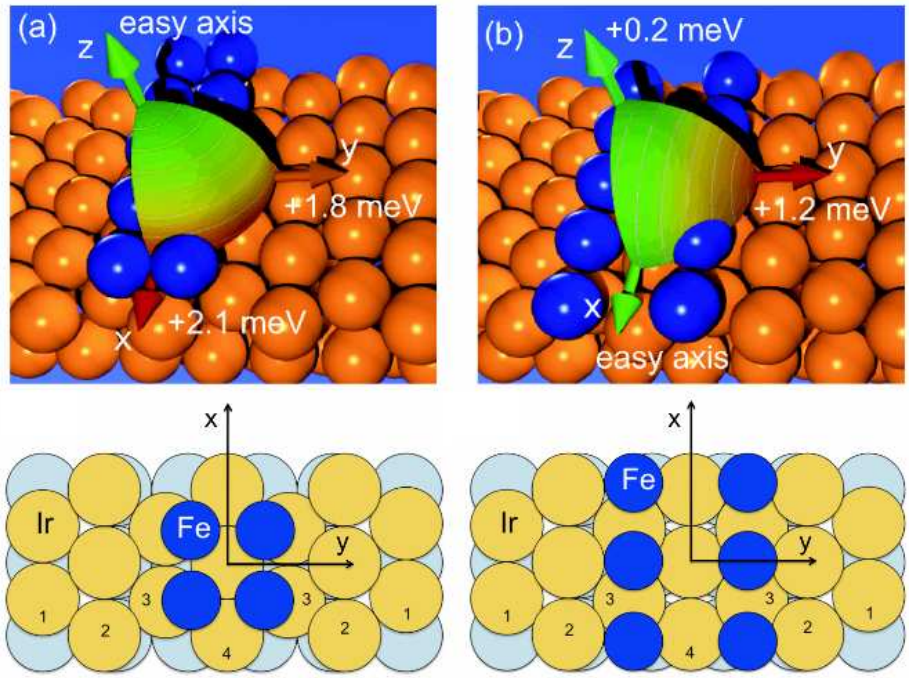


Figure 2. Structure and magnetic anisotropy of two types of biatomic Fe chains on a (5×1) reconstructed Ir(001) surface. The bottom Figures show the structure of the Fe chains of C1 and C4-type in the left and right panel, respectively. The chains extend periodically in x-direction, the separation in y direction is large enough to avoid interaction between the chains. From the calculations[10] it is seen that the two Fe strands are coupled ferromagnetically, along the x-direction the magnetic order is ferromagnetic in the C1-type and antiferromagnetic in the C4-type. The upper panels show the energy landscape as a function of the magnetization direction for the C1-type (a) and the C4-type (b).

C4-type chains than in the C1-type, be it as ground states or excitation. That long-ranged spin-spiral structures can be stabilized by relativistic effects has been demonstrated in the last years[2,3,12] and a related example will be shown in the next section.

3.2 Magnetic Domain Walls: Fe on W(110)

From the above discussion it can be seen that anisotropy and exchange interactions determine to a large extent the magnetic properties of low dimensional magnetic systems. A nice example, where the interplay of these interactions can be studied, is a magnetic domain wall, i.e. the region between two adjacent magnetic domains where the magnetization changes orientation. E.g. in a ferromagnetic system with out-of-plane anisotropy it changes from up to down. Magnetic domains are formed either due to long-range dipolar interactions which favour an antiferromagnetic alignment of domains or due to external magnetic fields that are used to write domain patterns in a material. For a high storage density of a magnetic data storage medium of course narrow domain walls are preferable.

In a simple model, the domain wall width is essentially determined by a compromise between the spin-stiffness, J, that characterizes the energy cost to tilt two neighbouring spins by an infinitesimally small angle and the magnetic anisotropy, K, that describes the energy cost to rotate a spin away from the easy axis. Then, the domain wall width, w, is given by $w = 2\sqrt{J/K}$, i.e. while the spin-stiffness favours wide domain walls where the magnetization changes only slightly from one atom to the next, the magnetic anisotropy favours sharp walls where most of the local magnetic moments are aligned along the easy axis. It should be noted, that in this model the magnetization rotates from the easy-axis via the medium-axis, i.e. it stays in the plane of lowest energy (in Figure 2 (a) this is the (z, y)-plane, while in Figure 2 (b) it is the (x, z)-plane).

The effect of different magnetic anisotropies can be clearly seen in the model system of Fe on W(110), where single monolayers show sharp domain walls, while a double layer Fe on this substrate shows a wide (about 7 nm) domain wall. Our calculations demonstrate that this is mainly due to the differences in anisotropy: 1 ML Fe/W(110) has an in-plane easy axis, but the energy to rotate the magnetization to out-of-plane or the other in-plane direction is large (about 2.5 meV). On the other hand, for the double layer the easy axis is out-of-plane, but the energy needed to orient the magnetization in an in-plane direction is rather small (about 0.5 meV). This, together with a rather small spin-stiffness in the monolayer system explains the observed difference in the domain-wall width[13].

Another interesting aspect in domain walls is its orientation: if two domains have to be formed, will they be separated by a wall that runs along a certain in-plane direction and, if yes, which one? Here, the directional dependence of the spin-stiffness comes into play. Taking the 2 ML Fe/W(110) system with its out-of-plane anisotropy as example, it might be easier to tilt two neighbouring spins in the (z, y)-plane than in the (z, x)-plane (x and y again indicate the two different in-plane directions). Experimentally, a clear preference for domain walls running along the $[1\bar{1}0]$ directions was observed. But since this preference was not found in the calculations of the spin-stiffness, other interactions that can account for this behaviour have to be included in the model.

The simple model introduced above treats the domain wall effectively as a non-collinear structure that locally resembles a spin-spiral and adds the effect of the magnetic anisotropy for all atoms in this spin-structure independently. This has the advantage that the parameters entering this model can be calculated independently from spin-spiral calculations and from relativistic calculations of a ferromagnetic film. Of course, such a model can not include effects arising from the relativistic treatment of spin-spirals, i.e. the Dzyaloshinskii-Moriya interaction. As outlined in section 2, this is a computationally demanding task and up to now was avoided in most calculations. Unfortunately, for the present problem this is no longer possible and the DMI has to be determined for these surfaces.

The consequences of the DMI for magnetic domain walls are quite fundamental: First, it can determine the type of the domain wall. Suppose in a wall the magnetization changes from up to down (w.r.t. the z-direction) as we move along the x-direction. Then the magnetization can rotate either in the (z, x)- or in the (z, y)-plane. The first case is called Néel-type wall, the second Bloch-wall. Secondly, the DMI can determine the sense of rotation in these planes, clockwise or anticlockwise. For the symmetry of the studied problem this is only possible in Néel-type walls. Finally, since DMI prefers Néel-type walls with a certain sense of rotation, the magnetic anisotropy is now indirectly coupled to the

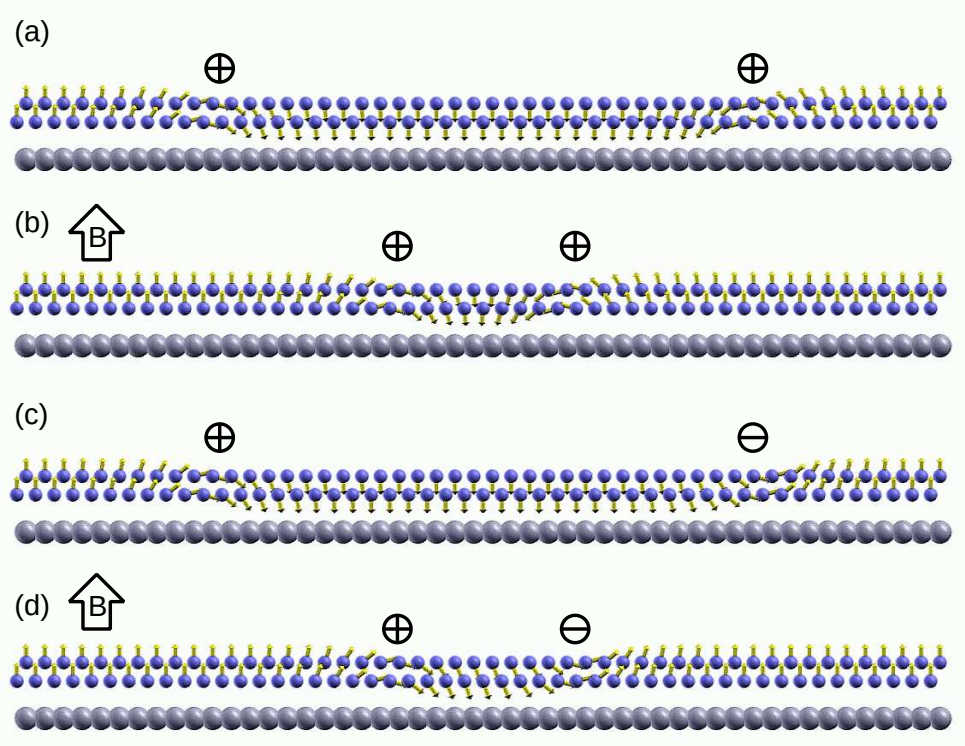


Figure 3. (a) Magnetic domain of a thin film bounded by two Néel-type domain walls of the same chirality. (b) If an magnetic field is applied, the domain with opposite orientation to the **B**-field shrinks and the domain walls move towards each other. Nevertheless, there is no magnetic torque in the centre of the domain that could destabilize this magnetic entity. (c) The same situation as in (a), but now with domain walls of opposite chirality. (d) If an magnetic field is applied, a torque acts throughout the domain and it will finally disappear, annihilating the $(+)$ and $(-)$ domain wall.

direction of the domain wall and the DMI can also influence this property. As it turns out, in the case of Fe double-layers on W(110) this is indeed the case and the experimental findings can be explained from relativistic, non-collinear DFT calculations[12].

As a consequence of this mechanism all domain walls have a certain rotational sense (sometimes called chirality) in this system as indicated in Figure 3 (a). Note, that this has also important implications for the stability of these domain walls with respect to external perturbations, e.g. an applied magnetic field. As this **B**-field increases, some domains will expand and some will shrink, causing a movement of the domain walls. The energy in the shrinking wall rises, as the structure gets increasingly non-collinear (Figure 3 b). In a continuum model we would say that energy piles up when the two domain walls are squeezed together, in this way protecting the middle domain from annihilation. On the other hand, if the domain walls are of opposite chirality as shown in Figure 3 (c) an applied magnetic field will be able to shrink and finally eliminate the oppositely oriented domain without this huge energy barrier (Figure 3 d). Interestingly, experiments on the Fe/W(110)

system using SP-STM reported the existence of residual domains of a few nanometers that were found to be stable in magnetic fields[14]. With our calculations we can identify the source of the homochirality of domain walls in this material, that leads to a kind of topological protection of domains in external fields.

Acknowledgements

The calculations described here would have been impossible without the grant of computer time provided by the VSR of the Research Centre Jülich, which is gratefully acknowledged.

References

1. G. Bihlmayer, P. Ferriani, S. Baud, M. Ležaić, S. Heinze, and S. Blügel, *Ultra-thin magnetic films and magnetic nanostructures on surfaces*, in: NIC Symposium 2006, vol. 32 of *NIC Series*, p. 151, Neumann Institute for Computing, Jülich, 2006.
2. M. Bode, M. Heide, K. v. Bergmann, P. Ferriani, S. Heinze, G. Bihlmayer, A. Kubetzka, O. Pietzsch, S. Blügel, and R. Wiesendanger, *Chiral magnetic order at surfaces driven by inversion asymmetry*, Nature, **447**, 190, 2007.
3. P. Ferriani, K. v. Bergmann, E. Y. Vedmedenko, S. Heinze, M. Bode, M. Heide, G. Bihlmayer, S. Blügel, and R. Wiesendanger, *Atomic-scale spin spiral with unique rotational sense: Mn monolayer on W(001)*, Phys. Rev. Lett., **101**, 027201, 2008.
4. For a program description see `http://www.flapw.de`.
5. Ph. Kurz, F. Förster, L. Nordström, G. Bihlmayer, and S. Blügel, Ab initio *treatment of non-collinear magnets with the full-potential linearized augmented planewave method*, Phys. Rev. B, **69**, 024415, 2004.
6. M. Heide, G. Bihlmayer, and S. Blügel, *Describing Dzyaloshinskii-Moriya spirals from first principles*, Physica B, **404**, 2678, 2009.
7. S. Baud, G. Bihlmayer, S. Blügel, and Ch. Ramseyer, *First-principles investigation of Co wires at Pt(111) step edges*, Surface Science, **600**, 4301, 2006.
8. A. Klein, A. Schmidt, L. Hammer, and K. Heinz, *Lateral nanoscale Fe-Ir superlattices on Ir(100)*, Europhys. Lett., **65**, no. 6, 830, 2004.
9. M. Menzel, *Wachstum und elektronische Eigenschaften von atomaren Eisen-Ketten auf Iridium(001)*, Diploma thesis, University Hamburg, 2006.
10. Y. Mokrousov, A. Thiess, and S. Heinze, *Structurally-driven magnetic state transition of biatomic Fe chains on Ir(001)*, submitted, 2009.
11. K. v. Bergmann, S. Heinze, M. Bode, G. Bihlmayer, S. Blügel, and R. Wiesendanger, *Complex magnetism of the Fe monolayer on Ir(111)*, New J. Phys., **9**, 396, 2007.
12. M. Heide, G. Bihlmayer, and S. Blügel, *Dzyaloshinskii-Moriya interaction accounting for the orientation of magnetic domains in ultrathin films: Fe/W(110)*, Phys. Rev. B, **78**, 140403(R), 2008.
13. M. Heide, *Magnetic domain walls in ultrathin films: Contribution of the Dzyaloshinsky-Moriya interaction*, Ph.D. thesis, RWTH Aachen, 2008.
14. O. Pietzsch, A. Kubetzka, M. Bode, and R. Wiesendanger, *Observation of Magnetic Hysteresis at the Nanometer scale by Spin-Polarized Scanning Tunneling Spectroscopy*, Science, **292**, 2053, 2001.

Pattern Formation: From the Macro- to the Nanoscale

E. A. Brener[1], M. Fleck[1], C. Hüter[1], H. Müller-Krumbhaar[1], D. Pilipenko[1], D. Temkin[1], and R. Spatschek[2]

[1] Institute for Solid State Physics IFF-III, Research Centre Jülich
52425 Jülich, Germany
E-mail: {e.brener, m.fleck, c.hueter, h.mueller-krumbhaar, d.pilipenko}@fz-juelich.de

[2] Interdisciplinary Center for Advanced Materials Simulation ICAMS, Ruhr University Bochum
Stiepeler Strasse 129, 44801 Bochum, Germany
E-mail: robert.spatschek@rub.de

We report on recent results for elastically influenced phase transition and pattern formation processes. A characteristic requirement for the modeling of these phenomena is to handle moving boundaries, which affect the solutions in a nonlocal and nonlinear manner. In order to avoid the tracking of the interfaces, where boundary conditions have to be applied, we use the phase field method. The arising partial differential equations are solved with parallel methods, and we demonstrate applications to crack propagation, solid-state transformations, melting processes and polycrystal formation.

1 Introduction

Phase transitions and their dynamics are essential ingredients for many features in nature, and their understanding is important to develop novel materials with superior properties. The mechanical behaviour of steels depends significantly on the transformation and cooling kinetics, with strong influence e.g. on dendrite formation and other solidification patterns. Elastic effects can become particularly important during solid-state transformations, which can occur with velocities up to values close to the sound speed, but also in diffusion-controlled processes. Often, deformations can become so severe that material failure through cracks happens, but also the formation of grain boundaries and crystallographic defects is essential for the hardening of the materials.

A thorough understanding therefore demands a description on a huge range of length- and timescales, which is only feasible with the help of highly parallelized supercomputers, using up to several thousand processors. In this article we present recent results for elastically influenced macroscopic processes, with a detailed modeling of diffusion-limited solidification, solid-state transformations and fracture on the mesoscale, with an outlook to new approaches down to a continuum modeling of individual defects with atomic resolution on the nanoscale.

2 Methods

The phase field method is nowadays one of the most important techniques to tackle problems with moving boundaries. It has the advantage that boundary conditions between the different components and phases of the system are automatically fulfilled, and therefore a complicated tracking of the advancing boundaries as in sharp interface methods is not

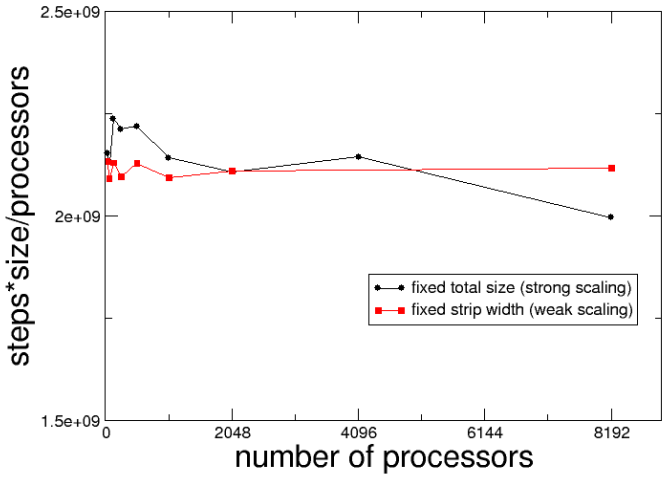


Figure 1. Scaling behaviour of the parallelized phase-field code. An ideal speedup would correspond to a constant horizontal line. One can see that the parallelization is extremely efficient.

required. Also, the undesired effect of lattice anisotropies can be well suppressed. Our programs go beyond the standard phase field models that are now available in the literature in the sense that they are derived variationally from the action instead of free energies, thereby successfully also capturing the effect of inertia.

Phase field methods are computationally demanding, as apart from the physical length-scales also the interface thickness has to be resolved. For a valid extraction of the sharp interface limit (for comparison with analytical results and other numerical methods as well as experiments, if the system is “mesoscopic”), the adjustable interface thickness has to be significantly smaller then the physical scales, which requires large-scale simulations. Fortunately, the phase field equations themselves are simple partial differential equations, which can be parallelized with a very high efficiency using MPI. For the numerical realization, we employ explicit representations of both the elastodynamic equations and the phase field dynamics, where the elastic displacements are defined on a staggered grid. Depending on the problem, we run our simulations on 20-2000 processors, although for special applications concerning crack coarsening we go to even higher processor numbers[1]. If we increase the system size with the number of processors (weak scaling), the speedup is about 98%. In the case of strong scaling, where the same computational domain is handled by an increasing number of processors, the speedup is still an excellent 92% if we use up to 8192 cpus. The benchmark results can be seen in Figure 1.

Additionally, we use eigenmode expansion methods, which are extremely efficient and precise for specific problems, but not as flexible as the phase field method. Since usually the solution of the elastic problem is the critical step, this leads to a major simplification, although the solution of the moving boundary problem still remains a nonlinear and non-

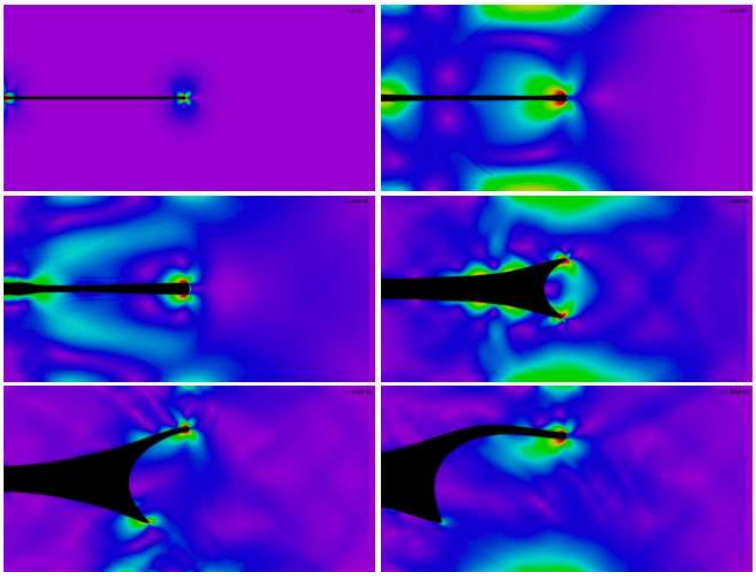


Figure 2. Unstable crack growth in a brittle medium for high driving force. The colour coding visualizes the elastic birefringence.

local problem that we solve by different minimization techniques.

Meanwhile, we developed also new codes for continuum modeling with atomistic resolution, based on phase field crystal and density functional descriptions. These codes are strongly accelerated using pseudospectral methods. All approaches, together with analytical and finite element methods, serve as a powerful basis for solving many pattern formation problems in solid-state physics.

3 Crack Growth in Brittle and Viscoelastic Media

The understanding of crack growth is highly important for materials science and engineering applications, but it still lacks convincing descriptions which predict generic features that are often observed in very different types of materials, but hardly understood. During the past years, we contributed to this field by a series of publications which propose a macroscopic description of fracture, based on pattern formation processes. Here, one of the central questions is related to the selection of a crack tip radius, because it is known that in a simple linear elastostatic description, an infinitely sharp crack tip should be the most favourable state, also known as the finite time cusp singularity of the Asaro-Tiller-Grinfeld[2] instability. We have demonstrated, that this theory can be naturally regularized by the inclusion of inertial effects, using theoretical arguments[3], multipole expansion methods[4] and fully dynamical phase field methods[5,6], see Figure 2.

Recently, we also included another important effect, which is bulk dissipation due to viscosity[7]. In the immediate vicinity of the crack tip, stresses are large and can generate additional heat apart from the dissipation directly at the crack front (see Figure 3). The clue

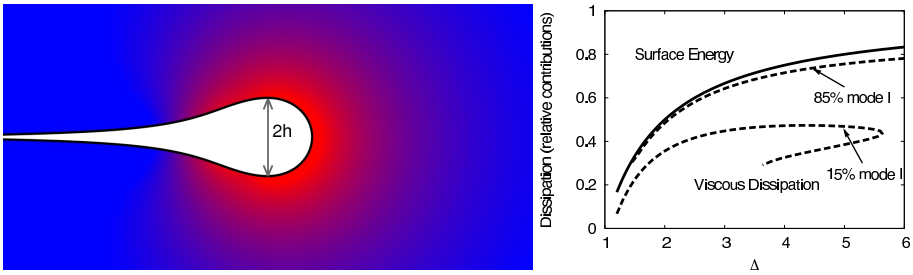


Figure 3. Left: Shape of a mode III crack in the steady state regime. The total incoming elastic energy flux is converted into surface energy, surface dissipation and viscous bulk dissipation, which is localized around the crack tip (visualized by the colour coding). The crack scale h is selected (together with the entire shape and the crack speed) self-consistently. Right: Distribution of energy consumption versus the crack extension force. The fraction above the solid curve is the relative contribution of surface energy generation, the part below the dashed line the viscous dissipation (shown here for two different admixtures of mode I loading; for 15% mode I also for the unstable branch). The remaining part between the curves is the surface dissipation, which obvioulsy plays an important role.

is that the viscosity introduces a new timescale to the problem, and therefore full selection becomes possible also in the low velocity regime. In particular, we investigated steady state crack propagation in viscoelastic media; this was done not only for pure tensile loading (mode I loading), but also additional shear stresses (mode III loading), and shed light on the important geometry dependence of fracture. We solved this complicated moving boundary problem by multipole expansion methods, which we adopted to include also the new source of dissipation.

It turns out that mode I and mode III situations exhibit strikingly different behaviours: In particular, mode III shows a pronounced transition towards unstable crack growth at higher driving forces, and the behaviour close to the Griffith point is determined entirely through crack surface dissipation, whereas in mode I the fracture energy is renormalized due to a remaining finite viscous dissipation (see Figure 3, right panel).

Intermediate mixed-mode scenarios allow steady state crack growth with higher velocities, leading to the conjecture that mode I cracks can be unstable with respect to a rotation of the crack front line. The results show that bulk dissipation is important for many crack processes and can significantly influence the behaviour.

4 Elastic Effects on Diffusion-Limited Solid-State Transformation

Transformations between melt, crystalline and amorphous states in nanovolumes at surfaces are basic processes during heterogeneous nucleation and growth. Due to the density differences between the phases, transformations in confined volumes are accompanied by internal stresses, which can be even increased by structural transformations or heterogeneities in the system. We investigated the growth regime where elastic effects naturally appear due to lattice strain.

Here we investigate the influence of elastic strain on dendritic growth in solids controlled by heat diffusion, using Green's function methods for either the thermal and elastic fields. By these means, we derive, as in the classical dendritic growth theory, a single

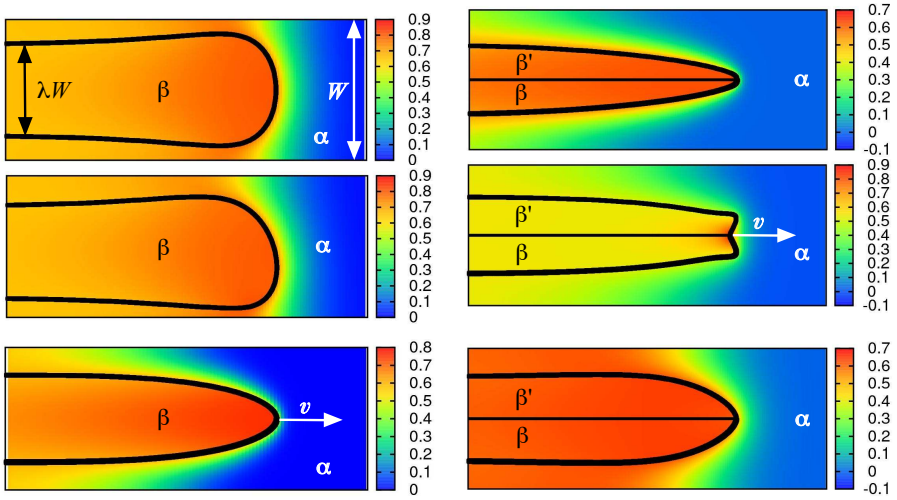


Figure 4. Steady state growth of a) a symmetrical finger and b) a parity-broken dendrite.

integro-differential equation for the shape of the interface which takes into account elastic effects. We consider two examples of dilatational and shear transformations in an infinite system. It turns out that in the case of a pure dilatation, the elastic effects lead only to a trivial shift of the transition temperature for the free growth. In contrast, for the growth in a channel, which is modeled by the phase field method, steady state solutions do exist in this case[8,9].

However, for the case of a mixture of dilatation and shear transitions in an infinite system, we find dendritic patterns even without anisotropy of surface energy, in contrast to classical dendritic growth theory. In this sense, the elastic effects serve as a new selection mechanism. Moreover, we find that the growth velocity becomes large, due to elastic effects, for small values of the undercooling, while in the classical dendritic growth, the velocity is always small, being controlled by tiny anisotropy effects. We performed numerical simulations for both the free growth in an infinite system (Green's function method) and in a channel or narrow strip (phase field approach). With the phase field simulations, we are able to investigate this process also in a channel geometry, and to look also at transient behaviours. Snapshots of different final growth patterns, consisting of symmetrical, asymmetrical and bicrystal patterns, are shown in Figure 4.

5 Kinetics of Dendritic Melting

In many heterogeneous systems several kinetic mechanisms together determine the growth of microstructures. Whereas bulk diffusion within solid phases is typically slow, diffusion along grain boundaries or heterogeneities can be much more efficient at ambient temperatures. Although dendrite propagation, also along grain boundaries, is possible if no

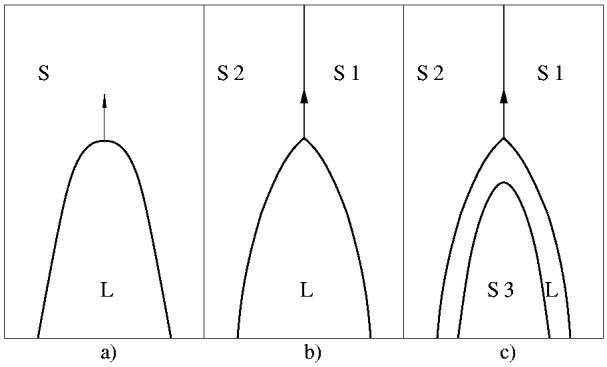


Figure 5. Schematic representation of symmetric melting structures. a) with only one melting front; b) with only one melting front during the growth along a grain boundary; c) combined motion of the melting front along the grain boundary and the solidification front $S3/L$ via the LFM mechanism.

component diffusion is required, the simultaneous motion of two-phase boundaries can be kinetically more efficient for example in peritectic or eutectic systems, if driven by coherency strains.

In our recent work we considered liquid film migration during the partial melting along a grain boundary (Figure 5b) . We have developed and analyzed a model for the melting kinetics along grain boundaries in alloys. The presence of the triple junction plays a central role in controlling the velocity of this process, since it generates a very strong perturbation of the solid-liquid interface. The anisotropy of surface tension is not important in such processes[10–12].

We solved this problem in lubrication approximation which allows to reduce the originally nonlocal problem to a local and even analytically tractable problem. The nucleation of the melt takes place at the grain boundaries of the mother solid phase. Then the thin liquid layer extends along the grain boundary.

The process with only one melting front (see Figure 5a) would be controlled by the very slow diffusion in the solid phase. In this case the growth velocity scales as in classical dendritic growth. If the nucleation of the melt takes place at the triple point where three grains meet together, the process proceeds as shown in Figure 5c.

6 Continuum Modeling on the Nanoscale

Phase field descriptions are specifically designed to model processes on the mesoscale, with solidification being the most prominent example. The introduction of the phase field interface thickness appears here traditionally as a numerical "tool" to avoid the complicated tracking of the interface motion in a sharp interface picture. It is important to mention that true interface thicknesses, i.e. the lengthscale over which crystalline ordering decays at a solid-liquid interface, is even for rough surfaces only on the nanometer scale; in contrast, the phase field interface thicknesses used for typical phase field simulations is rather on the scale of μm.

From a physical point of view the proximity of interfaces can lead to new phenomena,

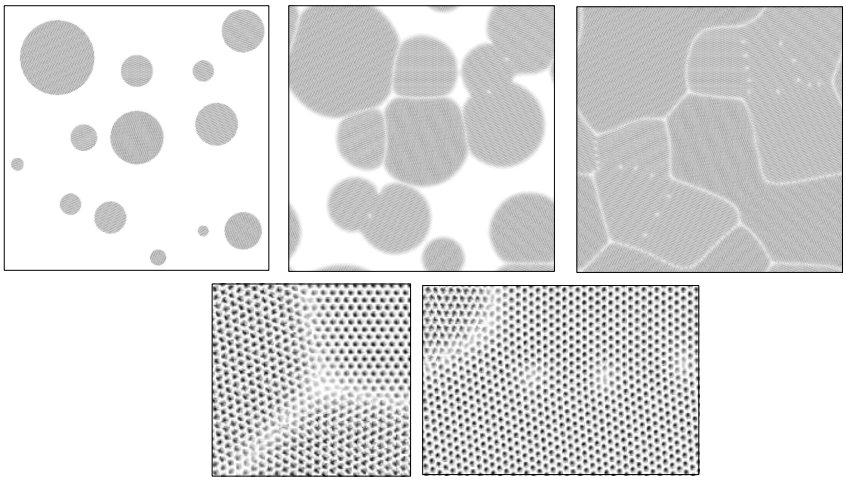


Figure 6. Growth of δ-Fe nuclei in liquid iron below the melting point. When the nuclei meet, they form grain boundaries. For small misorientations, they consist of isolated dislocations.

because the "overlap" of the profiles leads to changes of the energetic situation. In particular, if two grains, which were previously separated by a wide melt layer, come close to each other, they form a grain boundary, which consist of dislocations and therefore provoke additional long-range elastic deformations. This can lead to phenomena like grain boundary premelting, i.e. nanometric melting at the grain boundary already below the melting point, which is attributed to repulsive interactions between the neighbouring grains.

In order to go beyond the concept of phenomenological descriptions, we developed a Ginzburg-Landau theory for interface interaction, that is derived either from the phase field crystal model or density functional theory[13,14]. This theory, and the corresponding phase field crystal models, are strongly related to descriptions of block copolymer melts. It describes the complex amplitudes of the density waves, which appear here as fundamentally new entities, in contrast to conventional phase field models. This model allows then to study e.g. polycrystalline solidification with atomic resolution. In contrast to molecular dynamics simulations, it operates on diffusive timescales and therefore allows to simulate the entire pattern formation process.

7 Summary

Many problems in physics and materials science are related to moving boundaries, and we presented results of numerical and analytical calculations for applications in this field. By the development of suitable methods, based on sharp interface and phase field concepts, it is nowadays possible to model phenomena on a wide range of lengthscales and long timescales, which are currently not accessible to atomistic modeling approaching.

Acknowledgements

This project is supported by the German Research Foundation under grant SPP 1296 and the German-Israeli Foundation.

References

1. R. Spatschek, C. Müller-Gugenberger, and E. Brener, Effective Elastic Moduli in Solids with High Crack Density, accepted by Phys. Rev. B.
2. R. J. Asaro and W. A. Tiller, Interface Morphology Development During Stress Corrosion Cracking. Part 1. Via Surface Diffusion, Metall. Tran. **3**, 1789 (1972).
3. E. Brener and R. Spatschek, Fast crack growth by surface diffusion, Phys. Rev. E **67**, 016112 (2003).
4. D. Pilipenko, R. Spatschek, E. Brener, and H. Müller-Krumbhaar, Crack Propagation as a Free Boundary Problem, Phys. Rev. Lett. **98**, 015503 (2007).
5. R. Spatschek, C. Müller-Gugenberger, E. Brener, and B. Nestler, Phase Field Modeling of Fracture and Stress Induced Phase Transitions, Phys. Rev. E **75**, 066111 (2007).
6. R. Spatschek, M. Hartmann, E. Brener, and H. Müller-Krumbhaar, Phase Field Modeling of Fast Crack Propagation, Phys. Rev. Lett. **96**, 015502 (2006).
7. R. Spatschek, E. Brener, D. Pilipenko, Crack Growth by Surface Diffusion in Viscoelastic Media, Phys. Rev. Lett. **101**, 205501 (2008).
8. E. Brener, M. Fleck, D. Pilipenko, and R. Spatschek, Pattern formation during diffusion limited transformations in solids, accepted by Phil Mag.
9. E. A. Brener, G. Boussinot, C. Hüter, M. Fleck, D. Pilipenko, R. Spatschek, and D. E. Temkin, Pattern Formation during diffusional transformations in the presence of triple junctions and elastic effects, accepted by JCPM.
10. D. E. Temkin, Abstracts of ICASP, June 7-10 2005, Stockholm, Sweden (KTH, Stockholm, 2005).
11. E. A. Brener, C. Hüter, D. Pilipenko, and D. E. Temkin, Velocity Selection Problem in the Presence of the Triple Junction, Phys. Rev. Lett. **99**, 105701 (2007).
12. E. A. Brener, D. E. Temkin, Melting of alloys along the inter-phase boundaries in eutectic and peritectic systems, Acta Mater. **55**, 2785 (2007).
13. Kuo-An Wu, Alain Karma, Jeffrey J. Hoyt, and Mark Asta, Ginzburg-Landau theory of crystalline anisotropy for bcc-liquid interfaces, Phys. Rev. B **73**, 094101 (2006).
14. Kuo-An Wu and Alain Karma, Phase-field crystal modeling of equilibrium bcc-liquid interfaces, Phys. Rev. B **76**, 184107 (2007).

First Principles Approach for Vibrational Relaxation of Adsorbates on Semiconductor Surfaces

Sung Sakong and Peter Kratzer

Fakultät für Physik and Center for Nanointegration (CeNIDE), Universität Duisburg-Essen
Lotharstraße 1, 47048 Duisburg, Germany
E-mail: sung.sakong@uni-due.de

We develop a first-principles approach to calculate the lifetime of adsorbate vibrations on semiconductor or insulator surfaces. In the vibrational relaxation of adsorbates, quantum mechanical aspects are very important, since dissipation of the vibrational energy to smaller vibrational quanta and substrate phonons is the dominant relaxation mechanism. Numerically, the vibrational lifetime is determined by highly accurate calculations of vibrational energies and forces on coupled oscillators. With the help of high computer power, the energies and forces are mapped for realistic vibrational systems which are directly comparable to experiments. As examples, we study the relaxation of the stretching vibrations of CO on Si(100) and H on Ge(100), where lifetimes are in the range of nanoseconds.

1 Introduction

The damping of adsorbate vibrations is a subject of fundamental interest, since it epitomizes the intricacies of dissipation in open quantum systems, but also plays an important role in technology: Looking at surface passivation and nanolithography[1], "hot" chemistry at surfaces[2], or nanoscale friction[3], one realizes that, on the atomic scale, all these phenomena are governed by vibrational relaxation. For example, in laser-induced "hot" chemistry, it is crucial to understand the dissipation of energy from the quantized modes of adsorbates to the continuum of surface excitations in order to know for how long the energy pumped into a specific adsorbate vibration by the laser is available for a chemical reaction. Furthermore, isotopic effects on vibrational damping are important as well: The passivation of surfaces and defects in silicon-based semiconductor devices using deuterium rather than light hydrogen for more efficient vibrational damping (and hence more persistent passivation) is an example of a technological application[4].

In recent studies[5,6], we devised a first-prinicples approach to calculate the vibrational lifetime of adsorbates with high accuracy. While quantum dissipation, at least in principle, requires an open system with inifinitely many degrees of freedom, only a few degrees of freedom, namely the local modes of the adsorbate, govern the overall rate of energy decay of a specific adsorbate vibration into the phonon system. These relevant degrees of freedom are identified by means of total-energy calculations using density functional theory (DFT) for a finite system in a slab geometry. Since electronic excitations could be ruled out experimentally[7–9], ground-state DFT is suitable for the problem at hand. The coupling of the local modes to the phonon modes is described using a model Hamiltonian whose parameters are determined from the DFT calculations.

$$\hat{H} = \sum_{i=1} -\frac{\hbar^2}{2m_i}\frac{\partial^2}{\partial x_i^2} + V(x_1, x_2, \ldots) \tag{1}$$

The x_i are the Cartesian coordinates of the atoms moving on the adiabatic Born-Oppenheimer surface V, which is (in principle) accessible through DFT calculations. We transform this problem to a normal-mode representation, thereby introducing a system-bath decomposition[10].

$$\hat{H} = \hat{H}_{vib}(\partial/\partial q_j, q_j) + \hat{H}_{ph}(\partial/\partial Q_i, Q_i) + \hat{H}_{\lambda}(q_j, Q_i) \tag{2}$$

Here, $\hat{H}_{vib}$ is the system Hamiltonian depending on the *harmonic amplitudes* of local vibrations q_j ($j = 1, \ldots N$). These are related to small displacements of the adsorbate and possibly few substrate atoms to which it is bonded. The anharmonicity and couplings between the number of local modes, N, to be treated at a high level of accuracy using DFT, must be selected for each system according to physical considerations. The Q_i are normal mode coordinates of the phonon system which will be treated within harmonic approximation using a classic potential. The coupling between local vibrations and phonons is considered weak, and obtained from a Taylor expansion of v in the phonon coordinate $\hat{Q}_i$[10,11],

$$\hat{H}_{\lambda} = \sum_i \frac{\partial v}{\partial Q_i}\hat{Q}_i + \frac{1}{2}\sum_{i_1 i_2} \frac{\partial^2 v}{\partial Q_{i_1}\partial Q_{i_2}}\hat{Q}_{i_1}\hat{Q}_{i_2} + \cdots \tag{3}$$

Finally, rates for the transitions between the full normal modes mediated by emission or absorption of one phonon are calculated using Fermi's Golden Rule. The lifetime of mode n is inversely proportional to the sum of the rates over all possible relaxation channels m.

$$\tau_{vib}^{-1} = \sum_m \Gamma_m = \sum_m \frac{2\pi}{\hbar}|\langle m|\hat{H}_{\lambda}|n\rangle|^2 \int d\Omega\ \delta(\varepsilon_m - \varepsilon_n - \hbar\Omega) \tag{4}$$

To link this approach to DFT calculations, the anharmonic coupling within the adsorbate vibrations $\langle m|\hat{H}_{\lambda}|n\rangle$ must be accurately known, and therefore DFT calculations for the potential energy surface (PES) v for these local vibrational modes (q_j) must be carried out. For highly accurate PES calculations, a data set on a fine grid must be constructed. In general, the number of data points depends both on the number of mesh points per mode, q, and on the number of local modes considered, i.e. q^N DFT calculations are required. For example, CO/Si(100) requires 10^6 DFT calculations when we pick 10 mesh points for each of the six modes. This numerical effort requires immense computer power; however, combining physical intuition and the power of supercomputers enables us to calculate the PES for systems sufficiently large and complex so as to closely match experiment. In this contribution, we will demonstrate the calculations of vibrational lifetime of CO stretching on Si(100) and H(D) on Ge(100) surfaces.

2 Numerical Details

The DFT calculations for slab models of the semiconductor surfaces were carried out using the software VASP[12], employing the GGA-PBE functional[13] and the projector augmented-wave method[14]. The Si and Ge substrates are modeled by slabs of ten layers that are separated by 12 Å vacuum. The four uppermost Si and Ge layers are fully relaxed, and the dangling bonds at the lower surface of the slab are saturated with hydrogen atoms. The calculations have been done using a plane wave basis set up to the energy cutoff of 400 eV

for CO/Si(100) and 300 eV for H/Ge(100). For the Si(100) and Ge(100) surface, $c(4 \times 2)$ and $p(2 \times 1)$ unit cells are selected with a 4×8 and 4×4 **k**-point grids, respectively. After full relaxation including adsorbates, the dynamical matrix for the vibrations is calculated through the finite difference method.

In the golden-rule expression for the transition rate, the phonon density of state enters. Therefore we require a highly resolved phonon density of states, as well as the phonon eigenvectors for Si and Ge surfaces. The phonon properties are obtained from valence force field calculations using the General Utility Lattice Program (GULP)[15] with bond order potentials of the Tersoff[16] or Brenner[17,18] type.

3 Results

We are interested in calculating the lifetime of vibrational stretching modes of adsorbates typically lying far above the phonon continuum of the substrate bulk material. The dominant decay channel of the large quantum of the stretching vibration will be the decay into a few other energetically high-lying quanta, notably the hindered translational or rotational modes of the adsorbate. Due to overall energy conservation, the remainder of the vibrational energy *not* accommodated in local modes must be transferred to a substrate phonon, or likewise a substrate phonon (always present at finite temperature) may be absorbed to allow for the transition from the stretching mode into a slightly higher combination of other local modes. In the regime where a perturbative treatment is applicable, one would tend to believe that processes involving as few quanta as possible dominate the rate; however, a detailed treatment shows that the availability of suitable phonon modes and the size of transition matrix elements are important factors, too.

3.1 Damping of CO Stretching on Si(100)

We calculate the harmonic frequencies of CO on Si(100) by diagonalizing the Hessian matrix. We find that the hindered translations parallel to the surface correspond to low-lying modes (< 50 cm^{-1}). Since their quanta are very small, these modes are not relevant for the relaxation of the CO stretch vibration. The higher-lying modes have been classified according to the magnitude of the displacement at the C and O atoms. The four important vibrational modes of CO identified in this way are illustrated in Figure 1a. The C-O stretching mode is found at 2006 cm^{-1}, and the C-Si stretching (=shift) mode is at 507 cm^{-1}. Bending modes are in the plane of the Si dimer, or out-of-plane, with harmonic frequencies of 429 and 441 cm^{-1}, respectively. We use displacements $x = \sum_j \alpha_j u_j$ obtained from linear combinations of the eigenvectors u_j of the dynamical matrix to map out the PES. From the PES, the local potential $v(q)$ is determined in DFT accuracy. In Figure 1b, selected cuts through the multi-dimensional potential $v(q)$ are plotted. Deviations of the contour lines from perfect ellipses indicate an anharmonic coupling between the corresponding modes.

The system Hamiltonian $\hat{H}_{vib} = -\nabla^2 + v(q)$ is constructed for low quantum numbers and diagonalised. From the eigenvalues of the system Hamiltonian, the anharmonic frequencies of the stretching and the shift mode are obtained as 1978 cm^{-1} and 494 cm^{-1}, respectively. Thus, the stretching frequency obtained with the DFT-PBE functional is about 100 cm^{-1} smaller, and the shift mode frequency is 80 cm^{-1} larger than the corresponding

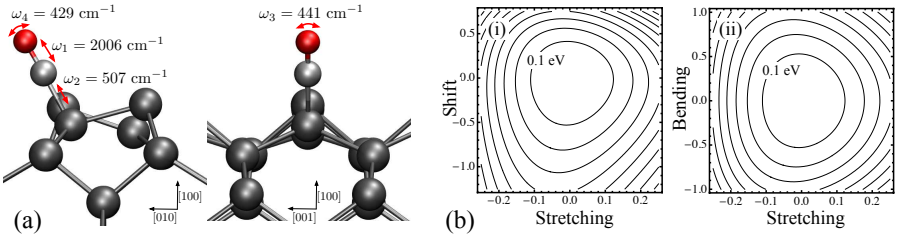


Figure 1. (a) Illustration of four selected vibrational modes of CO on $c(4 \times 2)$Si(100) with a CO coverage of 0.25. (b) Selected two-dimensional cuts through the potential energy surface of CO on Si(100) for the vibrational modes. The contour spacing is 0.1 eV.

experimental values. From the DFT-PBE calculations, the most favourable decay channels yield three shift/bending quanta plus one optical phonon. The lifetime τ_{vib} of the CO stretch mode is calculated to be 0.55 ns at 0 K and 0.50 ns at 100 K. Although harmonic frequencies in DFT-PBE have limited accuracy, as mentioned above, the lifetime calculated in this way is already in qualitative agreement with the experiment, but about four times shorter than the measured value of 2.3 ns at 100 K[7,8]. Because the phonon energy for the dominant transition channels is large, induced transitions due to thermally excited phonons are unimportant, and the temperature dependence of the lifetime is found to be small.

To remedy the inaccuracy of DFT-PBE in calculating vibrational frequencies, we consider alternative schemes where the system Hamiltonian is set up using improved values for the vibrational frequencies through rescaling the PES[5]. The best results are associated with anharmonic stretching and shift frequencies of 2075 cm^{-1} and 411 cm^{-1}, yielding a lifetime of 2.40 ns at 100 K. Employing this adjustment, we find a dominant decay channel into three shift quanta and one in-plane bending quantum, plus a phonon of 53 meV. The transition time for this decay channel is 3.44 ns at 0 K. Around 82 % of all excited CO molecules decay via this channel.

3.2 Damping of H/D Stretching on Ge(100)

A Ge(100):H surface saturated in the monohydride phase shows $p(2 \times 1)$ symmetry as Ge dimers form parallel to the surface, and each Ge atom is occupied by one hydrogen atom. We have calculated harmonic vibrational eigenmodes and frequencies from a Hessian matrix that involves the hydrogen atoms and the four top-most Ge layers. The leading interaction couples the vibrations of hydrogen atoms on the left and right sides of the same Ge surface dimer v_{L-R}, as depicted in Figure 2a. As a result, the degeneracy is broken and symmetric and antisymmetric Ge-H stretching modes are created. For H, the asymmetric mode is found 13 cm^{-1} lower in energy than the symmetric mode. As mentioned in Section 3.1, to avoid the misrepresentation of phonon energies, we introduce an empirical adjustment of the potential energy surface that yields improved vibrational frequencies. In Table 1, adjusted vibrational frequencies of adsorbates for the dimers saturated using pure H, pure D and isotopic mixture of H and D are listed. Using the vibrational modes, the PESs for the H saturated, D saturated and isotope mixture saturated dimers are mapped out.

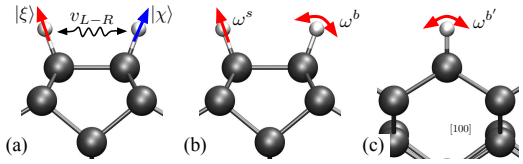


Figure 2. Illustration of vibrational excitations of adsorbates within a dimer on Ge(100). The weak coupling v_{L-R} between localized hydrogen stretch modes ($|\xi\rangle$, $|\chi\rangle$) is indicated in (a). Stretching (ω^s), in-plane bending (ω^b) and out-of-plane bending ($\omega^{b'}$) are depicted in (b) and (c).

Adsorbates		Vibrational frequencies (cm^{-1})					
H+H	Mode	ω_H^{s+}	ω_H^{s-}	ω_H^{b+}	ω_H^{b-}	$\omega_H^{b'+}$	$\omega_H^{b'-}$
	Adjusted PES	1991	1978	533	523	522	519
	Experiment	1992[19–21]	1978[19,20]	532[19]			
D+D	Mode	ω_D^{s+}	ω_D^{s-}	ω_D^{b+}	ω_D^{b-}	$\omega_D^{b'+}$	$\omega_D^{b'-}$
	Adjusted PES	1433	1423	384	377	376	374
	Experiment	1432[19,21]	1422[19,21]	388[19]			
H+D	Mode	ω_H^{s}	ω_H^{b}	$\omega_H^{b'}$	ω_D^{s}	ω_D^{b}	$\omega_D^{b'}$
	Adjusted PES	1986	526	523	1431	379	377
	Experiment	1987[9]			1433[9]		

Table 1. The vibrational frequencies of H(D)-Ge on the saturated Ge(100) surface from the calculated PES after frequency adjustment. With X={H, D}, ω_X^s denotes the X-Ge stretching mode and ω_X^b and $\omega_X^{b'}$ represent in-plane and out-of-plane bending modes, respectively. $+$ and $-$ signs represent symmetric and antisymmetric modes.

The vibrational excitations calculated by diagonalising the system Hamiltonian are demonstrated in Figure 3a. As shown in the mid pannel of Figure 3a, the vibrational energies of D are $\sqrt{2}$ times smaller than of H excitations (upper pannel), because of mass scaling. Interestingly, we find that the Ge-H stretching mode ω_H^s (1986 cm^{-1}) in a heterogeneously saturated dimer has a different relaxation scheme than in the homogeneously saturated cases, as depicted in the lower panel of Figure 3a. The combination of Ge-H and Ge-D vibrations results in a multitude of eigenstates scattered on the energy scale. In conclusion, we find that the relaxation scheme of vibrational energy is substantially altered with replacing an adsorbate by another isotope.

In the relaxation of the symmetric hydrogen stretch (ω_H^{s+}) on a purely H saturated dimer, the largest contribution comes from the transition to the state forming the lower shoulder of the $4\omega_H^b$ peak, which requires absorption of a (thermally excited) phonon of 122 cm^{-1}. This phonon belongs to the Ge dimer rocking modes (113 cm^{-1})[22], thus it couples to the adsorbate vibrations strongly. We note that the transition by absorbing a rocking mode phonon is energetically possible only because individual states forming the $4\omega_H^b$ peak are split over some energy range. This splitting is due to the coupling between the H bending modes at the left and right side of the Ge dimer, resulting in energetically different symmetric and antisymmetric modes. Alternative transitions $\Gamma^{(2)\downarrow}$ from the stretch mode into $3\omega_H^b$ are found to make only minor contributions to the overall rate, because *two* phonons need to be created in this process. The calculations show that the lifetime of

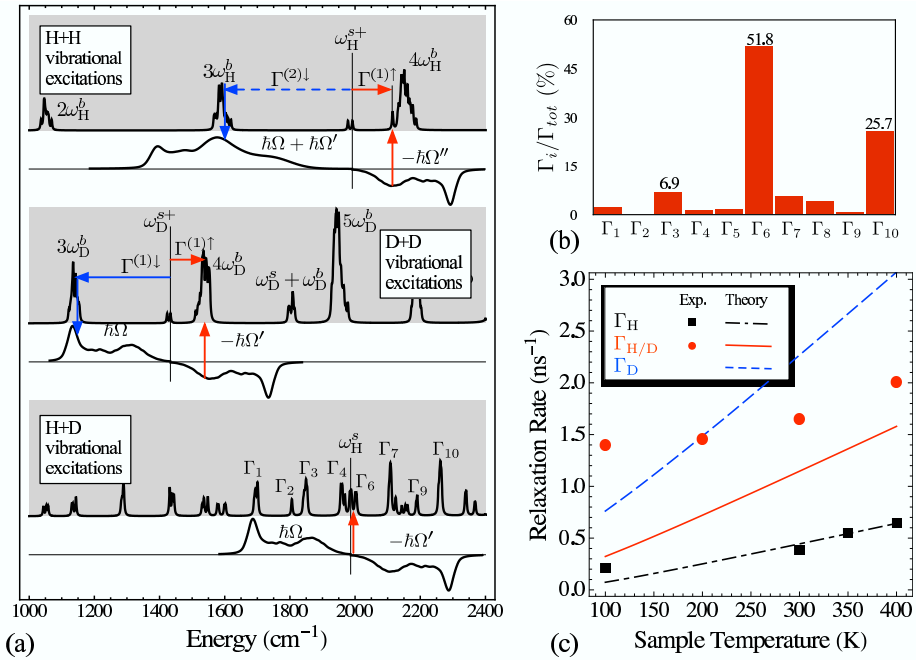


Figure 3. (a) Calculated relaxation scheme of the hydrogen stretch on Ge(100). (b) Contributions of each relaxation channel of the ω^s_{H} mode. (c) The relaxation rates for the decay of the hydrogen stretching (Γ_{H} for the relaxation of the ω^{s+}_{H} mode of pure H saturated dimer and $\Gamma_{\mathrm{H/D}}$ for the relaxation of the ω^s_{H} mode of the heterogeneously saturated dimer with H and D) and for the decay of the deuterium stretch mode ω^{s+}_{D} (Γ_{D}) on Ge(100).

the ω^{s+}_{H} mode is 1.58 ns at 400 K, in excellent agreement with the experiments yielding 1.56 ns[9].

In the case of the symmetric deuterium stretch (ω^{s+}_{D}), the transition to the $4\omega^b_{\mathrm{D}}$ states ($\Gamma^{(1)\uparrow}$ in the mid panel of Figure 3a) requires absorption of phonons in the range from 77 to 121 cm^{-1}. Hence, the energy is smaller than in the case of the hydrogen stretch. Thus, there are several transitions to $4\omega^b_{\mathrm{D}}$ states for which the remaining energy difference can be matched by adsorbing a rocking mode phonon. These transitions are highly efficient. Mostly for this reason, the transition rate for D is higher than for H, leading to a shorter lifetime for D compared to H. The alternative transitions to $3\omega^b_{\mathrm{D}}$ states, thereby emitting an optical phonon in the range from 279 to 312 cm^{-1}, make only minor contributions of around 25 % of the total rate at 300 K. The calculated lifetime of the D stretch vibration at 300 K is 0.43 ns. Thus, the relaxation at room temperature is predicted to be four times faster for D as compared to H.

When the dimer is heterogeneously saturated it is clear that several types of intermediate states for the ω^s_{H} decay are available that can be reached by various phonons. As the main relaxation channel of the ω^s_{H} mode we identify a group of transitions (labeled Γ_6) that involves absorption of a small acoustic phonon with an energy of about 7 cm^{-1}. Despite the low density of states of low-frequency acoustic phonons, the transitions Γ_6

dominate due to their large matrix elements of $\langle m|\hat{H}_\lambda|n\rangle$ (see Figure 3b). The major part of the energy of the $\omega_{\mathrm{H}}^{\mathrm{s}}$ mode ends up in the bending modes of the light hydrogen isotope, while the remainder is transferred to deuterium bending quanta. All relaxation channels together lead to a lifetime of the H stretching vibration of 0.63 ns in the heterogeneously saturated Ge dimers, a factor of 2.5 shorter than for pure H coverage at an assumed sample temperature of 400 K. Such a strong effect of the chemical environment on the lifetime of adsorbate vibration may seem surprising, but is actually corroborated by the experimental findings.

Figure 3c compares the relaxation rates obtained from theory and experiment[6] for both the pure H and the isotope-mixed adsorbates. For homogeneously saturated Ge(100):H, quantitative agreement between theory and experiment over a wide range of temperatures could be reached. Moreover, in both theory and experiments, the vibrational relaxation rate of the hydrogen stretch is considerably enhanced for heterogeneously saturated Ge(100) compared to pure Ge(100):H. The calculations enable us to identify the change in the main relaxation channel, as explained above, as the reason for this strong sensitivity of the relaxation rate to the chemical environment. However, theory and experiment find a different temperature dependence of the relaxation rate for the heterogeneously saturated dimer. One possible reason for the discrepancy could be the difficulty to accurately predict the small energy of the acoustic phonon involved in the dominant transition Γ_6.

4 Conclusions

We presented a theoretical description of the vibrational energy relaxation of adsorbates on semiconductor surfaces. This theory, based on first-principles calculations of coupling matrix elements, while allowing for an empirical fine-tuning of vibrational frequencies, quantitatively accounts for the observed lifetimes as obtained from pump-probe experiments. The theoretical analysis shows that vibrational lifetimes may depend in a subtle way on the quantized intermediate states involved in the vibrational relaxation, and on the number as well as the energy of the phonons required to reach them. With the help of our calculations, we can also trace back the remarkable finding of strongly different lifetimes in different chemical environments to the different intermediate states in the relaxation which become accessible due to symmetry breaking by mixed-isotope adsorption.

Acknowledgements

The authors acknowledge a grant of computer time by the John von Neumann Institute for Computing (NIC) in Jülich, Germany and financial support by the German Science Foundation (DFG) through the SFB 616 “Energy Dissipation at Surfaces”.

References

1. E. T. Foley, A. F. Kam, J. W. Lyding, and Ph. Avouris, *Cryogenic UHV-STM Study of Hydrogen and Deuterium Desorption from Si(100)*, Phys. Rev. Lett. **80**, 1336, 1997.
2. J. C. Tully, *Mode-selective control of surface reactions*, Science, **312** 1004, 2006.

3. R. J. Cannara, M. J. Brukman, K. Cimatu, A. V. Sumant, S. Baldelli, and R. W. Carpick, *Nanoscale friction varied by isotopic shifting of surface vibrational frequencies*, Science, **318** 780, 2007.
4. J. W. Lyding, K. Hess, and I. C. Kizilyalli, *Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing*, Appl. Phys. Lett. **68**, 2526, 1996.
5. S. Sakong, P. Kratzer, X. Han, K. Laß, O. Weingart, and E. Hasselbrink, *Density-functional theory study of vibrational relaxation of CO stretching excitation on Si(100)*, J. Chem. Phys. **129**, 174702, 2008.
6. S. Sakong, P. Kratzer, X. Han, T. Balgar, and E. Hasselbrink, *Isotope effects in the vibrational lifetime of hydrogen on germanium(100): Theory and experiment*, J. Chem. Phys. **131**, 124502, 2009.
7. K. Laß, X. Han, and E. Hasselbrink, *The surprisingly short vibrational lifetime of the internal stretch of CO adsorbed on Si(100)*, J. Chem. Phys. **123**, 051102, 2005.
8. K. Laß, X. Han, and E. Hasselbrink, *Dynamics of the C-O stretch mode on Si(100)*, Surf. Sci. **600**, 4275, 2006.
9. X. Han, T. Balgar, and E. Hasselbrink, *Vibrational dynamics of hydrogen on Ge surfaces*, J. Chem. Phys. **130**, 134701, 2009.
10. I. Andrianov and P. Saalfrank, *Theoretical study of vibration-phonon coupling of H adsorbed on a Si(100) surface*, J. Chem. Phys. **124**, 034710, 2006.
11. B. N. J. Persson and Ph. Avouris, *Local bond breaking via STM-induced excitations: the role of temperature*, Surf. Sci. **390**, 45, 1997.
12. G. Kresse and J. Furthmüller, *Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set*, Phys. Rev. B **54**, 11169, 1996.
13. J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized gradient approximation made simple*, Phys. Rev. Lett. **77**, 3865, 1996.
14. P. E. Blöchl, *Projector augmented-wave method*, Phys. Rev. B **50**, 17953, 1994.
15. J. D. Gale and A. L. Rohl, *The General Utility Lattice Program*, Mol. Simul. **29**, 291, 2003.
16. J. Tersoff, *Modeling solid-state chemistry: Interatomic potentials for multicomponent systems*, Phys. Rev. B **39**, 5566, Mar 1989.
17. D. W. Brenner, *Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films*, Phys. Rev. B **42**, 9458, 1990.
18. A. J. Dyson and P. V. Smith, *Extension of the Brenner empirical interatomic potential to C-Si-H systems*, Surf. Sci. **355**, 140, 1996.
19. L. Papagno, X. Y. Shen, J. Anderson, G. Schirripa Spagnolo, and G. J. Lapeyre, *Hydrogen adsorption on Ge(100) studied by high-resolution energy-loss spectroscopy*, Phys. Rev. B **34**, 7188, 1986.
20. Y. J. Chabal, *High-resolution infrared spectroscopy of adsorbates on semiconductor surfaces: Hydrogen on Si(100) and Ge(100)*, Surf. Sci. **168**, 594, 1986.
21. J.-K. Wang, C.-S. Tsai, C.-E. Lin, and J.-C. Lin, *Vibrational dephasing dynamics at hydrogenated and deuterated semiconductor surfaces: Symmetry analysis*, J. Chem. Phys. **113**, 5041, 2000.
22. W. Stigler, P. Pavone, and J. Fritsch, *Structural and dynamical properties of Ge(001) in comparison with Si(001) and C(001)*, Phys. Rev. B **58**, 13686, 1998.

Determination of Structure and Electronic Properties of Free, Supported and Ligand Protected Metal Clusters by Density Functional Theory

Michael Walter[1,2] **and Michael Moseler**[1,2,3]

[1] Modelling and Simulation of functional Nanosystems, Physics Department
University of Freiburg, Hermann-Herder-Str. 3, 79106 Freiburg, Germany
E-mail: Michael.Walter@fmf.uni-freiburg.de

[2] Freiburg Materials Research Center, Stefan-Meier-Str. 21, 79104 Freiburg, Germany

[3] Fraunhofer Institute for Mechanics of Materials, Wöhlerstrasse 11, 79108 Freiburg, Germany
E-mail: mos@iwm.fhg.de

Large scale density functional calculations are employed to determine the nuclear and electronic structure of I) sodium cluster anions Na_n^- in the size range $n = 20 - 40$, II) the nonamer palladium cluster on a thin magnesia film supported by silver: Pd_9/MgO(001)/Ag(001) and III) the doped thiol protected gold tetracosamer cluster: $XAu_{24}(SR)_{18}$ with X=Pd, Ag and Cd. Photoelectron spectra of gasphase clusters are used as structural and electronic fingerprints to verify that in I) the correct structures are found. As support and prediction for future experiments STM images of the supported palladium cluster and the electronic density of states of the ligand protected gold-alloy cluster are analyzed.

1 Introduction

Properties of clusters and nanoparticles can differ substantially from the corresponding bulk material. This makes them very attractive objects both for applications and for fundamental research. An indispensable prerequisite for an understanding of their properties is the knowledge of their exact geometric structure. Therefore, in the recent years substantial efforts have been invested into the structure determination of clusters[1]. Meanwhile, gasphase cluster structures are known for many materials, although in most cases this knowledge is restricted to small clusters with up to roughly 20 atoms or for larger particles with several hundred atoms[1]. The situation is less fortunate for clusters in contact with supporting surfaces or with protecting ligand shells. Here, little by little structural knowledge has been collected over the past years.

This article reports the structure determination for three different cluster systems varying in the respective cluster material (Na, Pd, Au) and in the environment (vacuum, supporting surface, passivating ligand shell). It will be demonstrated that large scale gradient corrected density functional theory (DFT) calculations on the Jülich massively parallel platforms can be successfully used to infer the structure of metallic clusters of various complexity ranging from gasphase clusters of simple metals to transition metal clusters in contact with other materials.

Small metal clusters exhibit unique size-dependent properties due to their large fraction of surface atoms as well as their distinct electronic structure[2]. Reducing the atom count below a few hundred, the bulk band structure breaks down and discrete energy levels separated by gaps emerge. These quantization effects are in parts responsible for the high catalytic activity of small metal particles as well as for a variety of optical and electronic

phenomena. In spite of this, electron quantization is difficult to access experimentally. So far, the most comprehensive picture has been deduced from photo electron spectroscopy on size-selected clusters in the gas-phase[3]. In section 3, we present results based on photoelectron spectra (PES) of sodium clusters Na_n^- for an intermediate size range n =20-40. This study completes our previous investigations of the Na cluster anions in the size range from 4 to 19 atoms[4] and from 39 to 350 atoms[5] making gasphase sodium clusters the first system for which structural information is available from the tetramer to the nanoparticle size regime.

Such PES experiments cannot account for substrate effects on the cluster electronic structure, which might, however, decisively influence the properties of particle-support systems. The scanning tunnelling microscope (STM) provides a powerful alternative to study the interplay between geometric and electronic properties of metal clusters at the local scale. Magnesia supported palladium clusters are amongst the most popular model catalysts for the investigation of prototypical catalytic reactions, such as the oxidation of carbon monoxide[6]. Determining their structure is an important preliminary for an understanding of their catalytic activity. In section 4, we predict experimental STM images of a Pd_9 cluster adsorbed on a thin MgO(001) film supported by Ag(001).

The considerable scientific and technological interest in Ligand protected (LP) metal clusters is driven by the fact that these clusters can be synthesized in macroscopic quantities as well as by their pronounced stability in different environments and over a broad temperature range. Prominent representatives are thiol or phosphine protected gold clusters. Recently, it has been suggested to substitute some of the atoms in the LP gold cluster by more reactive transition metal atoms[7]. This could be a versatile route to tailor their catalytic, magnetic, electronic or optical properties. However, for the time being it is not evident at all, that LP gold alloy clusters obey the same stabilization mechanism and exhibit the same structure as pure Au clusters. In section 5, we report an explorative DFT study of a thiol protected gold alloy cluster[8].

2 Methods

In order to find the structural ground states of the sodium cluster anions in section 3, density functional theory in the framework of the gradient corrected[9] local spin density approximation for the description of the electronic ground state Born-Oppenheimer (BO) surface is employed. The Kohn-Sham equations are solved for the sodium $3s$ valence electrons using the BO-molecular dynamics method[10] with a plane wave basis (10 Ry cutoff) and nonlocal pseudopotentials[11]. The BO surfaces of the clusters are scanned for the global minimum employing a modified single-parent genetic algorithm[12]. For the structures with the lowest energy, photoelectron spectra were then simulated by shifting the calculated elctronic density of states (DOS) so that the highest molecular orbital (HOMO) eigenvalue coincides with the calculated vertical detachment energy. For the determination of delocalized electron angular quantum numbers the overlap of the Kohn-Sham orbitals with cluster centred spherical harmonics has been calculated.

The other DFT calculations are performed with the grid-projector-augmented-wave method (GPAW)[13,14] using the same GGA as for the Na clusters[9]. A 0.2 Å grid-spacing was used for the representation of the smooth wave-functions. The STM images are simulated with the Tersoff-Hamann formalism[15].

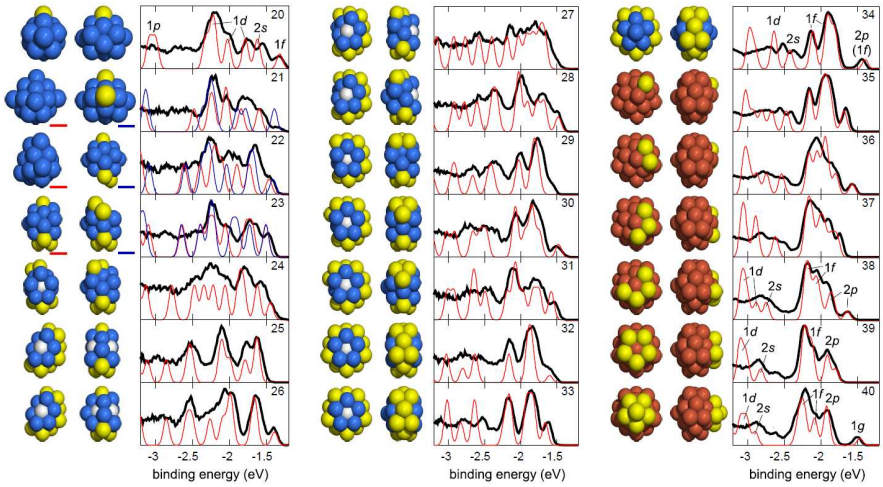


Figure 1. Ground state structures of Na_{20-40} and the comparison of the corresponding calculated DOS (red line) and the measured PES (thick black line). Two perpendicular views of the clusters are provided except for Na_{21-23} where two quasi degenerate isomers are depicted. For these three sizes, the DOS of the second lowest isomer is plotted with a blue line. Atoms added to the double icosahedron basic unit (outer atoms: blue, central atom: white) are coloured in yellow. For Na_{35-40}, atoms of the Na_{34}-core are coloured in red.

3 Photoelectron Spectra of Gasphase Sodium Cluster Anions

In two recent publications we have presented a combined experimental and theoretical study of the structure of sodium cluster anions Na_n^- in the size range $n = 4-19$ atoms[4] and $n = 39 - 350$ atoms[5]. Here we show results for the intermediate range $n = 20 - 40$. Experimental photoelectron spectra are displayed in Figure 1. Like the spectra of the smaller or larger clusters published earlier[4,5] they are strongly structured, indicating a highly discretized DOS. An electron shell structure is directly visible for some sizes. Within a model of delocalized electrons in a spherical container shell closures occur at electron numbers 20, 34, 40 and 58. This explains the appearance of a new shell with an extra single electron at the clusters with an additional electron, that is at Na_{20}^- , Na_{34}^- , Na_{40}^- and Na_{58}^-. For these and neighbouring sizes the angular momentum of the delocalized electrons have been assigned to the peaks visible in the calculated DOS. The geometries of the lowest energy structures from DFT are indicated in Figure 1. The DOS of the structures is shown as red and blue lines in the plots of the photoelectron spectra. As the DOS is very sensitive to the nuclear structure of the cluster, a good agreement between the DOS of the lowest energy isomer and the experimental spectrum gives strong evidence that the correct structure has been found.

For almost all of the clusters shown in Figure 1 the basic building block is a 19-atom double icosahedron, with the remaining atoms forming an equatorial belt around the waist of this core. The core double icosahedron does not necessarily stay intact in this build-up sequence; in many cases one or two of its tip atoms (the end atoms of its central linear tetramer) are transferred to the equatorial ring, which optimizes the average atom coor-

dination number and changes the overall cluster sphericity. For the same reason in some cases an atom out of one or two of the outer five-membered rings of the core is transferred to the waist position.

This growth pattern starts at size 20. While for the electronically closed shell Na_{19}^- the prolate double icosahedron structure is energetically unfavoured[4], the lowest energy structure found for Na_{20}^- is exactly this double icosahedron with one capping atom attached to its side. For Na_{21}^- our calculations revealed two quasi degenerate ground state structures with an energy difference of only $\Delta E = 20$ meV (left and right structure in Figure 1). The experimental spectrum can be well reproduced by a weighted superposition of the DOS of both isomers. The situation is similar for Na_{22}^- and Na_{23}^-. One isomer of Na_{22}^- is a triply capped double icosahedron, while the other one is irregular. For Na_{23}^- both isomers are capped double icosahedra; in the lower energy structure the four additional atoms and one of the tip atoms of the 19 atom core are placed onto the waist, while for the second isomer two core atoms (the tip atom and an adjoining atom from a five-membered ring) are transferred to the waist. Note that the resulting electronic densities of states exhibit significant differences, although the two structures differ only by the displacement of two atoms. As the DOS of the second isomer exhibits a peak at 1.9 eV, where the experimental spectrum shows a minimum, one can directly conclude that in spite of the calculated quasi-degeneracy the relative abundance of the second isomer must be small. For most of the following cluster sizes up to Na_{40}^- the good to very good agreement between the experimental and theoretical spectra provides strong evidence for the correct determination of the ground state structures. The growth scheme is very simple: the structure of a cluster Na_n^- in most cases is just the one of the previous cluster Na_{n-1}^- with one atom added. Only occasionally a rearrangement of a few other atoms occurs. A geometric shell closing is reached at size 34, which has D_{5h}-symmetry and consists of the 19 atom double icosahedron decorated by three five membered rings. More details can be found in Ref. 16.

4 STM Images of Palladium Clusters on Thin Silver Supported Magnesia Films

In this section, we present simulations of STM images of Pd_9 clusters on thin MgO films. The substrate with metal support is modeled by two MgO layers in registry with three Ag(001) planes of 4.14 Å lattice constant (Figure 2). Except for the two bottom Ag layers, all atoms are relaxed. Simulated STM images are shown in Figure 2. At negative bias (probing the occupied states), the brightest signal (i.e. the highest occurrence of the required density) comes from the region of the highest Pd atoms. This is what can naively be expected and in nice agreement with the result of a previous calculation without the supporting Ag(001)[17]. At positive bias, however, the long axis of the Pd_9 cluster and hence the highest atoms do not contribute as much as the side regions in the short cluster direction. Even more, there appears a "valley" in the constant electron density contour exactly on the axis of the highest atoms. This on a first glance rather counterintuitive result shows that (similar to the case of gold[18]) for supported Pd-clusters the STM images do not provide pictures of the nuclei, but map the properties of the electronic wave functions.

To get more insight in the reasons for the different behaviour of the electronic states below and above the Fermi energy, we examine the atom-projected local density of electronic states (LDOS), which is shown in Figure 2e. We concentrate here on the Pd derived

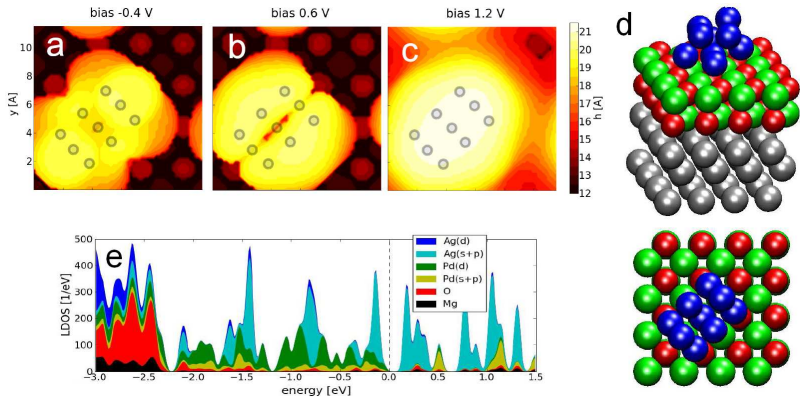


Figure 2. Simulated STM pictures for Pd_9. The circles indicate the positions of the Pd atoms. (panels a-c). d) structure of the cluster on the MgO(001)/Ag(001) support. e) shows the local density of electronic states (LDOS) relative to Fermi Energy (marked by the vertical dashed line).

states as these are the states responsible for the STM images. The states below the Fermi energy are derived from Pd(d) atomic orbitals mainly. The d-orbitals are local on the Pd atoms and hence the STM picture shows the atomic positions for negative bias. In contrast, we find the states above Fermi energy to be mainly derived from Pd(s) and Pd(p) atomic orbitals and therefore have a much more delocalized character. This character leads to the nodal structure on the longest and at the same time highest cluster axis. At bias 0.6 V, the state at 0.5 eV is displayed in the STM picture shown in Figure 2b. Increasing the bias to 1.2 V leads to a contribution of other delocalized states that obviously do not possess this node. Therefore the node is filled again as shown in Figure 2c.

5 Ligand Protected Gold Alloy Clusters

There has been a remarkable progress in the last years in the field of ligand protected (LP) clusters both experimentally as well as theoretically. Experimentally the crystal structure of two thiol-protected species that waited structural characterization for a long time was determined[19–21]. Theoretically it was demonstrated that the traditionally separated fields of thiol- and phosphine-protected gold clusters can be joined under a unified view and that the magic stability of these clusters can be explained by delocalized electron shell closings (DESC) suggesting a super-atom picture for the clusters[22]. All the magic clusters satisfy the simple counting rule

$$N_{\mathrm{del}} - M - q = n_s \ , \tag{1}$$

where N_{del} is the number of delocalized electrons, M the number of electron localizing ligands, q the cluster charge and n_s the DESC number (2, 8, 18, 34, ...)[3].

In this section we concentrate on replacing one Au atom in the well studied thiol protected gold cluster $(Au_{25}(SR)_{18})^-$. This compound was shown to be exceptionally stable and even survived an excess thiol environment. Only recently the structure of this cluster was elucidated[23]. There is experimental evidence that a similar cluster might also be stable

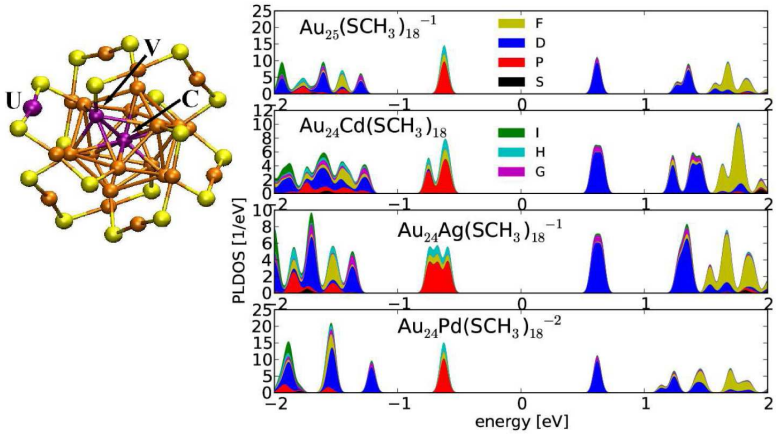


Figure 3. Structure and DOS of $(XAu_{24}(SCH_3)_{18})^{(q)}$ (X=Pd, Ag, Cd, $q = 1, 0, -1$). On the left: the structure of $Au_{25}(SR)_{18}$. The R groups are omitted for clarity. The purple atoms indicate the replacement positions for the substitutional metal atoms (X=Pd, Ag and Cd) resulting in the three different $XAu_{24}(SR)_{18}$ isomers **C**, **V** and **U**. Au atoms are drawn in orange or purple and S in yellow. On the right: the global Y_{lm} projected density of states (PDOS) for the core region of $Au_{25}(SCH_3)_{18}^{-}$, $Au_{24}Cd(SCH_3)_{18}$, $Au_{24}Ag(SCH_3)_{18}^{-}$ and $Au_{24}Pd(SCH_3)_{18}^{-2}$. The energy axis is defined relative to Fermi energy and the PLDOS is broadened by Gaussians of 0.03 eV width.

if one Au atom is substituted by a Pd[7]. In the following the stability of $XAu_{24}(SCH_3)_{18}$ with X=Pd, Ag and Cd will be investigated. The idea behind this study is that the substitutional atoms Pd, Ag and Cd contribute with 0, 1 and 2 delocalized electrons respectively, hence modifying the super-atom nature of the cluster substantially.

The structure of the $Au_{25}(SR)_{18}$ cluster is shown in Figure 3. It comprises a 13 Au atom core and 6 SR-Au-SR-Au-SR units. The Au_{13} core itself consists of a 12 atom cage around a central atom, where the cage atoms are all in nearly equivalent icosahedral positions, each of them connecting to a sulfur atom from the six units. The units are placed in nearly octahedral symmetry only slightly distorted by the asymmetric direction of the CH_3 ligand molecules (not shown). There are three distinct positions for a replacement of a single Au by another metal atom resulting in three different isomers (**C**, **V** and **U**). In **C** the central atom is replaced, in **V** one of the 12 vertex atoms and in **U** one of the Au atoms from the thiol-gold units.

According to this scheme, we have replaced one gold atom of the icosahedral $Au_{25}(SCH_3)_{18}$ and relaxed the $(XAu_{24}(SCH_3)_{18})^{(q)}$ (X=Pd, Ag, Cd, $q = 1, 0, -1$) to the next local minimum. The lowest energy isomer (ground state, GS) is found to be independent of q in all cases and is well separated from the higher lying isomers. In the case of Cd and Ag, the GS is **V** whereas the Pd atom prefers isomer **C** as GS.

Significant HOMO-LUMO gaps of 1.23 eV, 1.18 eV and 1.18 eV are found for X=Pd, Ag, Cd in the cluster charge states $q = -2, -1, 0$ respectively. These values compare very well to the 1.2 eV gap of $Au_{25}(SCH_3)_{18}^{-}$[23]. To study the electronic properties, we will concentrate on the GS and the charge states with the largest HOMO-LUMO gap, i.e. isomers **C**, **V**, **V** and charge states $q = -2, -1, 0$ for X=Pd, Ag and Cd respectively, in the following.

The angular momentum projected DOS of all three compounds (Figure 3) resemble the rare-gas super-atom picture known from $Au_{25}(SCH_3)^-_{18}$ also shown in the figure. The LUMO is of D and the HOMO of P symmetry showing again a DESC with $n_s = 8$ and explaining the large HOMO-LUMO gaps. The splitting seen in the D shell is a consequence of the octahedral symmetry of the $SR(AuSR)_2$ units around the cage. In the case of $Au_{25}(SCH_3)^-_{18}$ or X=Pd, where the Pd atom is located in the cluster centre, the P-shell shows a single peak only. This is not the case anymore for X=Ag and X=Cd since here the alloy atom is located on a vertex leading to symmetry breaking and hence to the small splitting of the P-shell.

The DESC for the different charge states shown in Figure 3 can be understood by a simple s-electron counting as given in Eq. (1): Cd donates two, Ag one and Pd no electron to the number of delocalized electrons in the cluster, hence $N_{\mathrm{del}} = 12, 13, 14$ in the case of X=Pd, Ag, Cd respectively. All clusters contain $M = 6$ electron localizing units and therefore a charge of $q = -2, -1, 0$ is needed to fulfil the 8 electron shell closing for Pd, Ag, Cd respectively.

6 Conclusions

In conclusion, we have shown for the sodium cluster anions in section 3 that DFT calculations in combination with state of the art experimental results are the ideal combination to obtain a deeper understanding of the structure of gasphase clusters. For the supported Pd nonamer cluster in section 4 and the the thiol protected doped Au tetracosamer cluster in section 5 we have trusted in the predictive power of DFT simulations in order to suggest possible future experiments and explain their possible outcome.

Acknowledgements

The three examples presented in this report demonstrate the usefulness of large scale quantum chemical calculations. However, it must be stressed that the availability of computational resources represents an important prerequisite to exploit the power of tools such as density functional theory. Therefore, the authors thank FZ Jülich and RZ Karlsruhe for providing computational resources. Funding from the Deutsche Forschungsgemeinschaft under the SPP 1153 is acknowledged.

References

1. F. Baletto and R. Ferrando, *Structural properties of nanoclusters: Energetic, thermodynamic, and kinetic effects*, Rev. Mod. Phys., **77**, 371, 2005.
2. W. Ekardt, (Ed.), *Metal clusters*, Chichester, Wiley, 1999.
3. Walt A. de Heer, *The physics of simple metal clusters: experimental aspects and simple models*, Rev. Mod. Phys., **65**, 611–676, 1993.
4. M. Moseler, B. Huber, H. Häkkinen, U. Landman, G. Wrigge, M. Astruc Hoffmann, and B. v. Issendorff, *Thermal effects in the photoelectron spectra of Na^-_N clusters $(N = 4 - 19)$*, Phys. Rev. B, **68**, 165413, Oct 2003.

5. O. Kostko, B. Huber, M. Moseler, and B. v. Issendorff, *Structure determination of medium-sized sodium clusters*, Phys. Rev. Lett., **98**, 043401, 2007.
6. U. Heiz and W.-D. Schneider, *Review article: Nanoassembled model catalysts*, J. Phys. D, **33**, 85, June 2000.
7. C. A. Fields-Zinna, M. C. Crowe, A. Dass, and R. W. Murray, submitted for publication 11/2008.
8. M. Walter and M. Moseler, *Ligand protected gold alloy clusters: doping the superatom*, J. Phys. Chem. C, **113**, 15834, 2009.
9. John P. Perdew, Kieron Burke, and Matthias Ernzerhof, *Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett., **77**, 3865–3868, 1996.
10. R. N. Barnett and U. Landman, *Born-Oppenheimer molecular-dynamics simulations of finite systems: Structure and dynamics of $(H_2O)_2$*, Phys. Rev. B, **48**, 2081–2097, 1993.
11. N. Troullier and José Luriaas Martins, *Efficient pseudopotentials for plane-wave calculations*, Phys. Rev. B, **43**, 1993–2006, 1991.
12. Ionel Rata, Alexandre A. Shvartsburg, Mihai Horoi, Thomas Frauenheim, K. W. Michael Siu, and Koblar A. Jackson, *Single-Parent Evolution Algorithm and the Optimization of Si Clusters*, Phys. Rev. Lett., **85**, 546–549, Jul 2000.
13. P. E. Blöchl, *Projector augmented-wave method*, Phys. Rev. B, **50**, 17953–17979, Dec 1994.
14. J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, *Real-space grid implementation of the projector augmented wave method*, Phys. Rev. B, **71**, 035109, 2005.
15. J. Tersoff and D. R. Hamann, *Theory and Application for the Scanning Tunneling Microscope*, Phys. Rev. Lett., **50**, 1998–2001, Jun 1983.
16. Bernd Huber, Michael Moseler, Oleg Kostko, and Bernd v. Issendorff, *Structural evolution of the sodium cluster anions Na_{20}^- - Na_{57}^-*, submitted 2009.
17. Bernd Huber, Pekka Koskinen, Hannu Häkkinen, and Michael Moseler, *Oxidation of magnesia supported Pd clusters*, Nature Mat., **5**, 45–47, 2006.
18. Michael Walter, Pentti Frondelius, Karoliina Honkala, and Hannu Häkkinen, *Electronic Structure of MgO-Supported Au Clusters: Quantum Dots Probed by Scanning Tunneling Microscopy*, Phys. Rev. Lett., **99**, 096102, 2007.
19. Pablo D. Jadzinsky, Guillermo Calero, Christopher J. Ackerson, David A. Bushnell, and Roger D. Kornberg, *Structure of a Thiol Monolayer-Protected Gold Nanoparticle at 1.1 ÅResolution*, Science, **318**, 430 – 433, 2007.
20. M. W. Heaven, A. Dass, P. S. White, K. M. Holt, and R. W. Murray, *Crystal Structure of the Gold Nanoparticle [N(C8H17)4][Au25(SCH2CH2Ph)18]*, J. Am. Chem. Soc., **130**, 3754–3755, 2008.
21. Manzhou Zhu, Christine M. Aikens, Frederick J. Hollander, George C. Schatz, and Rongchao Jin, *Correlating the Crystal Structure of A Thiol-Protected Au_{25} Cluster and Optical Properties*, J. Am. Chem. Soc., **130**, 5883–5885, 2008.
22. Michael Walter, Jaakko Akola, Olga Lopez-Acevedo, Pablo D. Jadzinsky, Guillermo Calero, Christopher J. Ackerson, Robert L. Whetten, Henrik Grönbeck, and Hannu Häkkinen, *A unified view of ligand-protected gold clusters as superatom complexes*, Proc. Nat. Acad. Sc., **105**, 9157–9162, 2008.
23. J. Akola, M. Walter, R. L. Whetten, H. Häkkinen, and H. Grönbeck, *On the Structure of Thiolate-Protected Au_{25}*, J. Am. Chem. Soc., **130**, 3756–3757, 2008.

Magnetic Properties of Co and Ni Multilayers on Diamond Surfaces

Bernd Stärk, Peter Krüger, and Johannes Pollmann

Institut für Festkörpertheorie, Westfälische Wilhelms-Universität Münster
48149 Münster, Germany
E-mail: {*bernd.staerk, kruger, pollman*}*@uni-muenster.de*

We present *ab initio* calculations of structural, electronic and magnetic properties of particular metal-semiconductor hybrid systems. They consist of Co multilayers on (111), as well as Co or Ni multilayers on (001) diamond surfaces. These systems are almost ideally lattice-matched. Our calculations for coverages of one up to six Co monolayers show that strong Co−C bonds are established across the interface giving rise to a large quenching of the magnetic moment at the Co interface layer. Yet, the spin polarisation at the Fermi level remains significantly large amounting to about $50\,\%$. This is a prerequisite for spintronic applications. In contrast, Ni adlayers on C(001) exhibit only a very small spin polarisation at the interface.

1 Introduction

Magnetism is a fascinating phenomenon that plays an important role in technological applications, as well as in basic research. The formation of magnetism in matter is a collective quantum mechanical effect. Generally speaking, the spins of the electrons align with respect to each other thus minimizing the total energy of the system. In case of parallel spin alignment the system becomes ferromagnetic.

Interestingly, magnetic properties of matter can strongly be influenced by reduced dimensions. Based on this fact, drastic improvements in information storage devices have been achieved recently by utilizing artificially structured magnetic materials[1–3]. Further substantial developments in computer technology seem to be accessible exploiting the fact that the carriers of electric currents have beside their charge also the spin as an additional degree of freedom. One example of so called spintronic devices is the spin field-effect transistor as proposed by Datta and Das[4]. This device uses the ability of ferromagnetic metals to generate spin-polarized currents which are transferred into the gate via ballistic transport. The role of the gate is to generate an effective magnetic field causing the spin to precess. By changing the voltage at the gate one can manipulate the orientation of the spins which can be detected at the ferromagnetic drain[3].

The gate is usually made of a semiconducting material. A thorough understanding of metal-semiconductor hybrid systems is of paramount importance, therefore. Of the multitude of conceivable hybrid systems[3,5–7] only very few are suitable for the aforementioned purposes. As has been shown previously, a useful hybrid system is required to have a good structural compatibility because a large lattice mismatch leads to diffusive transport processes[3]. These quickly destroy the magnetization of the currents induced by the source. A further prerequisite of a useful hybrid system is an energetic separation of the spin-up and spin-down states close to the Fermi level. Here, we discuss specific hybrid systems, namely Co and Ni adlayers on diamond surfaces, which are tailored to these particular needs and appear to be very promising, therefore.

Nowadays, there are powerful computational methods allowing for theoretical *ab initio* investigations[8] of such systems on high-performance computers. Using grants of computer time at the John von Neumann Institute for Computing (NIC) we have performed such studies of structural, electronic and magnetic properties of these systems.

After a brief overview of the employed computational methods in Section 2 we present some of our results for varying numbers of Co adlayers in Section 3. We conclude our discussion with a brief summary in Section 4.

2 Calculational Method

Our calculations are carried out in the framework of spin-density functional theory employing the generalized gradient approximation[9] for exchange and correlation. We use separable pseudopotentials[10,11] and include non-linear core-corrections[12] for a proper representation of structural and magnetic properties. Surfaces are simulated within the standard supercell approach using slabs containing six C substrate layers and up to six Co adlayers. Neighbouring supercells are decoupled by a vacuum region of at least 12 Å. Structure optimizations are carried out employing Hellmann-Feynman, as well as Pulay forces.

The wave functions are expanded in a basis set of atom-centred Gaussian orbitals of s, p, d and s^* symmetry. Since the d-component of the Co and Ni pseudopotentials is strongly localized, rapidly decaying Gaussian orbitals as well as a fine real-space mesh with up to one million points for an appropriate description of the investigated systems are used.

Integrations over the Brillouin zone are performed using special k-point sets[13]. To describe the spin polarization and magnetization properly a precise sampling of the Fermi surface is mandatory. Here, we use a $20 \times 20 \times 1$ **k**-point grid together with a Gaussian broadening of the density of states by 60 meV. For every **k**-point a generalized eigenvalue problem has to be solved which is the most time consuming part of the calculations. Since the diagonalizations for different **k**-points are independent of each other this step can however be massively parallelized.

Further very elaborate parts of our method are the calculation of the potential matrix elements and the charge density due to the fine real-space mesh. These evaluations are performed with very efficient algorithms, as suggested by Wieferink et al.[14]. For the charge density the procedure consists in first setting up the density matrix and subsequently evaluating the Gaussian basis functions on the mesh. In both steps the necessary calculations are independent for each pair of atoms. Building up the potential matrix elements also breaks up into independent calculations for atom pairs. In all cases a very high level of parallelization has been achieved leading to a very high overall scale-up of the program.

3 Results

3.1 Structural Properties

Cobalt is a ferromagnetic metal crystallizing in hcp and fcc structure depending on temperature and pressure. Under normal conditions the hcp structure is the standard modification. The fcc modification, on the other hand, can be stabilized by growing $Co_{0.92}Fe_{0.08}$ alloys[15]

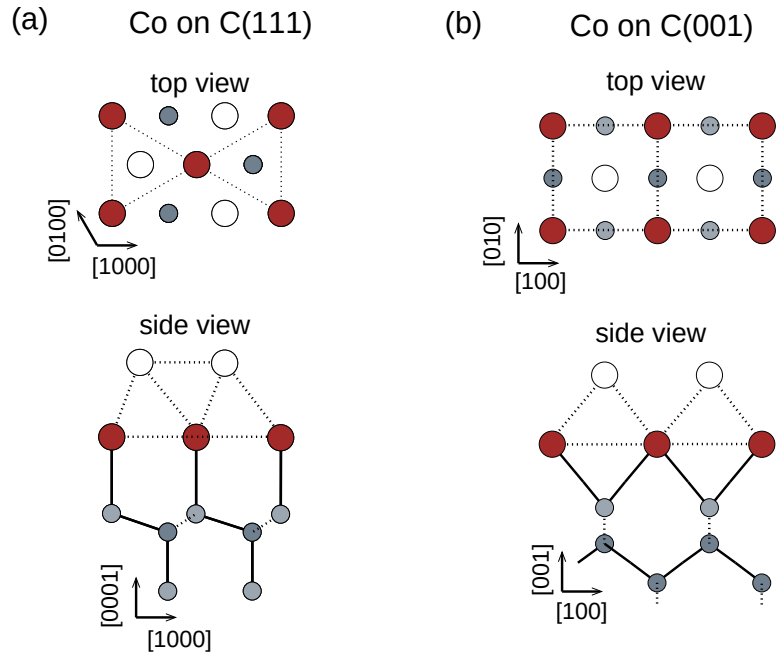


Figure 1. (a) Top and side view of one and two hcp Co adlayers on the C(111) surface. Large filled red circles mark Co atoms on the first and large open circles mark Co atoms on the second adlayer. Small blue circles indicate C substrate atoms. Full lines represent covalent sp^3 or pd bonds parallel to the drawing plane while dashed lines represent sp^3 bonds that form an angle with the drawing plane or metallic dd bonds. In the top view only dd bonds between Co atoms on the first adlayer are indicated by dashed lines. (b) Top and side view of one and two fcc Co adlayers on the C(001) surface.

and is furthermore the standard modification even at room temperature in Co nanoparticles[16].

The experimental lattice constants of Co are $a = 3.54\,\text{Å}$ for the fcc structure and $a = 2.51\,\text{Å}$ and $c = 4.07\,\text{Å}$ for the hcp structure. These lattice constants show an almost perfect matching with the surface lattice constants of $a_{\text{fcc}} = 3.57\,\text{Å}$ for C(001) and $a_{\text{hex}} = a_{\text{fcc}}/\sqrt{2} = 2.52\,\text{Å}$ for C(111). The lattice mismatch amounts to less than 1% in both cases. Therefore, the constituents of the metal-semiconductor hybrid system are highly compatible.

In the following, we discuss the structural properties of n Co layers adsorbed on a C(111) substrate surface, which will be referred to as n-Co:C(111). Starting with one Co adlayer, the energetically most favourable adsorption configuration is shown in Figure 1a. The Co adatoms are located above surface layer C atoms in *on top* positions. Thus, all substrate dangling bonds become saturated by the formation of one strong covalent Co−C bond per unit cell. This strong bond gives rise to a large binding energy of 1.9 eV per Co atom.

Since there is only a very small lattice mismatch between diamond and hcp cobalt, Co atoms of further adlayers adsorb in hcp bulk Co sites characteristic for a Co(0001) surface (large open circles in Figure 1(a)). The relaxed structures show some interesting trends in

the interlayer distances between neighbouring layers. First of all, the interlayer distances of the carbon layers show very little dependence on the number of Co adlayers. This implies a negligible impact of the Co layers on the structural properties of the substrate surface. In contrast, the Co layers in the slab show small contractions or expansions of their distances in an oscillatory manner, as compared to hcp bulk Co. This behaviour is in good agreement with the clean Co(0001) surface where similar oscillations occur. For thicker Co adlayers the interlayer distance in the middle of the adlayer converges toward the Co bulk interlayer distance, as one would expect. In total, we observe a fast and effective decoupling of all but the interface layer from the substrate.

The second kind of hybrid system investigated contains n layers of Co adsorbed on a C(001) substrate. For one Co adlayer the most stable adsorption site is a symmetric *bridge* site between neighbouring C surface atoms, as indicated in Figure 1b. In this case, the Co atoms reside in positions of an ideal diamond lattice on the first layer above the C(001) surface. They form two strong Co−C bonds which saturate all dangling bonds of the substrate surface. As a consequence, the binding energy of the Co adlayer is as large as 3.3 eV per unit cell.

Our structure optimization shows that Co atoms of further adlayers adsorb in fcc bulk Co sites (large open circles in Figure 1(b)). As in the case of Co on C(111), we observe an inward relaxation of the topmost Co adlayer together with an oscillatory behaviour of the interlayer spacings, as compared to the ideal fcc Co(001) surface. The general behaviour of the structural properties is very similar to that discussed above for Co adlayers on the C(111) surface.

A metal-diamond hybrid system with a small lattice mismatch can also be realized by replacing the Co atoms in the aforementioned structures by Ni atoms. Bulk Ni has a lattice constant a = 3.52 Å which is very close to that of fcc bulk Co. Our calculations show that the additional d electron per Ni atom has only a minor influence on the structural properties of nickel-diamond hybrid systems. The binding energy of a Ni adlayer is again very large with a value of 3.4 eV and the relaxations at the surface are comparable to those in the cobalt case.

3.2 Electronic and Magnetic Properties

The strong Co−C bonds at the interface, mentioned in the last section, have a large impact on the electronic structure for both systems. We start with a discussion of one Co adlayer on C(111). As noted above, the *on top* configuration of one Co monolayer on C(111) is characterized by the formation of one strong Co−C bond per unit cell. Figure 2 shows the spin-resolved band structures and densities of states (DOS) in the energy region near the Fermi level. We find seven bands in the projected band gap which result from the partially hybridized Co $4s$, Co $3d$ and C $2p$ states leading to a metallic surface. The spin-dependence of the bands is only very weak and does hardly change their dispersion. Based on a Mulliken analysis, we have marked the bands originating from both Co-$d_{3z^2-r^2}$ as well as C p_z states by thick lines in Figure 2. These orbitals give rise to the formation of an occupied bonding band B and a mostly unoccupied antibonding band B^*.

To further illustrate the character of these (anti-)bonding bands, Figure 3 shows charge density contours for the B and B^* states at the high-symmetry point K of the surface Brillouin zone. The bonding state B shows a large accumulation of charge density between

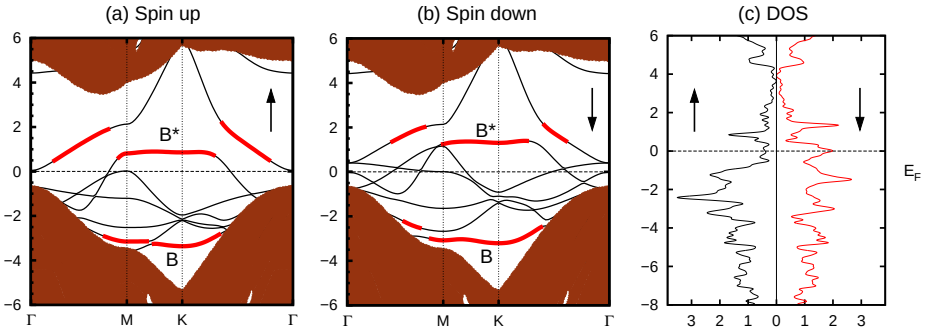


Figure 2. Band structures for (a) spin-up and (b) spin-down electrons and spin-resolved densities of states for one Co layer on the C(111) surface. The red shaded areas represent the projected band structure of bulk diamond. The band energies are referred to E_F as the zero energy.The band sections highlighted by thick lines originate from Co d states that have strong admixtures of C p states which are mainly localized at the atoms of the topmost C layer.

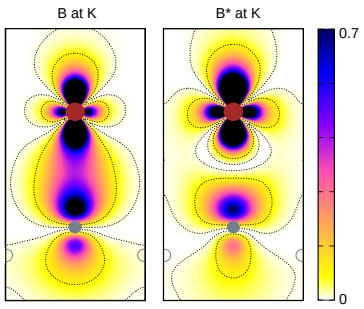


Figure 3. Charge density contours (in Å^{-3}) for the bonding (B) and antibonding (B*) state at the K point for the 1 Co:C (111) system. Large and small filled circles mark Co and C atoms, respectively, that are located in the drawing plane. The charge densities are plotted in the [1000]-[0001] plane containing the Co and C atom at the interface.

layer	1	2	3	4	5
μ_{Co6}					
μ_{Co5}					1.76
μ_{Co4}				1.78	1.60
μ_{Co3}			1.75	1.63	1.60
μ_{Co2}		1.76	1.51	1.56	1.53
μ_{Co1}	0.96	1.26	1.07	1.23	1.11
μ_{C1}	-0.09	-0.10	-0.13	-0.13	-0.13
μ_{C2}	0.02	0.03	0.03	0.03	0.03
ζ_{Co1}	-69	-51	-64	-50	-64
ζ_{C1}	-13	+6	+55	+31	+38
$\zeta^{\text{abs}}_{\text{Co1}}$	-1.56	-0.24	-1.09	-0.54	-0.84
$\zeta^{\text{abs}}_{\text{C1}}$	-0.01	-0.01	0.06	0.03	0.04

Table 1. Magnetic moment μ per atom and layer unit cell (in μ_B) and relative/absolute spin polarization $\zeta(E_F)$ per layer unit cell (in % resp. eV^{-1}) for n Co:C (111) hybrid systems.

the Co adlayer and C substrate top layer atom. In contrast, the charge density of the unoccupied state B^* has a nodal plane between the Co and C atom emphasizing its antibonding character.

Adding further Co adlayers to the system rapidly leads to an increase in the number of bands near the Fermi level so that the band structure becomes very congested. However, for all investigated hybrid systems with more than one adlayer they also turn out to be metallic. Furthermore we find the characteristic B and B^* bands near -3 and +1 eV originating from the Co−C interface bonds.

Layer-resolved magnetic moments are presented in Table 1 together with the respective

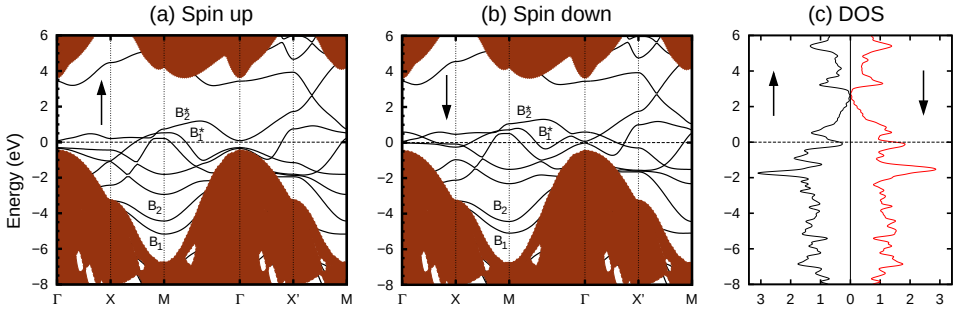


Figure 4. Band structures for (a) spin-up and (b) spin-down electrons and spin-resolved densities of states for one Co adlayer on the C(001) surface.

spin polarizations at E_F for varying numbers of adlayers. Some interesting features are to be noted. First of all, the formation of local magnetic moments is largely limited to the Co layers whereas the C layers remain essentially unmagnetized. Second, we observe that for each case the magnetic moment at the interface is considerably smaller than the respective value of a free-standing Co monolayer (μ=1.88 μ_B). This fact can be traced back to the formation of the B and B^* bands whose occupation is almost the same for both spin directions. As a result, the magnetic moment is drastically quenched at the interface, as compared to the free Co monolayer for which the occupation of the states in the flat region of the B^* band largely differs for the two spin directions. For more than one Co adlayer the local magnetic moments increase monotonously when going from the interface to the surface. At the topmost adlayer the magnetic moment per Co atom is close to that on a Co(0001) surface layer.

Another interesting quantity with regard to possible applications of these hybrid systems as part of a spintronic device is the relative spin polarization. It is defined via the spin resolved density of states as

$$\zeta(E) = \frac{N^{\uparrow}(E) - N^{\downarrow}(E)}{N^{\uparrow}(E) + N^{\downarrow}(E)} \tag{1}$$

and measures the surplus of spin-up states with respect to spin-down states at a certain energy. In Table 1 we show the relative spin polarisations at the Fermi energy for an increasing number of Co adlayers. A striking feature of the results is the substantial spin polarisation at the Co interface layer. The spin polarisation on the topmost C layer, on the other hand, is rather small and varies strongly for systems with different adlayer number.

For fcc cobalt layers on the C(001) surface the electronic structure becomes slightly more intricate since there are two Co$-$C bonds at the interface in this case. The band structure for one cobalt adlayer is presented in Figures 4a and 4b. The system is also metallic and the bands for spin-up and spin-down electrons are again rather similar. Yet the spin polarization at the Fermi level on the Co adlayer is about 50% also in this case. The interaction between the Co adlayer and the C(001) substrate now leads to two bonding (B_1 and B_2) and two antibonding (B_1^* and B_2^*) bands since there are two Co$-$C interface bonds per unit cell.

In Figure 5 we present associated charge densities at the $M-$point. The Co and C

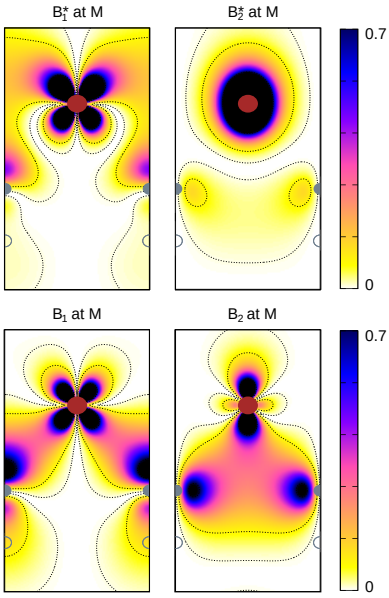


Figure 5. Charge densities at the M point of the surface Brillouin zone for the bonding (B_1 and B_2) and antibonding (B_1^* and B_2^*) states of one Co layer on C(001). The charge densities are plotted in the [100]-[001] plane containing the Co and the two C atoms at the interface.

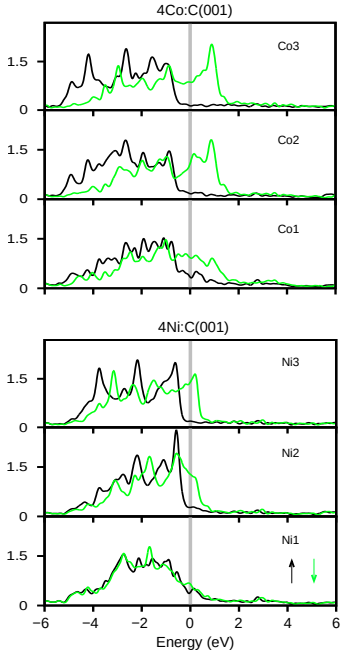


Figure 6. Layer-resolved densities of states for spin-up (black lines) and spin-down (green lines) electrons on the first three layers of systems containing four Co adlayers (upper panels) or four Ni adlayers (lower panels) on C(001), respectively.

orbitals form strong covalent bonds giving rise to considerably different spatial charge distributions for the B_1 and B_2 states. The double bonding leads to the large binding energy of 3.3 eV. The corresponding unoccupied antibonding states, on the other hand, exhibit a reduced screening in the spatial regions between the Co and C atoms.

Adding further layers of Co yields respectively more complex band structures. Here we only address layer-resolved densities of states (LDOS), therefore. As an example, we show in the upper three panels of Figure 6 the LDOS on the first three Co layers of a $4-\mathrm{Co{:}C}\,(001)$ system. One can immediately see that the spin-down LDOS on the second and third layer is shifted up by about 2 eV with respect to the spin-up LDOS near the Fermi level. This effect is strongly reduced for the first layer at the interface. On the first C layer at the interface, there is an appreciable LDOS within the gap energy region (not shown here), while on lower C layers the insulating behaviour of diamond is restored.

The two Co−C bonds per unit cell have significant consequences for the magnetic properties. The magnetic moment at the interface is only $0.42\mu_B$ and remains small for an increasing number of adlayers since the respective B_1, B_2 and B_1^*, B_2^* bands show almost no spin splitting. For more than one Co adlayer the magnetic moment of the top layer quickly converges to the value of the clean Co(001) surface.

The corresponding layer densities of states on the first three layers of a $4-\mathrm{Ni{:}C}\,(001)$

system are shown in the three lower panels of Figure 6. In general, the Co and Ni densities of states show a close resemblance. However, the spin splitting is drastically reduced for Ni as compared to the case of cobalt. This results from the additional d electron of each Ni atom occupying the spin-down states and shifting the respective LDOS to lower energies, therefore.

4 Concluding Remarks

In this article we have presented structural, electronic and magnetic properties of Co adlayers on C(111) and Co and Ni adlayers on C(001) as resulting from calculations in the framework of spin-density functional theory. The investigated systems are distinguished by an almost perfect matching of the lattice constants of their constituents. In all cases the strong covalent bonds established between adlayer and C substrate atoms give rise to a strong reduction of the local magnetic moment at the interface. Yet, for the systems with Co adlayers we find a high spin polarisation at the Fermi level even at the interface which is a prerequisite for spintronic applications. Ni adlayers, on the contrary, show only a very small spin polarisation at the interface.

Acknowledgements

We acknowledge financial support of this work by a grant of computer time on the JUMP computer of the John von Neumann Institute for Computing of the Forschungszentrum Jülich (Germany) under contract number HMS08-2961.

References

1. P. Grünberg, Rev. Mod. Phys. **80**, 1531 (2008).
2. A. Barthélémy, A. Fert, and F. Petroff, Handbook of Magnetic Materials, Vol. 12, edited by K. H. J. Buschow (Elsevier, Amsterdam, 1999).
3. I. Žutic, J. Fabian, and S. Das Sarma, Rev. Mod. Phys. **76**, 323 (2004).
4. S. Datta and B. Das, Appl. Phys. Lett. **56**, 665 (1990).
5. M. Hortamani, H. Wu, P. Kratzer, and M. Scheffler, Phys. Rev. B **74** , 205305 (2006).
6. L. Sacharow, M. Morgenstern, G. Bihlmayer, and S. Blügel, Phys. Rev. B **69**, 085317 (2004).
7. M. Zwierzycki, K. Xia, P. J. Kelly, G. E. W. Bauer, and I. Turek, Phys. Rev. B **67**, 092401 (2003).
8. R. Martin, *Electronic Structure*, (Cambride University Press, Cambridge, 2004).
9. J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
10. L. Kleinman, and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
11. D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).
12. S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
13. H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
14. J. Wieferink, P. Krüger and J. Pollmann, Phys. B **74**, 205311 (2006).
15. S. M. Shapiro and S. C. Moss, Phys. Rev. B **15**, 2726 (1977).
16. R. N. Grass and W. J. Stark, J. Mater. Chem. **16**, 1825 (2006).

Condensed Matter: From Deep Quantum to Classical Many-Body Systems

Kurt Binder[1] **and Alejandro Muramatsu**[2]

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

[2] Institut für Theoretische Physik III, Universität Stuttgart, 70550 Stuttgart, Germany
E-mail: mu@theo3.physik.uni-stuttgart.de

While the elementary building blocks of condensed matter systems are very well known, namely atoms with their nuclei and electrons, they offer a myriad of phenomena due to the complex collective states into which a macroscopic number of degrees of freedom can arrange. The understanding of such phenomena, a prerequisite for the prediction of new ones, may need full account of the quantum mechanical nature of the constituents, or in some limits, they can be understood within the realms of classical physics. In most of the problems in condensed matter systems, the interaction among the building blocks are determined by nature, as the Coulomb interaction for solid-state systems. However, in the new field of ultra-cold atoms, the interactions can be tailored by means of so-called Feshbach resonances, giving rise in some cases to three-body interactions, and hence, to new physics, beyond the frame known from solid-state systems. However, in contrast to elementary particle physics, where the nature of some of the basic interactions are still debated, the starting point of the theory of condensed matter is perfectly well defined, since the microscopic bare interactions are prefectly known. Hence, in principle, any phenomena in condensed matter could be theoretically understood if one were able to solve the many-body Schrödinger equation of electrons and nuclei (interacting with just the appropriate microscopic potentials) and use this solution as input to the formalism of (quantum) statistical mechanics.

In spite of the impressive increase in speed and memory of computers in the last years, as exemplified e.g. by latest generations of Petaflop computers, it will in general be impossible to solve the many-body Schrödinger equation due to the exponential increase of the number of states (the Hilbert space) with the number of constituents. Therefore, in parallel to the progess in hardware, an equally strong advance in efficient algorithms is necessary, in order to achieve significant progress in this field.

The articles that will follow in this section give an account of such developments in various areas in the field. Of course, these articles constitute only a small selection of problems in the science of condensed matter that are studied using the resources of NIC; in addition, we note that many articles that appear in the sections on chemistry, materials science, polymers and soft matter also deal with condensed matter, of course.

As a first example of algorithmic development along the lines discussed above is given by the article by F. B. Anders, where a numerical renormalization group (NRG) approach is presented to describe the steady-state transport properties of a nano-device with finite bias, i.e. transport out of equilibrium. Nano-devices (quantum dots) are at the above men-

tioned deep quantum limit, where the quantization of the electron charge dominates the charge transport through them. Furthermore, the spin degree of freedom of the electrons in the device may play a crucial role in the Coulomb-blockade regime, leading to correlated states, as in the Kondo-effect. This effect arises, when the interplay between the electron in the quantum dot and the conduction electrons in the leads brings them into a correlated state forming a many-body resonance (the Kondo resonance) that opens a new transport channel at low temperatures. The new algorithm implements Wilson's ideas for a numerical renormalization group extended to take into account scattering states that comply with the boundary conditions for a current carrying state. A parallelization of the algorithm is implemented that allows to efficiently perform the diagonalization of the Hamiltonian, that accounts typically for 35.000 to 40.000 states. The algorithm allows to calculate the conductance of the nano-device for arbitrary temperature and magnetic field.

Staying in the area of quantum dots, the article by Lorke *et al*. deals with their implementation for semiconductor lasers. A many-body treatment is necessary in order to take properly into account the consequences of quantum confinement and carrier interactions. The electronic excitations and optical properties are described using the single-particle density matrix, where the equations of motions for its elements have to be solved. This becomes a demanding task due to the couplings among the different matrix elements as a consequence of interaction processes and the presence of higher order correlation functions. Approximations are necessary in order to deal with the carrier-carrier and carrier-phonon correlations. They are treated in 2^{nd}. Born and random phase approximation, respectively. With the inclusion of many-body effects due to Coulomb interaction and the interaction with phonons, the calculation of the dynamics of interband transition amplitudes were obtained by solving a set of coupled integro-differential equations, where high-dimensional scattering integrals with non-Markovian contributions and the requirement of self-consistency render the calculation very costly. Such a heavy calculation load could only be solved by resorting to supercomputer resources. The results revealed the major limitations for quantum dots as laser sources as well as pointing out possible routes to improve their performance.

Still in the deep quantum limit but in the area of ultra-cold atoms, Bonnes and Wessel deal with strongly correlated polar molecules with three-body interactions. While Coulomb interactions are the ones determining the properties of solid-state and most other condensed matter systems, ultra-cold atoms allow for engineering new types of interactions that can be controlled by appropriate electric and magnetic fields. Specifically, polar molecules, i.e. molecules with a large dipole moment, generally due to a heteronuclear pair, can develop in the presence of an electric field and driven by a microwave field, strong three-body interactions. This opens the possibility of radically new states of matter, that are here studied in the case of bosonic molecules. A link to solid-state physics is established by considering the system in an optical lattice, that results from the standing waves of counterpropagating laser beams. The possible many-body ground-states were obtained by performing quantum Monte Carlo simulations based on the generalized directed loop algorithm within the stochastic series expansion (SSE) representation. Hence, apart from stochastic uncertainties, the results are exact. However, the systems studied show the tendency to form large superstructures, such that long autocorrelation times were observed. This is remedied by implementing quantum parallel tempering, in particular close to quantum phase transitions between different valence bond crystals. A fairly rich phase

diagram was obtained, where beyond a superfluid phase, valence bond crystals at different fractional filling were obtained. Such crystals are characterized not by particles sitting at the lattice sites but by particles resonating on the bonds, a purely quantum mechanical phenomenon.

Leaving the ground of models and entering the area of realistic many-body calculations, the article by Lazić *et al*. discusses the development of new functionals in the frame of density functional theory (DFT), that allow to treat the van der Waals interaction. Van der Waals interactions are ubiquitous in biological systems, with energy scales much lower than those typical of chemical bonding. However, their non-local microscopic nature could not be properly taken into account by *ab initio* calculations based on the local density approximation (LDA), or extensions of it. Recent developments for non-local functionals allowed to deal with the van der Waals interaction. The authors developed a massively parallel implementation of one such non-local functional. Interestingly, such an implementation is also able to deal efficiently with systems that were previously treated with extensions (generalized gradient expansion) of LDA, leading to results better in agreement with experiments for chemisorbed (small) molecules on metals. The major merit, however, is the possibility of dealing with the adsorbtion of large structures as graphene or large organic molecules on metal surfaces, as described in the article. In such cases, qualitative differences are obtained, making clear the advances achieved.

Finally, crossing over from the microscopic to the macroscopic world, it may be expected that quantum effects become less essential. On the other hand, at the meso-scale, microscopic details may be important. Such a case is presented in the article by Magiera and Wolf, where magnetic friction is simulated. In this case, if a system with large spin is considered, they can be treated classically, while taking them into account individually gives access to situations relevant for processes at the nanometer scale. While dissipation can be easily described in a macroscopic system, as done in undergraduate physics courses, its understanding at the microscopic level is much more demanding. The present article shows how, with phenomenological inputs, magnetic friction can be described dealing with the equations of motion of the individual spins in a classical anisotropic Heisenberg model. The microscopic treatment allows to determine the temperature dependence of the friction coefficient as well as the influence of anisotropy and substrate thickness. With such simulations of driven open systems we close the circle, where progress both in the classical as well as in the quantum world paves the pace towards establishing the grounds of a statistical theory of systems out of equilibrium with the tools of high performance computing.

Quantum Transport Through a Nano-Device: A Numerical Renormalization Group Approach to Open Quantum Systems

Frithjof B. Anders

Lehrstuhl für Theoretische Physik II, Technische Universität Dortmund
Otto-Hahn-Str. 4, 44221 Dortmund, Germany
E-mail: frithjof.anders@tu-dortmund.de

Scattering states fulfil the correct boundary conditions of a current-carrying open quantum system. Discretizing the energy continuum of these states allows the application of Wilson's numerical renormalization group approach to open quantum systems. We employ the time-dependent numerical renormalization group to evolve the known steady-state density operator for a noninteracting junction into the density operator of the fully interacting nanodevice at finite bias. The steady-state nonequilibrium Green functions are obtained by a recently developed algorithm. We calculate the current I as a function of bias and gate voltage for arbitrary temperature and magnetic field.

1 Introduction

Quantum-dots and single-molecular junctions have been considered as possible building blocks for nano-electronics and for quantum information[1]. The quantization of the electric charge dominates the charge transport of such devices at low temperatures leading to Coulomb-blockade physics: for voltages smaller than the charging energy $U = e^2/2C$, transport can only occur via tunneling of single electrons[1]. This blockade of charge transport is lifted in certain parameter regimes due to the Kondo effect[2]. Even when real charge fluctuations are suppressed, the remaining spin degree of freedom on a quantum-dot or a molecule with odd number of electrons can fluctuate due to virtual spin-flip transitions mediating charge transfer between the leads. As a consequence, a pinned many-body resonance at the chemical potential opens a new transport channel below a characteristic temperature T_K which is exponentially dependent on the charging energy.

The investigation of such devices is of fundamental importance for our understanding of open quantum systems out of equilibrium. Quantum impurity models are used in the theory of quantum transport on the molecular level. By now, the properties of such nano-devices are well understood at zero bias in the crossover regime from high to low temperatures using non-perturbative approaches such as Wilson's numerical renormalization group (NRG)[3]. Over the past 40 years, the Keldysh technique[4] has proven to be the most successful approach to nonequilibrium dynamics as it furbishes a perturbative expansion for correlation functions. It is applicable to problems having a well defined ground state in equilibrium such as in semi-conductor quantum optics[5], where the band gaps act as a cutoff for any divergency in perturbation theory. Quantum impurity models, however, often exhibit infrared divergencies[3], and therefore, perturbation theory fails in equilibrium. Furthermore, the Keldysh diagrammatic approaches are only applicable in certain parameter regimes usually at high temperature, large bias or weak interactions. In these regimes,

the infrared divergencies are cut off by temperature or voltage, both causing decoherence and a finite memory time. The combination of non-crossing approximation and Keldysh techniques has been a standard approach to calculate quantum transport[6] in strongly correlated systems with ballistic leads.

Recently, a large number of different methods[7–12] have been developed to tackle this difficult class of problems. In this paper, we report on the application of a scattering-states approach[7,9,13,14] to quantum-transport through a single-electron transistor[1]. The scattering-state numerical renormalization group[15] (SNRG) extends the NRG to current carrying ensembles of open quantum systems. The recently developed time-dependent NRG[16,17] was used to evolve a non-interacting current carrying system into to a fully interacting system at finite bias. The non-equilibrium spectral functions[18] of the device determine the $I(V)$ characteristics at arbitrary temperature, voltage and Coulomb repulsion.

Single-particle scattering states provide the exact solution for the Lippman-Schwinger equation for a non-interacting model. The states consist of independent left and right movers. The total current is given by the imbalance between the current flowing to the left and to the right due to a finite bias. Switching on a finite Coulomb interaction introduces backscattering: the scattering states are not eigenstates of the system, and the classification of a left and a right mover does not hold any more. This makes the numerical treatment of such a complicated many-body problem very costly. The NRG essentially is a two-channel problem with a lack of flavour conservation. For each remaining quantum number large matrices need to be diagonalized independently. Calculating the diagonalization efficiently was only possible with a large RAM and large number CPUs as provided by the NIC in Jülich.

2 The Scattering-States Numerical Renormalization Group (SNRG)

2.1 Model and Scattering States

Quantum impurity models (QIM) are used to describe quantum transport on the molecular level[2,6]. Their Hamiltonian $\mathcal{H} = \mathcal{H}_{imp} + \mathcal{H}_{bath} + \mathcal{H}_I$ consists of three parts: an impurity part $\mathcal{H}_{imp}$, modelling the interacting device with a finite number of degrees of freedom, one or several Bosonic or Fermionic baths $\mathcal{H}_{bath}$ and the coupling of these subsystems by $\mathcal{H}_I$. Examples of such quantum impurities are a simple two-level system, a localized spin, a superconducting qubit, a molecule or a quantum dot[3]. The leads provide two fermionic baths at different chemical potentials. We restrict ourselves to a junction with a single spinful orbital coupled to a left (L) and a right (R) lead

$$\begin{aligned}\mathcal{H} = &\sum_{\sigma\alpha=L,R}\int d\epsilon\,(\epsilon-\mu_\alpha)\,c^\dagger_{\epsilon,\sigma\alpha}c_{\epsilon,\sigma\alpha} + \sum_{\sigma=\pm1}\left[E_d+\frac{U}{2}-\frac{\sigma}{2}H\right]\hat{n}^d_\sigma \\ &+\frac{U}{2}\left(\sum_\sigma \hat{n}^d_\sigma - 1\right)^2 + \sum_{\alpha\sigma} V_{\alpha\sigma}\int d\epsilon\,\sqrt{\rho_\alpha(\epsilon)}\left(d^\dagger_\sigma c_{\epsilon\sigma\alpha} + c^\dagger_{\epsilon\sigma\alpha}d_\sigma\right). \qquad (1)\end{aligned}$$

Here $\hat{n}^d_\sigma = d^\dagger_\sigma d_\sigma$ measures the orbital occupancy, H denotes the external magnetic field and U the local charging energy if the quantum dot level is doubly occupied. The Hartree term has been partially absorbed into the single-particle energy $E_d \to E_d + U/2$. The different chemical potentials μ_α in both leads appear as a shift of the band centres and

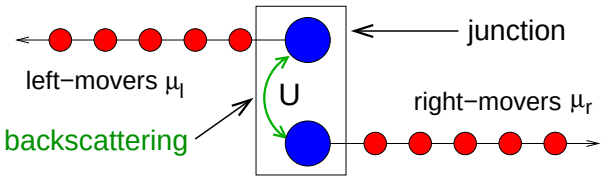


Figure 1. The $d_{\sigma\alpha}$ orbitals serves as start for a separate Wilson chain for each mover $\alpha = L, R$. The Coulomb repulsion violates flavour conservation and introduces backscattering between left and right-movers.

are functions of the external voltage $V = \mu_R - \mu_L$. We assume for both leads the same structureless density of states $\rho(\epsilon)$ characterized by the band width D.

In the absence of the local Coulomb repulsion $H_U = \frac{U}{2}\left(\sum_\sigma \hat{n}^d_\sigma - 1\right)^2$, the Hamiltonian in Eq. (1) is diagonalized exactly in the continuum limit by scattering states[14,15]. To avoid any contribution from bound states, we will implicitly assume a wide-band limit: $D \gg \max\{|\epsilon_d|, \Gamma, |V|, |H|\}$, where $\Gamma_\alpha = \pi V_\alpha^2 \rho(0)$ and $\Gamma = \Gamma_L + \Gamma_R$. $R = \Gamma_L/\Gamma_R$ measures the asymmetry of the junction couplings.

Hershfield[13] has shown that the density operator for such a non-interacting current-carrying quantum system retains its Boltzmannian form

$$\hat{\rho}_0 = \frac{e^{-\beta(\mathcal{H}_0^i - \hat{Y}_0)}}{\mathrm{Tr}\left[e^{-\beta(\mathcal{H}_0^i - \hat{Y}_0)}\right]}, \; \hat{Y}_0 = \sum_{\alpha\sigma} \mu_\alpha \int d\epsilon \, \gamma^\dagger_{\epsilon\sigma\alpha}\gamma_{\epsilon\sigma\alpha} \tag{2}$$

even for finite bias. $\gamma^\dagger_{\epsilon\sigma\alpha}$ creates a scattering state with energy ϵ, spin σ and direction α, and $\mathcal{H}_0^i = H - H_U$. The $\hat{Y}_0$ operator accounts for the occupation of the left- and right-moving scattering states, and μ_α for the different chemical potentials of the leads. Using this density-operator $\hat{\rho}_0$ in the calculation of the current expectation value[14] reproduces the standard result[19,6] for non-interacting devices. The knowledge of the analytical form of $\hat{\rho}_0$, however, makes this steady-state model accessible to the NRG[3,15].

The impurity operator $d^\dagger_\sigma$ is expanded into left- and right-mover contributions $d^\dagger_\sigma = r_R d^\dagger_{\sigma R} + r_L d^\dagger_{\sigma L}$. These two new operators[15] are defined as

$$d^\dagger_{\sigma\alpha} = \bar{V} \int d\epsilon \sqrt{\rho(\epsilon)} |G^r_{0\sigma}(\epsilon)| \gamma^\dagger_{\epsilon\sigma\alpha} \quad , \tag{3}$$

where $G^r_{0\sigma}(\epsilon)$ denotes the non-interacting retarded single-particle Green function.

The scattering-states continuum is discretized logarithmically in the usual way as in the standard NRG[3]. Then, the discretized version of the non-interacting Hamiltonian $\mathcal{H}_0^i$ is mapped onto a semi-finite Wilson chain whose tight binding matrix elements $t_{n\sigma\alpha}$ decay exponentially: $t_{n\sigma\alpha} \propto \Lambda^{-n/2}$. We use $d_{\sigma\alpha}$ defined in Eq. (3) as starting vector $f_{0\sigma\alpha} = d_{\sigma\alpha}$ for the Householder transformation[3]. The two Wilson chains are schematically depicted in Figure 1.

In order to include Coulomb interaction H_U, the local density operator $\hat{n}^d_\sigma = d^\dagger_\sigma d_\sigma$ must be expanded in the new orbitals $d_{\sigma\alpha}$. It consists of two contributions: a density term and a backscattering term $\hat{n}^d_\sigma = \hat{n}^0_\sigma + \hat{O}^{back}_\sigma$, where $\hat{n}^0_\sigma = \sum_\alpha r^2_\alpha d^\dagger_{\sigma\alpha} d_{\sigma\alpha}$ and the backscattering term is defined as $\hat{O}^{back}_\sigma = r_L r_R \left(d^\dagger_{\sigma R} d_{\sigma L} + d^\dagger_{\sigma L} d_{\sigma R}\right)$. The local Coulomb inter-

action term leads to a mixing of left- and right-movers since $\hat{O}_\sigma^{back}$ does not commute with Y_0.

2.2 Time-Dependent Numerical Renormalization Group (TD-NRG) Approach

The dynamics of a non-interacting model obtained via Hartee-Fock or LDA approximation can be solved exactly: all information is encoded in the density operator[13,14] $\hat{\rho}_0$, Eq. (2). We are interested in calculating the IV characteristics for an interacting device with finite charging energy U, for which the steady-state density operator is not explicitly known. We have developed a novel approach for calculating a faithful representation of $\hat{\rho}(U, V, t \to \infty)$ on a Wilson chain using the time-dependent NRG method[16,17]. We assume that the system approaches a steady state which is unique and independent of the initial conditions. We consider a sudden quench of the system after switching on the local Coulomb interaction U. Then, the time average of the density operator

$$\hat{\rho}_\infty = \lim_{T\to\infty} \frac{1}{T} \int_0^T dt \hat{\rho}(t) \tag{4}$$

projects out the steady-state contributions to the density operator $\hat{\rho}(t) = \exp(-i\mathcal{H}_f t)\, \hat{\rho}_0\, \exp(i\mathcal{H}_f t)$ even in a finite size system: only the energy diagonal terms contribute in accordance with the condition $[\mathcal{H}_f, \hat{\rho}_\infty] = 0$. Even though $\hat{\rho}_\infty$ remains analytically unknown, we can construct it systematically using the TD-NRG[16,17].

The I(V) curve is calculated using the bias-dependent spectrum of the retarded Green function $\rho_\sigma(\omega, V)$

$$I(V) = \frac{G_0}{e} \sum_\sigma \int_{-\infty}^{\infty} d\omega\ [f(\omega - \mu_L) - f(\omega - \mu_R)]\, \pi\Gamma \rho_\sigma(\omega, V) \tag{5}$$

where $f(\omega)$ denotes the Fermi function, and $G_0 = (e^2/h)4\Gamma_L\Gamma_R/\Gamma^2$. This expression[19] is valid for an interacting quantum-dot but requires the steady-state non-equilibrium spectral function $\rho_\sigma(\omega, V)$ at finite bias. As part of our new approach, a numerical challenging algorithm[18] can extract these spectral functions using the projected time-averaged steady-state density operator $\hat{\rho}_\infty$ defined in Eq. (4).

2.3 Computational Effort

The iterative diagonalization of the Hamiltonian in the NRG can be performed in parallel for the independent block matrices of each set of conserved quantum numbers. Furthermore, two almost independent runs are needed for the TD-NRG: one NRG run constructs the initial density matrix $\hat{\rho}_0$, the other run constructs the eigenstates of $\mathcal{H}_f$ needed for obtaining $\hat{\rho}_\infty$, which can run in parallel. Since the local Coulomb repulsion introduces additional backscattering, the flavour conservation of left and right movers is lost. The NRG calculation is essentially a two-channel calculation where only the total number of electrons Q and the z-component of the spin are used as quantum numbers. Therefore, the block matrices of each Hamiltonian subblock labelled by a (Q, S_z) are very large. At each iteration the Hamiltonian matrices are typically spanned by 35000-40000 many-body states, the program typically needs 20 GB of RAM at run time, which goes up to 80 GB if parallel z-averaging is used[16,17].

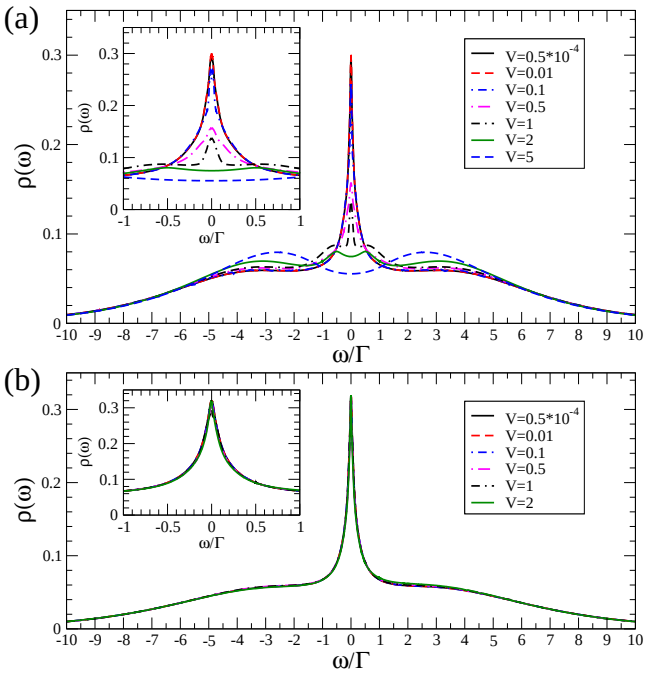


Figure 2. Nonequilibrium spectral functions[15] for (a) a symmetric junction $R = 1$ and different value of the bias V, and (b) for a strongly asymmetric junction $R = 1000$. The charging energy is set to $U/\Gamma = 8$ and the single-particle level at $E_d = -U/2$ and $T \to 0$. The insets show the evolution of the Kondo resonance.

3 Results

All energies as well as the bias voltage are measured in units of the single-particle charge fluctuation scale Γ. For simplicity, we used symmetric structureless leads characterized by a constant DOS $\rho_0 = 1/(2D)$ on the interval $[-D : D]$ where $D/\Gamma = 20$ is chosen.

The non-equilibrium spectral functions obtained with the SNRG and diagrammatic Keldysh approaches are identical for small values of $U/\Gamma \leq 1$[20]. The major contribution stems from the Hartree-term which causes a single-particle level shift. The line shape remains almost identical to the $U = 0$ solution: the additional scattering introduced by the finite value of U is much smaller than the resonant level width Γ. The diagrammatic approaches predict a slightly larger dephasing than the SNRG leading to a stronger decrease of the spectral function around $\omega = 0$. Nevertheless, the overall agreement of the all approaches is remarkable. This provides a reliable benchmark for the SNRG.

The spectral functions look quite different for larger values of U. A temperature dependent many-body resonance at $\omega = 0$ emerges when lowering the temperature below a characteristic low-energy scale T_K depending exponentially[3] on the charging energy U. In addition, the single-particle resonance is replaced by broadened atomic peaks with fractal weights at E_d and $E_d + U$. Results for the bias dependence of the spectral functions for $U/\Gamma = 8$ in the quantum-point contact regime ($R = 1$) and in the tunneling regime ($R = 1000 \gg 1$) are depicted in Figure 2. The spectral function changes significantly

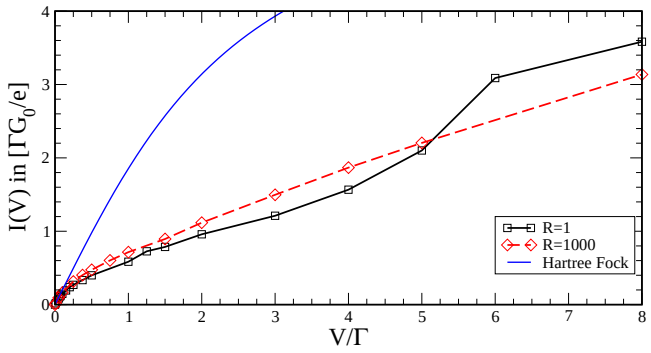


Figure 3. Current versus voltage for the point contact ($R = 1$) and the tunneling junction ($R = 1000$) calculated using the spectral functions of Figure 2. Note that the current has been normalized to $G_0 = (e^2/h)(4\Gamma_L\Gamma_R/\Gamma^2) = (e^2/h)4R/(1+R)^2$. The blue curve indicates the Hartree-Fock solution.

with V in the point-contact junction: the many-body resonance at $\omega = 0$ decreases with increasing voltage on the scale of T_K. An onset of a splitting of the resonance is also observed for $|V| > \Gamma$. As expected for the tunneling regime, the spectrum remains almost unaltered as shown in Figure 2(b): a Kondo effect develops only from the interaction of the quantum-dot with the lead which is strongly coupled, while the weakly coupled lead serves as a probe.

The IV characteristic was calculated for $U/\Gamma = 8$ by inserting the spectra of Figure 2 into Eq. (5). The normalized currents are shown for $R = 1$ and $R = 1000$ in comparison with the Hartree-Fock results in Figure 3. The normalization of the current eliminates the trivial current suppression by the prefactor G_0. In the quantum-point contact regime, the many-body resonance is destroyed with increasing voltages, and spectral weight is transfered to higher frequencies. As a consequence, the normalized current is initially lower than in the tunneling regime. The spectral weight loss accumulates at intermediate frequencies in the vicinity of the original atomic charge peaks. Hence, the current is enhanced once the voltage approaches the charging energy U. The normalized current will saturate at 2π for very large voltages independent of U. At zero bias and odd-integer fillings of the quantum-dot, the differential conductance is alway given by values close to G_0 for $T \to 0$ due to the formation of a local Fermi-liquid. For larger Coulomb interaction U, however, the current increases with bias significantly slower than predicted by the Hartree-Fock solution. Therefore, effective single-particle theories such as Hartree-Fock are unsuitable to describe interacting nano-devices and molecular junctions in the coherent regime at low temperature. They predict incorrect IV curves in parameter regimes governed by strong correlations as can be seen in Figure 3.

The numerically obtained differential conductance dI/dV is shown for different values of R in Figure 4(a). We also added dI/dV curves for finite magnetic field. Figure 4(b) depicts the differential conductance for the point-contact regime and three different values of the external magnetic field H. The zero-bias maximum in the dI/dV curves split into two peaks separated roughly by twice the Zeemann energy. The peak height decreases with increasing magnetic field due to the destruction of the many-body resonance with increasing field and bias voltage.

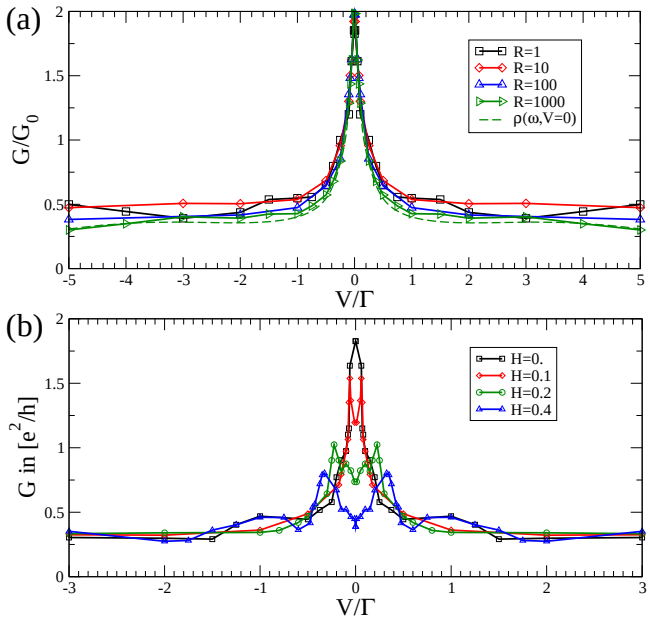


Figure 4. The differential conductance $G = dI/dV$ as a function of the bias voltage (a) for different asymmetry factors R and $H = 0$, (b) for different magnetic field $H = 0, 0.1, 0.2, 0.4$ and $R = 1$.

4 Conclusion

A novel non-perturbative approach to quantum-transport has been used to calculate the IV characteristics of a quantum-dot out of equilibrium. It is an extension of Wilson's numerical renormalization group to open quantum systems. It includes the crossover from the Coulomb-blockade regime at high temperature to coherent transport at temperatures smaller than T_K. Our approach uses a scattering-state description for the single-particle state to include the correct boundary condition for a current-carrying open quantum system. Knowing the analytical form of the steady-state density operator for a non-interacting device at arbitrary bias[13], makes it possible to employ Wilson's numerical renormalization to steady-state problems. The time-dependent TD-NRG provides all information on the steady-state density matrix of the interacting system needed to calculated the steady-state spectral function at finite bias for arbitrary temperatures.

Acknowledgements

I acknowledge stimulating discussions with S. Schmitt and financial support by the Deutsche Forschungsgemeinschaft under AN 275/6-1. All calculations were performed with a grant for computer time provided by the NIC at the Research Centre Jülich.

References

1. M. A. Kastner, *The single-electron transistor*, Rev. Mod. Phys., **64**, 849, 1992.

2. D. Goldhaber-Gordon, H. Shtrikman, D. Mahalu, D. Abusch-Magder, U. Meirav, and M. A. Kastner, *Kondo effect in a single-electron transistor*, Nature, **391**, 156, 1998.
3. R. Bulla, T. A. Costi, and T. Pruschke, *The numerical renormalization group method for quantum impurity systems*, Rev. Mod. Phys., **80**, 395, 2008.
4. L. V. Keldysh, Sov. Phys. JETP, **20**, 1018, 1965.
5. G. Khitrova, H. M. Gibbs, F. Jahnke, M. Kira, and S. W. Koch, *Nonlinear optics of normal-mode-coupling semiconductor microcavities*, Rev. Mod. Phys., **71**, 1591, 1999.
6. N. S. Wingreen and Y. Meir, *Anderson model out of equilibrium: Noncrossing-approximation approach to transport through a quantum dot*, Phys. Rev. B**49**, no. 16, 11040, 1994.
7. P. Mehta and N. Andrei, *Nonequilibrium Transport in Quantum Impurity Models (Bethe-Ansatz for open systems)*, Phys. Rev. Lett. **96**, 216802, 2006.
8. S. Kehrein, *Scaling and Decoherence in the Nonequilibrium Kondo Model*, Phys. Rev. Lett. **95**, 056602, 2005.
9. J. E. Han and R. J. Heary, *Imaginary-Time Formulation of Steady-State Nonequilibrium: Application to Strongly Correlated Transport*, Phys. Rev. Lett. **99**, 236808, 2007.
10. S. Weiss, J. Eckel, M. Thorwart, and R. Egger, *Iterative real-time path integral approach to nonequilibrium quantum transport*, Phys. Rev. B**77**, no. 19, 195316, 2008.
11. Peter Schmitteckert, *Nonequilibrium electron transport using the density matrix renormalization group method*, Phys. Rev. B**70**, 121302(R), 2004.
12. H. Schoeller, *A perturbative nonequilibrium renormalization group method for dissipative quantum mechanics*, Eur. Phys. J. Special Topics, **168**, 179–266, Feb. 2009.
13. S. Hershfield, *Reformulation of steady state nonequilibrium quantum statistical mechanics*, Phys. Rev. Lett. **70**, 2134, 1993.
14. Akira Oguri, *Mixed-state aspects of an out-of-equilibrium Kondo problem in a quantum dot*, Phys. Rev. B**75**, 035302, 2007.
15. Frithjof B. Anders, *Steady-State Currents through Nanodevices: A Scattering-States Numerical Renormalization-Group Approach to Open Quantum Systems*, Phys. Rev. Lett. **101**, no. 6, 066804, 2008.
16. F. B. Anders and A. Schiller, *Time-dependent Numerical Renormalization Group Approach to non-Equilibrium Dynamics of Quantum Impurity Systems*, Phys. Rev. Lett. **95**, 196801, 2005.
17. F. B. Anders and A. Schiller, *Spin precession and real-time dynamics in the Kondo model: Time-dependent numerical renormalization-group study*, Phys. Rev. B**74**, 245113, 2006.
18. Frithjof B. Anders, *A Numerical Renormalization Group approach to Non-Equilibrium Steady State Green's Functions for Quantum Impurity Models*, J. Phys.: Condens. Matter **20**, 195216, 2008.
19. Yigal Meir and Ned S. Wingreen, *Landauer formula for the current through an interacting electron region*, Phys. Rev. Lett. **68**, 2512, 1992.
20. Sebastian Schmitt and Frithjof B. Anders, *Comparison between scattering states numerical renormalization group and the Kadanoff-Baym-Keldysh approach to quantum transport: Crossover from weak to strong correlations*, p. arXiv:0909.5555 and submitted to Phys. Rev. B, 2009.

Van der Waals Interaction of Molecules on Surfaces from First Principles

Predrag Lazić, Nicolae Atodiresei, Vasile Caciuc, and Stefan Blügel

Institut für Festkörperforschung (IFF) and Institute for Advanced Simulation (IAS)
Forschungszentrum Jülich, D-52425 Jülich, Germany
E-mail: {*p.lazic, n.atodiresei, v.caciuc, s.bluegel* }*@fz-juelich.de*

In the past few decades, in many branches of the natural sciences, in geology, materials science and even engineering, the density-functional theory (DFT) developed to the tool of choice to determine the ground-state and excited state properties of molecules, clusters, solids, materials, and material systems from the basic laws of quantum mechanics, i.e. *ab initio*. However, until recently there existed a large class of systems, those in which the van der Waals interaction is important, that were poorly if not falsely described by the current approximations to the DFT. The reason lies in the origin of such an interaction, which is the non-local electron correlation, that was not included in the local or semi-local functionals that were used so far. Recently, the first *ab initio* non-local functional appeared that opened new vistas for DFT. We have developed a massively parallel implementation of the new functional, that was made available to the public, and we have shown that such a functional did not only bring van der Waals systems within the grasp of DFT applications, but made a major improvement in general.

1 Introduction

Many biological systems and processes are fundamentally driven by the van der Waals (vdW) interaction. Even though it results in attraction and binds two objects together, it has a great advantage in the living world as compared to chemical bonding. The van der Waals interaction is frequently of the same energy scale as the entropy at room temperature giving rise of the soft matter behaviour of these systems. Further, two objects bonded together by vdW forces are easily separated into their original units, because no chemical bond was formed. A gecko[1] and some other animals are examples that make use of this mechanism by being able to climb on flat and even smooth surfaces employing vdW forces that occur between their feet and a surface. Even though in this example the gecko is a macroscopic object, vdW forces have a microscopic, i.e. atomic or molecular origin. The origin of the vdW interaction is a charge-charge correlation, which occurs via charge fluctuation and polarization of electronic charges at distances. Thus, it is intrinsically a non-local electron correlation phenomenon. Van der Waals forces and the Casimir effect[2] are two names for the same phenomena, only that in the description of van der Waals effect the speed of light is considered to be infinite. The same type of attraction is crucial when studying organic crystals, polymers or the adsorption of organic molecules on metal surfaces, the latter of which are a promising set-up towards nanotechnological applications. Also in the case of graphene, another very promising nanoelectronic material, the vdW interaction plays a major role when sheets of graphene bind between themselves forming graphite or when graphene sheets bind on metal surfaces.

Although in the past few decades, density-functional theory (DFT) in the local density (LDA) or in semi-local approximations such as the generalized gradient approximations

(GGA) to the exchange and correlation energy developed to the standard model to determine the properties of molecules, clusters, solids, materials, and material systems from the basic laws of quantum mechanics, i.e. *ab initio*, the van der Waals interaction was poorly if not falsely described. This was an important road block for the expansion of the DFT into a huge class of interesting systems. The advent of the first *ab initio* non-local density functional, called vdW-DF[3–5], in which the electron correlation treated by the local density approximation is replaced by a non-local one, together with its first applications to realistic systems which showed results of unprecedented accuracy, have pushed the doors wide open for the exploration of van der Waals systems from *ab initio*.

Besides DFT as a computational tool, there are some other methods that naturally take non-local electron correlation into account, e.g. the Møller-Plesset perturbation theory of second order (MP2)[6] and the random-phase approximation (RPA)[7]. Because of their wavefunction nature, the computational scaling with the size of the studied system is quite unfavourable, limiting their applicability only to the smallest systems. On the other hand density-functional theory relies only on the electron density and shows thus a much better scaling. Until the appearance of the non-local vdW-DF functional[3–5], practically only the local LDA or semi-local GGA functionals were in use, so that the calculation of the energy scaled linearly with the number of grid points, N, at which the charge density is given. The evaluation of non-local functionals requires intrinsically a double-space integral which implies an N^2 scaling. Even with such a quadratic scaling, the evaluation of a non-local energy functional is still much faster than the use of wavefunction based methods. Indeed, exploiting the excellent parallelization properties of that space integral, we were able to implement a massively parallel program, named JUNOLO[8,9], which enables us to treat systems of the same size as in the local or semi-local approaches.

For systems with well-separated identifiable subunits (typical for vdW systems), it is well justified, rather practical and accurate to replace the calculation of the van der Waals forces using the *ab initio* approach by a semi-empirical substitute[10]. However, the *ab initio* non-local functional has the great advantage that it is derived in a seamless fashion. That means it is applicable as a general functional substituting previously used local or semi-local ones at all ranges of electron density. We have demonstrated the importance of this property by applying the functional to a famous yet unresolved DFT problem in surface science, known as the CO-puzzle problem[11]. It summarizes the observation that *ab initio* calculations employing the local or semi-local approximation to the DFT give the wrong prediction for the adsorption site preference of a CO molecule on various transition-metal surfaces, in particular on Pt(111). We could show in Ref. 12 that this problem is essentially resolved by employing the vdW-DF functional and that the non-local correlation has a significant influence on the energetics even at length scales at which the semi-local functionals (GGA) were expected to work well.

The paper is organized as follows: In the next section we introduce our implementation of the vdW-DF functional for DFT calculations together with some basics foundations of it. In section 3 we show results for the adsorption of a CO molecule, organic molecules and graphene on metallic surfaces pointing out significant differences introduced by non-local correlations as compared to the (semi-)local ones. In the last section we give conclusions.

2 JUNOLO - A Massively Parallel Program to Calculate Non-Local Correlations

The JUNOLO program[8,9] is written in Fortran F90 and is specially tailored to run on a huge number of CPUs (tested up to 65536 CPUs on a Blue Gene/L system). The main purpose of the program is to calculate the non-local correlation energy as given by the vdW-DF functional from the given electron density presented in real space. Thus, the program is used as a post-processing tool having the advantage to be used for densities obtained from the user's preferred DFT code. The main task the program performs is the evaluation of a six-dimensional, i.e. double-space integral, which follows from the vdW-DF theory which we explain briefly. The origin of the vdW interaction is a charge-charge correlation which occurs via charge fluctuation and polarization. Following the adiabatic-connection formula, Dion *et al.*[3] have derived an expression for the non-local correlation energy based on the electron density only. The key assumption of the vdW-DF theory is that the response of the electron density can be described within a plasmon-pole approximation. Besides this mathematical approximation for the response function, the functional itself is free of any empirical parameter and is of true *ab initio* nature. Even though the initial intention of the authors of the vdW-DF functional was mostly to make vdW systems treatable in DFT, as the name of the functional suggests, they have taken care to construct it in a seamless fashion so that in principle one can legitimately treat any given system with this new functional. Regarding this, one should keep in mind that electron correlation is only one part of the electron-electron interaction beyond the Hartree approximation, the other part being the exchange. Since the exchange in DFT was treated semi-locally until now, the question of a suitable new exchange functional to accompany the vdW-DF correlation is under debate with an answer probably provided by a very recent paper of Cooper[13]. Here we will not discuss the issue of the exchange functional any further. With details given in the original paper, the vdW-DF functional is applied as a post-processing procedure, expressing the total energy with non-local correlation by replacing the correlation energy obtained in standard DFT calculation (GGA),

$$E_{\rm tot}^{\rm NL} = E_{\rm tot}^{\rm GGA} - E_{\rm c}^{\rm GGA} + (E_{\rm c}^{\rm LDA} + E_{\rm c}^{\rm NL}) \tag{1}$$

by the new term $E_{\rm c}^{\rm NL}$ given by the double-space integral:

$$E_{\rm c}^{\rm NL} = \frac{1}{2} \int \int d^3\mathbf{r} d^3\mathbf{r}' n(\mathbf{r}) \phi(\mathbf{r}, \mathbf{r}') n(\mathbf{r}'). \tag{2}$$

This integral reduces to a double sum on a discrete real-space grid on which the charge density $n(\mathbf{r})$ is given:

$$E_{\rm c}^{\rm NL} = \frac{1}{2} \sum_{\rm i} \sum_{\rm j} n(r_{\rm i}) \phi(r_{\rm i}, r_{\rm j}) n(r_{\rm j}). \tag{3}$$

The execution of the double summation is the main task of the JUNOLO program. Subtle details of the parallelization are given in the program manual[14]. The parallelization is carried out efficiently as long as each processor used in the calculation has more grid points than some critical number at which the calculation time becomes comparable to communication time. The efficiency of the parallelization is shown in Figure 1.

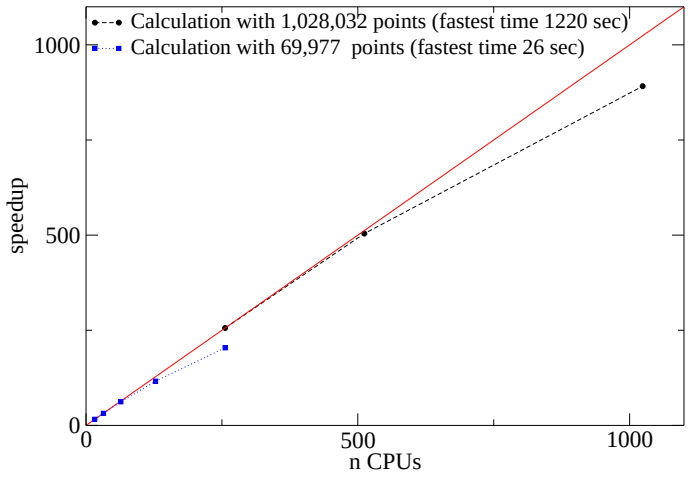


Figure 1. Speedup curves for the JUNOLO program. Full line represents the ideal speedup, linear in the number of CPUs used. Shortest time is the duration of the calculation when the largest number of CPUs is used, i.e. at the last point of the corresponding curve.

3 Applications of the JUNOLO Code

Efficient parallelization of the code and availability of a Blue Gene class supercomputer made it possible to treat systems that are considered huge even in terms of the currently most popular semi-local DFT implementations. To our surprise not only even a small molecule strongly chemisorbed on a metal surface proved to be much better described by the vdW-DF functional that was originally designed to treat typical van der Waals systems. Therefore, we show here as a first example a solution to a long standing DFT problem, known as *CO puzzle*[11]. In a second subsection, we show a few of the largest calculations that were performed with our code, that would have taken typically 10,000 CPU days (around 30 CPU years) on a single 1 GHz CPU.

3.1 Small Molecules on a Surface - CO Puzzle

The CO puzzle ascribes a DFT problem of correctly predicting the FCC adsorption site as energetically favourable as compared to the TOP site on the Pt(111) surface (and few other metals). It is common belief that a better, more appropriate exchange functional is crucial for the solution of this problem. After applying the vdW-DF functional[3–5] to this problem we have shown that in fact (non-local) correlation is decisive for a proper description of this strongly chemisorbed system. Details of the solution are given in Ref. 12, but the essence of the problem can be described briefly. Using semi-local GGA functionals (like PBE[15]) for strongly chemisorbed systems gives reasonable or even good agreement with experiment, but almost always one gets too large adsorption energies. The same is true for the CO puzzle problem, which additionally suffers from a wrong site preference prediction and because of this, it was considered to be a qualitative problem of the GGA functional with complicated explanations regarding the HOMO-LUMO gap[16]. Our solution to this

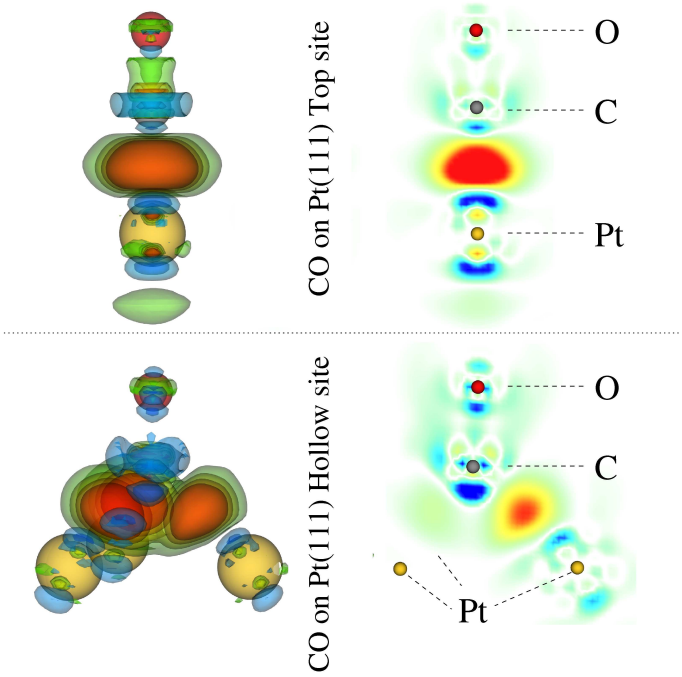


Figure 2. Overbinding PBE correlation energy. Difference of PBE and vdW-DF correlation binding energy density is shown. Red colour shows PBE overbinding while blue colours show underbinding compared to vdW-DF. TOP and FCC adsorption sites are shown.

puzzle lies in the observation that a (semi-)local and a non-local functional exhibit a very different behaviour in the spatial distribution of the correlation binding energy with respect to the coordination number of the atoms (see Figure 2). Specifically comparing the spatial distribution of correlation binding energy density between TOP and FCC site adsorption, we notice that for the case of the TOP site there is only one region in which binding energy appears (where the chemical bond is formed) while in the FCC site there are three such regions. Knowing that semi-local functionals overbind it is clear how the FCC site tends to be preferred in such situations.

3.2 Large Molecules on a Surface

For larger molecules adsorbed on metal surfaces we discovered an interesting competition between different parts of the molecule trying to bind to surface. The bonds in the molecule itself are usually rather strong and molecules cannot be stretched easily, which leads to a trade-off between chemical and van der Waals bonding. As an instructive example we show in Figure 3 the adsorption of a graphene layer on the Ir (111) surface. Since the lattice constants of these two structures do not match it is impossible that every carbon atom adsorbs on the most ideal position of the surface. Thus carbon atoms that are on favourable positions for adsorption (on top of Ir atoms) are attracted towards the surface tending to

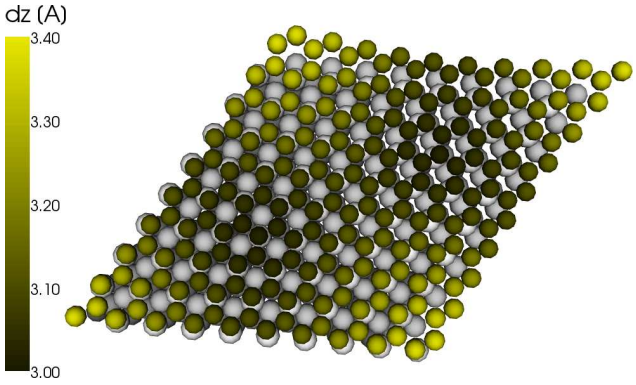


Figure 3. Calculated adsorption of a 10×10 graphene unit cell that matches a 9×9 Ir (111) surface unit cell. Distance of carbon atoms from Ir surface is shown in colour bar.

make a chemical bond pulling by means of strong chemical bonds the neighbouring carbon atoms also towards the surface, but those atoms approach the surface at unfavourable sites. Due to Pauli repulsion these atom remain at higher positions, at the same time disabling the atom at TOP position to go close enough to form a strong chemical bond. The result is that the graphene sheet comes close to the metal surface due to van der Waals interaction and is corrugated strongly due to the mismatch of the lattice parameters of the two structures. Using semi-local (GGA) functionals, this system shows no binding at all. Inclusion of the vdW interaction gives binding (∼165 meV/C atom) resulting in a corrugation of the graphene sheet.

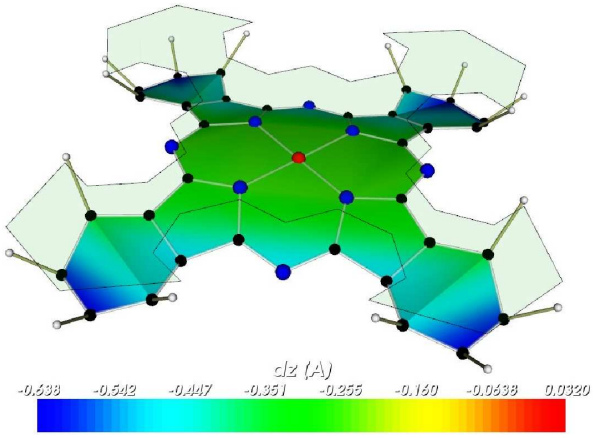


Figure 4. Adsorption of the cobalt phthalocyanine (CoPc) molecule on the FeW (110) surface. Transparent, flat green surface shows the adsorbed position of the molecules when only a semi-local (PBE) functional is used. Balls and sticks model shows state of the art result with vdW interaction included. Colour bar shows the difference of those two solutions regarding the distance from the underlying FeW (110) surface (not shown).

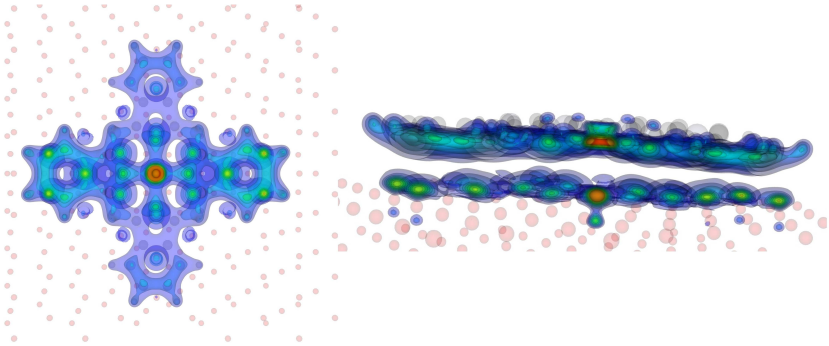


Figure 5. Non-local correlation binding energy for copper phthalocyanine (CuPc) adsorption on Cu(110) surface. Blue regions correspond to small binding values while red regions correspond to larger ones.

An analogous mechanism is at work for large molecules. One such example is the cobalt phthalocyanine (CoPc) molecule adsorbed on an Fe covered W(110) surface. Using GGA, the molecules remain flat and weakly bonded to the surface. But with the proper description with the non-local correlation, due to vdW forces it comes closer to the surface and bonds strongly. The energy trade-off of chemical bond formation, Pauli repulsion and van der Waals attraction governs such a complicated adsorption geometry as depicted in Figure 4. This was noticed also for some smaller molecules and it was named vdW triggered chemisorption[17]. This mechanism can be visualized by showing the non-local correlation binding energy, see Figure 5, which shows that there is some binding energy dispersed all around the molecule and it is of typically vdW origin (hence the name *dispersive forces*), but there are some specific places where this contribution is much more concentrated indicating the formation of the chemical bond, i.e. occurrence of charge transfer.

4 Conclusions

We have carried out large-scale *ab initio* calculations based on the density-functional theory for the adsorption of large molecules on metal surfaces which included the van der Waals interaction by treating electron correlation not by the commonly used local functional of the density but by a non-local functional as introduced by Dion *et al.*[3]. Compared to the interest in these molecules, relatively few such molecules on metal surfaces had been studied from first principles, a consequence of the expected errors introduced by the neglect of a significant amount of interaction energy coming from vdW interaction, which was not included in the commonly used approximation to the DFT. As we have shown, the scenario of the adsorption of large molecules is governed by three main elements: (i) Dispersive vdW forces always appear and they tend to bind a molecule to a surface. (ii) Within a molecule, some atoms (or regions) tend to form chemical bonds with the surface and by doing so, they pull neighbouring atoms closer to the surface. (iii) Some of those atoms are pulled into regions that are not energetically favourable, i.e. locations where Pauli repulsion takes place. The interplay of these three effects determines the final adsorption of the molecule that results in specific adsorption sites and often in a deformation of the

adsorbed molecule. Without treatment of the vdW interaction, the proper description of such a complicated scenario of adsorption is impossible, in particular as GGA functionals simply yield no adsorption – in clear contradiction to experiment. In summary, we have shown that the inclusion of non-local electron correlation in DFT calculations did not only allow a consistent study of new and interesting van der Waals systems that were not accessible before, but also a significantly improved treatment of systems that were believed to be describable by proper semi-local correlation functionals, but not developed so far. The new vdW-DF functional is the first of its kind and even if it suffers from some shortcomings, it outlines a route for all future functionals to follow paving the path to the theoretical design of nanodevices.

Acknowledgements

Predrag Lazić acknowledges support from the Alexander von Humboldt Foundation. The computations were performed under the auspices of the VSR-committee at the Gauss Centre for Supercomputing petaflop supercomputer JUGENE, a IBM Blue Gene/P installation operated by the Jülich Supercomputing Centre at the Forschungszentrum Jülich.

References

1. K. Autumn, M. Sitti, Y. A. Liang, A. M. Peattie, W. R. Hansen, S. Sponberg, Th. W. Kenny, R. Fearing, J. N. Israelachvili, R. J. Full, PNAS **99**, 12252 (2002).
2. H. B. G. Casimir, and D. Polder, Phys. Rev. **73**, 360 (1948); H. B. G. Casimir, Proc. Kon. Nederland. Akad. Wetensch. B**51**, 793 (1948).
3. M. Dion, H. Rydberg, E. Schröder, D. C. Langreth and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401, (2004); **95**, 109902(E) (2005).
4. D. C. Langreth, M. Dion, H. Rydberg, E. Schröder, P. Hyldgaard and B. I. Lundqvist Int. J. Quantum Chem. **101** 599 (2005).
5. T. Thonhauser, V. R. Cooper, S. Li, A. Puzder, P. Hyldgaard and D. C. Langreth, Phys. Rev. B **76**, 125112, (2007).
6. C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
7. P. Nozières and D. Pines, Phys. Rev. **111**, 442 (1958).
8. P. Lazić, N. Atodiresei, M. Alaei, V. Caciuc, S. Blügel, and R. Brako, accepted for publication by Comp. Phys. Comm.
9. `http://www.fz-juelich.de/iff/src/th1/JuNoLo/`.
10. S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
11. P. J. Feibelman, B. Hammer, J. K. Nørskov, F. Wagner, M. Scheffler, R. Stumpf, R. Watwe, and J. Dumesic, J. Phys. Chem. B **105**, 4018 (2001).
12. P. Lazić, M. Alaei, N. Atodiresei, V. Caciuc, S. Blügel, and R. Brako, accepted for publication by Phys. Rev. B.
13. V R. Cooper, arXiv:0910.1250v1.
14. `http://www.fz-juelich.de/iff/src/th1/JuNoLo/vdw_manual.pdf`.
15. J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
16. A. Stroppa, G. Kresse, New J. Phys. **10**, 063020 (2008).
17. N. Atodiresei, V. Caciuc, P. Lazić, and S. Blügel, Phys. Rev. Lett. **102**, 136809 (2009).

On the Potential of Semiconductor Quantum Dots as Active Media for Future Semiconductor Lasers

M. Lorke[1], P. Gartner[1], F. Jahnke[1], and W. W. Chow[2]

[1] Institute for Theoretical Physics, University of Bremen
28334 Bremen, Germany
http://www.itp.uni-bremen.de/ag-jahnke
E-mail: frank.jahnke@itp.uni-bremen.de

[2] Sandia National Laboratories
Albuquerque, New Mexico 87185-0601, USA

Presently semiconductor quantum dots (QDs) are intensively studied for scientific interest and as key elements for the next generation of optoelectronic devices. The much debated question is whether the actual extent of potential advantages will eventually lead to QD lasers with overall performance significantly better than what is achievable with the presently used mature technology based on quantum wells (QWs). Our investigations address this question by examining intrinsic behaviour and underlying physics: quantum confinement and carrier interactions in reduced dimensions. We show that inhomogeneous broadening due to growth fluctuations of current QDs is the main obstacle. The improved threshold behaviour is confronted with a larger sensitivity to saturation effects.

1 Introduction

Semiconductor lasers are central components of optical technologies, most prominently used in optical data storage on CDs, DVDs, or BDs and in the fiber-based optical communication underlying the internet. Moreover, semiconductor lasers entered almost any branch of or daily life: from laser printers to laser bar-code scanners. The success of semiconductor lasers was driven by an increasing performance to cost ratio, realized by several seminal developments. The heterostructure design provides an efficient confinement of the pumped charge carriers to the active region and the double heterostructure design, independently proposed by Kroemer as well as Alferov and Kazarinov[1], yields an additional confinement of the optical mode. The integration of QWs as active material[2] increased the carrier confinement with the possibility of bandgap engineering. The main results are improved emission properties and a larger range of available emission frequencies. The use of microcavity resonators[3] opens a new level of miniaturization and allows a better coupling to fiber optics.

The implementation of QDs as active material[4,5] promises even greater benefits. It is generally agreed that advantages of QDs are linked to the discrete nature of an energy spectrum associated with localized carrier states. This has led to predictions of further increased efficiency for laser applications, which are based on less temperature-dependent operational parameters[6], further reduced threshold currents and higher differential gain[7], higher modulation bandwidth[8] and reduced anti-guiding effects[9] as well as a reduced sensitivity to material defects. Lasing from QD microcavities has been demonstrated[10,11] and the combination of a three-dimensional photon confinement[12,13] with QDs opens novel applications in quantum information technologies.

Will QDs become the active media of choice for future semiconductor lasers? The answer impacts on the research direction for the next generation of semiconductor lasers and is of interest for scientific understanding. Crucial to the discussion of QD advantages is knowledge of the intrinsic QD behaviour and insight into the physical mechanisms involved. Extracting the information experimentally is challenging because sample inhomogeneities make the separation of intrinsic from extrinsic properties difficult. While it is presently possible to produce QW lasers with performance determined by intrinsic properties, the early developmental stage of QD growth technology is such that QD laser performance will, for the foreseeable future, remain limited by inhomogeneous broadening from QD dimension or composition variations.

Attempts to circumvent the sample imperfection problem by seeking an answer theoretically had been hindered by the lack of a predictive (i.e., without free or fitting parameters) QD gain theory. Essential to any predictive gain theory is a rigorous description of scattering leading to dephasing. Polarization dephasing directly influences the magnitude and spectral broadening of absorption and gain in all semiconductor lasers. In more general context, it plays a decisive role in limiting all coherent and quantum entanglement processes. A serious weakness of typical semiconductor gain calculations is the use of phenomenological descriptions of scattering effects. Not only is the predictive capability severely compromised by having the dephasing rate as a free parameter, but experimental spectral shapes and excitation dependences are not accurately reproduced. Differences between the scattering processes in QD and QW active materials should be taken into account precisely in the comparison.

Primary sources for dephasing are carrier-carrier and carrier-phonon scattering. The manner they act, and how they need to be described, differs for discrete and continuous states. As a result one has to carefully choose a treatment that is sufficiently general to encompass both situations. Successful many-body approaches have been established, that are based on the 2nd-order Born approximation for carrier-carrier scattering and the random phase approximation for the carrier-phonon scattering[14,15]. For bulk and QW gain structures, a perturbative evaluation (which can be improved using a quasi-particle approximation) in combination with a Markov approximation (describing energy-conserving scattering processes) accurately predict experimental gain spectra over wide ranges of experimental conditions and material systems[16]. However, the same approach when applied to QDs leads to inconsistencies. In particular, the perturbation theory resulted in the prediction of strongly inhibited carrier longitudinal-optical (LO) phonon scattering, which in turn gave rise to the belief that QD device performance would be seriously degraded because of a phonon bottleneck problem[17]. Such a bottleneck is not observed in laser experiments.

The solution came from a nonperturbative treatment within the polaron picture for carrier-LO-phonon scattering[18–20]. This development together with a non-Markovian treatment of polarization dephasing due to carrier-carrier scattering within a QD system allow the completion of a truly predictive theory for QD optical response[21]. The resulting formalism contains detailed accounting of electronic structure effects and a microscopically consistent treatment of the many-body interaction. The many-body effects include the renormalizations (band-gap shrinkage and Coulomb enhancement) which contribute to shifts and amplitude modifications of absorption and emission resonances, as well as polarization dephasing, which is responsible for the intrinsic (homogeneous) broadening of these resonances and additional energy shifts. With such a microscopic theory one can

progress beyond generalities when debating the relative merits of QD versus QW lasers. The ability to calculate intrinsic optical response makes quantitative projection of ultimate laser performance possible.

2 Many-Body Theory for the Optical Properties of Quantum Dots

The starting point for the theory development are the single-particle states for a QW layer containing QDs and a wetting layer. The determination of these states needs to consider material-specific effects like band-mixing or strain as well as the quantum confinement in the heterostructure design. In a second step interaction matrix elements for the dipole-coupling to the optical field, the Coulomb interaction between excited electrons and holes as well as the carrier-phonon interaction need to be determined for the single-particle states. Dipole-selection rules provide information about possible interband transitions between the single-particle states. However, electronic excitations in the system are coupled by the Coulomb-interaction, and only the interacting states are visible in optical experiments.

To describe electronic excitations and optical properties of the systems, the single-particle density-matrix elements can be used, which contain the occupation probabilities $f_n^{e,h}$ of the single particle states n for electrons and holes, respectively, as well as interband-transition amplitudes p_{nm} between the states n and m with the single-particle energies ε_{nm}. Equations of motion for the density-matrix elements are coupled by various interaction processes among themselves and to higher-order correlation functions. Methods of many-body theory are used to obtain closed equations[14–16].

The optical properties of the system can be determined from the microscopic interband transition amplitudes. Their equations of motion have the structure[16,21]

$$\frac{dp_{nm}}{dt} = -i\omega_{nm}p_{nm} - i\Omega_{nm}(f_{en} + f_{hm} - 1) + S_{nm}^{c-c} + S_{nm}^{c-p}. \tag{1}$$

Here the renormalized transition and Rabi frequencies, ω_{nm} and Ω_{nm}, contain the single-particle transition frequency $(\varepsilon_n + \varepsilon_m)/\hbar$ and the Rabi energy of the optical field $\mu_{nm}E/\hbar$, respectively, as well as Hartree-Fock Coulomb-interaction contributions. Additionally, S_{nm}^{c-c}, S_{nm}^{c-p} describe the complex carrier-carrier and carrier-phonon correlations, respectively. Similar to QW and bulk calculations, the carrier-carrier and carrier-phonon interaction remains treated in the 2nd-order Born and random phase approximation, respectively. However, because of the discrete energy levels in QDs, two factors make numerical calculations rather demanding. First, it is now more important to properly evaluate the scattering integrals by using self-consistently renormalized energies rather than the more readily available single-particle ones. Second, memory effects are non-negligible, so that the Markov approximation should no longer be used. Of particular importance is a non-perturbative approach based on the polaron picture for the carrier-LO phonon scattering. The extra effort involves solving for the time dependent polaron Green's function, in addition to keeping track of memory effects and energy renormalizations, as in the Coulomb correlation case.

With the many-body contributions the set of equations (1) takes the form of integro-differential equations. The high-dimensional scattering integrals are additionally complicated by non-Markovian contributions and the requirement of self-consistency. Numerical solutions of microscopic models on this level are only possible with the resources provided at the John von Neumann Institute for Computing.

3 Optical Spectra of the Quantum-Dot Active Material

For studying the optical properties of the highly-excited QD active material, we consider the linear optical response to a weak laser probe field E. This allows to determine absorption and gain spectra[21] which directly influence the laser performance. The calculations also provide access to derived quantities and effects like the modulational response, refractive index changes, the anti-guiding factor or linewidth enhancement effects[22]. Near the threshold of a laser device, the occupation probabilities of excited carriers $f_n^{e,h}$ are to a good approximation Fermi-Dirac functions which are determined by the total density of excited carriers.

The solution of Eq. (1) provides the intrinsic QD gain. In Figure 1 calculated gain spectra (negative absorption) are shown, which include many-body correlations due to carrier-carrier scattering, and interaction with LO-phonons, both evaluated with self-consistently renormalized single-particle energies. As in the result for dephasing due to Coulomb interaction, we obtain the bleaching and red-shift of the resonances due to many-body interactions and also the saturation of the QD gain.

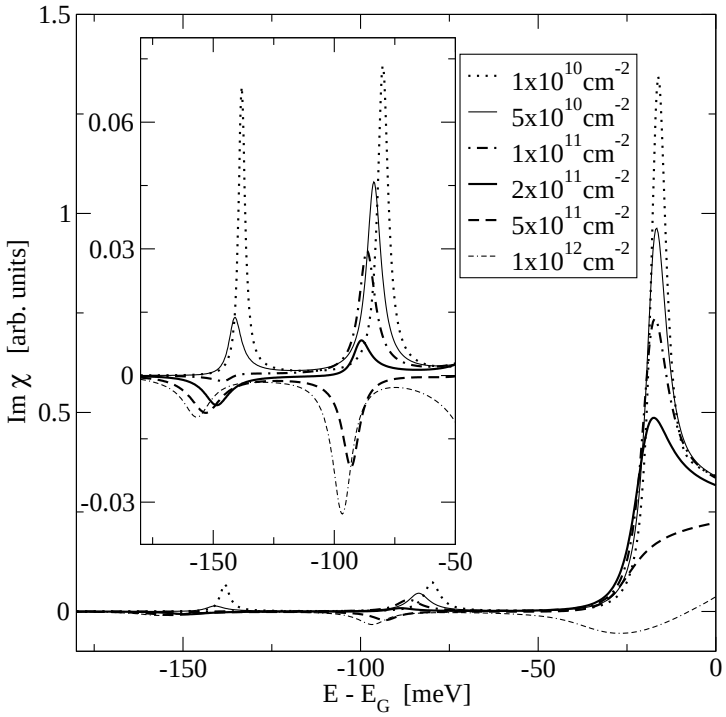


Figure 1. Imaginary part of the optical susceptibility, which determines the optical absorption (Im $\chi > 0$) and gain (Im $\chi < 0$), for the combined QD-WL system including interaction-induced dephasing and line shifts due to Coulomb interaction and carrier-phonon interaction. The results are displayed for various total carrier densities in the system at room temperature. The inset shows a scale up of the QD resonances. From Ref. 21.

The influence of the QD confinement potential and inhomogeneous broadening on the peak gain is demonstrated in Figure 2. Calculations are performed for $In_{0.3}Ga_{0.7}As$ and InAs QDs. In each case, the QDs are embedded in GaAs QWs and the actual QD shape is approximated by a disk. For the $In_{0.3}Ga_{0.7}As$ QDs, a $2nm$ height and $18nm$ diameter are chosen to give a simple electronic structure, consisting of only one localized electronic state and one localized hole state. For the InAs QDs, the height and diameter of $3nm$ and $20nm$, respectively, are chosen to obtain a ground-state emission wavelength at around 1.5μm, which is interesting for optical fiber communication. The two structures enable us to examine the two very different scenarios of shallow and deep quantum confinement. The computed QD spectral resonances for both structures exhibit noticeable increase in linewidth with increasing carrier density, indicating strong dephasing rate dependence on excitation. Additionally, energy renormalizations give rise to red shifts of the QD resonances and refractive index spectra, not predicted by an atom-like theory.

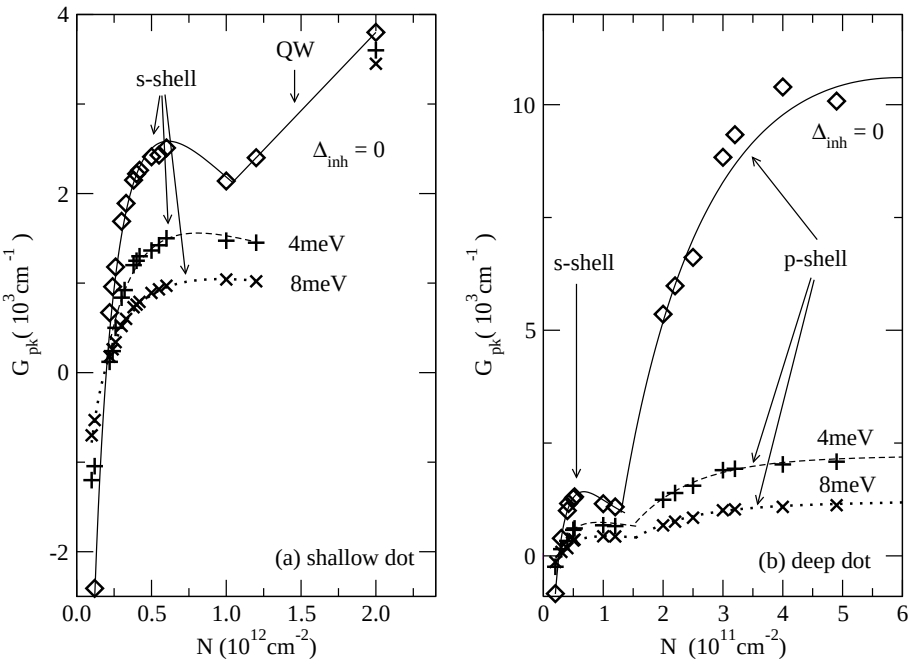


Figure 2. Peak gain versus carrier density for different inhomogeneous broadening Δ_{inh} due to QD size and composition fluctuations. The left picture is for a shallow QD with one confined s-shell. In the right picture, a deeper QD with confined s- and p-shells is considered. From Ref. 22.

Gain saturation is of critical importance for laser applications, because it limits the useful range of excitation densities. As in QW and bulk structures a contribution to gain saturation is the limitation of state filling according to the Pauli principle. However, for

QDs, there is an additional contribution from the appreciable increase in the dephasing rate with excitation.

As a standard quantity to characterize semiconductor lasers, the α-factor describes both, the linewidth enhancement as well as antiguiding effects due to the interplay of excitation-induced refractive-index and gain changes[23]. As an example, large values for $|\alpha|$ lead to declining beam quality of edge emitters. For the above-discussed QD model systems, calculations of the α-factor have been performed. In the range of small optical gain we find small values of the α-factor, while in the regime of gain saturation $|\alpha|$ increases drastically. An advantage over QW lasers is only expected near the laser threshold.

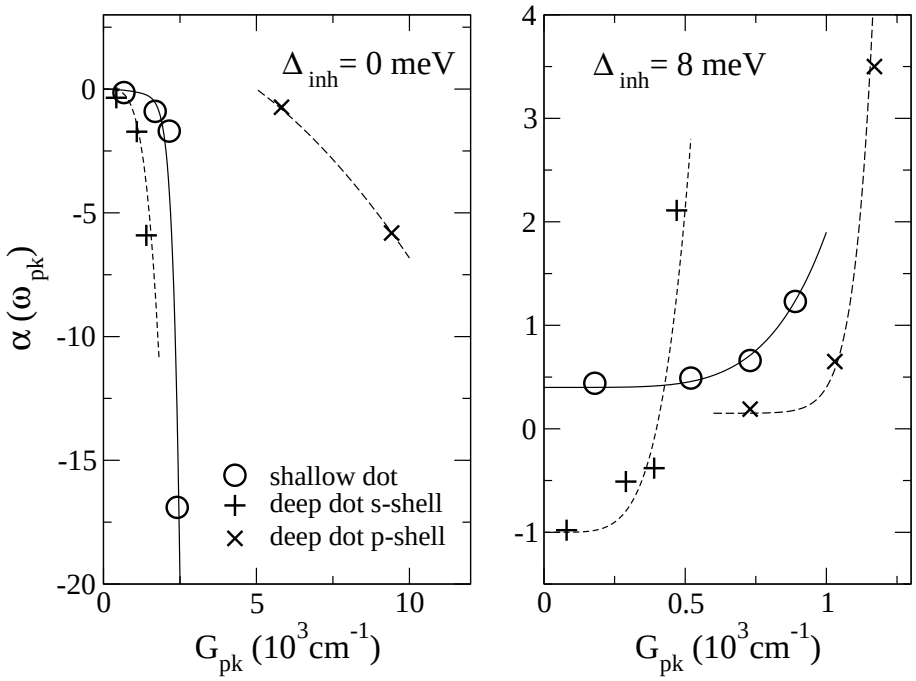


Figure 3. α-factor versus peak gain for deep and shallow dot and (a) without inhomogeneous broadening (b) with inhomogeneous broadening of 8 meV. From Ref. 22.

In summary, the developed microscopic theory and its evaluation with supercomputers allows to explore the interplay of intrinsic QD physics and extrinsic effects due to sample imperfections. A microscopic theory with a quantum-kinetic treatment of scattering is necessary because little is known quantitatively about the dephasing rates. The QD active material shows a large sensitivity to saturation effects. A large QD density and improvements in the reduction of the inhomogeneous broadening due to QD fluctuations in present samples are necessary to obtain superior performance in comparison to QW lasers[24].

Acknowledgements

The work is supported by the Deutsche Forschungsgemeinschaft. We gratefully acknowledge a grant for CPU time from the John von Neumann Institute for Computing.

References

1. H. C. Casey and M. B. Panish, *Heterostructure Lasers - Part A: Fundamental Principles*, Academic Press, New York, 1978.
2. R. Dingle and C. H. Henry, Quantum effects in heterostructure lasers. U.S. Patent 3982207, 1976.
3. P. L. Gourley, *Microstructured semiconductor lasers for high-speed information processing*, Nature, **371**, 571, 1994.
4. N. N. Ledentsov, V. M. Ustinov, A. Y. Egorov, A. E. Zhukov, M. V. Maksimov, I. G. Tabatadze, and P. S. Kop'ev, *Optical properties of heterostructures with InGaAs-GaAs quantum clusters*, Semiconductors, **28**, 832–834, Aug. 1994.
5. N. Kirstaedter, N.N. Ledentsov, M. Grundmann, D. Bimberg, V.M. Ustinov, S.S. Ruvimov, M.V. Maximov, P.S. Kop'ev, Zh.I. Alferov, U. Richter, P. Werner, U. Gosele, and J. Heydenreich, *Low threshold, large T_o injection laser emission from (InGa)As quantum dots*, Electronics Letters, **30**, no. 17, 1416–1417, Aug 1994.
6. Y. Arakawa and H. Sakaki, *Multidimensional quantum well laser and temperature dependence of its threshold current*, Appl. Phys. Lett., **40**, 939, 1982.
7. M. Asada, Y. Miyamoto, and Y. Suematsu, *Gain and the threshold of three-dimensional quantum-box lasers*, IEEE J. Quantum Electron., **QE-22**, 1915, 1986.
8. S. M. Kim, Y. Wang, M. Keever, and J. S. Harris, *High-frequency modulation characteristics of 1.3 μm InGaAs quantum dot lasers*, IEEE Photonics Technology Letters, **16**, 377, 2004.
9. T. C. Newell, D. J. Bossert, A. Stintz, B. Fuchs, K. J. Malloy, and L. F. Lester, *Gain and linewidth enhancement factor in InAs quantum-dot laserdiodes*, IEEE Photonics Technology Letters, **11**, 1527, 1999.
10. M. Nishioka, R. Schur, M. Kitamura, H. Watabe, and Y. Arakawa, *Light emission from vertical-microcavity quantum dot laser structures*, PHYSICA B, **227**, no. 1-4, 404–406, 1996.
11. Hideaki Saito, Kenichi Nishi, Ichiro Ogura, Shigeo Sugou, and Yoshimasa Sugimoto, *Room-temperature lasing operation of a quantum-dot vertical-cavity surface-emitting laser*, Applied Physics Letters, **69**, no. 21, 3140–3142, 1996.
12. K. Vahala, *Optical Microcavities*, Nature, **424**, 839, 2003.
13. P. Lodahl, A. Floris van Driel, I. S. Nikolaev, A. Irman, K. Overgaag, D. Vanmaekelbergh, and W. L. Vos, *Controlling the dynamics of spontaneous emission from quantum dots by photonic crystals*, Nature, **430**, 654, 2004.
14. H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport & Optics of Semiconductors*, Springer, Berlin, 1. edition, 1996.
15. W. Schäfer and M. Wegener, *Semiconductor Optics and Transport Phenomena*, Springer-Verlag, Berlin, 1. edition, 2002.
16. W. W. Chow and S. W. Koch, *Semiconductor-Laser Fundamentals*, Springer, Berlin, 1999.

17. H. Benisty, C. M. Sotomayor-Torres, and C. Weisbuch, *Intrinsic mechanism for the poor luminescence properties of quantum-box systems*, Phys. Rev. B, **44**, 10945, 1991.
18. T. Inoshita and H. Sakaki, *Density of states and phonon-induced relaxation of electrons in semiconductor quantum dots*, Phys. Rev. B, **56**, 4355, 1997.
19. K. Kral and Z. Khas, *Electron self-energy in quantum dots*, Phys. Rev. B, **57**, 2061, 1998.
20. J. Seebeck, T. R. Nielsen, P. Gartner, and F. Jahnke, *Polarons in semiconductor quantum-dots and their role in the quantum kinetics of carrier relaxation*, Phys. Rev. B, **71**, 125327, 2005.
21. M. Lorke, T. R. Nielsen, J. Seebeck, P. Gartner, and F. Jahnke, *Influence of carrier-carrier and electron-phonon correlations on optical absorption and gain in quantum-dot systems*, Phys. Rev. B, **73**, 085324, 2006.
22. M. Lorke, F. Jahnke, and W. Chow, *Excitation Dependence of Gain and Carrier Induced Refractive Index Changes in Quantum Dots*, Appl. Phys. Lett., **90**, 051112, 2007.
23. H. Haug and S. W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, World Scientific Publ., Singapore, 4. edition, 2004.
24. W. W. Chow, M. Lorke, and F. Jahnke, unpublished.

Simulation of Magnetic Friction

Martin P. Magiera and Dietrich E. Wolf

Faculty of Physics and CeNIDE, University of Duisburg-Essen
Lotharstr. 1, D-47048 Duisburg, Germany
E-mail: m.magiera@uni-duisburg.de

In this project, the occurence and phenomenology of magnetic friction is studied, using the classical anisotropic Heisenberg model. We consider a magnetic tip, moved parallel to a surface of atomistic spins, analogous to a magnetic force microscope tip scanning a surface or the reading head of a hard disk. The system is simulated dynamically by solving the stochastic Landau-Lifhsitz-Gilbert equation (LLG). Benchmark tests illustrate the excellent scaling behaviour of our code. Results from supercomputer simulations provide deeper understanding of the temperature dependence of magnetic friction.

1 Motivation

The physics of friction is one of the oldest scientific problems studied. Dissipative processes are omnipresent and responsible for important phenomena. While on the macroscopic scale their phenomenology is well known, still actual studies on the micron and nanometer scale yield interesting new aspects[1]. On the other hand magnetic materials which nowadays can be controlled down to the nanometer scale become more and more important. Improvements in magnetism are usually directly applicable (e.g. the giant magneto resistance effect in the data storage industry), hence progress in this field is additionally stimulated by economy. The trend towards smaller and smaller devices makes them sensitive to thermal influences, e.g. heating of magnetic memory devices leads to data loss. Because friction is usually connected with heat generation, friction in magnetic materials will gain importance.

However, for a long time dissipative processes in magnetic materials have been neglected, they attract interest only recently[2–4]. For this reason, experimental studies are rare. An energy dissipation effect has been measured recently by Wiesendanger[5] using *exchange force microscopy*, where forces between an oscillating tip and antiferromagnetic surface atoms are measured in an atomistic resolution. But the dissipative effect could not be quantified and its origin has not been clarified yet, hence a theory is not established. One possible explanation assumes the generation of spinwaves excited by the tip motion. To clarify such assumptions, computer simulations are necessary. In recent studies dissipation has been detected in Ising solids with relative motion along a contact plane[3], and for special cases it has been analytically solved[6]. The authors found a fluctuation-induced friction force, which does not depend on the velocity (like Coulomb friction) and vanishes at zero temperature.

As the Ising model is treated by Monte-Carlo simulations it does not contain intrinsic dynamics – flip events are caused by thermal fluctuations, thus only strongly anisotropic materials can be modelled that way. In order to study the weakly anisotropic case, we consider in this paper the atomistic Heisenberg model, which evolves in time by the stochastic Landau-Lifshitz-Gilbert equation (LLG[7,8]) and thus possesses intrinsic dynamics. Besides,

as the Ising model can be regarded as the strong anisotropic limiting case of the Heisenberg model, a comparative study is desirable. To get a setup which is reproducible by experimentalists, we model a tip which is moved parallel to the solid's surface, analogous to the scanning tip of a magnetic force microscope or the reading head of a hard disk drive.

2 Simulation Model

Basically the studied system consists of two parts; the substrate and the tip, moved parallel to the substrate. The substrate is modeled by magnetic moments (normalized "spins" $\mathbf{S}_i{=}\boldsymbol{\mu}_i/\mu_s$, μ_s denoting the saturation magnetization) on a lattice $L_x{\times}L_y{\times}L_z$ with lattice constant a. They are allowed to change their orientation in any of the three spin dimensions, but not their position. The atomistic character prohibits a change of the length of the spins. The substrate spins underly the nearest neighbour exchange coupling, expressed by the Heisenberg Hamiltonian

$$\mathcal{H}_{\mathrm{ex}} = -J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \tag{1}$$

where the angular brackets represent a summation over nearest neighbours and J the exchange constant (for ferromagnets $J{>}0$, for antiferromagnets $J{<}0$). An anisotropy term is necessary due to the solid's crystalline structure. The simplest one can be expressed by

$$\mathcal{H}_{\mathrm{ani}} = -d_x \sum_i S_{i,x}^2 - d_z \sum_i S_{i,z}^2, \tag{2}$$

where the magnetization prefers either an *easy axis* if the corresponding anisotropy constant $d_{x/z}$ is positive, or an *easy plane* if it is negative. Parallel to the substrate surface, in this work two lattice constants above the surface, the tip is positioned and moved parallel to the substrate with scanning velocity $\mathbf{v}$ (cf. Figure 2). It is modeled by a magnetic dipole moment $\mathbf{S}_{\mathrm{tip}}$ with a fixed magnetization, chosen about 1000 times larger than the substrate spins. It interacts with each substrate spin by long-range dipole-dipole interaction

$$\mathcal{H}_{\mathrm{tip}} = -w \sum_{i=1}^{N} \frac{3\,(\mathbf{S}_i \cdot \mathbf{e}_i)(\mathbf{S}_{\mathrm{tip}} \cdot \mathbf{e}_i) - \mathbf{S}_i \cdot \mathbf{S}_{\mathrm{tip}}}{R_i^3}, \tag{3}$$

where $R_i = |\mathbf{R}_i|$ denotes the norm of the distance vector $\mathbf{R}_i = \mathbf{r}_i - \mathbf{r}_{\mathrm{tip}}$, and $\mathbf{e}_i$ its unit vector $\mathbf{e}_i = \mathbf{R}_i/R_i$. $\mathbf{r}_i$ and $\mathbf{r}_{\mathrm{tip}}$ are the position vectors of the substrate spins and the tip respectively. w quantifies the dipole-dipole-coupling between the substrate spins and the tip, and is set to $w{=}0.01J$ in this work. The total Hamiltonian $\mathcal{H}$ consists of the sum of these three contributions. While the tip magnetization is fixed, each substrate spin is evolved in time with respect to local fields $\mathbf{h}_i$, defined below, due to the LLG equation

$$\frac{\partial}{\partial t}\mathbf{S}_i = -\frac{\gamma}{(1+\alpha^2)\mu_s}\left[\mathbf{S}_i \times \mathbf{h}_i + \alpha\, \mathbf{S}_i \times (\mathbf{S}_i \times \mathbf{h}_i)\right], \tag{4}$$

where $\gamma{=}1.76 \times 10^{11}(\mathrm{Ts})^{-1}$ is the absolute value of the gyromagnetic ratio. The first cross product in Eq. (4) describes the spin precession ("Lamor precession"), which can be derived from Heisenberg's equation of motion in the classical limit. The second product represents the relaxation of the moments towards their local fields ("Gilbert damping")

weightened by the phenomenlogical damping constant α, which denotes the coupling to a heat bath. The local fields are defined by $\mathbf{h}_i = -\partial\mathcal{H}/\partial\mathbf{S}_i + \boldsymbol{\zeta}_i(t)$ and include the coupling of each substrate spin to a heat bath through the stochastic vector $\boldsymbol{\zeta}_i(t)$. It represents white noise with average and correlator

$$\langle \boldsymbol{\zeta}_i(t) \rangle = 0 \quad \text{and} \tag{5}$$

$$\langle \zeta_i^{\kappa}(t)\zeta_j^{\lambda}(t') \rangle = 2\frac{\alpha\mu_s}{\gamma}\, k_B T \delta_{i,j}\delta_{\kappa,\lambda}\delta(t-t'), \tag{6}$$

where T is the temperature, $\delta_{i,j}$ means the absence of correlation between lattice sites and $\delta_{\kappa,\lambda}$ among different coordinates. This nonlinear stochastic differential equation can be solved using the Heun integration scheme[9]. A friction force can be calculated from energy dissipation rates by $F{=}v^{-1}dE_{\mathrm{diss}}/dt{=}v^{-1}\mathcal{P}_{\mathrm{diss}}$. Accordingly we deal with energy rates (powers $\mathcal{P}$) in the following. From time derivatives of $\mathcal{H}$ two powers can be distinguished[10]. The *pumping* power term represents the energy pumped by the tip motion into the substrate, and can be calculated from the derivative $\mathcal{P}_{\mathrm{pump}}{=}\mathbf{v}\cdot\partial H_{\mathrm{tip}}/\partial\mathbf{r}_{\mathrm{tip}}$. The second term represents the dissipation of energy inside the substrate, i.e. the transfer to the heatbath. This *dissipated power* is given by

$$\mathcal{P}_{\mathrm{diss}} = \frac{\gamma\alpha}{\mu_s(1+\alpha^2)} \sum_{i=1}^{N} (\mathbf{S}_i \times \mathbf{h}_i)^2 \tag{7}$$

for $T{=}0$. This term cannot be used to calculate the friction force for $T{>}0$, because its extension has severe numerical drawbacks, as explained in Ref. 10. We present this term to stress that the occurence of energy dissipation is always connected with a misalignment between the substrate spins and their local fields, in other words the continuously driving of spins out of equilibrium. Equilibrium is reached when each spin points in the direction of its local field and thus freezes. Such a scenario cannot occur in a driven system.

3 Computational Details and Performance

In the past we have shown that neglecting internal long-range interactions, namely the stray field energy, is justified for our tasks. Accordingly we were able to implement a very efficient *domain decomposition* parallelization (DDC), where only the domains' borders are communicated between nearest neighbours. This technique leads to a linear scaling at JSC supercomputers (cf. Figure 1). To keep simulation time in a reasonable range, we additionally implemented a "conveyor belt" technique (for details see Ref. 10). However, the greatest bottleneck of the task is caused by the implementation of thermal effects for several reasons. On one hand, the calculation of the Gaussian distributed random numbers consumes more than 60% of computing time, although we use optimized techniques like the Marsenne twist generator and the Ziggurat method[11]. On the other hand stochastic simulations require several statistically independent simulation runs to get proper statistics. To reduce simulation time, we developed the *random number sharing* (RNS) technique. When a parameter set of e.g. several velocities is simulated, each simulation run may use the same random numbers without loss of statistics. Cashing of random numbers is not possible because a large amount of them is required (each timestep $k{=}3N$ when N is the total number of spins). Using RNS, several parameter sets are started independently. Every p^{th} timestep

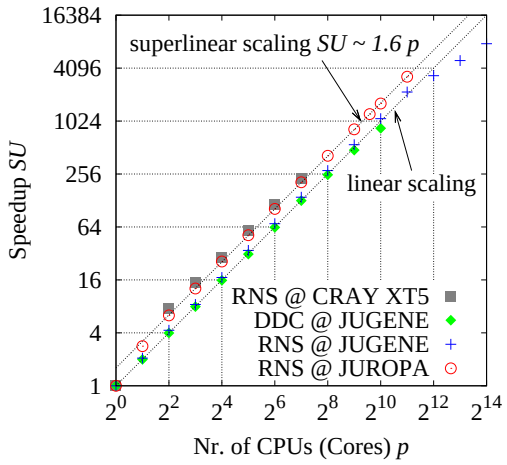


Figure 1. Scaling tests for the random number sharing technique (RNS) and the domain decomposition technique (DDC). The lowest dotted straight represents linear scaling, above superlinear scaling is present.

each of the p processors calculates k random numbers which are communicated globally. In a naive ansatz we assume that each processor needs a time $t_{\rm gen}k/p$ for generation of the random numbers ($t_{\rm gen}$ is the effort to generate *one* random number), $t_{\rm com}(p-1)k/p$ for receiving, $t_{\rm com}k/p$ for sending and $t_{\rm calc}k$ for all the rest of the calculations. Comparing the serial $p{=}1$ case (with $t_{\rm com}{=}0$) to the parallel case, the ks cancel out and the resulting speedup is

$$SU = p\,\frac{t_{\rm gen}+t_{\rm calc}}{\frac{t_{\rm gen}}{p}+t_{\rm com}+t_{\rm calc}} \overset{p\to\infty}{=} p\,\frac{t_{\rm gen}+t_{\rm calc}}{t_{\rm com}+t_{\rm calc}} = p\,a. \tag{8}$$

Accordingly, the RNS technique is suitable if the generation of random numbers accounts for the major part of computing time and $t_{\rm com}$ is short compared to $t_{\rm gen}$. Then a superlinear speedup with a constant $a{>}1$ is conceivable. In fact scaling tests verify such a behaviour with $a{\sim}1.6$ at *JuRoPA* and a *CRAY XT5* (cf. Figure 1). In other words: When many CPUs are used for the simulations, less net computing time is used than in the serial case. In practice we use hybrid types of RNS and DDC to make simulations of large systems accessible, where additionally memory access times gain importance.

4 Magnetic Friction: From Thin Films to Slabs

Our first studies of magnetic friction dealt with monolayers of spins ($50{\times}30{\times}1$), representing thin magnetic films. By studying these systems, we could develop a model for energy dissipation[10]. Both, simulations at $T{=}0$ and simulations at $T{>}0$ yield a viscous friction force for velocities $v{<}0.2\frac{\gamma J a}{\mu_s}$, i.e. friction raises linearly with the velocity. While at $T{=}0$ the corresponding friction coefficient is a linear function of the damping constant α, at $T{>}0$ it has a more complicated behaviour (cf. Figure 3). From the model it is known that the viscosity is caused by each substrate spin adiabatically following its local field (cf.

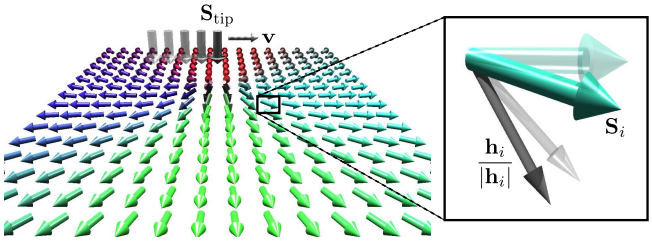


Figure 2. System stup: The magnetic tip is moved parallel to a magnetic substrate, here a monolayer. The substrate spins align due to the local fields, whose main contributions come from the dipolar tip-field. Dissipation is caused by the lagging of the substrate spins behind their local fields, illustrated in the magnification. Transparent objects denote orientations in the past.

Figure 2). It can be shown that the dissipated energy is proportional to the lagging of the substrate spins behind their local fields. How fast the substrate spins can follow depends on relaxation times. While at $T{=}0$ the only relaxation time is given by $\frac{\mu_s}{\gamma J \alpha}$, at $T{>}0$ an additional thermal relaxation time exists, resulting in temperature dependent friction coefficients[12]. The limiting case is the so-called high temperature regime, where friction does not depend on α anymore. In our previous work[12] the characterics of the decrease with rising temperature could not be clarified. Therefore, we extend our research to systems with the same width and length ($50{\times}30$), but different thicknesses and anisotropies.

4.1 Dependence on the Anisotropy for Monolayers

The thermal properties of magnetic materials are influenced by anisotropy, especially when a thin film, i.e. a quasi-two-dimensional structure, is considered. While a two-dimensional Heisenberg ferromagnet with an easy axis possesses a twofold groundstate and therefore long-range order – and undergoes a ferro-/paramagnetic phase transition at the Curie temperature (which depends on the anisotropy constant) – its isotropic pendant possesses an infinitely degenerated groundstate without long-range order and therefore no ordered phase (Mermin-Wagner theorem). The two-dimensional Heisenberg ferromagnet with strong easy plane anisotropy undergoes a Kosterlitz-Thouless transition, and therefore does not possess a ferromagnetic groundstate either. Because these three cases of anisotropy describe very different physical systems, it is astonishing that all of them yield the same friction coefficients in the high temperature regime as shown in Figure 3. The "transition temperatures" were identified with the maximum position of the specific heat, and are about $k_B T_0 {\sim} 0.7J$ for *all* anisotropies.

All systems have the paramagnetic phase in common. The universal behaviour for $T{>}T_0$ is comprehensible, when the tip is treated as a perturbation of the paramagnet, which has forgotten about its anisotropy. To make this more clear, we rewrite the local field term:

$$\mathbf{h}_i = \boldsymbol{\zeta}_i(t) - \partial_{\mathbf{S}_i}\mathcal{H}_{\mathrm{ex}} - \partial_{\mathbf{S}_i}\mathcal{H}_{\mathrm{ani}} - \partial_{\mathbf{S}_i}\mathcal{H}_{\mathrm{tip}}. \tag{9}$$

In the paramagnetic regime, the exchange term vanishes. Furthermore, it is intuitive that for high temperatures the large amplitude of the thermal field lets the anisotropy term become negligible. Then only the tip term can locally induce order into the paramagnet[12], giving rise to magnetic friction. The higher the temperature, the less order can be induced.

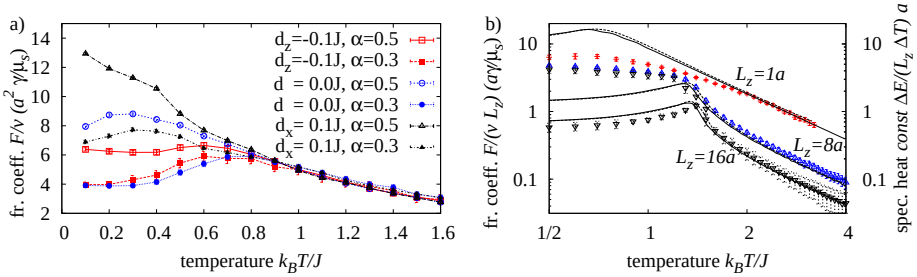


Figure 3. a) Friction coefficients for three different types of anisotropy (easy x-axis with d_x=$0.1J$, easy x/y-plane with d_z=$-0.1J$ and no anisotropy d=0) for a monolayer. While for low temperatures the anisotropy impact on the friction coefficient is high, in the high temperature regime universal behaviour is present.
b) Lines: Specific heat, calculated from aniosotropic Heisenberg systems *without a tip* with the size $16^2\times L_z$ (solid) and from the substrate-tip-systems (dashed), multiplied by a constant (the same for all L_z) and normalized by substrate thickness in a logarithmic scale. The peak, denoting the ferro-/paramagnetic transition temperature, is shifted to higher temperatures for thicker systems. Points: Friction coefficients, normalized by the substrate thickness. In the low temperature regime the coefficients depend nearly linearly on the substrate thickness. In the high temperature regime (T>T_0) they decrease as power laws of temperature, for L_z>a independent of L_z.

The specific decay of friction coefficients vs. temperature will be discussed in the next chapter. The ordering by the tip is the origin of the universal transition temperature. The tip induces well defined groundstates for the isotropic and easy-plane systems, which would possess an infinite degeneracy and no ordered phase otherwise. This remains even true irrespective of spin anisotropy in the transition region. Once a global spin configuration is stabilized, the friction mechanism is a different one. The tip has to perform work against the magnetization. The anisotropy can promote this process (e.g. easy plane anisotropy and α=0.5) and lower the resulting friction force, or inhibit it (e.g. easy axis anisotropy) and enhance the friction force. In the following we assume d_z=$-0.1J$ and α=0.5.

4.2 Dependence on the Substrate Thickness

As an indicator for the transition into the paramagnetic state, the specific heat can be studied, which peaks at the transition temperature. Thus we calculated the specific heat from the energy for both, a substrate-tip system and a substrate without a tip. For the tipless system we chose the size $(16a)^2\times L_z$. The specific heat for both systems has the same characteristics: The transition temperature is slightly shifted to a higher temperature for the substrate-tip systems, and the high temperature slopes coincide exactly (cf. Figure 3b)). When the substrate thickness is increased, the transition temperature approaches the known critical bulk temperature k_BT_c=$1.44J$ – a direct consequence of the rising impact of the exchange term in Eq. (9) due to the increasing amount of effectively interacting neighbours. First the characteristics of the friction coefficient in the high temperature phase are discussed. In Figure 3 b) the friction coefficients are plotted normalized by the substrate thickness. They collapse to one universal curve for high temperatures. The coefficients decrease with power laws above T_0. The corresponding exponents equal intriguingly the exponent, which describes the decrease of the specific heat in the paramagnetic regime. The coincidence of these completely independently measured quantities leads to the con-

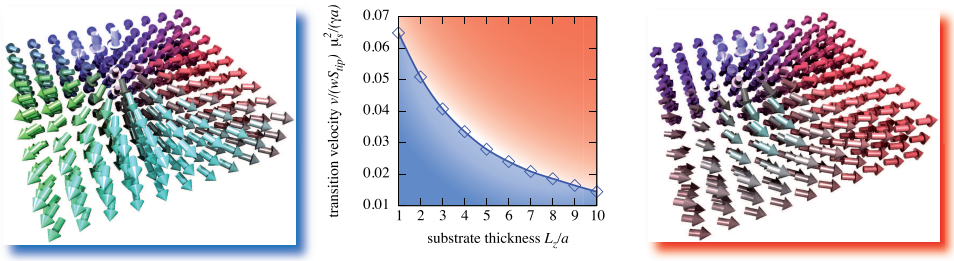


Figure 4. Phase diagram and two states observed in the simulations (for $50\times30\times L_z$-systems). For a strong tip field, low velocities or thin systems, the (blue) cylindrically symmetric state is adopted, due to the large impact of the dipole interaction. When the tip field is smaller or the substrate thickness is increased, as well as when velocity is raised, the "strained" (red) ferromagnet is adopted. This state leads to friction coefficients, which are at least one order of magnitude weaker than those in the cylindrically symmetric state.

clusion that magnetic friction in the paramagnet, which is a fluctuation induced dissipation, is directly connected with the specific heat.

In the low-temperature regime, the friction coefficients increase approximately linearly with the thickness, resulting in collapsing curves in Figure 3b). This rise in the friction force is plausible, because the thicker the system is, the more spins are exposed to the tip field and the more spins contribute to friction force. At $L_z{=}16a$ two steady states can be distinguished, which are explained in the following. When the system consists of only one monolayer, the dominant interaction is the (cylindrically symmetric) dipole-dipole interaction. Accordingly the system adopts a cylindrically symmetric state, as shown in Figure 4. When the tip magnetization is lowered towards zero, the impact of the dipole field decreases until the system switches into the *strained ferromagnet state*: Zero field would lead to a ferromagnetic state (due to the open boundaries and the strongly confined size we have even for $L_z{=}a$ a pseudo-longrange-order), while a small tip field disturbs the ferromagnet, which is then *strained*. When the tip field is strong, but the scanning velocity is so high that the system can not align in the dipole field any more, the strained state is adopted as well. The transition velocity is proportional to the tip magnetization. Now let us assume that the system thickness is enhanced: The exchange interaction gains more importance (this is a direct consequence of the larger amount of nearest neighbours), so the transition velocity gets lower. This effect can be observed in the phase diagramm in Figure 4. For the simulated velocities (usually $v{\leq}0.2\frac{\gamma Ja}{\mu_s}$) the transition becomes apparent for a thickness of $L_z{=}16a$. Because in the strained ferromagnet the influence of the tip is strongly localized, the friction coefficients are at least one order of magnitude lower. In Figure 3b) the transition becomes apparent through the existence of two friction coefficient branches, where the stronger one is measured at small velocities and the weaker one at higher velocities. In the paramagnetic regime, the exchange term in Eq. (9) becomes effectively zero, hence the strained ferromagnet state disappears at $T{\sim}T_0$.

Finally we want to stress the importance of supercomputing for these results. To get the presented signal-noise ratio for the $T{>}T_0$ data of the $L_z{=}8a$ system, we occupied one *JUMP* node for 20 days in SMP mode (30 seeds, corresponding to 35.000 CPUh). Results obtained from 20 different seeds were not sufficient to obtain the presented power law, accordingly we can conclude that the application of supercomputing was crucial to gain the relationship between specific heat and friction force.

5 Conclusion

In this work we reported the successful parallelization of our code. Thanks to the use of supercomputers, we could present measurements of magnetic friction for different anisotropies in two-dimensional systems. We have shown that friction in the paramagnet is independent of the anisotropy. Second, measurements of temperature dependent friction coefficients in three-dimensional systems have been shown. Besides a rise in the transition temperature we found an explanation for the characteristics and origin of friction in the paramagnet. The coefficient drops with the same exponent as the specific heat. Enhancing the thickness of the substrate stabilizes a new steady state for higher velocities, which minimizes the exchange energy.

Acknowledgements

We thank Lothar Brendel, Fred Hucht and Ulrich Nowak for several fruitful discussions. Financial support was granted by the German Research Foundation (DFG) via SFB 616. Computing time at the John von Neumann Institute for Computing (NIC) is gratefully acknowledged.

References

1. B. N. J. Persson, *Sliding Friction*, Springer, Berlin, Heidelberg, New York, 1998.
2. J. Ortín and J. Goicoechea, *Dissipation in quasistatically driven disordered systems*, Phys. Rev. B, **58**, 5628–5631, 1998.
3. D. Kadau, F. Hucht, and D. E. Wolf, *Magnetic Friction in Ising Spin Systems*, Phys. Rev. Lett., **101**, 137205, 2008.
4. C. Fusco, D. E. Wolf, and U. Nowak, *Magnetic friction of a nanometer-sized tip scanning a magnetic surface*, Phys. Rev. B, **77**, 174426, 2008.
5. R. Schmidt, C. Lazo, H. Holscher, U. H. Pi, V. Caciuc, A. Schwarz, R. Wiesendanger, and S. Heinze, *Probing the Magnetic Exchange Forces of Iron on the Atomic Scale*, Nano Letters, **9**, 200–204, 2009.
6. F. Hucht, *Non-equilibrium phase transition in an exactly solvable driven Ising model with friction*, submitted to Phys. Rev. E, arXiv:0909.0533, 2009.
7. L. D. Landau and E. M. Lifshitz, *On the Theory of the dispersion of magnetic permeability in ferromagnetic bodies*, Phys. Z. Sowjetunion, **8**, 153–169, 1935.
8. T. L. Gilbert, *A Phenomenological Theory of Damping in Ferromagnetic Materials*, IEEE Trans. Magn., **40**, 3443–3449, 2004.
9. J. L. García-Palacios and F. J. Lázaro, *Langevin-dynamics study of the dynamical properties of small magnetic particles*, Phys. Rev. B, **58**, 14937, 1998.
10. M. P. Magiera, L. Brendel, D. E. Wolf, and U. Nowak, *Spin excitations in a monolayer scanned by a magnetic tip*, EPL (Europhysics Letters), **87**, 26002, 2009.
11. D. B. Thomas, W. Luk, Ph. H.W. Leong, and J. D. Villasenor, *Gaussian random number generators*, ACM Comput. Surv., **39**, 11, 2007.
12. M. P. Magiera, D. E. Wolf, L. Brendel, and U. Nowak, *Magnetic Friction and the Role of Temperature*, IEEE Trans. Magn., **45**, 3938–3941, 2009.

Three-Body Strongly Correlated Polar Molecules in Optical Lattices

Lars Bonnes and Stefan Wessel

Institut für Theoretische Physik III, Universität Stuttgart
Pfaffenwaldring 57, 70550 Stuttgart, Germany
E-mail: wessel@theo3.physik.uni-stuttgart.de

We employ large-scale quantum Monte Carlo simulations to study the properties of ultra-cold bosonic polar molecules on a two-dimensional optical lattice of hexagonal symmetry. Following a recent proposal by Büchler et. al. [Nat. Phys. **3**, 726 (2007)], such a system is described by an extended Bose-Hubbard model of hard-core bosons, that includes both extended two- and three-body repulsions. Using quantum Monte-Carlo simulations, exact finite cluster calculations and the tensor network renormalization group, we explore the rich phase diagram of this system, resulting from the strongly competing nature of the three-body repulsions on the honeycomb lattice. We develop effective descriptions of the various phases in terms of emerging valence bond crystal states and quantum dimer models.

1 Three-Body Repulsions on Optical Lattices

Competing interactions together with quantum fluctuations are known to allow for exotic phases to emerge in strongly correlated quantum systems. This route towards novel phases of matter has been explored in recent years, also in the context of ultra-cold quantum gases on optical lattices, where one gains a high degree of control of the interaction strength[1]. The dominant inter-particle potentials in such systems are often well described by two-body interaction and exchange terms. It is, however, also fascinating to explore realistic set-ups of many-body quantum systems that are dominated – via engineered interaction potentials – by multi-body interaction terms of e.g. three-particle type. Recently, polar molecules have been proposed as promising candidates towards realizing such many-body systems. Driven by significant progress towards producing degenerate gases of polar molecules[2–6], various proposals have been put forward, how to drive ultra-cold polar molecules into regimes of strong many-body interaction effects[7–11]. In a recent work by Büchler et. al.[10] an effective interaction potential was derived for polar molecules on an optical lattice in the presence of static electric and microwave fields. It was found, that upon appropriately tuning the external fields, the interactions between the polar molecules become characterized by extended strong two- as well as three-body interactions.

In a parameter regime, where the leading first-order contribution to the two-body terms has been tuned to zero, the residual interactions take the form

$$V_{\mathrm{eff}} = \frac{1}{2}\sum_{ij} V_{ij} n_i n_j + \frac{1}{6}\sum_{ijk} W_{ijk} n_i n_j n_k, \tag{1}$$

where $V_{ij} = V/r_{ij}^6$ and $W_{ijk} = W/(r_{ij}^3 r_{jk}^3 + \mathrm{perm.})$, and W and V can be tuned to be of similar strength, $V \gtrsim W$[10]. Here, r_{ij} denotes the spatial separation between particles on lattice sites i and j, and n_i the local density at site i.

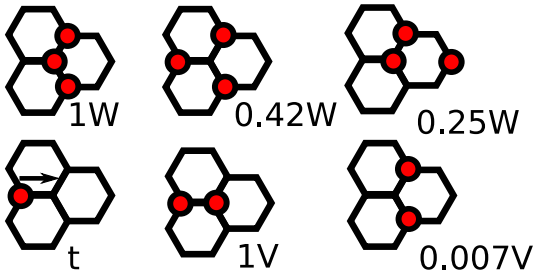


Figure 1. Top row: Leading three-body repulsion terms and their relative strengths on the honeycomb lattice. Bottom row: Illustration of the nearest neighbour hopping term, and the relative strengths of the leading two-body repulsion terms on the honeycomb lattice.

In the following, we consider in particular the case of bosonic polar molecules. In this case, the suppressed tunneling of a particle to an already occupied site needs to be accounted for by a hard-core constraint on the bosonic occupations[10]. Since the extended interactions decay rapidly with the inter-particle separations (c.f. Figure 1), we truncate the interactions after the leading terms, resulting thus in an extended Bose-Hubbard model with two- and three-body nearest neighbour interactions only,

$$H = -t \sum_{\langle ij \rangle} \left(b_i^\dagger b_j + \text{h.c.} \right) + V \sum_{\langle ij \rangle} n_i n_j + W \sum_{\langle ijk \rangle} n_i n_j n_k - \mu \sum_i n_i. \qquad (2)$$

Here, b_i and $b_i^\dagger$ denote boson annihilation and creation operators respectively, and the local density operator $n_i = b_i^\dagger b_i$ has eigenvalues 0 and 1 in the hard-core limit. Furthermore, t denotes the nearest-neighbour hopping matrix element, and μ is a chemical potential, allowing to control the filling (i.e. the density) of the system between $n = 0$ (empty) and $n = 1$ (full).

In previous works, models of hard-core bosons with extended two- and three-body interactions as effective models of ultra-cold polar molecules were studied for the case of a one-dimensional optical lattice[12] and the two-dimensional square lattice geometry[13]. In the one-dimensional case, an incompressible phase at a filling $n = 2/3$ was established for dominant three-body interactions, stabilizing both change-density wave (CDW) and bond-order wave (BOW) long-ranged correlations, apart from conventional CDW phases appearing at half-filling ($n = 1/2$) in the presence of two-body interactions. On the square lattice, several solid phases at fractional fillings as well as supersolid phases were found in a semi-classical approximation, some of which could also be verified by full numerical simulations. Here, we extend such systematic explorations to the case of the honeycomb lattice, where nearest-neighbour three-body repulsions lead to characteristic effects of strong frustrations.

2 Quantum Monte Carlo

In order to explore the physics of this system, we used a combination of quantum Monte Carlo (QMC) simulations based on the generalized directed loop algorithm within the

stochastic series expansion (SSE) representation[14–16], as well as exact finite cluster calculations and the tensor network renormalization group approach[17] for the classical ($t = 0$) limit of the above Hamiltonian. In the following, we discuss the results of our calculations on the ground-state phase diagram in the absence of two-body interactions (i.e. for $V = 0$), thus focusing on the main aspects of three-body repulsions on the honeycomb lattice. We observe states with large degeneracies in the classical limit, and the emergence of complex valence bond crystal (VBC) phases, to be described in detail below.

In order to perform the SSE QMC simulations for the current model, we employed a cluster decomposition of the Hamiltonian in terms of trimers of nearest-neighbour sites, such that each cluster carries one of the three-body interaction terms. In the SSE directed loop construction, we thus used a doubly-linked list of 6-leg vertices. While the algorithm performed well for large values of the hopping t, we were not able to reach significantly below $t/W \approx 0.1$ due to the dynamical freezing in the Monte Carlo configurations, once the competing diagonal interaction terms dominate the Hamiltonian (which is a general issue of the SSE for Hamiltonians dominated by large diagonal terms). In terms of the simulation temperature T, we find that an inverse temperature $\beta = 1/T = 20 - 50W$ ($k_B = 1$) provided an optimal trade-off between the finite temperature incoherence and the algorithmic performance (the SSE algorithmic cost scales linearly in β). Furthermore, in order to be commensurate with all superstructures identified in the incompressible phases, the linear system size L is required to be an integer multiple of 6 (the total number of lattice sites being $N = 2L^2$, as the honeycomb lattice contains two sites per unit cell, forming the two sublattices A and B). Within the above temperature regime, we were able to simulate systems up to $L = 24$, in some cases up to $L = 36$ in linear extent.

In phases with large superstructures – described below – we find that the autocorrelation times of the bosonic structures increase such that the algorithm is not able to describe tunnelling between different realizations of the ordering pattern even within several 10^6 QMC sweeps, but resides in one particular sector of the ground-state manifold (which however varies upon performing independent runs with different random-number streams for a given set of model parameters). During a first stage of the thermalization process, we annealed the system in a cyclic way by heating it up and cooling it down slowly such that it is able to relax globally into one of the equivalent ground-states. A fixed temperature thermalization was then performed in the second stage of the thermalization process. This annealing approach yields better performance than parallel tempering over an extended temperature range, since we are mainly interested in the ground-state phase diagram of the system, where β is large.

We furthermore employed quantum parallel tempering[18] (in μ or t) at fixed low temperatures in certain regions of the phase diagram, in particular in order to study the quantum phase transitions between neighbouring VBC phases.

3 The Valence Bond Crystal at Filling 9/16

The ground state phase diagram in Figure 2 summarizes the results from our analysis. It exhibits at low values of t/W a variety of incompressible phases of different (unconventional) fillings $n = 9/16$, $5/8$ $(= 10/16)$, $2/3$ and $3/4$ $(= 12/16)$. Due to the incompressible nature of these incompressible phases, they lead to finitely extended plateaus in the μ-dependence of the density n, as shown e.g. for fixed $t/W = 0.3$ in the inset of Figure 2.

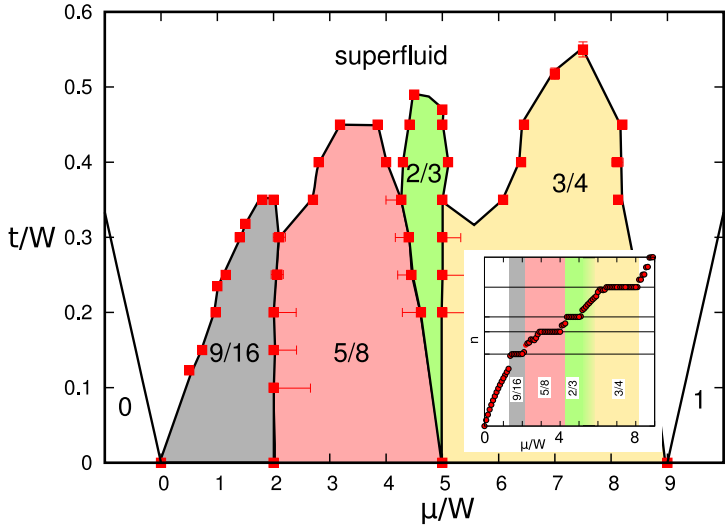


Figure 2. Ground-state phase diagram of hard-core bosons on the honeycomb lattice for $V = 0$ in terms of μ/W and t/W. The incompressible phases at fillings $n = 9/16$, $5/8$, $2/3$ and $3/4$ are labeled by n and underlayed by different colours. Uncertainties on the estimated phase boundaries are indicated by error bars. The inset shows the density n as a function of the chemical potential μ/W for fixed $t/W = 0.3$, linear system size $L = 12$ and an inverse temperature of $\beta = 20W$.

For larger values of $t/W \gtrsim 0.4$, the system is eventually driven by its kinetic energy from these solid phases via first-order quantum melting transitions into a uniform superfluid phase with a finite superfluid density ρ_s. In the QMC simulations, the superfluid density is obtained as $\rho_s = \langle w^2 \rangle / (\beta t)$ from measuring the bosonic winding number w fluctuations in the standard way[19] (here, $\beta = 1/T$ ($k_B = 1$) denotes the inverse temperature). The first-order nature of the melting transitions follows from pronounced jumps that are observed upon crossing the transition lines in both the density and the superfluid density.

In the following, we only discuss the nature of the $n = 9/16$ phase in more detail. The first density plateau that is encountered upon filling the lattice has a density $n = 9/16$ and extends between $0 \leq \mu/W \leq 2$ in the classical limit ($t = 0$). It has an potential energy per lattice site (equal to the internal energy for $t = 0$) of $E_{pot}^{(9/16)} = -9/16\mu$, and corresponds to the closest packing of hard-core bosons on the honeycomb lattice without introducing any three-body repulsions. From geometrical considerations in the classical limit, this bosonic structure is obtained by covering the lattice with equilateral triangles with side length $4\sqrt{2}a$ (where a is the distance between two lattice sites), each covering 16 lattice sites. These triangles are filled by nine bosons in a staggered (checkerboard) arrangement in order to obtain the overall filling $9/16$. Neighbouring triangles differ in the placement of the checkerboard pattern. This leads to domain walls along the edges of the triangles, where pairs of bosons reside. For $V = 0$, this does, however, not lead to a potential energy penalty. For an illustration of this structure, see the left panel of Figure 3. This structure allows for a denser packing of the particles (and thus a higher filling) than

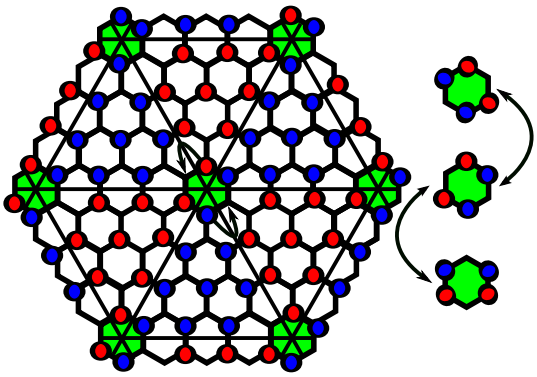


Figure 3. Illustration of the classical configuration in the $n = 9/16$ plateau (left panel). The bosons are shown in two different colours, in order to underline the different checkerboard patterns within neighbouring triangles. The green hexagons illustrate those plaquettes, where the bosons are allowed to change positions as illustrated in the right panel, without changing the potential energy. For finite hopping $t > 0$, this leads to resonances within these hexagonal plaquettes (right panel).

the overall checkerboard state of filling $n = 1/2$, without introducing any three-body energy terms. On every forth hexagon (such as the hexagons indicated in the left panel of Figure 3), six triangles share a common corner. In the classical limit, the energy of the system remains unchanged, if two particles change their positions along such a hexagon by an angle of $2\pi/6$, as indicated in Figure 3. This local move results in a classical ground state degeneracy W, that scales linearly with the system size, $W \propto N$. Thus, the ground state entropy is not extensive, and the entropy per site $S = (\ln W)/N$ vanishes in the thermodynamic limit.

For finite values of t, the local moves allowed in the classical configurations lead to resonances on the hexagonal plaquettes, corresponding to second-order hopping processes of the bosons, effectively rotating a hexagon by an angle of $2\pi/3$, as illustrated in the right panel of Figure 3. In this way the system is able to gain kinetic energy from these tunneling processes. Such resonances also provide the dominant quantum fluctuations on this density plateau, and stabilize a VBC phase with a superstructure of checkerboarded triangles, linked by the resonating hexagons. We can indeed identify these peculiar features of the $9/16$ plateau phase from the QMC simulations. In Figure 4, we show the local density $\langle n_i \rangle$, along with the kinetic energy density per bond $\langle K_{ij} \rangle$, where $K_{ij} = b_i^\dagger b_j + b_j^\dagger b_i$ and sites i and j belong to a nearest neighbour bond on the honeycomb lattice, for a representative point within the $n = 9/16$ phase. We find the local density is close to $\langle n_i \rangle = 1$ and 0, thus exhibiting only few fluctuations, in a staggered (checkerboard) pattern within triangular structures. On the other hand, the density around a subset of the hexagons – those, where the triangular checkerboard patterns meet – is within an intermediate range, and it is along the bonds of these hexagons, where the kinetic energy is mainly located. Both observations indicate a residual density dynamics in this incompressible phase, located along these hexagons. The resonant nature of the corresponding hopping events is reflected in the bond-bond correlation function $\langle K_{ij} K_{kl} \rangle$, shown in Figure 5. We identify the main contribution from the hexagonal resonances, which lead to the enhanced bond-bond corre-

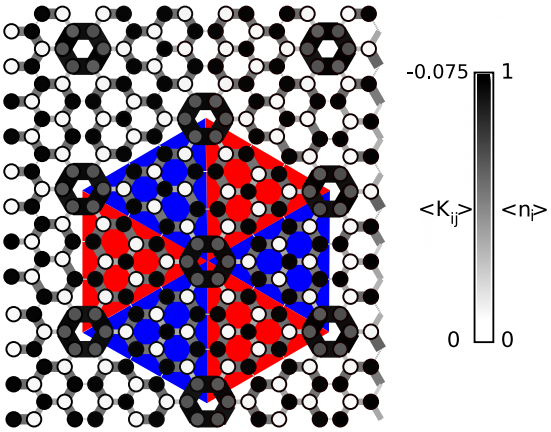


Figure 4. QMC data of the local density $\langle n_i \rangle$ (shading) and the kinetic energy density along the nearest-neighbour bonds $\langle K_{ij} \rangle$ (line thickness and shading) for bosons on the honeycomb lattice in the $n = 9/16$ VBC phase at $V = 0$, $t/W = 0.2$, $\mu/W = 1$, and a system with $L = 12$ at $\beta = 20$. For clarity, the equilateral triangles, together forming the unit cell of the VBC superstructure, are indicated as well.

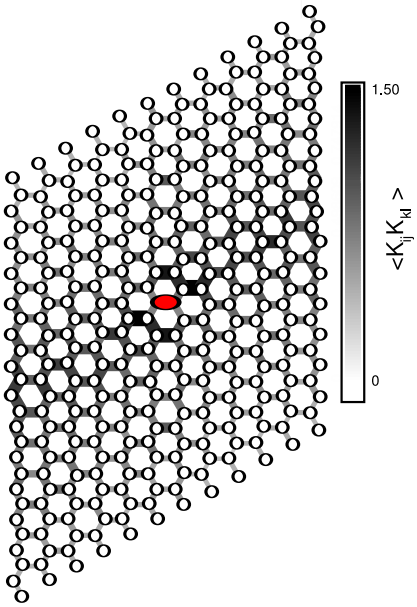


Figure 5. QMC data for the bond-bond correlations in the kinetic energy $\langle K_{ij} K_{kl} \rangle$ along the nearest-neighbour bonds for bosons on the honeycomb lattice in the $n = 9/16$ VBC phase at $V = 0$, $t/W = 0.2$, $\mu/W = 1$, and a system with $L = 12$ at $\beta = 20$. The reference bond $\langle ij \rangle$ is indicated by the red ellipse.

lation between the reference bond (marked by an ellipse) and the bonds atop and below the reference bond in Figure 5. Furthermore, correlations of the same strength are visible between the reference bond and its next-nearest neighbour bonds to the left and right. These result from residual quantum fluctuations that lead to the finite kinetic energy distribution inside the checkerboarded triangular structures in Figure 4. We do not observe long ranged bond-bond correlations, thus the hexagonal resonances evolve essentially independently of each other.

4 Conclusions

We studied a model of hard-core bosons with strong three-body repulsions on the honeycomb lattice using quantum Monte Carlo simulations. The system's ground state phase diagram exhibits besides a superfluid region several complex valence bond crystal phases at fractional fillings $9/16$, $5/8$, $3/2$ and $3/4$. We presented details on the nature of the $9/16$ phase, here. The quantum phase transitions between neighbouring valence bond crystal phases are furthermore consistent with first-order transitions (not shown here). With regard to a possible experimental realization based on cold polar molecules in appropriately tuned external electric and microwave fields, we also included in addition to the three-body repulsions also nearest neighbour two-body interactions, which in a realistic set-up are at least of the same strength as the three-body interactions. Considering the competition between the different interaction terms, we then obtained a cascade of intermediate incompressible plateaus as the two-body interaction strength increases.

A further step towards modeling cold polar molecules on the honeycomb lattice would be to take into account the full long-ranged nature of both the three-body and two-body interactions[10]. Treating the longer ranged contributions appropriately appears to require the use of different calculational techniques, since already the leading contributions to both interaction sectors provided a challenge for the quantum Monte Carlo approach.

Acknowledgements

We should like to thank K. P. Schmidt, A. Läuchli, R. Moessner, and A. Muramatsu for interesting discussions, and acknowledge the allocation of CPU time on the high performance computers at HLRS Stuttgart and NIC Jülich, where the numerical calculations have been performed. L. B. furthermore acknowledges support from the Studienstiftung des Deutschen Volkes, and S. W. from the DFG within the SFB/TRR 21.

References

1. I. Bloch, J. Dalibar, and W. Zwerger. Many-body physics with ultracold gases. *Rev. Mod. Phys.* 80:885, 2008.
2. J. M. Sage, S. Sainis, T. Bergeman, and D. DeMille. Optical production of ultracold polar molecules. *Phys. Rev. Lett.* 94:203001, 2005.
3. K.-K. Ni, S. Ospelkaus, M. H. G. deMiranda, A. Pe'er, B. Neyenhuis, J. J. Zirbel, S. Kotochigova, P. S. Julienne, D. S. Jin, and J. Ye. A high phase-space-density gas of polar molecules. *Science* 322:231, 2008.

4. J. Deiglmayr, A. Grochola, M. Repp, K. Morlbauer, C. Gluck, J. Lange, O. Dulieu, R. Wester, and M. Weidemüller. Formation of ultracold polar molecules in the rovibrational ground states. *Phys. Rev. Lett.* 101:133004, 2008.
5. F. Lang, K. Winkler, C. Strauss, R. Grimm, and J. Hecker Denschlag. Ultracold triplet molecules in the rovibrational ground state. *Phys. Rev. Lett.* 101:133005, 2008.
6. J. G. Danzl, E. Haller, M. Gustavsson, M. J. Mark, R. Hart, N. Bouloufa, O. Dulieu, H. Ritsch, and H.-C. Nägerl. Quantum gases of deeply bound ground state molecules. *Science* 321:1062, 2008.
7. S. Kotochigova and E. Tiesinga. Controlling polar molecules in optical lattices. *Phys. Rev. A* 73:041405, 2006.
8. A. Micheli, G. K. Brennen, and P. Zoller. A toolbox for lattice-spin models with polar molecules. *Nature Phys.* 2:341, 2006.
9. D.-W. Wang, M. D. Lukin, and E. Demler. Quantum fluids of self-assembled chains of polar molecules. *Phys. Rev. Lett.* 97:180413, 2006.
10. H. P. Büchler, A. Micheli, and P. Zoller. Three-body interactions with cold polar molecules. *Nature Physics* 3:726, 2007.
11. T. Lahaye, C. Menotti, L. Santos, M. Lewenstein, and T. Pfau. The physics of dipolar bosonic quantum gases. *Rep. Prog. Phys.* Report arXiv:0905.0386.
12. B. Capogrosso-Sansone, S. Wessel, H. P. Büchler, P. Zoller, and G. Pupillo. Phase diagram of one-dimensional hard-core bosons with three-body interactions. *Physical Review B (Condensed Matter and Materials Physics)* 79(2):020503, 2009.
13. K. P. Schmidt, J. Dorier and A. M. Laeuchli. Solids and supersolids of three-body interacting polar molecules on an optical lattice. *Phys. Rev. Lett.* 101:150405, 2008.
14. A. W. Sandvik. Stochastic series expansion method with operator-loop update. *Phys. Rev. B* 59:R14157, 1999.
15. O. F. Syljuåsen and A. W. Sandvik. Quantum monte carlo with directed loops. *Phys. Rev. E* 66:046701, 2002.
16. F. Alet, S. Wessel, and M. Troyer. Generalized directed loop method for quantum monte carlo simulations. *Phys. Rev. E* 71:036706, 2005.
17. M. Levin and C. P. Nave. Tensor renormalization group approach to two-dimensional lattice models. *Phys. Rev. Lett.* 99:120601, 2007.
18. P. Sengputa, A. W. Sandvik, D. K. Campbell. Bond-order-wave phase and quantum phase transitions in the one-dimensional extended Hubbard model. *Phys. Rev. B* 65:155113, 2002.
19. E. L. Pollock and D. M. Ceperley. Path-integral computation of superfluid densities. *Phys. Rev. B.* 36(16):8343–8352, 1987.

Computational Soft Matter Science

Kurt Kremer

Max Planck Institute for Polymer Research, 55021 Mainz, Germany
E-mail: kremer@mpip-mainz.mpg.de

Soft Matter is the generic term for all complex materials, which have elastic constants much smaller than conventional (crystalline) materials. Such small elastic constants go in hand with characteristically small cohesive energy density and/or characteristic mesoscopic length scales. Thus typical examples of soft matter are biological and synthetic polymers and membranes but also colloidal suspensions. The characteristic energy scale is the thermal energy k_BT and thus energetic and entropic contributions typically contribute roughly equally to most macroscopic properties. To obtain such properties conformational and morphological fluctuations have to be considered properly. Due to the complicated chemical structure of most soft materials (colloidal systems can be different under certain circumstances) analytic theory is only able to treat rather idealized limiting cases and computer simulations often are the only satisfactory link between theory and experiment. Theoretically it is extremely difficult to account for the complicated correlations originating from the macromolecular character of the systems beyond certain limiting cases (i.e. small concentrations) and experimentally it is almost impossible to provide systems so well defined and controlled that a direct comparison with theory is possible. In this context computer simulations on a variety of levels of detail are absolutely crucial and an extensive supply of CPU power is a most important precondition. For this the CPU time and support provided by the John von Neumann Institute for Computing is instrumental for many researchers in Germany and beyond. In most cases mid size systems have to be followed by means of simulations for a long time. There, computers of the JUMP class are the tool of choice. However, for certain applications, where significant finite size effects are faced and the system itself is not too complicated on a local scale also Blue Gene like architectures are excellent sources of CPU power. In the following we will find examples for both kinds of problems.

The first contribution deals with the statistical mechanics of **bottle-brush polymers** by Binder and co-workers. Bottle-brush polymers are conceptually rather simple models of branched polymeric systems, however, already display a huge complexity in behaviour. Since it became possible to synthesize this kind of systems they provide perfect test cases and model systems also for applications. The reason for that is that their conformational properties depend very sensitively on e.g. solvent properties which makes them ideal test candidates for molecular based nano-actuators. In the present simulation studies systems on a rather generic level are investigated. The back bone as well as the side chains which are grafted to the backbone in order to achieve a typical bottle-brush configuration are modeled by either simple lattice or bead spring chains. This reduction is however needed in order to be able to investigate in a satisfactory way the statistical properties of such bottle-brush polymers. As a function of grafting density and solvent quality the conformational properties are investigated in detail and predictions for experiments are given in terms of

the competition of side chain and backbone properties. A special example is the limiting case of a rigid backbone in marginal and poor solvent. Though this might look a bit artificial at the beginning such systems are relatively close to typical polymers employed in organic electronics where aliphatic side chains are needed in order to make these systems processable at all. The conformational properties of theses systems and the characteristic pearl-necklace like structure might give valuable hints how these objects can arrange as a function of solvent quality.

In a second contribution by Daoulas and Müller on **Diblock Copolymers on Patterned Surfaces** an even coarser model to simulate polymers is chosen in order to be able to treat almost macroscopic systems. Basis for their study is a Gaussian bead spring model for polymer chains which contain a block of A-beads and a block of B-beads. In order to account for the interaction of these beads with the surrounding, the average concentration of beads of other chains is accounted for by a background density function. By this the complicated many chain interactions are reduced to a single chain problem resulting in a very efficient and highly parallel single chain simulation. By matching the characteristic block length and densities they are able to map their parameters onto typical experimental systems. In this context the authors study the kinetics of self assembly on a surface and in particular the competition between order due to characteristic length scales of the underlying surface patterning and the intrinsic length scales of the typical block copolymer morphology. The examples are symmetric diblock copolymers which in bulk would form nice lamellar structures. Due to the pattern of a substrate where the interaction of the substrate is different for the two different blocks, the competition of the different length scales can lead to rather new morphologies and functionalities at the surface. In extension of this methodology first studies on rheological properties have also been performed leading the way to a rather extensive surface characterization of complex block copolymer systems at patterned surfaces.

Another class of very important soft matter systems are biological membranes. In the contribution of West and Schmid on **Membrane-Protein Interactions in Lipid Bilayers** the two authors study the situation between a simplified molecular picture and the typical elastic continuum theories. Bio-membranes are very interesting structures which result from a physical self-assembly process. Embedded in these two-dimensional liquid structures are membrane proteins and their function is controlled to a certain extent by the lateral organization of these proteins in the membrane. This makes it a very interesting problem for physicists. It allows for studies on a rather generic level in order to understand basic principles which govern the spatial order of embedded proteins, treated as simple membrane inclusions. One important aspect of this is the so-called hydrophobic mismatch, which occurs when the hydrophilic and the hydrophobic parts of the membrane proteins do not match the hydrophilic and hydrophobic parts of the membrane. This leads to an elastic distortion of the membrane which mediates long range interactions between these proteins. Such long range interactions are best studied by generic coarse grained lipid models such as the one developed in the group of F. Schmid. Since the essential properties of the lipid molecules which form the membrane are still taken into accounts characteristic deviations of protein-protein interactions compared to continuum elastic theories can be studied and analyzed in detail. Such simulation approaches then can be used to couple long range effects to more specific investigations on a local molecular basis.

So far, three different contributions have been introduced which make best use of moderately parallel computers containing fairly powerful compute cores.

The last contribution of this chapter by Gompper et al. requires much larger systems while the typical interactions are fairly simple. This is an optimal kind of problem to be dealt with by massively parallel computers such as the JUGENE in Jülich. The study of the **dynamics of semi-dilute polymer solutions under strong shear**, where full hydrodynamic interactions have to be taken into account, is only possible if systems are treated where the overall box size is significantly larger than the fully extended polymer chains. That means that we typically have to deal with system sizes one or two orders of magnitude larger than for conventional polymer simulations. The present contribution employs a multi-particle collision dynamics method to study the hydrodynamic interactions of semi-dilute polymer solutions under high shear rates. With this the stretching process of originally overlapping chains has been studied in detail as well as the characteristic mean the tilt angle of the extended polymer with respect to the shear plain. By varying the Weissenberg number for the simulation the authors could show that the minimal systems size to study experimentally realistic effects definitely require the use of massively parallel computing resources. This however opens the way to a direct comparison of experiment and theory via extended computer simulations for complex flow problems which so far was far beyond any computer simulation possibility.

All presented simulations provide significant new insight into important problems of current soft matter science. The first two chapters on bottle-brush polymers in solution and block copolymers on patterned surfaces show a level of maturity which already allows a direct comparison with experiment. They are able to provide insight and guidance for further experimental studies and in certain cases even for technologically relevant problems. The latter two studies on interactions of membrane proteins which can be characterized by a hydrophobic mismatch and the dynamics of semi-dilute polymer solutions under strong shear flow are still at an earlier stage. They point towards directions of simulation research in soft matter, where also experiments are much less developed and a precise material characterization is quite difficult. In all cases the support by the NIC was instrumental for the success of the investigations.

Phase Transitions and Relaxation Processes in Macromolecular Systems: The Case of Bottle-Brush Polymers

Hsiao-Ping Hsu[1], Wolfgang Paul[1,2], Panagiotis E. Theodorakis[1], and Kurt Binder[1]

[1] Institute for Physics, Johannes Gutenberg University Mainz
55099 Mainz, Germany
E-mail: {hsu, wolfgang.paul, theodora, kurt.binder}@uni-mainz.de

[2] Institute for Physics, Martin-Luther University Halle-Wittenberg,
06120 Halle, Germany
E-mail: wolfgang.paul@physik.uni-halle.de

As an example for the interplay of structure, dynamics, and phase behaviour of macromolecular systems, this article focuses on the problem of bottle-brush polymers with either rigid or flexible backbones. On a polymer with chain length N_b, side-chains with chain length N are endgrafted with grafting density σ. Due to the multitude of characteristic length scales and the size of these polymers (typically these cylindrical macromolecules contain of the order of 10000 effective monomeric units) understanding of the structure is a challenge for experiment. But due to excessively large relaxation times (particularly under poor solvent conditions) such macromolecules also are a challenge for simulation studies. Simulation strategies to deal with this challenge, both using Monte Carlo and Molecular Dynamics Methods, will be briefly discussed, and typical results will be used to illustrate the insight that can be gained.

1 Introduction

The so-called "bottle-brush polymers" consist of a long macromolecule serving as a "backbone" on which many flexible side chains are densely grafted[1,2]. Since the chemical synthesis of such complex polymeric structures has become possible, these molecular bottle-brushes have found much interest for various possible applications: the structure reacts very sensitively to changes of solvent quality (due to change of temperature of the solution, pH value, etc.), enabling the use of these molecules as sensors or actuators on the nanoscale[3,4]. For some conditions these bottle-brush molecules behave like stiff cylindrical rods, and hence they can serve as building blocks of supramolecular aggregates, or show orientational order in solution as nematic liquid crystals do. On the other hand, these molecular bottle-brushes are also very "soft", i.e, they show only very small resistance to shear deformation; bottle-brush molecules of biological origin such as aggrecan which occurs in the cartilage of mammalian (including human!) joints are indeed held responsible for the excellent lubrication properties (reducing frictional forces) in such joints[5].

However, for being able to control the function of these complex macromolecules one must be able to control their structure, i.e. one must understand how the structure depends on various parameters of the problem: chain length N_b of the "backbone", chain length N of the side chains, grafting density σ of the side chains along the backbone, solvent quality, just to name the most important ones of these parameters. This is the reason why computer

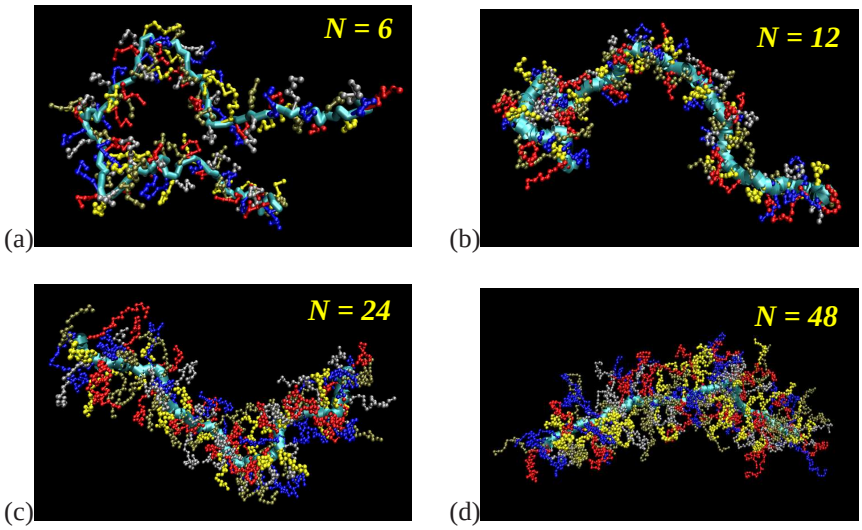


Figure 1. Snapshot pictures of simulated bottle-brush polymers with a flexible backbone containing $N_b = 131$ effective monomers. At each backbone monomer one side chain is grafted (grafting density $\sigma = 1$) and side chain lengths are $N = 6$ (a), $N = 12$ (b), $N = 24$ (c) and $N = 48$ (d). The chains are described by the bond fluctuation model on the simple cubic lattice (see Section 2) and the only interaction considered is the excluded volume interaction between effective monomers.

simulations are needed in order to understand such bottle-brushes: even in the case of a good solvent, the structure of these objects is very complicated as the snapshot pictures[6,7] (Figure 1) of the simulated models show, and there occur a multitude of characteristic length scales (Figure 2)[6,7] that are needed to describe the structure; it is very difficult to extract this information from experiment, and hence simulations are valuable here as they can yield a far more detailed picture.

Of course, the large size of these bottle-brush polymers (even the coarse-grained models described in the following sections contain of the order of several thousand or even more than 10000 effective monomers per polymer) is a serious obstacle for simulation, too; in fact, developing efficient models and methods for the simulation of macromolecular systems is a longstanding and important problem[8,9]. We hence shall address this issue of proper choice of both suitable models and efficient algorithms in the next sections.

2 Bottle-Brush Polymers with Rigid Backbones in Good Solvents

If one assumes the backbone of the bottle-brush to be completely stiff, the problem is reduced to grafting side chains of length N to a straight line. Although this limiting case seems somewhat artificial, from the point of view that one wishes to model real systems, it is a very useful test case: there is no reason to assume that approximations that already fail in this rather simple limit become accurate for the more complicated case of flexible backbones; moreover, this case is rather simple to simulate, and the analysis of the simulation data is relatively straightforward.

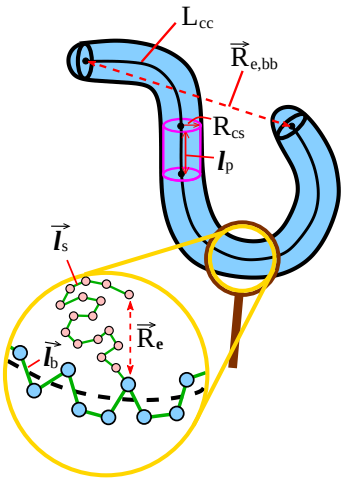


Figure 2. Multiple length scales are needed for the structural characterization of a bottle-brush polymer: a coarse-grained view describes this object as a flexible spherocylinder of length L_{cc} and cross-sectional radius $R_{\rm cs}$, which is locally straight over a length ℓ_p ("persistence length"). A further length scale of interest is the end-to-end distance $\vec{R}_{e,bb}$ of the backbone chain. This coarse-grained view is experimentally obtained by atomic force microscopy imaging of bottle-brush polymer adsorbed on substrates, for instance. A "microscope" with resolution on the scale of atoms would see the monomers of the backbone chains (connected by backbone bond vectors $\vec{\ell}_b$) and monomers of the side chains (connected by bond vectors $\vec{\ell}_s$). A mesoscopic length of interest then is the end-to-end distance $\vec{R}_e$ of the side chains.

For the study of this case, polymers were simply represented as standard self-avoiding walks on the simple-cubic lattice (i.e. beads of the chains are occupied lattice sites, connected by nearest-neighbour links on the lattice, and multiple occupancy of sites is forbidden[8,9]).

As a simulation method, pruned-enriched Rosenbluth methods (PERM)[10,11] were used[12]. In the Rosenbluth method, all side chains are grown simultaneously step by step, choosing only from sites which are not yet occupied, and the statistical weight of the polymer configuration is computed recursively. In the PERM algorithm, one does not grow a single polymer at a time, but a large "population" of equivalent chains is grown simultaneously. From time to time configurations with very low statistical weight are killed, and configurations with large statistical weight are "cloned" ("go with the winners" strategy[10]). The advantage is that (unlike dynamic Monte Carlo algorithms[8,9]) this method does not suffer from "critical slowing down" when N gets large, and data for all N (up to the maximum value studied) are obtained in a single simulation (see Ref. 11 for more details).

Now one hypothesis popular in most experimental studies (e.g. Ref. 13, 14) is the factorization approximation[15] for the structure factor $S(q)$ (that describes the small angle scattering intensity of neutrons, X-rays or light from dilute solutions of bottle-brushes polymers) into a contribution due to the backbone ($S(q)$) and a contribution due to the side chains $(S_s(q))$

$$S(q) \approx S_b(q) S_s(q) \approx S_b(q) S_{\rm cs}(q) \tag{1}$$

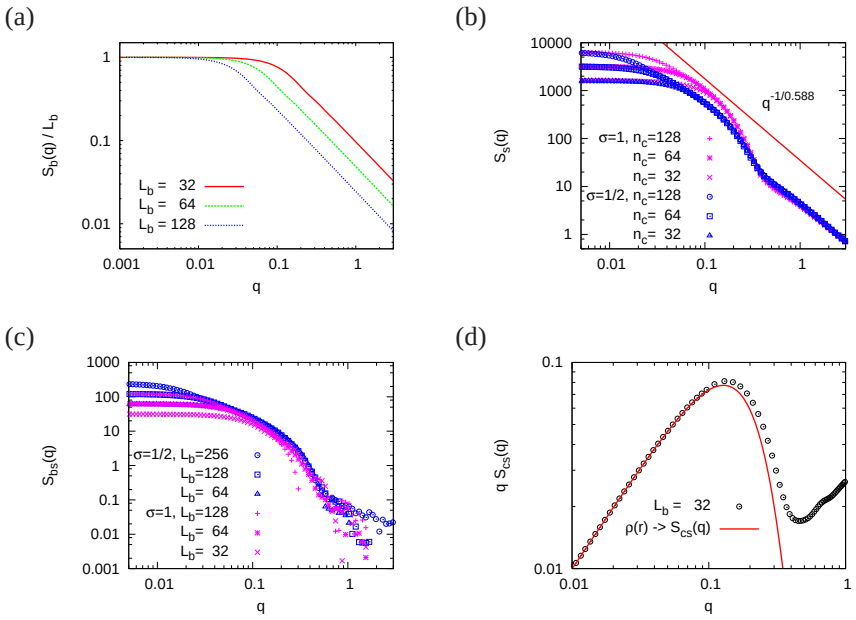


Figure 3. a) Log-log plot of the normalized scattering function of the backbone $S_b(q)/L_b$, L_b being the backbone length, $L_b = N_b a$ where $a(= 1)$ is the lattice spacing. Three choices of L_b are shown. b) Log-log plot of the scattering from all the side chain monomers, for $N = 50$, and two choices of the grafting density c) Log-log plot of the scattering due to interference contributions from monomers in side chains and the backbone (for the same parameters as in b). d) Log-log plot of $qS_{\rm cs}(q)$ vs. q in the range $0.01 \leq q \leq 1$, comparing the actual scattering due to the side chains with the prediction resulting from Eq. (2), using $\rho(r)$ as actually recorded in the simulation[12].

where q is the absolute value of the scattering wave vector. In the last step the side chain scattering was approximated by $S_{\rm cs}(q)$. This cross sectional scattering is in turn approximated related to the radial density $\rho_{\rm cs}(r)$ perpendicular to the cylinder axis at which the side chains are grafted as (the constant c ensures proper normalization[12])

$$S_{\rm cs}(q) = c^{-1} \langle | \int d^2\vec{r} \rho_{\rm cs}(r) \exp(i\vec{q}\cdot\vec{r})|^2 \rangle. \qquad (2)$$

Obviously, Eq. (1) neglects interference effects due to correlations between the monomer positions in the side chains and in the backbone. Writing the scattering due to the monomers in the side chains as the average of a square, Eq. (2) ignores correlations in the occupation probability in the z-direction along the bottle-brush axis. All such correlations do contribute to the actual structure factor, of course, when it is computed from its definition,

$$S(q) = \frac{1}{\mathcal{N}_{\rm tot}} \sum_{i=1}^{\mathcal{N}_{\rm tot}} \sum_{j=1}^{\mathcal{N}_{\rm tot}} \langle c(\vec{r}_i) c(\vec{r}_j) \rangle \sin(q|\vec{r}_i - \vec{r}_j|)/(q|\vec{r}_i - \vec{r}_j|) \quad . \qquad (3)$$

Here $c(\vec{r}_i)$ is an occupation variable, $c(\vec{r}_i) = 1$ if the site i is occupied by a bead, and

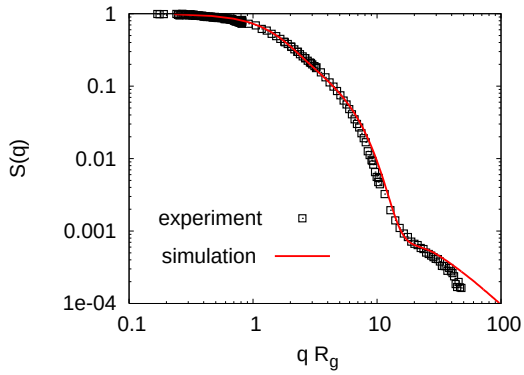


Figure 4. Structure factor $S(q)$ of a bottle-brush polymer obtained from the scattering experiment[13] mapped to the simulated model (cf. text) by requiring that the total gyration radius R_g is matched, to fix the translation factor for the length scale[7].

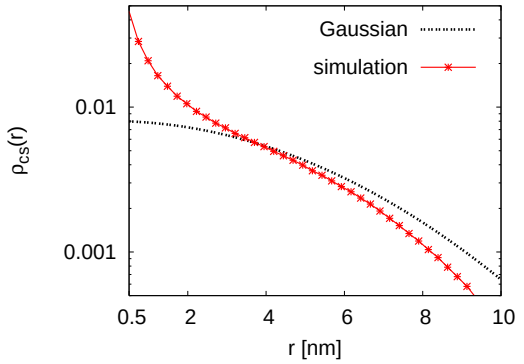


Figure 5. Log-log plot of the density profile $\rho_{\rm cs}(r)$ vs. r, for the systems shown in Figure 4. For fitting the experimental data, a Gaussian form for $\rho_{\rm cs}(r)$ was assumed[13].

zero otherwise, and the sums extend over all $\mathcal{N}_{\rm tot}$ monomers (in the side chains and in the backbone).

In the simulation, one can go beyond experiment by not only recording the total scattering $S(q)$, but also the contributions $S_b(q)$, $S_s(q)$ and $S_{\rm cs}(q)$ individually (Figure 3). One can see that the approximation $S(q) \approx S_b(q)S_s(q)$ leads to a relative error of the order of a few % only (cf. the difference in ordinate scales between parts b) and c), while the assumption $S_s(q) \approx S_{\rm cs}(q)$ [with Eq. (2)] leads to appreciable errors at large q (see part d)). Therefore the use of Eqs. (1), (2) to analyze experiments can lead to appreciable errors, when one wishes to predict[12] the radial density distribution $\rho(r)$.

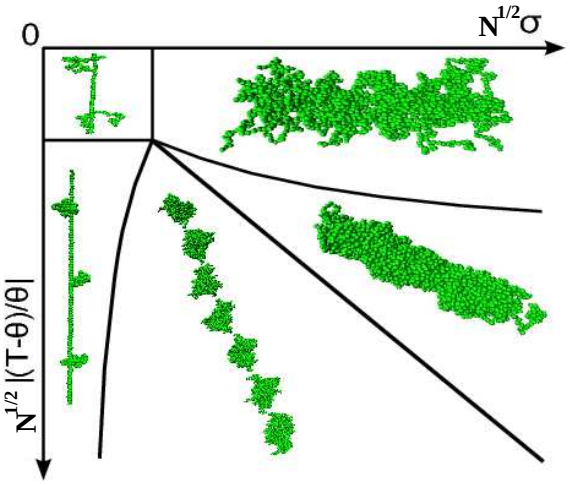


Figure 6. Schematic diagram of states of a bottle-brush polymer with a rigid backbone under poor solvent conditions in the plane of variables $x = N^{1/2}\sigma$ and $y = N^{1/2}(1 - T/\theta)$, as proposed by Scheiko et al.[19]. The lines indicate (smooth) crossovers between the different states[19], which we have characterized by snapshots from our simulations[20]. For further explanations see text.

3 Bottle-Brushes with Flexible Backbones in Good Solvents

In the case of flexible backbones, the PERM algorithm no longer provides an efficient sampling of configuration space; there "dynamic" Monte Carlo algorithms are used, but with "unphysical moves" allowing chain intersections and large configurational changes in a single step, as achieved by the Pivot algorithm[8,9]. Using the bond fluctuation model[8,9], the "L26 algorithm"[16] was used[17] for local moves allowing bond intersection and nevertheless respecting excluded volume. In this way well-equilibrated data for systems with up to $N_b = 259$ monomers at the backbone and up to $N = 48$ monomers in the side chains could be studied (for $\sigma = 1$). Examples of snapshot pictures have already been given in Figure 1.

Adjusting the physical meaning of the lattice spacing to correspond to 0.263 nm, the structure factor of the simulated model matches almost perfectly an experimental result[13] (for slightly different numbers of chemical monomers, $N_b^{\rm exp} = 400$, $N^{\rm exp} = 62$; this difference is irrelevant, since there is no one-to-one correspondence between covalent bonds and the "effective bonds" of the model), see Figure 4[7]. From the simulation one can directly extract the cross-sectional density profile $\rho_{\rm cs}(r)$ and compare it[7] to the approximate experimental result (Figure 5), which was obtained from fitting $S(q)$ to Eqs. (1), (2)[13]. One sees that the analysis of the experiment could predict roughly correctly the distance on which the profile $\rho_{\rm cs}(r)$ decays to zero, but does not account for its precise functional form. A further interesting finding[7,17] is the result that the persistence length ℓ_p (cf. Figure 2) depends strongly[7] on N_b (at least if one uses the textbook definitions[18]), and hence is not a useful measure of local chain stiffness.

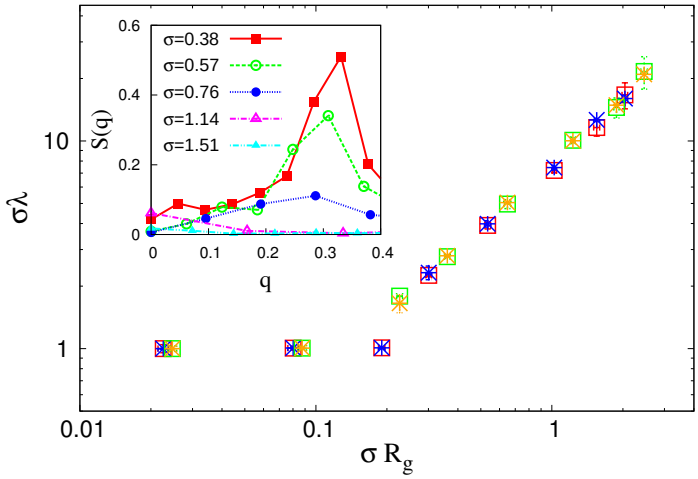


Figure 7. Log-log plot of wavelength λ (dimensionless by normalization with σ) versus grafting density σ (dimensionless by normalization with the radius of gyration of the side chains, R_g). The wavelength was extracted either from the z-dependence of a normalized correlation function at radial distance $r \approx 3$ Lennard-Jones units σ (stars) or extracted from the peak position of its Fourier transform $S(q)$ (squares) [inset]. All data refer to temperature $T = 1.5$ (in Lennard Jones units). Blue and red symbols refer to $N = 35$, orange and green ones to $N = 50$. Inset shows $S(q)$ vs. q for $N = 35$ and grafting densities $\sigma \geq 0.38$.

4 Bottle-Brushes in Poor Solvents

While isolated single chains in dilute solution collapse to dense globules[18] when the temperature T is lowered below the Theta temperature $T = \theta$, for a bottle-brush the constraint of the chains being grafted to a backbone leads to an interesting diagram of states, considering $1 - T/\theta$ and σ as variables (Figure 6)[19]. Near $T = \theta$ and $\sigma N^{1/2} \to 0$ the side chains collapse to dense globules when $N^{1/2}(1 - T/\theta) \gg 1$. For $T \approx \theta$ and $N^{1/2}\sigma \gg 1$ one has a cylindrical structure (as in the previous sections) which collapses to a dense cylinder when $N^{1/2}(1 - T/\theta) \gg 1$. However, interestingly at intermediate grafting densities and $N^{1/2}(1-T/\theta) > 1$ a laterally inhomogeneous "pearl necklace-structure" was predicted[19].

This problem can neither be studied efficiently by the PERM method nor by the bond fluctuation algorithm of Section 3 - both methods get very inefficient for dense polymer configurations. Thus, instead Molecular Dynamics simulations of the standard Grest-Kremer-type bead-spring model[8,9] were carried out[20]. Figure 7 shows[20] characteristic wavelength λ plotted vs. grafting density σ. One finds that for small σ a trivial periodicity given by $1/\sigma$ occurs, while for $\sigma R_g \geq 0.2$ ($R_g = \langle R_g^2 \rangle^{1/2}$) the periodicity is essentially independent of σ, given by about that $\lambda \approx 8R_g$. Of course, mapping out the phase behaviour precisely for a wide range of N, σ and T is a formidable problem, due to very long relaxation times of the collapsed states in Figure 6.

5 Concluding Remarks

The above examples have illustrated that polymers with complex architecture pose challenging problems with respect to their structure and thermodynamics. Investigation of their relaxation behaviour is another interesting problem, to be tackled in the future. The above examples have also illustrated the need to carefully adjust both the model and the algorithm to the problems one wants to deal with.

Acknowledgements

One of us (H.-P. Hsu) received support from the DFG (SFB625/A3), another (P. E. Theodorakis) from the MPG via a Max-Planck Fellowship of the MPI-P. We are grateful to the NIC Jülich for computer time at the JUMP and SoftComp cluster.

References

1. M. Zhang and A. H. E. Müller, J. Polym. Sci., Part A, Polym. Chem. **43**, 3461, 2005.
2. S. S. Sheiko, B. S. Sumerlin, and K. Matyjaszewski, Prog. Polym. Sci. **33**, 759, 2008.
3. T. Stephan, S. Muth, and M. Schmidt, Macromolecules **35**, 9857, 2002.
4. C. Li, N. Gunari, K. Fischer, A. Janshoff, and M. Schmidt, Angew. Chem. Int. Ed. **43**, 1101, 2004.
5. J. Klein, Science **323**, 47, 2009.
6. H.-P. Hsu, W. Paul, and K. Binder, INSIDE (2009, in press).
7. H.-P. Hsu, W, Paul, S. Rathgeber and K. Binder, preprint.
8. K. Binder (ed.) *Monte Carlo and Molecular Dynamics Simulations in Polymer Science* (Oxford Univ. Press, New York, 1995).
9. N. Attig, K. Binder, H. Grubmüller, and K. Kremer (eds.) *Computational Soft Matter: From Synthetic Polymers to Proteins* (NIC, Jülich, 2004).
10. P. Grassberger, Phys. Rev. E **56**, 3682, 1997.
11. H.-P. Hsu, W. Paul, and K. Binder, Macromol. Theory Simul. **16**, 660, 2007.
12. H.-P. Hsu, W. Paul, and K. Binder, J. Chem. Phys. **129**, 204904, 2008.
13. S. Rathgeber, T. Pakula, A. Wilk, K. Matyjaszewski, and K. L. Beers, J. Chem. Phys. **122**, 124904, 2005.
14. B. Zhang, F. Gröhn, J. S. Pedersen, K. Fischer, and M. Schmidt, Macromolecules **39**, 8440, 2006.
15. J. S. Pedersen and P. Schustenberger, Macromolecules **29**, 7602, 1996.
16. J. P. Wittmer, P. Beckrich, H. Meyer, A. Cavallo, A. Johner, and J. Baschnagel, Phys. Rev. E **76**, 011803, 2007.
17. H.-P. Hsu, K. Binder, and W. Paul, Phys. Rev. Lett. (2009, in press).
18. A. Yu Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, New York, 1994).
19. S. S. Sheiko, O. V. Borisov, S. A, Prokhorova, and M. Möller, Eur. Phys. J. E. **13**, 125, 2004.
20. P. E. Theodorakis, W. Paul, and K. Binder, preprint.

Studying Directed Assembly of Diblock Copolymers on Patterned Substrates by Soft, Coarse-Grained Models

Marcus Müller and Kostas Ch. Daoulas

Institut für Theoretische Physik, Georg-August Universität Göttingen
Friedrich-Hund-Platz 1, 37077 Göttingen, Germany
E-mail: {mmueller, daoulas}@theorie.physik.uni-goettingen.de

1 Introduction

Copolymer materials self-assemble into periodic microphases with a characteristic length scale of tens of nanometers. The periodicity is dicated by the end-to-end distance $R_{\rm eo}$ of the amphiphilic molecules and can be tuned by their molecular weight. Since the free-energy differences between different morphologies or the free-energy costs of defects in soft materials are small, the self-assembly of diblock copolymers often results in morphologies that contain many defects and the equilibration times are often protracted. Sometimes it may even be difficult to reach equilibrium on the largest length scale considered.

One strategy to avoid defects and fabricate surface patterns with defect-free, long-range order consists in guiding the self-assembly by a lithographically prepared surface pattern, which perfectly matches the bulk morphology of the diblock copolymer in symmetry and length scale[1]. In this report, we describe a soft, coarse-grained model study of structure formation in these fascinating materials, and highlight the relation between the dynamics of individual molecules and the kinetics of structure formation.

2 Soft, Coarse-Grained Model for Multi-Component Polymer Melts

Being interested in the universal behaviour of multi-component polymer melts, we restrict ourselves to a bead-spring model, which is a computationally efficient representation of a Gaussian polymer chain. Generalizations to more complex molecular architectures are straightforward. In our minimal, soft, coarse-grained model the bonded interactions are given by the Edwards Hamiltonian:

$$\frac{\mathcal{H}_{\rm b}[\mathbf{r}_i(s)]}{k_BT} = \sum_{s=1}^{N-1} \frac{3(N-1)}{2{R_{\rm eo}}^2} \left[\mathbf{r}_i(s) - \mathbf{r}_i(s+1)\right]^2 , \tag{1}$$

where N denotes the number of effective interaction centres per molecule and $\mathbf{r}_i(s)$ the coordinate of the $s^{\rm th}$ segment of chain i. The non-bonded interactions take the form

$$\mathcal{H}_{\rm nb}[\hat{\phi}_A, \hat{\phi}_B] = \rho_{\rm o} k_BT \int {\rm d}^3\mathbf{r} \, \left(\frac{\kappa_{\rm o}}{2} \left[\hat{\phi}_A + \hat{\phi}_B - 1\right]^2 - \frac{\chi_{\rm o}}{4} \left[\hat{\phi}_A - \hat{\phi}_B\right]^2 \right) \tag{2}$$

where the microscopic density of A-segments $\hat{\phi}_A(\mathbf{r})$ depends on their explicit positions $\{\mathbf{r}_\alpha(s)\}$ and in the case of a diblock copolymer melt is formally given by

$$\hat{\phi}_A(\mathbf{r}) = \frac{1}{\rho_\mathrm{o}} \sum_{\alpha=1}^{n} \sum_{s=1}^{Nf} \delta(\mathbf{r} - \mathbf{r}_\alpha(s)) \tag{3}$$

where f is the asymmetry of the diblock defined via the number of the effective interaction centres in the A block N_A as $f = N_\mathrm{A}/N$. The normalization is chosen such that $(1/V) \int \mathrm{d}^3\mathbf{r}\ \hat{\phi}_A(\mathbf{r})$ equals the average composition $\bar{\phi}_A = f$ of the system. Here V, n, and ρ_o denote the volume, the number of AB copolymers, and the segment number density, respectively. A similar definition holds for the microscopic density of the B-segments $\hat{\phi}_B(\mathbf{r})$.

In order to advance, one has to introduce a microscopic cut-off. Either one regularizes the δ-function that appears in Eq. (3) by a smoothing function of finite support or one employs a collocation grid with spacing ΔL to assign to it the local densities. The former method leads to soft, DPD-like models and, since the interactions are translationally invariant, one can easily calculate the pressure and perform simulations with variable sizes and shapes of the simulation cell. The latter technique, however, is chosen in the following because it significantly speeds up the calculation of non-bonded interactions. Let $\mathbf{c}$ denote a point of the collocation grid, we define

$$\hat{\phi}_A(\mathbf{c}) = \frac{1}{\Delta L^3} \int \mathrm{d}^3\mathbf{r}\ \Pi(\mathbf{r}, \mathbf{c}) \hat{\phi}_A(\mathbf{r}) = \frac{1}{\rho_\mathrm{o} \Delta L^3} \sum_{\alpha=1}^{n} \sum_{s=1}^{Nf} \Pi(\mathbf{r}_\alpha(s), \mathbf{c}). \tag{4}$$

$\Pi(\mathbf{r}, \mathbf{c})$ denotes the assignment function onto the lattice. Often, a linear assignment of the segment position $\mathbf{r}$ onto the point $\mathbf{c}$ of the collocation grid[2] is used *i.e.* $\Pi(\mathbf{r}, \mathbf{c}) = w(r_x - c_x) w(r_y - c_x) w(r_z - c_z)$ with $w(d) = 1 - \frac{|d|}{\Delta L}$ for $|d| < \Delta L$ and zero otherwise. Here, d denotes the distance between the grid point and the segment position along a Cartesian direction. Similar schemes are used for particle-in-cell techniques in plasma physics[3] or particle-mesh methods in electrostatics[4] in order to assign a particle-based density/charge distribution onto a lattice. The spatial integrals in Eq. (2) are evaluated via the collocation grid.

The discretized Edwards-Hamiltonian in Eq. (1) and the non-bonded, pair-wise interactions in Eq. (2) completely specify a soft, coarse-grained, particle-based model for a multi-component polymer system[2,5,6]. It is characterized by four coarse-grained parameters R_eo, $\chi_\mathrm{o} N$, $\kappa_\mathrm{o} N$ and $\bar{\mathcal{N}} \equiv (\rho_\mathrm{o} {R_\mathrm{eo}}^3/N)^2$ which set the molecular length scale, the strength of the relevant interactions, and density. The coarse-grained parameter $\bar{\mathcal{N}}$ is denoted as invariant degree of polymerization and it controls the strength of fluctuation effects. It dictates a multitude of important properties ranging from the correlation hole of the intermolecular pair-correlation function and the concomitant long-range corrections to the Gaussian chain conformations in a melt, to the capillary-wave broadening of interfaces, or the effect of fluctuations on phase transitions. Without resorting to any approximation, one can study the equilibrium properties of this coarse-grained, particle-based model by a variety of Monte-Carlo simulation techniques, which can be either chosen to faithfully represent the single-chain dynamics[7] or to explore the configuration space of the model most efficiently.

Unlike the pair-wise interactions of coarse-grained models with harsh excluded-volume interactions, the potentials in this coarse-grained model are soft, *i.e.* effective segments can strongly overlap. This is quite a natural property because each segment represents the centre of mass of a group of chemical repeat units. By the same token the liquid of coarse-grained segments does not exhibit pronounced density-density correlations (*i.e.* packing effects).

A significant advantage of soft, coarse-grained models consists in their ability to study systems with an experimentally relevant value of the invariant degree of polymerization: The increase of the molecular density ρ_o/N is much more efficient in achieving experimentally relevant, high invariant degrees of polymerization $\bar{\mathcal{N}}$ than the increase of the chain discretization N. Consider for example that one wants to describe a small patch of the lamellar phase with cubic geometry $(L = 5R_{eo})^3$, which is comprised of 3 lamellar sheets of a symmetric diblock copolymer with $\bar{\mathcal{N}} = 10\,000$. In a particle-based model with hard-core repulsion the segment density is limited by $\rho_o b^3 \lesssim 1$, where $b \equiv R_{eo}/\sqrt{N-1}$ denotes the statistical segment length. For higher densities the liquid of segments freezes into a crystal or it arrests in a glassy structure. Since $\bar{\mathcal{N}} = (\rho_o b^3)^2 N$, the only way to achieve $\bar{\mathcal{N}} = 10\,000$ is to consider chains comprised of $N = 10^4$ beads. The total number of interaction centres in a box of size $(5R_{eo})^3$ is $n = N\sqrt{\bar{\mathcal{N}}}(L/R_{eo})^3 = 1.25 \cdot 10^8$. Since these molecules in a dense melt reptate, the single chain relaxation time scales like $\tau = \tau_o N^3 = \tau_o 10^{12}$, where τ_o is a microscopic time scale that characterizes the motion of an individual segment. The total number of segment motions to propagate the system by one single-chain relaxation time is $n\tau/\tau_o \approx 10^{20}$. In a soft, coarse-grained model without harsh repulsion between segments, one can use a chain discretization of $N = 32$ in order to faithfully represent the Gaussian chain architecture. Using $\rho_o b^3 = \sqrt{\bar{\mathcal{N}}/N} \approx 19$ one achieves the value $\bar{\mathcal{N}} = 10^4$. The system of size $(5R_{eo})^3$ is only comprised of $n = 4 \cdot 10^5$ segments. More importantly, since the single-chain dynamics obeys Rouse behaviour, only $\tau/\tau_o = N^2$ segment movements are required to relax a chain conformation. Thus the total effort to simulate the system amounts to $4 \cdot 10^8$ segment motions which is more than 10 orders of magnitude less than for models like the bond fluctuation model or Lennard-Jones bead-spring models.

3 Kinetics of Self-Assembly

Using nearly symmetric PS-PMMA diblock copolymers, the Nealey-group observed that ordering on a stripe-pattern with a periodicity of $L_o = 48$nm occurs in 6 hours[1]. At 3 hours, however, one observes on the top surface of the film a pattern of spots that align with the stripe-pattern. Earlier simulations have revealed that the morphology at intermediate times signals the breaking up of the initial ordering during which the stripe-morphology is established at the chemically patterned, bottom surface and the negative pattern is created at the non-preferential, top surface. Thus, the stripe pattern at the top surface is displaced by half a period with respect to the bottom morphology. The top morphology is eliminated not by lateral motion of defects but, rather, by "pushing out" the misaligned domains at the top surface. The breaking-up of the misaligned stripes at the top surface gives rise to the lines of spots in the experiment[1].

In order to match the time scale of the simulation and the experiment, we use the longest single-chain relaxation time, *i.e.* the time it takes a molecule in the disordered

phase to diffuse a distance $L_{\rm o}$. There are significant uncertainties in this assignment because (i) the self-diffusion coefficient of the two blocks differ and we assume that the time scale of structure formation is dictated by the slowest component and (ii) the confinement into a thin film may affect the dynamics even in the disordered phase, *i.e.* the glass transition temperature of the components depends on the film thickness. Notwithstanding these caveats, such a mapping results in the estimate that 40 Monte-Carlo steps of a random-local-displacement algorithm corresponds to roughly 1 second real time. In the simulations of the soft, coarse-grained model, perfect ordering on the length scale of several lamellae occurs in $t = 0.85\tau$, where τ denotes the time it takes a molecule to diffuse a distance $R_{\rm eo}$ in the disordered bulk. This corresponds to roughly 8 minutes in the experiment, which is about a factor 45 too fast[1].

In addition to the caveats mentioned above, an additional reason for the too fast dynamics in the simulation consists in the crossability of the molecules in the course of the simulation. By virtue of the soft, non-bonded interactions and the unconstraint bond lengths, the single-chain motion corresponds to Rouse-dynamics for all but the first 10 Monte Carlo steps. In the experiment, however, the single-chain motion of long macromolecules in a melt is strongly affected by the non-crossability of the molecules, which gives rise to a slithering motion along the molecular contour (reptation). The single-chain motion in a disordered and a lamellar phase will be differently affected by the crossability constraint because, in the latter system, the reptation motion displaces the junction point between the two blocks from the AB interface – a process which is enthalpically unfavourable.

In a soft, coarse-grained model non-crossability cannot be enforced by a combination of finite bond length and excluded volume[7]. Either this constraint has to be explicitly incorporated in the simulation algorithm[8], which is feasible but prohibitively expensive, or it has to be included in an effective way.

An effective description of the effect of topological constraints consists in "tethering" the chain contour to the primitive path. This method has successfully been employed by Schieber[9] and Likhtman[10] to describe the dynamics of homogeneous melts of linear chains and it is also used for the description of polymer networks. In order to restrict motion perpendicular to the primitive path, each molecule i is tethered via $N_{\rm sl}$ slip-links to its primitive path. The $t^{\rm th}$ slip-link on polymer i is anchored in space at position $\mathbf{a}_i(t)$. Its other end is harmonically attached to segment $s_i(t)$ and it is free to "slip" along the chain contour by hoping from segment $s_i(t)$ to a neighbouring one. In our model, two slip-links must not be attached to the same segment.

Within this framework, one considers an expanded configuration space that is comprised of chain conformations $\{\mathbf{r}_i(s)\}$, anchor points $\{\mathbf{a}_i(t)\}$ and segment indices $\{s_i(t)\}$, to which the $t^{\rm th}$ anchor on the $i^{\rm th}$ molecule is attached to. The energy due to the slip-links is given by

$$\frac{\mathcal{H}_{\rm sl}[\mathbf{r}_i(s),\mathbf{a}_i(t),s_i(t)]}{k_BT} = \frac{3(N-1)}{2(\alpha_a R_{\rm eo})^2}\sum_{t=1}^{N_{\rm sl}}[\mathbf{a}_i(t)-\mathbf{r}_i(s_i(t)]^2 + \sum_{t_1,t_2=1}^{N_{\rm sl}}\ln\left[1-\delta_{s_i(t_1),s_i(t_2)}\right] \tag{5}$$

where the first term describes the harmonic interaction between segments and slip-links and the last factor guarantees the self-avoidance of slip-links along a molecule. The parameter α_a characterizes the strength of the tether or the width of the tube.

The partition function is obtained by integrating over the expanded configuration space.

$$\mathcal{Z} \propto \frac{1}{n!}\prod_{i=1}^{n}\int \mathcal{D}[\mathbf{r}_i(s)]\prod_{t=1}^{N_{\rm sl}}\sum_{s_i(t)=1}^{N}\int \mathrm{d}^3\mathbf{a}_i(t)\; e^{-\frac{\mathcal{H}_{\rm b}+\mathcal{H}_{\rm nb}+\mathcal{H}_{\rm sl}}{k_BT}} \tag{6}$$

The probability distribution of chain conformations $\mathcal{P}[\mathbf{r}_i(s)]$ is obtained after averaging over the degrees of freedom, $\{\mathbf{a}_i(t)\}$ and $\{s_i(t)\}$, associated with the slip-links

$$\mathcal{P}[\mathbf{r}_i(s)] \propto \frac{1}{n!}\left\langle \exp\left(-\frac{\mathcal{H}_{\rm b}+\mathcal{H}_{\rm nb}+\mathcal{H}_{\rm sl}}{k_BT}\right)\right\rangle_{a_i(t),s_i(t)} \propto \frac{1}{n!}\prod_{i=1}^{n} e^{-\frac{\mathcal{H}_b[\mathbf{r}_i(s)]+\mathcal{H}_{nb}[\hat{\phi}_A,\hat{\phi}_B]}{k_BT}} \tag{7}$$

and one recovers the Boltzmann distribution for the chain conformations, *i.e.* equilibrium properties are not affected by the presence of slip-links.

To sample the expanded configuration space by Monte-Carlo simulation, we move polymer segments with the local SMC-algorithm, where the force includes contributions from bonded and non-bonded interactions as well as slip-links. Additionally, we give slip-links the chance to move to a neighbouring segment along the chain. The trial moves of slip-links are accepted by a Metropolis criterion. Since slip-links represent the topological constraints between two chain portions, they occur in pairs. Similar to Likhtman's model[10], spatial correlations between pairs of slip-links are ignored. If one slip-link is located at an end-segment and the trial move makes it leave the chain contour, then the slip-link and its partner are eliminated, and we create a new pair of slip-links. We choose one end-segment and an arbitrary segment along the same chain and attach two new slip-links. If $\mathbf{r}_i(s_i(t))$ is the position of the bead to which the $t^{\rm th}$ slip-link is attached, then the position of the corresponding anchor point $\mathbf{a}_i(t)$ will be chosen according to

$$P(\mathbf{a}_i(t)|\mathbf{r}_i(s_i(t))) \propto \exp\left(-\frac{3(N-1)}{2(\alpha_a R_{\rm eo})^2}\sum_{t=1}^{N_{\rm sl}}\left[\mathbf{a}_i(t)-\mathbf{r}_i(s_i(t)\right]^2\right). \tag{8}$$

This process mimics the disentanglement (tube renewal) at the chain ends and the concomitant constraint release. We define a Monte-Carlo step (MCS) by the time during which each polymer segment and each slip-link had the change to by moved once on average.

The slip-link model introduces three new parameters $N_{\rm sl}, \alpha_a$ and the relative mobility of slip-links with respect to the dynamics of polymer segments. While none of these parameters alters the equilibrium properties, they do affect the details of the dynamics. Using a very similar model, Likhtman presented a careful investigation of the role of the different parameters and their correlation for a homogeneous polymer melt[10]. Our choice $N/N_{\rm sl} = 4$ and $\alpha_a = 1$ is a compromise. If we decreased $N_{\rm sl}/N$ further or increased α_a, the tube would become "tight" and the Rouse-like motion inside the tube could not be observed. In this limit, the dynamics approaches the slithering-snake dynamics. In the opposite limit, the chains would only be weakly entangled and the dynamics would resemble the local, unconstrained dynamics. Likewise, if the attempt frequency of the slip-link moves tended to zero, the chains would become localized in space. With our choice, the dynamics of the slip-links is comparable to the chain dynamics, which is a plausible choice because in a melt the topological constraints are created by identical molecules in the surrounding. In the following, we adopt the rather pragmatic point of view that the

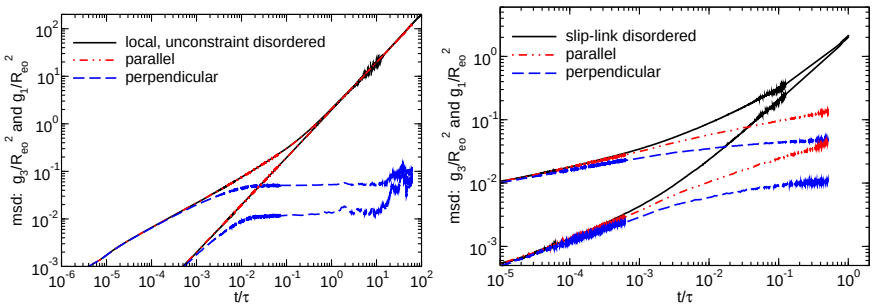


Figure 1. Mean-square displacements of segments $g_1(t)$ and mean-square displacements of the centre of mass $g_3(t)$ as a function of time in the lamellar phase ($\chi_o N = 80$) for local, unconstrained dynamics (a,left) and slip-link dynamics (b, right). Parallel and perpendicular displacements are shown and the results of the disordered phase. "Time" is measured in units of the Rouse time τ. From Ref. 7.

slip-link algorithm in conjunction with SMC-moves gives rise to a microscopic dynamics. Using the single-chain dynamics on large time scales, we identify the parameters of the tube model and investigate to what extent the dynamics on smaller time and length scales agrees with the predictions of the tube model.

A limiting case of the slip-link model is the slithering snake algorithm, where one tries to remove a segment at a randomly chosen end of a molecule and to re-attach it at the opposite end. This algorithm corresponds to the idealized limit that the tube is extremely tight and there is no constraint release.

4 Results

In the following we use $\bar{\mathcal{N}} = 128^2$ and $N = 128$. In Figure 1 we show the mean-square displacement of segments g_1 and of the centre-of-mass g_3 in a lamellar phase using the local SMC-algorithm (a) and the slip-link model (b). The local SMC-algorithm gives rise to Rouse-dynamics in the disordered system and we observe that the dynamics parallel and perpendicular to the interfaces decouples. The motion parallel to the interfaces is identical to the dynamics in the disordered system, which is included in the graph for comparison. Only the dynamics perpendicular to the interfaces is slowed down. At intermediate times, the mean-square displacement exhibits a plateau, which indicates the localization of the macromolecules inside a lamella. At very long-times, one observes flip-flops between neighbouring lamellae which give rise to a very slow diffusive motion perpendicular to the interface at very long times.

Figure 2 depicts different stages of the structure formation of a symmetric diblock copolymer on a patterned surface. The simulations employed local SMC-moves, the slip-link algorithm and slithering-snake motion of the chains. "Time" is indicated in units of the Rouse-time $\tau = {R_{\rm eo}}^2/D$ in the disordered bulk. One observes that the ordering proceeds via similar morphologies indicating that the morphology is chiefly determined by the free-energy landscape. The ordering time, however, increases by a factor 1.3 and 8.8 for the weakly entangled slip-link model and the slithering-snake dynamics. Thus the simulations confirm that non-crossability slows down the dynamics in the lamellar phase compared to the behaviour in the disordered system.

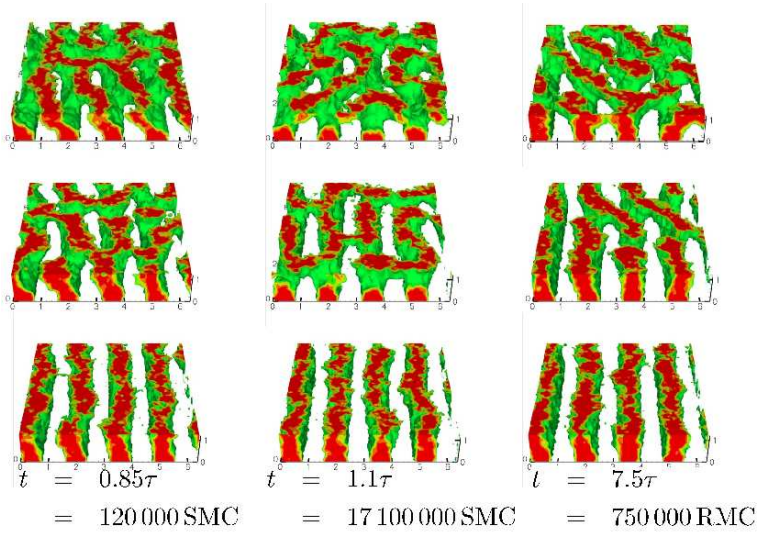


Figure 2. Structure formation of a symmetric diblock copolymer ($\chi_o N = 80$) on a stripe-patterned surface. The three columns correspond to local SMC-moves (a, left), slip-link simulations (b, middle), and the slithering-snake algorithm. The three images depict morphologies at $t/3$, $2t/3$ and the time t, at which defect-free ordering is observed. The ordering time t is given in units of the bulk relaxation time τ and the number of attempted displacements per segment.

5 Concluding Remarks

The simulations highlighted the role of the single-chain dynamics on the collective kinetics of structure formation. The specific example illustrated the role of non-crossability and explored a simple yet computationally efficient way to incorporate the effect of entanglements in a soft, coarse-grained model. The slip-link model does not predict entanglements but it only incorporates their effects including constraint release. The strength of the entanglement effects is set by the number and strength of the slip-links. The question arises how to chose these parameters for a specific experimental system. We suggest to employ the empirical relation between the packing length p and the plateau modulus $G_N^{(0)}$ [11, 12]

$$\frac{G_N^{(0)}}{G_0} = \frac{0.00226 k_B T/p^3}{k_B T N\sqrt{\bar{\mathcal{N}}}/{R_{\rm eo}}^3} = 0.00226\frac{\bar{\mathcal{N}}}{N} \qquad \text{with} \quad \frac{G_{\rm o}}{N} = k_B T\frac{\sqrt{\bar{\mathcal{N}}}}{{R_{\rm eo}}^3} \tag{9}$$

which has be corroborated by experiments and simulations. For the parameters of the slip-link model, we roughly find $\frac{G_N^{(0)}}{G_0} \approx 0.14(5)$, which leads to the relation $\bar{\mathcal{N}} \approx 62N$. We emphasize that this is a purely dynamic relation because for the equilibrium properties the chain discretization N is immaterial. Thus, in order to simultaneously describe equilibrium fluctuations and entangled dynamics for $\bar{\mathcal{N}} = 14\,883$, we have to increase the chain discretization to $N = 240$.

Alternatively, one can try to apply a primitive-path analysis to the multi-chain configurations and identify entanglements from the static structure. This procedure has the

advantage that it can predict the spatial variation of the entanglements as the chain conformations adapt to their spatially inhomogeneous environment. The combination of slip-link model and primitive-path analysis will lead to a self-consistent description of the dynamics of inhomogeneous, multi-component polymer systems.

Acknowledgements

Computing time at the Jülich Supercomputing Centre (JSC) is gratefully acknowledged and we thank the Volkswagen foundation for financial support.

References

1. E. W. Edwards, M. P. Stoykovich, M. Müller, H. H. Solak, J. J. de Pablo, and P. F. Nealey, *Mechanism and kinetics of ordering in diblock copolymer thin films on chemically nanopatterned substrates*, J. Polym. Sci. B: Polymer Physics, **43**, 3444–3459, 2005.
2. M. Müller and K. C. Daoulas, *Calculating the free energy of self-assembled structures by thermodynamic integration*, J. Chem. Phys., **128**, 024903, 2008.
3. J. M. Dawson, *Particle simulation of plasmas*, Rev. Mod. Phys., **55**, 403–447, 1983.
4. J. W. Eastwood, R. W. Hockney, and D. N. Lawrence, *P3M3DP - The 3-dimensional periodic particle-particle-particle-mesh program*, Computer Physics Communications, **19**, 215–261, 1980.
5. K. Ch. Daoulas and M. Müller, *Single-Chain-in-Mean-Field Simulations: Quasi-instantaneous field approximation and quantitative comparison with Monte Carlo simulations*, J. Chem. Phys., **125**, 184904, 2006.
6. M. Müller, K. Ch. Daoulas, and Y. Norizoe, *Computing free energies of interfaces in self-assembling systems*, Phys. Chem. Chem. Phys., **11**, 2087, 2009.
7. M. Müller and K. Ch. Daoulas, *Single-chain dynamics in a homogeneous melt and a lamellar microphase: A comparison between Smart-Monte-Carlo dynamics, slithering-snake dynamics, and slip-link dynamics*, J. Chem. Phys., **129**, 164906, 2008.
8. J. T. Padding and W. J. Briels, *Time and length scales of polymer melts studied by coarse-grained molecular dynamics simulations*, J. Chem. Phys., **117**, 925–943, 2002.
9. C. C. Hua and J. D. Schieber, *Segment connectivity, chain-length breathing, segmental stretch, and constraint release in reptation models. I Theory and single-step strain predictions*, J. Chem. Phys., **109**, 10018–10027, 1998.
10. A. E. Likhtman, *Single-chain slip-link model of entangled polymers: Simultaneous description of neutron spin-echo, rheology, and diffusion*, Macromolecules, **38**, 6128–6139, 2005.
11. L. J. Fetters, D. J. Lohse, S. T. Milner, and W. W. Graessley, *Packing length influence in linear polymer melts on the entanglement, critical, and reptation molecular weights*, Macromolecules, **32**, 6847–6851, 1999.
12. R. Everaers, S. K. Sukumaran, G. S. Grest, C. Svaneborg, A. Sivasubramanian, and K. Kremer, *Rheology and microscopic topology of entangled polymeric liquids*, Science, **303**, 823–826, 2004.

Membrane-Protein Interactions in Lipid Bilayers: Molecular Simulation versus Elastic Theory

Beate West[1] **and Friederike Schmid**[2]

[1] Fakultät für Physik, Universität Bielefeld, D-33615 Bielefeld, Germany

[2] Institut für Physik, Universität Mainz, D-55118 Mainz, Germany
E-mail: friederike.schmid@uni-mainz.de

Many processes in life are controlled by membrane proteins. We study generic interaction mechanisms between such proteins by simulations of highly simplified coarse-grained models. Specifically, we focus on a type of interaction that has been intensely discussed in the literature, the "hydrophobic mismatch" interaction. We analyze it quantitatively in detail and compare our results with the predictions of a popular continuum theory, the elastic model.

1 Introduction

Biomembranes are central components of all living matter. They delimitate cells, they structure them by separating different compartments from each other, they play an important role in controlling the transport of molecules and information within the cell and between cells[5].

Remarkably, these vitally important structures are mainly the result of a physical self-assembly process. Their main constituent is a bilayer of lipids, which aggregate into sheets because this is the most efficient way for the hydrophilic molecule parts ("head groups") to shield the hydrophobic molecule parts (hydrocarbon "tails") from the water environment while remaining themselves immersed in water. The sheets contain large amounts of proteins and other macromolecules, which are the biochemically active components and have a variety of functions: transport, enzyme activity, signal transmission, cell connection, cell-cell recognition etc.

The function of the proteins is to a large extent controlled by their lateral organization, which in turn depends on their mutual interactions. Interactions between membrane proteins are therefore a topic of great interest. Apart from obvious direct interactions, the membrane environment also plays an important role in organizing proteins. For example, in recent times the so-called "raft hypothesis" has been discussed intensely, according to which proteins are sorted and transported by mobile small stiff domains ("rafts") in the membranes, which supposedly result from a demixing process between membrane lipids of different type. Even simple one-component membranes already mediate substantial interactions between membrane proteins. This has been recognized and investigated theoretically for decades, but only recently have computers become powerful enough that it can also be studied at a molecular level by computer simulations[15, 20, 22, 24].

Theories of membrane-mediated protein interactions of course make assumptions. For example, analytical approaches often model the membrane as one elastic sheet, or two coupled sheets with certain elastic properties. Such "elastic theories" become questionable on length scales which are comparable to the membrane thickness. Nevertheless, they make predictions on precisely these length scales. Simulations offer a way to test such

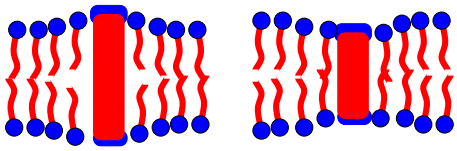


Figure 1. Sketch of hydrophobic mismatch. The hydrophobic thickness of the inclusion (red part) does not match the membrane thickness. Left: positive mismatch, Right: Negative mismatch.

predictions and thus give insight into the validity of the elastic model and its interpretation of the interaction mechanisms.

In the present paper, we describe a comparative study of so-called "hydrophobic mismatch" interactions in an elastic model and in molecular simulations. Hydrophobic mismatch is encountered if the hydrophobic portion of a membrane protein does not fit perfectly into a membrane, *i.e.* the membrane is locally distorted by the protein (see Figure 1). Two mismatched proteins interact with each other *via* these local distortions. This has been one of the first membrane-mediated interaction mechanism between protein discussed in the literature[17,18,16], it has been established experimentally (at a qualitative level)[19,7,6] and discussed theoretically for many decades (see Ref. 24 for a brief overview).

2 Model and Method

In our molecular simulations, we use a "generic" coarse-grained lipid model[21] where lipids are represented by chains of tail beads of diameter σ_t with one slightly larger head bead, and the solvent is a fluid of single beads. Tail beads attract each other, all other beads are purely repulsive, and solvent beads do not interact at all with each other (to make the model computationally efficient). The details of the model are described elsewhere[4,11,21]. We carry out Monte Carlo simulations at constant pressure in a simulation box of varying volume and shape (boundary conditions). The membranes are tensionless.

At the temperature used here, the lipids self-assemble spontaneously into a stable fluid bilayer. The ratio of area per lipid and bilayer thickness (squared) roughly corresponds to that of dimetylphosphatidylcholine (DPPC), one of the most common lipids in biomembranes. The thickness can be used to assign "real" SI-units to our model length unit σ_t ($1\ \sigma_t \sim 6$Å). Both our model membranes and DPPC bilayers undergo a marked phase transition to a more ordered "gel" phase at lower temperatures[12]. By matching the transition temperatures of this "main transition", we can also identify our energy scale.

The "protein" model is even simpler. They are modeled as cylinders with diameters corresponding to that of a β-helix, with a central hydrophobic section of length L. They have translational freedom in all spatial directions, but their orientation is confined to the z-direction. Tilt is forbidden, because the "proteins" in the elastic theory, which we use for comparison, are also untilted cylinders with a given hydrophobic length. This gives us two parameters to play with: The hydrophobic length L and the hydrophobic strength of the proteins ϵ_{pt} (which is, in our case, the strength of the attractive force between the protein and the tail beads).

We compare with an elastic theory for hydrophobic mismatch interactions originally developed by Aranda-Espinoza *et al*[1], which has recently been extended by Brannigan

Parameter	Our model	All-Atom	Experiment	Unit
Bending rigidity k_c	2.2 ± 0.2	4	5-20	10^{-20} J
Compressibility parameter k_A/d_0^2	3.6 ± 0.9	1.1	6	10^{-20} J/nm^4
Monolayer curvature c_0	-0.08 ± 0.03	-0.05 ... -0.04	-0.04	nm^{-1}

Table 1. Elastic constants (selection) of our model membrane as obtained from a fit of the fluctuation spectra to the elastic theory, and from the pressure profiles, compared to values from all-atom simulations of DPPC[10,14] and experiments on DPPC bilayers[13].

and Brown[2]. This theory is chosen here, because it fits the fluctuation spectra of pure membranes in an excellent way[24] and also works reasonably well for single proteins (see below). We have also compared the simulation data with an alternative, similarly popular Landau-type type theory, but the results were much worse and shall not be shown here[24]. In the elastic theory, the membranes are described as two coupled elastic monolayer sheets with fixed volume, which can be compressed and bent and have each a favourite curvature. The elastic parameters (compressibility, bending rigidity, spontaneous curvature etc.) can be extracted from the fluctuation fits and the pressure profiles (Table 1). Considering the simplicity of our model, they are in remarkably close agreement with the actual values for DPPC.

3 Membrane-Protein Interactions

We first discuss the interactions between a single protein and the membrane. Figure 2 illustrates the membrane distortion in the vicinity of proteins with different hydrophobic thickness L and hydrophobic strength ϵ_{pt}. Proteins with hydrophobic thickness $L = 6\sigma_t$ are hydrophobically matched, $L = 8\sigma_t$ corresponds to positive mismatch (too thick), and $L = 4\sigma_t$ to negative mismatch (too thin). At low hydrophobic strength ϵ_{pt}, the membrane thickness is reduced in the vicinity of the protein in all three cases: The protein merely repels the lipids. In order to pin the hydrophobic membrane thickness to the value L at the protein, the hydrophobic strength must exceed a certain value ($\epsilon_{pt} \sim 4 - 5$ in our units).

This "critical" hydrophobic strength can be directly related to the free energy needed for inserting the protein into the membrane from solution. The latter has been determined by a thermodynamic integration method, where the interaction strength between protein and lipids was gradually turned off[25]. The result for a hydrophobically matched protein is shown in Figure 3. It is positive for weakly hydrophobic proteins and negative for strongly hydrophobic proteins, and crosses zero roughly at the point where the protein starts to pin the membrane. Hence proteins that are effectively attracted by the membrane also pin it.

The solid lines in Figure 2 show the fits of the profiles to the elastic theory. They are very good. It should, however, be noted that in order to produce such good fits, we had to modify the theory such that the boundary conditions at the protein surface became adjustable[3]. One of the boundary conditions is the bilayer distortion at the protein, the other one is related to the monolayer spontaneous curvature c_0, which is known (Table 1).

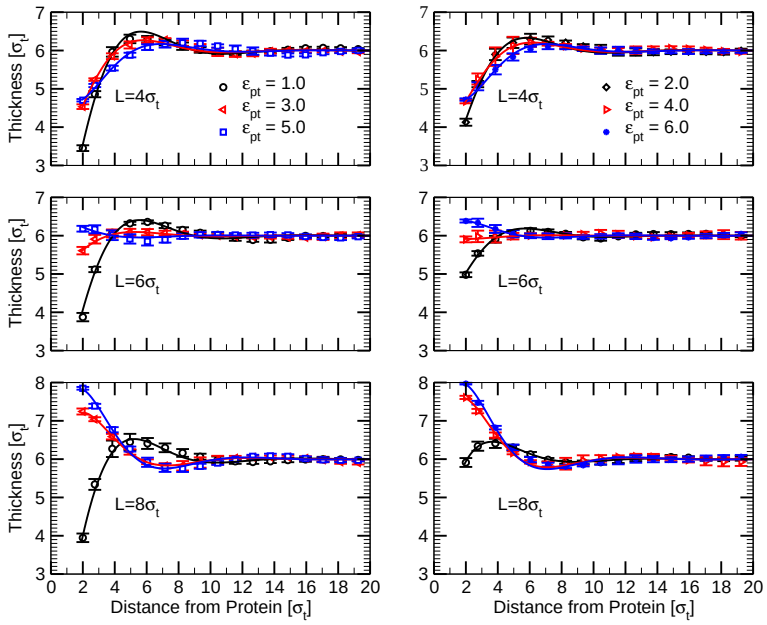


Figure 2. Radial profiles of bilayer thickness around proteins with different hydrophobic strength ϵ_{pt} and hydrophobic thickness L as indicated. Symbols: simulation data. Lines: Fits to the elastic theory (see text for explanation).

However, using the values of Table 1, we could not produce any reasonable fit[24]. We had to "renormalize" c_0. One can rationalize such a renormalization by assuming that intrinsic membrane properties (*e.g.* the hydrophobic density or the local nematic order parameter) change in the vicinity of the surface, which results in an effective boundary condition[24]. The thus modified elastic theory works very well and even accounts for details of the profiles such as their slightly oscillatory shape.

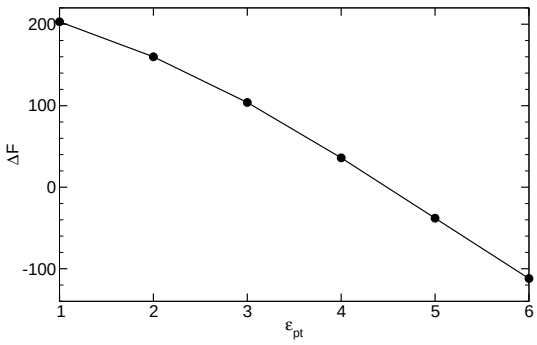


Figure 3. Free energy ΔF of inserting a single hydrophobically matching protein in the membrane versus hydrophobic strength ϵ_{pt} of the protein.

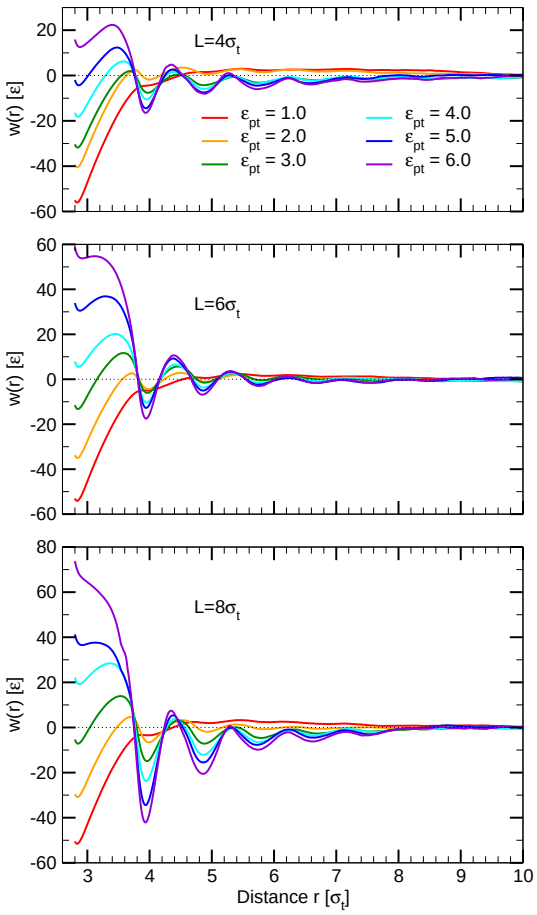


Figure 4. Effective interaction between two proteins for different hydrophobic strength ϵ_{pt} and/or hydrophobic thickness L as indicated.

We are now ready to look at the effective interactions between two proteins (determined using an appropriate umbrella sampling technique). They are shown in Figure 4 and compared to the prediction of the elastic theory in Figure 5. At first sight, the most pronounced feature is a strongly oscillatory shape, which can definitely not be explained by elastic interactions. It results from local rearrangements of the lipids. These "packing interactions" have also been discussed in the literature[8,9], and they represent a source of effective membrane-mediated interactions not mentioned so far. On top of them, hydrophobically mismatched proteins experience an additional attractive force which is absent for hydrophobically matched proteins. The sign of this force does not depend on the sign of the mismatch. When comparing this smoother contribution with the elastic theory (Figure 5), we find that the elastic theory seems to fail. At small distances, it predicts attraction, in accordance with the simulations. At larger distances, however, it predicts a broad peak at and a repulsive regime, which has no counterpart in the simulation data.

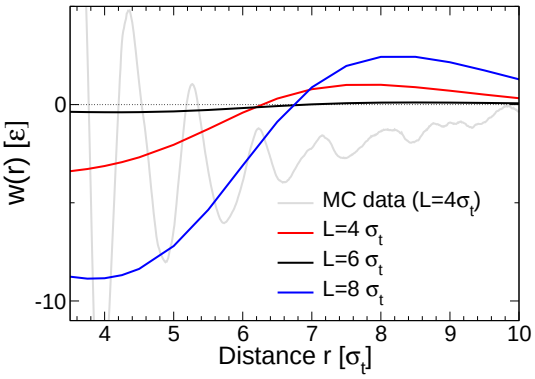


Figure 5. Effective interaction between two proteins for different thickness L according to the elastic theory. Light grey line shows simulation data from Figure 4 ($\epsilon_{pt} = 6$) for comparison.

To analyze this discrepancy, we inspect the thickness profiles in the two-dimensional membrane plane close to the two proteins. The most pronounced difference between theory and simulation is observed for distances where the theoretical interaction potential crosses zero. Here, the theory predicts a thickness overshoot between the two proteins, whereas in the simulations, the thickness relaxes towards its unperturbed value (Figure 6). This finding hints at a possible reason for the failure of the theory: It assumes that the boundary conditions at the protein (the membrane distortion and the "effective curvature") are fixed. They are constant everywhere on the surface of the protein, and independent of the distance between the proteins. On the other hand, we have already noted that the boundary conditions, most notably the effective curvature, are renormalized by structural changes of the membrane in the vicinity of the proteins. One would expect these changes to be affected by the presence of a second protein, and to depend on the distance of the two proteins.

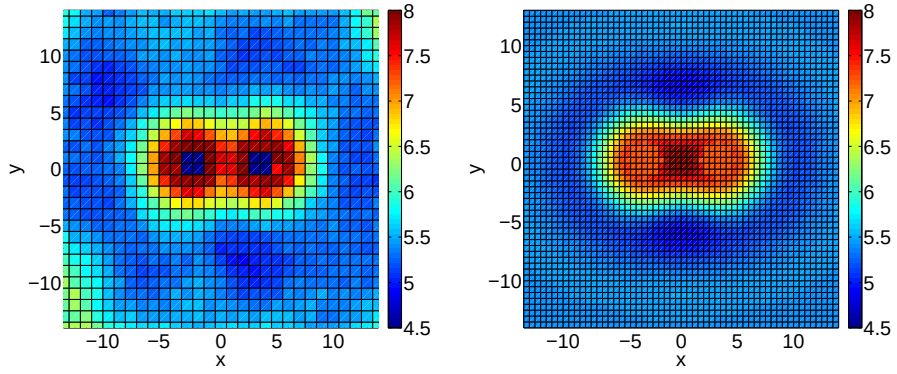


Figure 6. Thickness profiles in the vicinity of two proteins with positive hydrophobic mismatch ($L = 8\sigma_t$) according to simulation (left) and theory (right).

4 Summary and Outlook

We have studied the hydrophobic mismatch interaction between proteins and lipid bilayers, with a focus on comparing molecular simulation data with an established elastic theory. Over all, the elastic theory performs well, if one accounts for the fact that local structural membrane changes affect the effective boundary conditions at the surface of the proteins. Membrane distortions around single proteins can be fitted very nicely. The prediction for the effective interactions between two proteins is not yet satisfactory, but this can probably be remedied with a more sophisticated Ansatz for the boundary conditions. At this point, it is somewhat discouraging that new fit parameters have to be introduced at every level of complexity (going from the pure membrane to membranes with one inclusion to membranes with two inclusions), and the *predictive* power of the elastic theory of hydrophobic mismatch interactions has yet to be established. Nevertheless, it seems likely that the elastic model does capture the essential physics of the hydrophobic mismatch interaction.

We have studied interaction mechanisms in one-component membranes as a first step. As mentioned in the introduction, the heterogeneity of multicomponent membranes is believed to play an essential role for the lateral organization of membrane proteins. This will be the focus of our future work.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft within the Sonderforschungsbereich SFB 613 and SFB 625. The computations were carried out at the NIC Jülich. We thank J. Neder for useful discussions.

References

1. H. Aranda-Espinoza, A. Berman, N. Dan, P. Pincus, and S. Safran, *Interaction between Inclusions Embedded in Membranes*, Biophysical Journal **71**, 648, 1996.
2. G. Brannigan and F. L. H. Brown, *A Consistent Model for Thermal Fluctuations and Protein-Induced Deformations in Lipid Bilayers* Biophysical Journal **90**, 1501, 2006.
3. G. Brannigan and F. L. H. Brown, *Contributions of Gaussian Curvature and Nonconstant lipid Volume to Protein Deformation of Lipid Bilayers* Biophysical Journal **92**, 864, 2007.
4. D. Düchs, and F. Schmid, *Phase behavior of amphiphilic monolayers: Theory and simulations*, J. Phys: Cond. Matt. **13**, 4853, 2001.
5. R. B. Gennis, *Biomembranes - Molecular Structure and Function*, Springer Verlag, New York, 1989.
6. T. A. Harroun, W. T. Heller, T. M. Weiss, L. Yang, and H. W. Huang, *Experimental evidence for hydrophobic matching and membrane-mediated interactions in lipid bilayers containing gramicidin*, Biophysical Journal **76**, 937, 1999.
7. J. A. Killian, *Hydrophobic mismatch between proteins and lipids in membranes*, Biochimica and Biophysica Acta **1376**, 401, 1998.
8. P. Lagüe, M. M. Zuckermann, and B. Roux, *Protein inclusion in lipid membranes: A theory based on the hypernetted chain integral equation*, Faraday Disc. **111**, 165, 1998.

9. P. Lagüe, M. M. Zuckermann, and B. Roux, *Lipid-mediated interactions between intrinsic membrane proteins: A theoretical study based on integral equations*, Biophys. J. **79**, 2867, 2000.
10. E. Lindahl and O. Edholm, *Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations*, Biophysical Journal **79**, 426, 2000.
11. O. Lenz, and F. Schmid, *A simple computer model for liquid lipid bilayers*, J. Mol. Liquids **117(1-3)**, 147, 2005.
12. O. Lenz, and F. Schmid, *Structure of Symmetric and Asymmetric Ripple Phases in Lipid Bilayers*, Phys. Rev. Lett. **98**, 058104, 2007.
13. D. O. Marsh, *Elastic curvature constants of lipid monolayers and bilayers*, Chemistry and Physics of Lipids **144**, 146, 2006.
14. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tielemann, and A. H. de Vries, *The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations*, Journal of Physical Chemistry B **111**, 7812, 2007.
15. F. J. -M. de Meyer, M. Venturoli, and B. Smit, *Molecular Simulations of Lipid-Mediated Protein-Protein Interactions*, Biophysical Journal **95**, 1851, 2008.
16. O. G. Mouritsen, and M. Bloom, *Models of lipid-protein interactions in membranes*, Quarterly Reviews of Biophysics: Biomol. Struct. **22**, 145, 1993.
17. J. C. Owicki, M. W. Springgate, and H. M. McConnell, *Theoretical study of protein-lipid interactions in bilayer membranes*, PNAS **75**, 1616, 1978.
18. J. C. Owicki and H. M. McConnell, *Theory of protein-lipid and protein-protein interactions in bilayer membranes*, PNAS **76**, 4750, 1979.
19. M. R. R. de Planque and J. A. Killian, *Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring (review)*, Molecular Membrane Biology **20**, 271, 2003.
20. B. J. Reynwar, G. Illya, V. A. Harmandaris, M. M. Müller, K. Kremer, and M. Deserno, *Aggregation and vesiculation of membrane proteins by curvature-mediated interactions*, Nature **447(7143)**, 461, 2007.
21. F. Schmid, D. Düchs, O. Lenz, and B. West, *A generic model for liquid monolayers, bilayers, and membranes*, Comp. Phys. Comm. **177(1-2)**, 168, 2007.
22. U. Schmidt, G. Guigas, and M. Weiss, *Cluster Formation of Transmembrane Proteins Due to Hydrophobic Mismatching*, Phys. Rev. Lett. **101**, 128104, 2008.
23. M. Venturoli, B. Smit, and M. M. Sperotto, *Simulation Studies of Protein-Induced Bilayer Deformations, and Lipid-Induced Protein Tilting, on a Mesoscopic Model for Lipid Bilayers with Embedded Proteins*, Biophysical Journal **88**, 1778, 2005.
24. B. West, F. L. H. Brown, F. Schmid, *Membrane-Protein Interactions in a Generic Coarse-Grained Model for Lipid Bilayers*, Biophysical Journal **96**, 101, 2009.
25. B. West, Dissertation Universität Bielefeld, 2009.

Semidilute Polymer Solutions under Shear Flow

Godehard Sutmann[1], **Chien-Cheng Huang**[2], **Roland G. Winkler**[2], **and Gerhard Gompper**[2]

[1] Jülich Supercomputing Centre, Forschungszentrum Jülich
52425 Jülich, Germany
E-mail: g.sutmann@fz-juelich.de

[2] Institut für Festkörperforschung, Forschungszentrum Jülich
52425 Jülich, Germany
E-mail: {c.c.huang, r.winkler, g.gompper}@fz-juelich.de

A newly developed massively parallel hybrid simulation program, MP^{2}C, combining molecular dynamics simulations and the multiparticle collision dynamics approach, is exploited to study the structural and dynamical properties of semidilute polymer solutions under shear flow. Results are presented for the density and shear-rate dependence of the polymer conformations, orientations, and orientation-angle distributions.

1 Introduction

The dynamical behaviour of dilute and semidilute polymer solutions is strongly affected or even dominated by hydrodynamic interactions[1–3]. From a theoretical point of view, scaling relations predicted by the Zimm model at infinite dilution, e.g. the dependence of dynamical quantities on the length of the polymer, are, in general, accepted and confirmed[1,4,5]. However, we are far from a similar understanding of the dynamics of semidilute polymer solutions under equilibrium or non-equilibrium conditions. Insight into the behaviour of such systems is of fundamental importance in a wide spectrum of systems ranging from biological cells, where transport appears in dense environments, to turbulent drag reduction in fluid flow. Moreover, in semidilute solutions of long polymers, viscoelastic effects play an important role. Due to the long structural relaxation time, the internal degrees of freedom of a polymer cannot relax sufficiently fast under non-equilibrium conditions and an elastic restoring force tries to push the system towards its original state.

Recent advances in experimental single-molecule techniques provide insight into the dynamics of individual polymers under equilibrium and non-equilibrium conditions[6,7], and raise the need for a quantitative theoretical description in order to determine molecular parameters such as diffusion coefficients and relaxation times. The complex molecular interactions in semidilute solutions hamper an analytical treatment, but a theoretical understanding can be achieved by computer simulations. The large length- and time-scale gap between the solvent and macromolecular degrees of freedom requires a mesoscale simulation approach in order to assess their structural, dynamical, and rheological properties[3]. Here, we apply a hybrid simulation approach, combining molecular dynamics simulations (MD) for the polymers with the multiparticle collision dynamics (MPC) method describing the solvent[3,8].

Experiments[6,7],theoretical studies[9], and simulations[10,11] of individual polymers under shear flow conditions exhibit large deformations and a strong alignment of the polymers. Moreover, a large overlap is present in semidilute solution of long polymers. A typical

simulation requires $10^5 - 10^7$ monomer and $10^7 - 10^8$ fluid particles. Hence, despite the adopted mesoscale approach, large systems can only be studied on a massively parallel computer architecture.

Here, we present results of large-scale simulations of semidilute polymer solutions under shear. The simulations were performed with our program MP^{2}C (massively parallel multiparticle collision dynamics), which exhibits excellent scaling behaviour on the massively parallel architecture of the IBM Blue Gene/P computer[12].

2 Simulation Method and Polymer Model

Within the multiparticle collision dynamics approach, the fluid is represented by point particles of mass m, which interact with each other by a stochastic process. The algorithm consists of alternating streaming and collision steps[2,3,8]. In the streaming step, the N_s fluid particles move ballistically and their positions change according to

$$\vec{r}_i(t) = \vec{r}_i(t-h) + h\vec{v}_i(t-h), \tag{1}$$

$i = 1, \ldots, N_s$ in the time interval h, which we denote as collision time. In the collision step particles are sorted into cubic cells of side length a and their relative velocities, with respect to the centre-of-mass velocity of every cell, are rotated around a randomly oriented axis by a fixed angle α. This imposed stochastic process represents the effect of many real collisions. In a collision step mass, momentum, and energy are conserved, which leads to the build up of correlations between the particles and gives rise to hydrodynamic interactions. Hence, the velocity of a particle changes according to

$$\vec{v}_i(t+h) = \vec{v}_i(t) + (\mathbf{R}(\alpha) - \mathbf{E})\,(\vec{v}_i(t) - \vec{v}_{cm}(t)), \tag{2}$$

where $\vec{v}_i(t)$ is the velocity before the collision, $\mathbf{R}(\alpha)$ is the rotation matrix, $\vec{v}_{cm} = \sum_{j=1}^{N_c} \vec{v}_j / N_c$ is the centre-of-mass velocity of the particles contained in the cell of particle i, and N_c is the total number of fluid particles in that cell. $\mathbf{E}$ is the unit matrix. To insure Galilean invariance, a random shift of the lattice of collision cells is performed at every collision step[13].

A polymer chain is introduced into the system by adding N_m point particles each of mass M, which are connected linearly by bonds with the potential

$$U_B = \frac{\kappa}{2} \sum_{i=1}^{N_m - 1} \left(|\vec{r}_{i+1} - \vec{r}_i| - l\right)^2, \tag{3}$$

where l is the bond length[5]. To account for excluded-volume interactions, the monomers interact via the repulsive, truncated Lennard-Jones potential

$$U_{RLJ} = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon, & r < \sqrt[6]{2}\,\sigma \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

The dynamics of the chain monomers is determined by Newton's equations of motion between the collisions with the solvent. These equations are integrated by the velocity Verlet algorithm with time step h_p.

The monomer dynamics can easily be coupled to that of the solvent by incorporating them in the collision step[5,14]. For a collision cell with N_c fluid particles and N_m^c

monomers, which may belong to different polymers, the centre-of-mass velocity is given by

$$\vec{v}_{cm}(t) = \left[\sum_{i=1}^{N_c} m\vec{v}_i(t) + \sum_{\nu} \sum_{k}^{N_m^c} M\vec{v}_k^{\nu}(t) \right] / \left[mN_c + MN_m^c \right]. \tag{5}$$

Here, ν and k denote those polymers and monomers, which are within the considered collision cell. This results in an exchange of momentum between the solvent and polymer degrees of freedom. The new monomer velocities are then used as initial conditions for the MD simulation of the polymers.

A shear flow is imposed by the opposite movement of two confining walls. The walls are parallel to the xy-plane and periodic boundary conditions are applied in the x- and y-directions. The equations of motion of the fluid particles are then modified by the wall interactions[15]. We will assume no-slip boundary conditions, which we realize by the bounce-back rule, i.e. the velocity of a fluid particle is reverted when it hits a wall ($\vec{v}_i \rightarrow -\vec{v}_i$). The same rule is applied for the monomers when colliding with a wall[3].

The program MP^{2}C is implemented in module-oriented Fortran 90[12]. Message passing between processors is realized with the MPI standard. The parallel algorithm is based on a three-dimensional domain-decomposition approach, where particles are sorted onto processors according to their spatial coordinates. The present version of the program is implemented on a 2^n-subdivision of processors, where every processor is responsible for a sub volume $V_P = V/N_p$, where $V = a^3 \prod_\alpha n_\alpha$ and n_α is the number of collision cells in each Cartesian direction $\alpha \in \{x, y, z\}$. Volumes and boundaries of spatial domains are not altered during a simulation. The domains are chosen such that the ratio of surface/volume is minimized.

The following simulation parameters are adopted. The simulation box size is $L_x = 450a$, $L_y = L_z = 75a$, with an average of $\langle N_c \rangle = 10$ particles in a collision cell. The rotation angle is set to $\alpha = 130^o$. Length and time are scaled according to $\tilde{r}_\beta = r_\beta/a$ and $\tilde{t} = t\sqrt{k_BT/(ma^2)}$, which corresponds to the choice $k_BT = 1$, $m = 1$, and $a = 1$, where T is the temperature and k_B the Boltzmann constant. The collision time is $\tilde{h} = 0.1$ and the MD time step $\tilde{h}_p = 2 \times 10^{-3}$. These values yield the viscosity $\eta = 8.7(mk_BT/a^4)^{1/2}$. The polymer parameters are $l = a$, $\sigma = a$, $\epsilon/k_BT = 1$, and $\tilde{\kappa} = 5 \times 10^3$. Results are presented for the chain length $N_m = 250$ and the number of polymers $N_p = 50$, 100, and 3000. The segmental concentration for the largest polymer number is significantly above the overlap concentration.

3 Results

The conformational and dynamical properties of polymers exposed to a shear flow strongly depend on the applied shear rate $\dot{\gamma}$, as illustrated in Figure 1. A convenient quantity to characterize the conformational changes is the average gyration tensor, which is defined as

$$G_{\alpha\beta}(\dot{\gamma}) = \frac{1}{N_m} \sum_{i=1}^{N_m} \langle r_{i,\alpha} r_{i,\beta} \rangle, \tag{6}$$

where $r_{i,\alpha}$ is the position of monomer i relative to the centre-of-mass of the polymer.

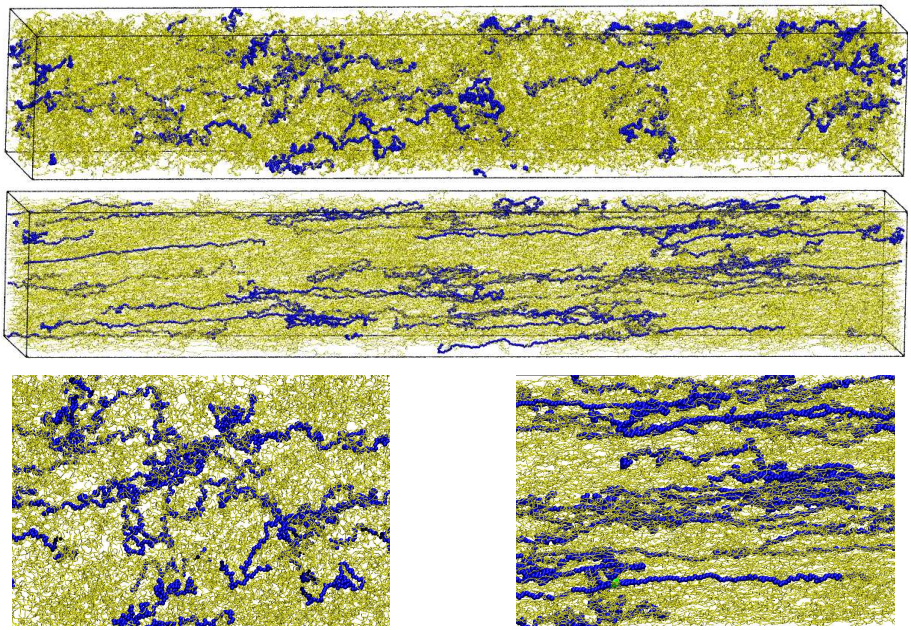


Figure 1. Snapshots of a system with $N_p = 800$ polymers for the Weissenberg numbers $\mathrm{Wi} = 11$ (top, bottom left) and $\mathrm{Wi} = 340$ (middle, bottom right). For illustration, some of the chains are highlighted in blue. The bottom images show closeup views.

The diagonal components $G_{\alpha\alpha}$ are shown in Figure 2 as a function of the Weissenberg number $\mathrm{Wi} = \dot{\gamma}\tau$, with τ the longest relaxation time, for various polymer concentrations. The polymer radius of gyration has been predicted[1] to shrink in semidilute solutions according to $r_g^2(c) = (r_g^{(0)})^2 (c/c^*)^{(2\nu-1)/(1-3\nu)}$, when the segment density c exceeds the value at the overlap concentration $c^* = 3N_m/(4\pi(r_g^{(0)})^3)$. The radius of gyration at zero concentration is approximately given by $(r_g^{(0)})^2 = l^2 N^{2\nu}/6$. Based on results of previous simulations of (shorter) polymers in dilute and semidilute solutions, we find a scaling exponent of $\nu \approx 0.64$ for our chain model[5]. Taking this estimate of the radius of gyration, we scale the components $G_{\alpha\alpha}$ by $G^0_{\alpha\alpha}(c) = r_g^2(c)/3$. The density dependence of the radius of gyration of our polymer system needs to be confirmed, which requires equilibrium simulations. Similar, the longest relaxation times need to be determined from such equilibrium simulations. The simulations of shorter polymers in dilute solutions show that an estimate of the relaxation time is obtained from the relation[1] $\tau = 0.325\eta l^3 N_m^{3\nu}/(k_B T)$. The relaxation time also depends on the polymer concentration. In Ref. 16 the relation $\tau(c) = \tau(c/c^*)^{3(1-\nu)/(3\nu-1)} = \tau\mu(c)$ is provided in the very high overlap regime. Since we are in a moderate overlap regime only, it is not *a priori* evident that the relation for μ applies to our systems. We therefore adjust μ such that we achieve a scaling behaviour for G_{xx} and the alignment angle, which is displayed in Figure 3.

The gyration tensor component in the flow direction increases with increasing shear rate, whereas the components in the gradient and vorticity directions decrease. The simulations of polymers for various concentrations in the dilute regime yield the same values of

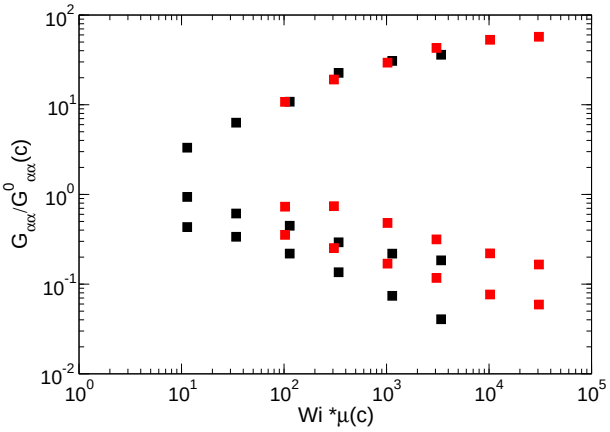


Figure 2. Normalized components of the average gyration tensor along the flow (top), gradient (bottom), and vorticity (middle) directions, respectively. The black squares (■) are for a dilute system, the red squares (■) are for a semidilute solution with $c/c^* \approx 13.5$, i.e., $\mu(c) = 9$, which corresponds to 3000 polymers in the simulation box.

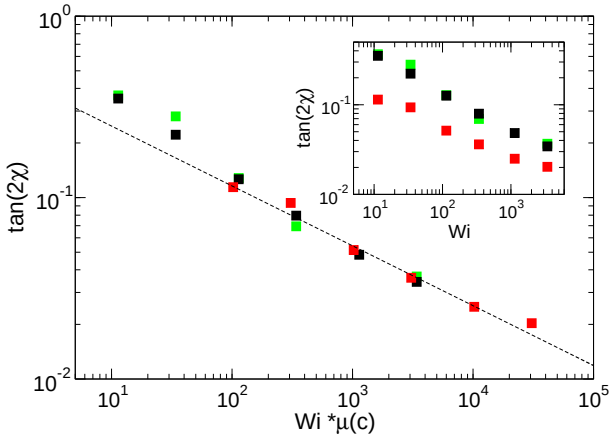


Figure 3. Shear rate dependence of the orientation angle for the number of polymers $N_p = 50$ (green), 100 (black), and 3000 (red) corresponding to $c/c^* = 0.23,\ 0.45,\ 13.5$, respectively. The dashed line illustrates the dependence $\mathrm{Wi}^{-1/3}$. The inset displays the same data without the concentration dependent factor μ.

$G_{\alpha\alpha}$ for all considered shear rates. Similar, the polymer deformation along the flow direction seems to be independent of the concentration. However, the transverse components G_{yy} and G_{zz} exhibit a smaller shrinkage at higher concentrations. This could be related to the fact that the polymers are already more compact at higher concentrations, and therefore their relative compression is smaller. Interestingly, without the scale factor $\mu(c)$, the components G_{yy} and G_{zz} are very similar for the various concentrations.

The alignment of the polymers is characterized by the orientation angle χ, which is the angle between the eigenvector of the gyration tensor with the largest eigenvalue and the

flow direction. It is obtained from the components of the radius of gyration tensor via

$$\tan(2\chi) = 2G_{xy}/(G_{xx} - G_{yy}). \tag{7}$$

Figure 3 shows that systems of low and high polymer concentrations follow a universal curve. The orientation obeys the theoretical prediction[9] $\tan(2\chi) \sim \mathrm{Wi}^{-1/3}$ valid for $\mathrm{Wi} \gg 1$, which has been also found in single chain simulations[11]. At small Weissenberg numbers, the dependence $\mathrm{Wi}^{-\gamma}$, with $\gamma \approx 1/2$, is consistent with experiments and simulations[11]. There is a pronounced dependence of the alignment on the concentration as illustrated by the inset of Figure 3, when the Weissenberg number in dilute solution is considered only. This dependence, however, is related to the change in the polymer relaxation time. From our fitted value $\mu = (c/c^*)^{\hat{\nu}} \approx 9$, with $c/c^* \approx 13.5$ we obtain a scaling exponent $\hat{\nu} \approx 0.85$. This value is smaller than the theoretical value[16] $\hat{\nu} = 1.17$. The reason could be that we have not yet reached the asymptotic limit for the applicability of the scaling law, which is valid for large overlap concentrations. This aspect deserves further considerations.

Further insight into the orientational behaviour of an individual polymer is obtained by the distribution functions $P(\theta)$ and $P(\phi)$, where θ is the angle between the end-to-end vector and its projection onto the flow-gradient plane and ϕ is the angle between this projection and the flow direction[9,7]. As shown in Figure 4, the distribution function for the angle ϕ displays a significant shear rate dependence. At zero shear, no angle is preferred[9]. With increasing shear rate, a peak develops at finite ϕ, which shifts to smaller values with increasing $\dot{\gamma}$. At the same time, the width of the distribution decreases. Theoretical calculations[9] show that the angle at the maximum of the distribution equals the angle χ presented in Figure 3. The probability distribution $P(\phi)$ depends on the polymer concentration. The inset of Figure 4 displays the distribution functions for a dilute solution and a concentrated solution for approximately the same product of shear rate and relaxation time. Evidently, the distribution of the more concentrated system is broader, but the maxima of the distributions appear at the same angle, which is consistent with the scaling presented in Figure 3. We like to emphasize that the distribution function at the larger concentration was smaller and the maximum was shifted, if the factor $\mu(c)$ would be left out and systems of the same Weissenberg number were compared.

The probability distributions of the angle θ are displayed in Figure 5 for various Weissenberg numbers. Theoretical calculations[9] predict a cross-over from a Gaussian shape of the distribution function to a power-law decay according to $P(\theta) \sim \theta^{-2}$,within a certain range of angles, with increasing shear rate, which is confirmed by the simulations. We find a gradual change in the dependence of the distribution functions on the concentration. At low shear rates, the distribution functions for low and high concentrations are different. However, for the two largest Wi, we find that they agree well with each other. This behaviour is strongly related to the deformation of the polymers. For lower Weissenberg numbers, the polymers at higher concentrations are more strongly deformed along the flow direction than those in a dilute solution. Hence, they exhibit a narrower distribution in θ. In the high shear-rate regime of our simulations, the polymers exhibit approximately the same deformation independent of concentration, which is reflected in the similarity of the distributions. This seems not to apply to the distribution $P(\phi)$, as is evident from the difference in the alignment angle (cf. Figure 3).

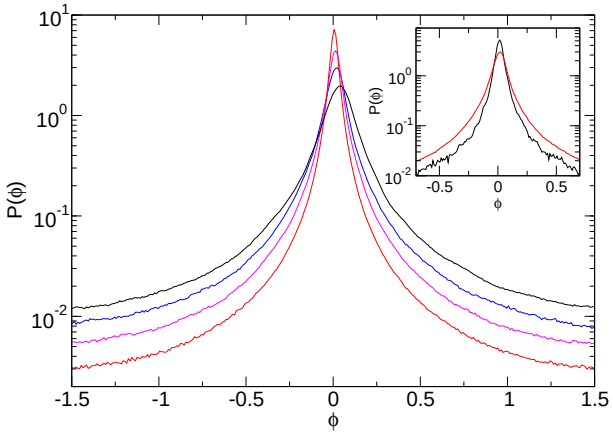


Figure 4. Probability distributions of the angle ϕ between the projection of the polymer end-to-end vector onto the shear plane and the flow direction for the Weissenberg numbers Wi $= 34$ (black), Wi $= 114$ (blue), Wi $= 340$ (magenta), and Wi $= 3400$ (red) of the system with $N_p = 3000$ polymers. The inset compares the distribution function of a dilute system (Wi ≈ 1140 (black)) with that of a concentrated system ($N_p = 3000$) with a comparable product of shear rate and relaxation time (Wi$\mu(c) \approx 1025$ (red)).

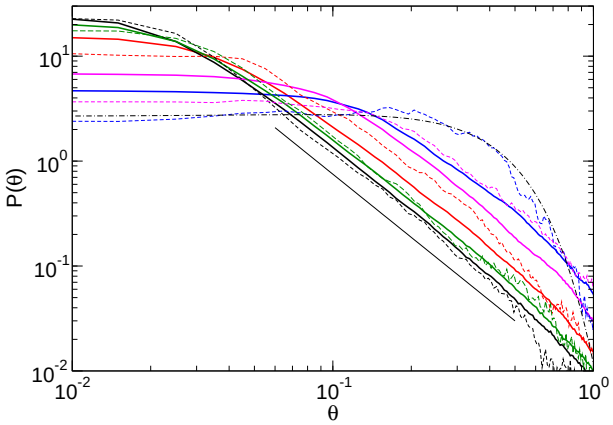


Figure 5. Probability distributions for the angle θ between the end-to-end vector and its projection onto the flow-gradient plane for the Weissenberg numbers Wi $= 11$ (blue), Wi $= 34$ (magenta), Wi $= 340$ (red), Wi $= 1100$ (green), and Wi $= 3400$ (black) of a dilute system (dashed lines) and for $N_p = 3000$ (solid lines). The dashed-dotted line is a Gaussian function with non-zero mean fitted to the distribution function for Wi $= 11$. The straight solid line has a slope of -2.

4 Conclusions

We have presented results for the conformational and structural properties of semidilute polymer solutions in a shear flow. The asymptotic (scaling) regime of polymers under shear is only reached for large Weissenberg numbers, which corresponds to long polymers and large overlap concentrations. Thus, the simulation of relevant polymer sizes under shear flow are only possible on massively parallel computers.

Acknowledgements

The simulations have been performed on JUGENE of the Research Centre Jülich within a grant provided by VSR.

References

1. M. Doi and S. F. Edwards *The Theory of Polymer Dyanmcis* (Clarendon Press, Oxford, 1986).
2. R. Kapral, *Multiparticle collision dynamics: Simulation of complex systems on mesoscales*, Adv. Chem. Phys. **140**, 89, 2008.
3. G. Gompper, T. Ihle, D. M. Kroll, and R. G. Winkler, *Multi-particle collision dynamics – a particle-based mesoscale simulation approach to the hydrodynamics of complex fluids*, Adv. Polym. Sci. **221**, 1, 2009.
4. P. Ahlrichs and B. Dünweg, *Simulation of a single polymer chain in solution by combining lattic Boltzmann and molecular dynamics*, J. Chem. Phys. **111**, 8225, 1999.
5. K. Mussawisade, M. Ripoll, R. G. Winkler, and G. Gompper, *Dynamics of polymers in a particle based mesoscopic solvent*, J. Chem. Phys. **123**, 144905, 2005.
6. D. E. Smith, H. P. Babcock, and S. Chu, *Single-polymer dynamics in steady shear flow*, Science **282**, 1724, 1999.
7. S. Gerashchenko and V. Steinberg, *Statistics of tumbling of a single polymer molecule in shear flow*, Phys. Rev. Lett. **96**, 038304, 2006.
8. A. Malevanets and R. Kapral, *Mesoscopic model for solvent dynamics*, J. Chem. Phys. **110**, 8605, 1999.
9. R. G. Winkler, *Semiflexible polymers in shear flow*, Phys. Rev. Lett. **97**, 128301, 2006.
10. M. Ripoll, R. G. Winkler, and G. Gompper, *Star polymers in shear flow*, Phys. Rev. Lett. **96**, 188302, 2006.
11. C. M. Schroeder, R. E. Teixeira, E. S. G. Shaqfeh, and S. Chu, *Dynamics of DNA in the flow-gradient plane of steady shear flow: Observations and simulations*, Macromolecules **38**, 1967, 2005.
12. G. Sutmann, R. G. Winkler, and G. Gompper, *Simulating hydrodynamics of complex fluids: Multi-particle collision dynamics coupled to molecular dynamics on massively parallel computers*, J. Comput. Phys., submitted for publication.
13. T. Ihle and D. M. Kroll, *Stochastic rotation dynamics: A Galilean-invariant mesoscopic model for fluid flow*, Phys. Rev. E **63**, 020201(R), 2001.
14. A. Malevanets and J. M. Yeomans, *Dynamics of short polymer chains in solution*, Europhys. Lett. **52**, 2231, 2000.
15. R. G. Winkler and C.-C. Huang, *Stress tensors of multiparticle collision dynamics fluids*, J. Chem. Phys. **130**, 074907, 2009.
16. M. Adam and M. Delsanti, *Viscosity and longest relaxation time of semidilute polymer solutons. I. Good solvent*, J. Physique **44**, 1185, 1983.

Earth and Environment

Ulrich Hansen

Institut für Geophysik, Westfälische Wilhelms-Universität Münster
Corrensstraße 24, 48149 Münster, Germany
E-mail: hansen@earth.uni-muenster.de

Earth and Environment can best be understood as one integrated system. Numerous studies are clearly revealing this message. Climatic changes may have been triggered not only by influences from the exterior, for example by the sun. Processes deep in the Earth's body can lead to intense phases of volcanism and such cause strong impacts on the Earth's atmosphere. The Earth has a magnetic field and this protects us against particle bombardment and radiation from the sun and from space. Variations in the Earth's magnetic field can change the conditions. The term Space Weather is by now commonly used. Not the fluctuations in temperature, air pressure, etc. in the Earth's atmosphere are meant. Instead Space Weather describes the fluctuation in Space, due to changes in solar radiation, leading to changes in the solar wind, a flux of charged particles originating from the Sun. When Space Weather reaches Earth it runs into the Earth's magnetic field and the outcome is a result of complex interactions between the flux, as emitted from the Sun and the magnetic field, as generated in the Earth's core. In fact, the Earth's magnetic field is generated by a dynamo process, deep in the molten Outer core of the Earth at a depth of more than 3000 km. I find it important to realize that the conditions which we meet on the Earth are the result of the interaction of processes on the Sun and in the deep Interior of the Earth. In my view this perfectly underlines the interconnectedness of processes which is anything but clear at a first glance. Understanding of Earth and Environment depends critically on an understanding of the individual processes and their relevant interactions.

An approach, well-balanced between data and models, seems to be the most promising strategy to understand the structure and evolution of Earth and Environment. In each of the contributions to the Earth and Environment section of this NIC- Symposium, different weight is given to either data or models. The work of Schuh et al. aims at determining the Earth's gravity field from satellite measurements (GOCE-mission). While the foundation of that type of analysis goes back to the times of Gauss, the new technologies allow to collect data, which need sophisticated processing techniques, in order to retrieve the essential information. In this case, two different sources of data are considered: One is the low frequency part of the gravity field as being measured by conventional Satellite to Satellite Tracking (SST), the other one is a new method, employing the principle of Satellite Gravity Gradiometry (SGG), in order to measure the high frequency components of the gravity fields. Combining the two sources poses a high dimensional data fitting problem, facing irregularly distributed data. Well designed algorithms and significant computer power is needed to solve this problem. Such an analysis will allow for a precise determination of mass anomalies within - and - on the surface of the Earth and will thus allow a sharper picture of the planets structure and even its dynamics.

Prediction lies at the heart of the work as presented by Elbern et al. In the atmospheric sciences, predictive power is usually gained from forward modeling, i.e. from integrating the partial differential equations governing the evolution of momentum, temperature, pressure and chemical constituents, in space and time. Deterministic models, however, suffer from uncertainties in the initial conditions, and from poor data quality. Thus the principally well posed initial value problem becomes, under realistic conditions, a severely underdetermined problem. Additional information, either from observations or often from the model itself has to be taken into account. Here the information can be of statistical character, or can be based on knowledge about the physics of the system and thus be of dynamical nature. After all, the whole task, to determine the state of the atmosphere form an inverse problem, in which temperature, pressure and other key-quantities are estimated from observed data and from model results. In a novel approach, Elbern et al. employ methods from variational calculus to develop new methods of data assimilation and powerful inversion techniques.

At a first glance, the contribution by Stellmach and Hansen deals with a purely fundamental problem: The dynamics of a vigorously convecting fluid, subject to strong rotation. This study is motivated by the wish to better understand the dynamics of the Earth's core. The Earth's core is essentially a body of liquid metal, whose motion is strongly influenced by the planet's rotation. Up to now, it was not clear under which circumstances the rotation dominates over turbulent convection or vice versa. Simulating the dynamics of the Earth's core under realistic condition seems even today an elusive goal. Still it is possible with the power of today's supercomputers to unravel at least certain aspects of the dynamics. The present study shows that the boundary layer behaviour of the system seems to determine its state, rather than a global force balance. Applied to the Earth core, one would expect the flow within the core dominated by rotation, however, much less than estimated in previous studies. Further extrapolated, a core, being at the border between rotation-dominated and turbulence-dominated might temporarily be within the one or the other regime, and this basically could explain sudden changes in the dynamics of the core. If this has any bearing on the reversals of the Earth's magnetic field, can not be answered by now and remains a speculative but likewise tantalizing issue.

IMACCO: Virtual Institute for Inverse Modelling of the Atmospheric Chemical Composition

Hendrik Elbern[1,2], Achim Strunk[1], Lars Nieradzik[1,2], Nadine Goris[1,2], Jörg Schwinger[1], and Ketevan Kasradze[1]

[1] Rhenish Institute for Environmental Research at the University of Cologne
Aachener Str. 209, 50931 Köln, Germany
E-mail: {he, as, ln, ng, js, kk}@riu.uni-koeln.de

[2] Forschungszentrum Jülich, ICG-2, 52425 Jülich, Germany

Today, it is widely recognised that the present changes in the earth's environment challenges our ability to detect and understand the various atmospheric processes and to assess the resulting impacts. The urgently needed knowledge about the present state of the atmosphere as part of the earth system, its atmospheric trace gas and aerosol budgets and their future variability requires reliable analyses, resting on information from observations of the governing processes. Even increasingly accurate and comprehensive measurements only provide sparse snapshots of the instantaneous state of the system. Numerical models produce complete space-time data sets based on our present yet incomplete knowledge of the underlying processes. Hence, the results of pure model calculations are of limited and, more importantly, often unknown validity. Improved conclusions must be inferred from processing procedures, that are able to combine observational data with models in a consistent and synergistic manner. The requested techniques can only be provided by advanced spatio-temporal data assimilation and inversion methods. The mission, objective, and work program of IMACCO are aligned to this basic principle. A key technique adopted here is the variational calculus, resting on the tangent-linear and adjoint model derivatives.

1 Introduction

The combined task to analyse and predict the atmospheric chemical composition – given such diverse, yet error affected information sources like deterministic models, emission inventories, *in situ* observations and space borne remote sensing data – requires a methodological shift beyond traditional simulations by numerical integration of partial diffential equations. For a proper system understanding, we need to know the chemical system state depending on observability and predictability. Observability encompasses the available extent of observing network with respect to known constraints by chemical laws, while predictability involves sensitivity to parameter modifications compared with our precision of knowledge. The virtual institute IMACCO seeks to address these issues by developing and operating a wide range of novel inverse modelling extensions, which supplement the forecast system. Here, advanced spatio-temporal data assimilation and inverse modelling techniques deal with observability while targeted observation techniques cope with predictability. In IMACCO regional tropospheric air quality inverse modelling as well as global middle atmospheric data assimilation is addressed. In all applications, the tangent-linear and adjoint model versions play crucial roles and had to be developped. The computational problems include meteorological and chemistry transport model integrations and optimisation procedures with a degree of freedom of order $\mathcal{O}(10^7)$. Consequently, the algorithms applied are computationally extremely demanding. Several novel algorithmic

techniques had to be implemented to render these tasks feasible. The access to the NIC JSC compute platforms is prerequisite to achieve this task.

2 Tropospheric Inverse Modelling

The core of all tropospheric inverse modelling techniques described in this chapter is the EURAD-IM (EURopean Air pollution and Disperation - Inverse Modelling) chemistry-transport model (CTM). It simulates the evolution of tropospheric trace gases and aerosols in time and space by taking into account transport, diffusion, chemical transformation, wet and dry deposition, and sedimentation.

2.1 Optimisation of Initial Values and Emission Factors for Gas-Phase Chemistry

Chemical weather forecasting in the troposphere is governed by several physical and chemical processes and controlled by parameters which are not known with sufficient precision. 4-dimensional data assimilation (4D-var) admits for the optimisation of further parameters than initial values only. Generally, those parameters should be optimised that are poorly known and to which the system's evolution in time is very sensitive. In tropospheric chemical data assimilation initial values and emission rates are a suitable set of optimisation parameters.

The basic idea of 4D-var is to minimise a scalar cost-function $\mathcal{J}$, that measures the distance between a CTM model run and the observations within a predefined time-span (also referred to as assimilation interval) on the one hand, and an appropriate background field on the other:

$$\begin{aligned}\mathcal{J}(\delta\mathbf{x}(t_0),\delta\mathbf{e}) =& \frac{1}{2}(\delta\mathbf{x}(t_0))^T\mathbf{B}^{-1}\delta\mathbf{x}(t_0) + \frac{1}{2}\int_{t_0}^{t_F}(\delta\mathbf{e})^T\mathbf{K}^{-1}\delta\mathbf{e}dt + \\ & \frac{1}{2}\int_{t_0}^{t_F}\left(\mathbf{d}(t)-\mathbf{H}(t)\delta\mathbf{x}(t)\right)^T\mathbf{R}^{-1}(\mathbf{d}(t)-\mathbf{H}(t)\delta\mathbf{x}(t))dt\end{aligned} \tag{1}$$

Here $\delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}_b$ and $\delta\mathbf{e}(t) = \mathbf{e}(t) - \mathbf{e}_b(t)$ are the deviation of the background chemical state $\mathbf{x}_b$ and the emission inventory $\mathbf{e}_b(t)$, whereas $\mathbf{x}_b$ is usually obtained from a preceeding forecast or assimilation run. $\mathbf{d}(t) := \mathbf{y}(t) - \mathbf{H}(t)\mathbf{x}_b(t)$ is the observation minus model discrepancy at time t ($\mathbf{y}(t)$ is the vector of available observations, $\mathbf{H}(t)$ calculates the model equivalent to each observation). The error covariance matrices grant a proper weighting of the reliability of information and are denoted as follows: $\mathbf{B} \in \mathbb{R}^{N\times N}$ with N the number of model variables for the first guess or background values, $\mathbf{K} \in \mathbb{R}^{E\times E}$ with E the number of emitting grid points times emitted species for the emission factors and $\mathbf{R} \in \mathbb{R}^{M(t)\times M(t)}$, with $M(t)$ the number of available observations at time t for observation errors. For a more detailed description see for example Elbern et al., 1997[1] and 2007[2].

The gradient of $\mathcal{J}$ with respect to the joint chemical state and emission rate variable $\mathbf{z} = (\delta\mathbf{x}(t_0), \delta\mathbf{e})^T$ has to be determined for minimisation. The adjoint model operator, formally integrating from time t_F backward in time to the initial time t_0 is needed for this purpose. The optimisation problem can be solved by a quasi–Newton minimisation procedure, e.g. the limited memory BFGS[6].

A both crucial and critical part of data assimilation is the generation of the background error covariance matrix $\mathbf{B}$ as introduced in Eq. (1). A costly but very accurate way to generate $\mathbf{B}$ is to set up an ensemble with perturbed emissions or initial values so that $\mathbf{B}$ can be derived via

$$\mathbf{B}_{ij} = \frac{1}{K} \sum_{n=1}^{K} \left(x_i^{(n)} - \overline{x}_i \right) \left(x_j^{(n)} - \overline{x}_j \right), \tag{2}$$

where $\mathbf{B}_{ij}$ is the background error covariance between boxes i and j, K is the number of ensembles, $x_l^{(n)}$ the model state of a species of ensemble member n in grid cell l and $\overline{x}_l$ the ensemble mean of a species.
A key problem in applying algorithms with the whole matrix $\mathbf{B}$ is, that excessively high storage capacities are needed. To circumvent the necessity of storing $\mathbf{B}$ in Eq. (1) Weaver and Courtier[4] proposed a generalised diffusion operator to replace $\mathbf{B}$ while granting the statistical properties of covariance matrices, say positive definiteness and symmetry. The fundamental solution $\psi(z,t)$ of the diffusion equation at location z and for time t, for clarity reading in one-dimensional form, reads

$$\frac{\partial \psi}{\partial t} - \kappa \frac{\partial^2 \psi}{\partial z^2} = 0, \tag{3}$$

which is given by the convolution of $\psi(z,0)$ with a Gaussian

$$\psi(z,t) = \frac{1}{\sqrt{4\pi\kappa t}} \int \exp\left(-\frac{(z-z')^2}{4\kappa t} \right) \psi(z',0)\, dz'. \tag{4}$$

After normalization a valid correlation function can be defined with

$$2\kappa t =: L^2 \tag{5}$$

acting as the square of a correlation length-scale. Under the assumption of the model acting Gaussian and taking into account that $\mathbf{B}_{ij}$ can be determined for two grid cells via Eq. (2) and r being the distance between the centres of these cells, one can define

$$\mathbf{B}(r) = \mathbf{B}(0)\, exp\left(-\frac{r^2}{2L^2} \right) \quad \Rightarrow \quad L = \left(2ln\left(\frac{\mathbf{B}(0)}{\mathbf{B}(1)} \right) \right)^{-\frac{1}{2}}, \tag{6}$$

where r is expressed in scales of grid-cell spacing and $\mathbf{B}(0) = \frac{1}{2}(\mathbf{B}_{ii} + \mathbf{B}_{jj})$. Only adjacent cells are being considered resulting in $r = 1$. The κ derived via Eq. (5) (one κ for each direction North, East, Top) are then stored instead of $\mathbf{B}$, resulting in a memory demand of the order $\mathcal{O}(n)$ instead of $\mathcal{O}(n^2)$.

As an example of the outcome of the described techniques, the impact of emission rate optimisation is demonstrated in Figure 1. It shows SO_2, CO, NO_2, and xylene emission factors over the integration domain of a 2 km grid. Since Berlin is a large urban island within a rural environment, sulfur emissions are confined to the greater metropolitan area. The upper left panel of Figure 1 clearly indicates a nearly overall reduced emission rate over the densely populated area. In the case of CO similar effects can only be claimed for the area of east Berlin. The effects vary moderately for NO_2, while xylene appears to be underestimated by the emission inventory, with an amplification factor of about 1.2. In all exhibited cases, the inversion results remain well within the error limits of the inventory. Emission rate optimisation of SO_2 and CO is mainly based on concentration observations

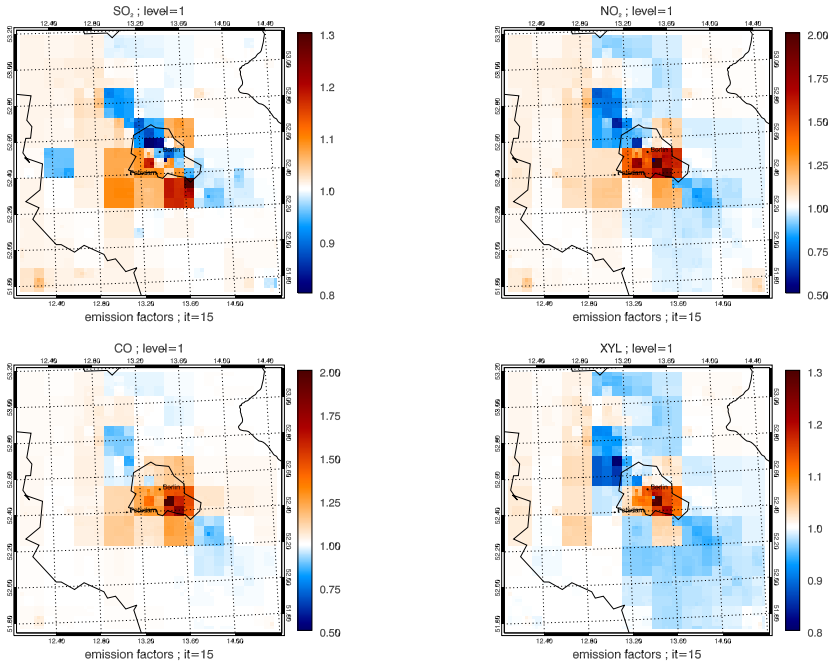


Figure 1. SO_2, CO, NO_2, and xylene optimisation factors over the integration domain of the finest grid (2 km resolution) over the greater Berlin area for the surface layer. Coloured squares indicate the impact areas of the individual nests, ranging from 54 km to 2 km resolution, involving 4 nest levels.

of these species. In the case of other emissions, which are rarely observed, inference can only rest on measured product constituents, most prominently ozone.

2.2 Aerosol Assimilation

The aerosol component of the EURAD-IM model system includes inorganic and secondary organic aerosol modules. The Modal Aerosol Dynamics model for Europe (MADE) has been developed for the EURAD-IM model to allow for a detailed treatment of aerosol effects. Due to its complexity no full adjoint is available yet. Therefore, a 3D–var algorithm has been implemented, which, in contrast to 4D–var, only involves the adjoint of the observation operator $\mathbf{H}$, i.e. only a spatial assimilation is accomplished at several times and no emission factors are optimised[7].

Aerosols play a very important role in atmospheric chemistry modelling. They have direct impact on the global radiative transfer balance and, thus, on climate as well as on local air quality and weather. Since aerosols are not, as gas-phase species, simply part of the ambient air, but instead a rather arbitrary agglomeration of suspended particles like soot or mineral dust and/or an aqueous solution of SO_4^{2-}, NO_3^- and NH_4^+ mainly. The sources of these particles often are non-periodic events like wild fires, volcanic eruptions or dust outbreaks which are difficult to predict. Furthermore, in contrast to gas-phase species measurements, aerosol observations are generally given in integrated units, like PM_{10} (in-

tegrated particulate matter with a diameter of less than 10 μm in $\left[\frac{\mu g}{m^3}\right]$) or, as from several satellite instruments or retrieval algorithms, as aerosol optical thickness (AOT). Figure 2 (upper left panel) shows predicted PM_{10} concentrations for July 14, 2003 11:00 UTC that are rather moderate to low. The image in the bottom row displays *in situ* measurements (diamonds) and SYNAER PM_{10} retrievals (ashlars) at the same time, showing much higher values from the Iberian Peninsula to the North Sea deriving from vast wild fires in Spain and Portugal in that period. The upper right image shows the analysis based on all available measurements and the background field after assimilation.

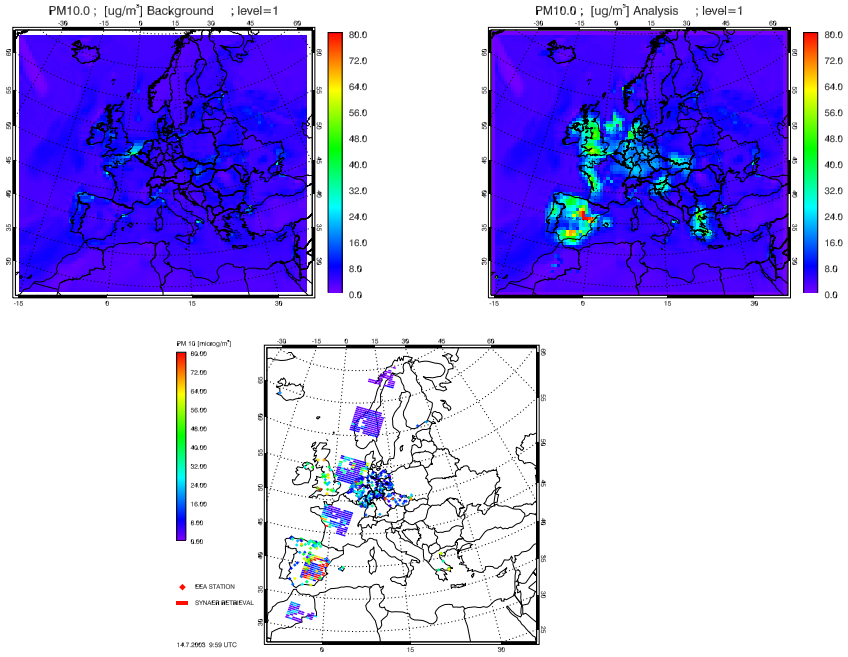


Figure 2. PM_{10} concentrations [μg/m^3] on July 14, 2003, 11:00 UTC, as forecasted without assimilation (upper left), assimilation results based on all observations (upper right) and available observations, both *in situ* and SYNAER retrievals (bottom row).

2.3 Targeted Observations

Adaptive observations taken in well defined targeted areas can reduce the initial condition uncertainty and decrease forecast errors. Reasonable target areas are unstable system modes, which can be detected by means of singular vectors. The singular vectors of a numerical model identify the directions of fastest pertubation growth over a finite time interval and are therefore associated with dynamically sensitive system states. Within IMACCO the main attention is paid to targeted observations of chemical species. Here, not only the identification of sensitive regions is of special interest, but furthermore the specification of sensitive species. Since this problem is concerned with determining the evolution of initial

errors, the ratio between perturbation magnitudes at final time t_F and initial time t_I is used to define a measure of error growth

$$\mathcal{E}\left(\delta\mathbf{c}(t_I)\right) = \frac{\|\delta\mathbf{c}(t_F)\|^2_{A_F}}{\|\delta\mathbf{c}(t_I)\|^2_{A_I}}, \tag{7}$$

with $\delta\mathbf{c}(t)$ denoting the uncertainties of chemical species at time t and $\|\delta\mathbf{c}(t)\|_{A_{I/F}} = <\delta\mathbf{c}(t), \mathbf{A}_{I/F}\,\delta\mathbf{c}(t)>$ defining an inner product by a positive definite matrix $\mathbf{A}_I$ and a semipositive matrix $\mathbf{A}_F$. Assuming the initial uncertainty to evolve linearly, the final uncertainty $\delta\mathbf{c}(t_F)$ can be expressed in terms of the initial uncertainty:

$$\mathcal{E}\left(\delta\mathbf{c}(t_I)\right) = \frac{\|\mathcal{L}_{t_I,t_F}\,\delta\mathbf{c}(t_I)\|^2_{A_F}}{\|\delta\mathbf{c}(t_I)\|^2_{A_I}} = \frac{\delta\mathbf{c}(t_I)^T {\mathcal{L}_{t_I,t_F}}^T \mathbf{A}_F\,\mathcal{L}_{t_I,t_F}\,\delta\mathbf{c}(t_I)}{\delta\mathbf{c}(t_I)^T\,\mathbf{A}_I\,\delta\mathbf{c}(t_I)}. \tag{8}$$

The operators $\mathcal{L}_{t_I,t_F}$ and ${\mathcal{L}_{t_I,t_F}}^T$ are denoted as the tangent linear and adjoint model respectively. The vectors that maximize the ratio $\mathcal{E}$ are the singular vectors $\mathbf{s}_i(t_I)$ with largest singular values σ_i^2:

$${\mathcal{L}_{t_I,t_F}}^T \mathbf{A_F}\,\mathcal{L}_{t_I,t_F}\,\mathbf{s}_i(t_I) = \sigma_i^2\,\mathbf{A_I}\,\mathbf{s}_i(t_I). \tag{9}$$

The solutions $\mathbf{s}_i(t_I)$ depend on the selection of the $\mathbf{A}$-norm. Using different matrices $\mathbf{A}$ allows local and chemical projection as well as scaling by typical concentrations to gain a relative error growth.

The forward (tangent linear) and backward (adjoint) evolution of the chemical uncertainties in space and time is calculated with the EURAD-IM model system, which provides transport, diffusion, and gas-phase transformation as forward and adjoint model. The tangent linear model for each process has been developed. The eigenvalue problem in Eq. (9) can be solved efficiently using the software package PARPACK.

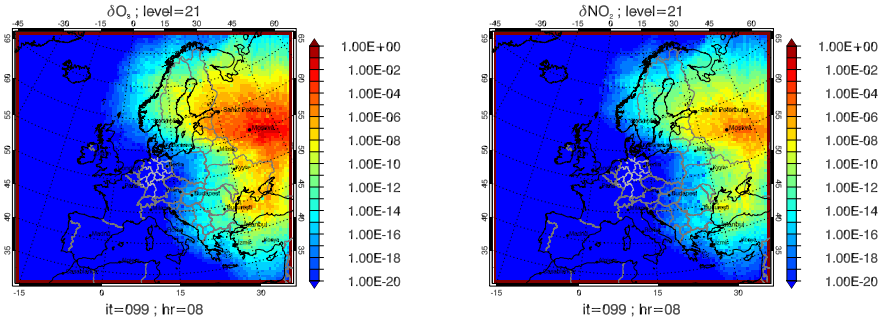


Figure 3. 1^{st} projected singular vector (O_3 and NO_2 entries in altitude range from 8.8 to 10 km).

Figure 3 shows the initial fractions of O_3 and NO_2 (in altitude range from 8.8 to 10 km) of the directions of largest relative error growth. The 10h-simulation started on October 10, 2008, 8.00 UTC, with initial projection on all species and final projection on O_3. No local projection was applied. Since the directions of largest error growth are dimensionless, the

vectors were set to unit length. The singular vectors contain information for each species on each grid point, they demonstrate how sensitive each location and each species is. Figure 3 points out that for the given configuration O_3 is more sensitive than NO_2 in altitude range from 8.8 to 10 km and that both species are more sensitive in the east than in the west.

3 Stratospheric Data Assimilation with SACADA

SACADA is a 4D-Var assimilation system for stratospheric trace gas observations[3] which makes use of the diffusion approach (see Section 2.1). The German Weather Service's global forecast model GME[8] is integrated in the system in order to compute the meteorological fields at the same temporal and spatial locations where they are needed by the chemistry module. Storing the meteorological states for each time step and recoverage for forward and adjoint integration of the chemistry module avoids temporal and spatial interpolation errors of meteorological parameters. GME has been selected because of its icosahedral grid design (Figure 4), which provides a nearly homogeneous distribution of grid points over the globe, avoiding singularities at the poles and crowding of grid points due to meridional convergence at high latitudes. Given the fact that most computational burden is associated with the solution of the stiff ordinary differential equations of chemistry at each grid point, a high numerical efficiency is realized by this grid design. To obtain rectangular data structures, two adjacent icosahedral triangles are combined to form a diamond, partitioning the grid into ten logically rectangular sub-grid domains as marked by the bold lines in Figure 4. The model setup is parallelized using a diamond-wise domain decomposition. Each processor works on one portion of each diamond. This is a simple yet effective strategy to achieve a good load balancing between processors. Each domain has a halo of two rows and columns of grid points that have to be exchanged among processors twice each time-step using the Message Passing Interface (MPI) standard.

Within IMACCO, the region of the upper troposphere/lower stratosphere is of special interest. It is characterized by a strong spatial and temporal variability of the dynamic

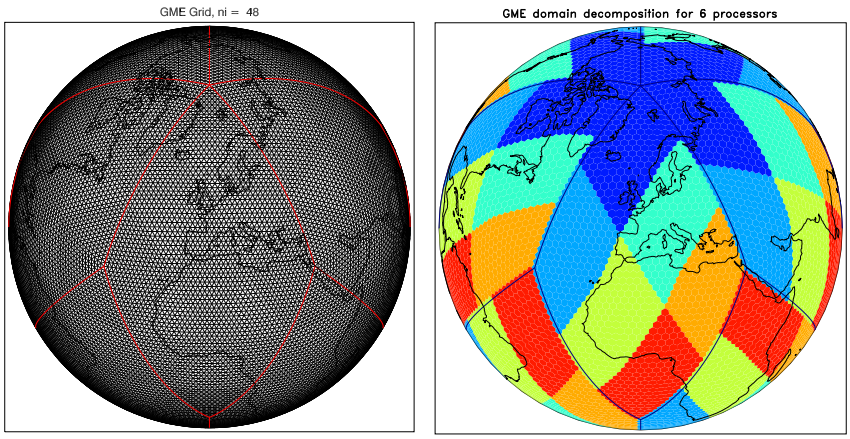


Figure 4. Left hand side: Icosahedral grid with a resolution of 147 km. Right hand side: Domain decomposition for six processors. Each colour indicates a region that is assigned to one processor.

structures. This variability spreads by transport and transformation processes to the trace gas distribution in this region. For this reason the grid of the assimilation system SACADA was refined. The horizontal grid point distance was reduced from 250 km to 147 km and the vertical distance of the model layers up to a height of 22 km has become less than 1 km. According to the specific tropospheric chemistry the stratospheric chemical mechanism of the SACADA assimilation system was extended and contains additional 12 photolysis, 25 gasphase and 3 heterogeneous reactions. These changes gave way to the proper assimilation of small-scale aircraft measurements of CRISTA-NF.

Acknowledgements

The authors are indebted to their IMACCO-partners: the Institute for Chemistry and Dynamics of the Geosphere I/II at the Forschungszentrum Jülich GmbH, the Institute for Meteorology and Climate Research at the Forschungszentrum Karlsruhe GmbH, the Institute for Meteorology and Climate Research at the University of Karlsruhe and the Institute for Geophysics and Meteorology at the University of Cologne. Compute facilities were granted from the Jülich Supercomputing Centre of the Research Centre Jülich.

References

1. H. Elbern and H. Schmidt and A. Ebel, *Variational data assimilation for tropospheric chemistry modeling*, J. Geophys. Res. **102**, 15,967, 1997.
2. H. Elbern, A. Strunk, H. Schmidt and O. Talagrand, *Emission rate and chemical state estimation by 4-dimensional variational inversion*, Atmos. Chem. Phys. **7**, 3749, 2007.
3. H. Elbern, J. Schwinger and R. Botchorishvili, *Chemical State Estimation for the Middle Atmosphere by 4-Dimensional Variational Data Assimilation, Part I: System Configuration*, J. Geophys. Res. , submitted, 2009.
4. A. Weaver and P. Courtier, *Correlation Modelling on the Sphere Using a Generalized Diffusion Equation*, Q. J. R. Meteorol. Soc. **127**, 1815, 2001.
5. H. J. Eskes and K. F. Boersma, *Averaging kernels for DOAS total-column satellite retrievals*, Atmos. Chem. Phys **3**, 1285, 2003.
6. D. C. Liu and J. Nocedal, *On the limited memory BFGS method for large scale optimization*, Math. Programming **45**, 503, 1989.
7. L. P. Nieradzik and H. Elbern, *Variational assimilation of combined satellite retrieved and in situ aerosol data in an advanced chemistry transport model*, Proceedings of the ESA Atmospheric Science Conference, Frascati, 2006.
8. D. Majewski, D. Liermann, P. Prohl, B. Ritter, M. Buchhold, T. Hanisch, G. Paul, W. Wergen and J. Baumgardner, *The Operational Global Icosahedral-Hexagonal Gridpoint Model GME: Description and High-Resolution Tests*, Month. Weath. Rev. **130**, 319-338, 2001.

Turbulent Rotating Rayleigh-Bénard Convection

Stephan Stellmach and Ulrich Hansen

Institut für Geophysik, Westfälische Wilhelms-Universität Münster
Corrensstrasse 24, D-48149 Münster, Germany,
E-mail: {stellma,hansen}@earth.uni-muenster.de

Convectively driven, rotating turbulence is ubiquitous in geo- and astrophysical systems. Here, we investigate the rotating Rayleigh-Bénard system by means of direct numerical simulations. Our simulation results allow us to establish scaling relationships for the heat transport which can be readily applied to a broad class of geophysical and astrophysical systems. In particular, we identify two different dynamical regimes: one in which rotation dominates and where the Nusselt number Nu scales approximately as $Nu \sim Ra^{6/5}$, and one in which the heat transfer is not affected by rotation and where hard-turbulence behaviour with $Nu \sim Ra^{2/7}$ is observed. We further hypothesize that the transition from the rotation dominated regime to the hard turbulence regime is tied to boundary layer dynamics, in sharp contrast to widely used global force balance arguments.

1 Introduction

Turbulent convective flows under the influence of strong Coriolis forces occur in many geophysical and astrophysical bodies. They are important in ocean and atmosphere dynamics and they drive the strong turbulence observable in the outer layers of the sun and of the gas giants Jupiter, Saturn and Neptune. Rotating convection also occurs deep in planetary interiors where the resulting flows of electrically conducting fluids are thought to generate planetary magnetic fields.

The ubiquity of rotating convection in geo- and astrophysical systems poses challenging problems for a thorough understanding of these systems. For example, an exact estimation of the heat transfer resulting from highly turbulent convective motions is crucial for typical thermal evolution models. Other applications require a detailed knowledge of the flow characteristics. The large scale magnetic fields generated by planets and stars for example are believed to be a consequence of certain spatial and temporal flow correlations deep in their convective interiors[1]. These correlations are thought to be induced by Coriolis forces, and it is important to predict how strong the background rotation needs to be in order to organize the flow accordingly.

In this paper, we consider the special case of rotating Rayleigh-Bénard convection, a system that has received considerable attention during the last years[2–4]. While being geometrically simple, this configuration retains most of the basic ingredients characteristic for a broad class of geophysical and astrophysical flows.

2 Model Description

The model consists of a rapidly rotating fluid layer that is subject to a uniform vertical gravity field. A constant temperature difference between the bottom and top bound-

aries causes local density variations which drive convective motions. In the Boussinesq-Approximation, the system behaviour is governed by the equations

$$\frac{\partial \mathbf{u}}{\partial t} - Pr\nabla^2\mathbf{u} + \frac{Pr}{E}\mathbf{e}_z \times \mathbf{u} = -\nabla P + PrRaT\mathbf{e}_z - (\nabla \times \mathbf{u}) \times \mathbf{u} \ , \tag{1}$$

$$\frac{\partial T}{\partial t} - \nabla^2 T = -\mathbf{u} \cdot \nabla T \ , \tag{2}$$

$$\nabla \cdot \mathbf{u} = 0 \quad , \tag{3}$$

where $\mathbf{u}, P$ and T denote the dimensionless velocity, modified pressure and temperature. $\mathbf{e}_z$ is the vertical unit vector. The layer height L has been used as the fundamental length scale and time has been scaled by the thermal diffusion time L^2/κ, where κ denotes the thermal diffusivity. The temperature difference ΔT between the top and the bottom boundary has been used to scale temperature. The three arising dimensionless similarity parameters

$$Ra = \frac{\alpha \, \Delta T \, g_0 L^3}{\kappa\nu}, \quad E = \frac{\nu}{2\Omega L^2}, \quad Pr = \frac{\nu}{\kappa} \tag{4}$$

are the Rayleigh number Ra, the Ekman number E and the Prandtl number Pr. Here, α denotes the thermal expansion coefficient, g_0 is the acceleration due to gravity, ν is kinematic viscosity and Ω denotes angular velocity. We use fixed temperature boundary conditions $T(z=0) = 1, T(z=1) = 0$ and focus on no slip boundary conditions for the velocity field, $\mathbf{u}(z=0) = \mathbf{u}(z=1) = 0$.

For the numerical solution, the above equations are expressed in terms of toroidal and poloidal scalars. The resulting coupled system of partial differential equations is then solved by a pseudo-spectral method. All unknowns are expanded into Fourier series in the horizontal direction. In the vertical direction, a Chebyshev expansion on a Gauss-Lobatto grid is used. Fast transform algorithms can then be applied to switch between physical and transform space. A semi-implicit multistep method is employed for time integration. All linear terms, including the Coriolis force, are treated implicitly by a second order Backward-Differencing scheme, while a second order Adams-Bashforth scheme is applied to the nonlinear terms. An implicit treatment of rotational forces is beneficial at small E since it prevents the time step size from being restricted by the characteristic time scale of inertial waves. A thorough description of the numerical algorithm can be found elsewhere[5].

Our simulation code uses a purpose-build transpose-based parallel FFT algorithm, following an approach which had perviously been shown to yield good performance on the Blue Gene architecture[6]. As an example, Figure 1 shows the speedup for fixed problem sizes as a function of the number of tasks. Even for moderate problem sizes of 256^3 grid points, much smaller than those typically used in highly resolved simulations, the code scales up to $O(10^4)$ processors.

3 Investigated Parameter Space

The parameter values have been chosen in order to cover an area of parameter space as large as possible. Figure 2 gives a graphical sketch of the investigated parameter space. Six Ekman numbers, $E = \infty, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$ and 10^{-6} have been considered,

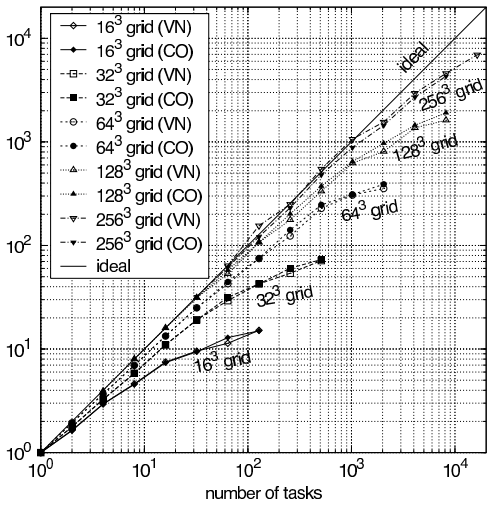


Figure 1. Scaling of our model on the Blue Gene. Here, the so-called *strong scaling* properties are shown. For fixed problem sizes ranging from 16^3 to 256^2, the speedup is shown as a function of the number of MPI tasks. Since for high resolution the local memory is insufficient to run the code on a single CPU, the speedup results were obtained by assuming an ideal speedup for the lowest possible number of CPUs.
This plot contains results for the Blue Gene/L system, which can operate in two modes, the *virtual node* (VN) mode and the *communication coprocessor* (CO) mode. Both CPUs can be used for computation in VN mode, whereas one CPU is dedicated exclusively to communication in CO mode. The plot shows that the differences in speedup are small. This is also the case for the total execution time, which means that our code is able to exploit both processors of each node efficiently. On the Blue Gene/P machine, a DMA engine has been introduced which frees the cores from managing data communications. This leads to a slightly better scaling on the Blue Gene/P supercomputer.

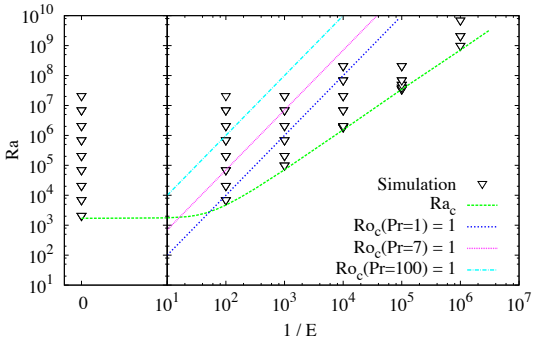


Figure 2. The parameter space covered by this study. Triangles mark numerical simulations. Typically, for each combination (E, Ra), computations where performed for $Pr = 1$, $Pr = 7$ and $Pr = 100$, adding up to roughly 120 simulations in total. Also shown are the critical Rayleigh numbers for the onset of convective instability as well as lines corresponding to unity convective Rossby number for the three Pr values considered.

covering non-rotating as well as rapidly rotating cases. It is a well known result of linear stability theory that rotation inhibits the onset of convective instability, with the critical Rayleigh number Ra_c scaling as $E^{-4/3}$ for small Ekman number. For each of the Ekman number values given above, we increased the Rayleigh number from weakly supercritical values to the point where it becomes impossible to resolve the solution at reasonable computational cost. We further varied the Prandtl number in order to investigate in how far the molecular diffusivities influence the dynamics. Three Prandtl numbers with $Pr = 1$, typical for a gas like air, $Pr = 7$, typical for water at room temperature and $Pr = 100$, typical for viscous silicon oils, have been considered.

Care has been taken to guarantee that the assumed horizontal periodicity does not overly constrain the dynamics. It is important in this context to choose a sufficiently large aspect ratio Γ, defined as the ratio of the maximum allowed horizontal wavelength to the layer depth. We used $\Gamma = 4$ for the non-rotating case. At lower Ekman number, smaller aspect ratios suffice since the dominant spatial scales decrease with increasing rotation. All simulations were started from a static state $\mathbf{u} = 0$, with a small random perturbation in the temperature field. After an initial transient, the system settles into a statistically stationary state, from which well defined averages can be obtained.

4 Results

Convective flows are traditionally classified by the vertical heat flux caused by the fluid motions. A convenient non-dimensional measure of heat transfer efficiency can be obtained by normalizing the total heat transport by the heat transfer that would occur in a solid body with the same thermal properties. The resulting ratio is called the Nusselt number Nu, and in general Nu is a function of Ra, Pr and E.

The time averaged heat transport computed from our simulations is shown in Figure 3. For non-rotating flows ($E = \infty$), we find that for all investigated Prandtl numbers the Nusselt number finally approaches a power law scaling of the form $Nu \sim Ra^{\alpha}$. The exponent α is close to $2/7$, consistent with the so-called "hard" turbulence regime of convection[7]. If the system rotates, the onset of convection is delayed by the action of Coriolis forces. When convective instabilities finally set in, Nu increases much more rapidly with Ra than in the non-rotating case. The simulations in this regime suggest a scaling exponent α that considerably exceeds one, roughly consistent with $Nu \sim Ra^{6/5}$, a scaling law that had been suggested recently for convection in rotating spherical shells[8,9].

The different scaling behaviour found in both regimes reflects differences in the spatial structure of the flow field. Figure 4 illustrates this by means of visualizations of the temperature field, showing both a typical non-rotating case and a rapidly rotating example in which the flow is strongly influenced by Coriolis forces. If the system is not rotating, the flow field exhibits a disordered cellular structure, with the kinetic energy being distributed among a broad range of spatial and temporal scales. In contrast, strong rotation tends to organizes the flow field into narrow columnar vortices connecting both boundaries. Most of the heat transport occurs within these so-called Taylor columns.

For strong enough thermal forcing, the constraining influence of rotation finally vanishes, leading to similar behaviour in both rotating and non-rotating cases. This is reflected in Figure 3 by the fact that the heat transport for the rotating cases finally approaches the non-rotating hard turbulence scaling. The same effect is observed in laboratory experiments of Rayleigh-Bénard convection where heat transport measurements are possible at

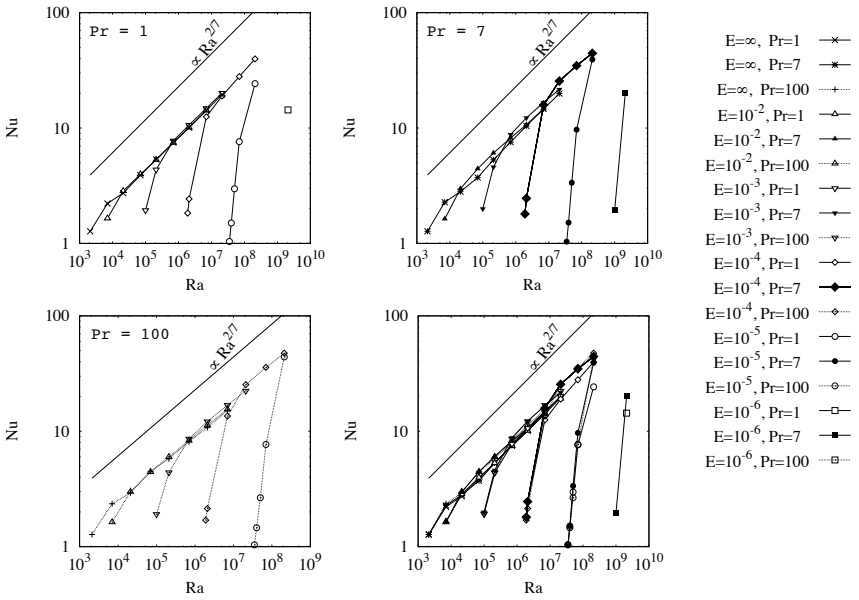


Figure 3. Functional dependence of the Nusselt number Nu on the Rayleigh number for three different Prandtl and five different Ekman numbers. For moderately super-critical Rayleigh numbers, Nu increases steeply with Ra. At high enough thermal forcing, all curves finally approach the classical hard-turbulence scaling law $Nu \propto Ra^{2/7}$.

Rayleigh numbers higher than those achievable in simulations[10]. Note that before rotating and non-rotating cases finally exhibit similar heat transport scaling laws, an intermediate regime is observed where rotation can effectively *increase* the heat transport as compared to non-rotating flows. This effect can be explained by so-called *Ekman pumping*[11], i.e. by secondary flows driven by the mechanical boundary layers.

From a geophysical or astrophysical perspective, we are interested in predicting the importance of rotation for specific natural systems, i.e. for given sets of control parameters. We may define the transition point between rotation dominated and essentially non-rotating convection by the point of intersection between their respective scalings. Equating the two, we find that the above results suggest a transitional Rayleigh number $Ra_t \sim E^{-7/4}$ and a transitional Nusselt number $Nu_t \sim E^{-1/2}$. When $Ra < Ra_t$, heat transport is affected by rotation, while for $Ra \gg Ra_t$, the heat transfer scaling is similar to the non-rotating case.

It is interesting to compare this result with previously published *a priori* estimates of the dynamical effects of rotational forces in convective flows[2]. A frequently used argument is based on the force balance in the interior of the convective region. The Rossby number $Ro = EPr^{-1}U$, where U denotes a typical dimensionless flow velocity, is a dimensionless measure of the importance of inertial forces relative to Coriolis forces. If we assume that inertia scales with buoyancy and employ the free-fall assumption $U \approx \sqrt{RaPr}$, the Rossby number becomes the so-called *convective Rossby number* $Ro_c = \sqrt{RaE^2/Pr}$. If

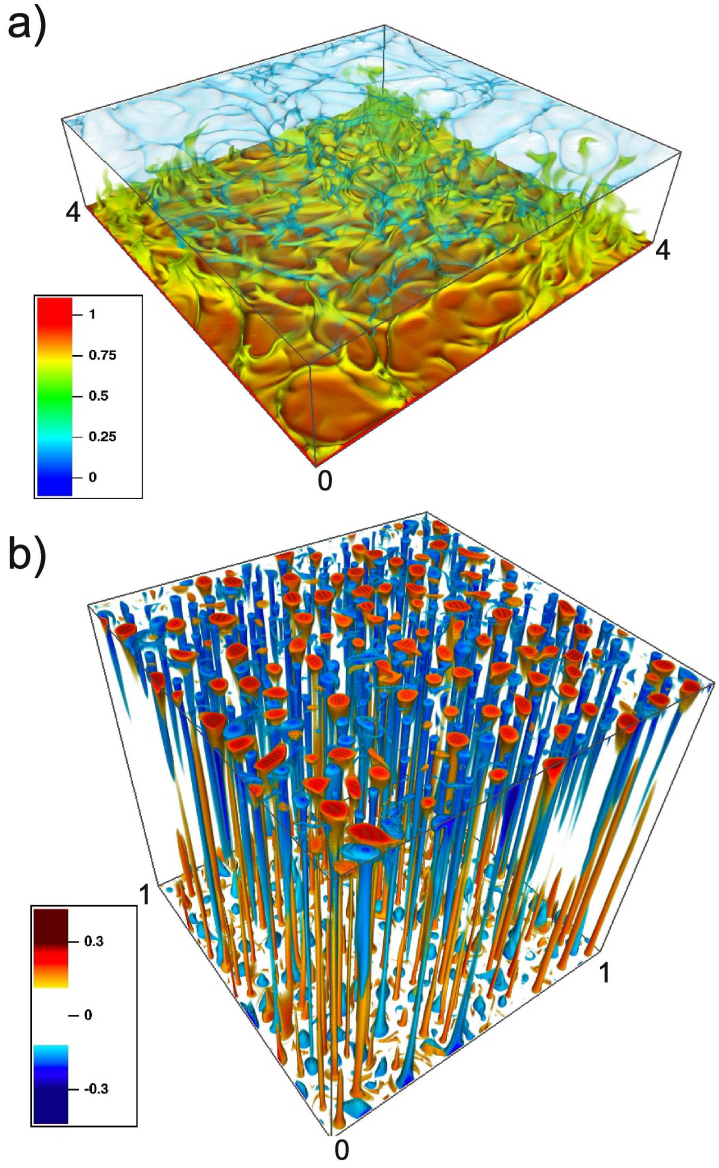


Figure 4. Visualization snapshots from two simulations illustrating characteristic features of turbulent non-rotating and rapidly rotating convection. **a)** Temperature field in a numerical simulation of turbulent non-rotating ($E = \infty$) convection at $Ra = 2.1 \times 10^7$, $Pr = 1$ in a $4 \times 4 \times 1$ computational box. The flow field has a disordered cellular structure, with the turbulent kinetic energy distributed among a broad spectrum of spatial scales. **b)** Illustration of the temperature field in rapidly rotating convection at $Ra = 7 \times 10^9$, $Pr = 100$ and $E = 10^{-6}$ in a $1 \times 1 \times 1$ computational box. This time, the temperature anomaly $T - \langle T \rangle$ is shown, where $\langle ... \rangle$ denotes the horizontal average. The flow field consists of narrow, highly vortical columnar structures with preferred horizontal length scales of order $E^{1/3}$. Note that the simulation domain in this case covers a horizontal area that is sixteen times smaller than that shown in *a)* but nevertheless hosts a multitude of convection cells.

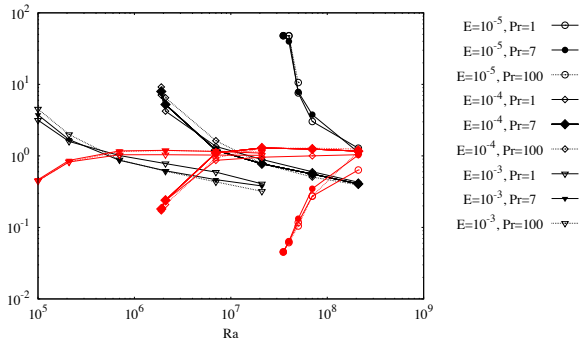


Figure 5. Influence of boundary layers on heat transfer scaling for $10^{-3} \leq E \leq 10^{-5}$. The red symbols show the Nusselt number Nu, normalized by the respective non-rotating scaling law $Nu = 0.16Ra^{2/7}$. The black symbols represent the relative thickness of thermal and Ekman boundary layers, δ_T/δ_E. Note that the heat transfer data crosses unity approximately where the thickness ratio also crosses unity, supporting the hypothesis that the relative boundary layer thickness controls the heat transfer transition.

the global ratio of inertial to Coriolis forces controls the transition from rotation dominated to non-rotating heat transport, we would expect it to occur at $Ro_c = 1$, corresponding to a transitional Rayleigh number $Ra_t \sim E^{-2}$. This differs from the scaling $Ra \sim E^{-7/4}$ derived from our numerical results, thus suggesting that the global force balance argument does not adequately describe the dynamics.

A detailed analysis of our simulation data and a careful comparison with laboratory data seems to suggest another hypothesis based on boundary layer dynamics[10]. In the bulk of the convective domain, molecular diffusion typically plays only a minor role in the convection dynamics. It, however, becomes important in thin regions close to the boundaries, called the boundary layers, where the interior flow adjusts to the applied boundary conditions. In the rapidly rotating case, the important viscous boundary layer is the Ekman layer, with a characteristic thickness $\delta_E \sim E^{1/2}$. The temperature field adjusts to the boundary conditions in a thin thermal boundary layer, whose thickness usually depends on the Nusselt Number, $\delta_T \sim Nu^{-1}$. If we hypothesize that the transition occurs at $\delta_E \approx \delta_T$, we would expect $Nu_t \sim E^{-1/2}$, exactly as found from our data.

To test the above hypothesis, we computed the respective boundary layer thicknesses from the simulation output. We define the Ekman boundary layer thickness as the distance from the boundaries at which the maximum horizontal root-mean-square velocity occurs. The thermal layer depth is defined by using the maximum value of the temperature variance. Figure 5 shows the Nusselt number, normalized by the non-rotating scaling, as a function of Ra. The ratio δ_T/δ_E of the boundary layer thicknesses is also shown against Ra in this plot. The heat transfer transition for each Ekman number occurs where $Nu/Nu_{\text{non-rotating}}$ crosses unity. Figure 5 shows that this happens exactly when δ_T becomes smaller than δ_E, thus supporting our boundary layer argument for a broad range of E and Pr.

5 Conclusions

The scaling laws found in our study allow *a priori* predictions of the dynamical importance of Coriolis forces for a given system. Application to the Earth's core for example suggests that it is controlled by rotation, but much closer to the transition point than previously expected[10]. Future studies, incorporating additional dynamical effects such as fluid compressibility, low Prandtl number, magnetic fields and more complicated geometries are necessary to account for the individual characteristics of specific natural systems.

Acknowledgements

We are grateful to the John von Neuman Institute for Computing for providing computing time. Many results presented in this paper were published[10] in a joint work with Eric King and Jonathan Aurnou (UCLA), who performed laboratory experiments on rotating Rayleigh-Bénard convection. The visualizations shown in Figure 4 were made using VAPOR (`http://www.vapor.ucar.edu/`), a product developed by the National Center for Atmospheric Research.

References

1. H. K. Moffatt, *Magnetic field generation in electrically conducting fluids*, Cambridge University Press, Cambridge, 1978.
2. K. Julien, S. Legg, J. McWilliams, and J. Werne, *Rapidly rotating turbulent Rayleigh-Bénard convection*, J. Fluid Mech., **322**, 243–272, 1996.
3. S. Stellmach and U. Hansen, *Cartesian convection-driven dynamos at low Ekman number*, Phys. Rev. E, **70**, 056312, 2004.
4. R. Kunnen, H. Clercx, and B. Geurts, *Enhanced Vertical Inhomogeneity in Turbulent Rotating Convection*, Phys. Rev. Lett., **101**, 17, 2008.
5. S. Stellmach and U. Hansen, *An efficient spectral method for the simulation of dynamos in Cartesian geometry and its implementation on massively parallel computers*, Geochem. Geophys. Geosyst., 2008.
6. M. Eleftheriou, B. G. Fitch, A. Rayshubskiy, T. J. C. Ward, and R. S. Germain, *Scalable framwork for 3D FFTs on the Blue Gene/L Supercomputer: Implementation and early performance measurements*, IBM J. RES. & Dev., **49**, 457–464, 2005.
7. B. Castaing, G. Gunaratne, F. Heslot, L. Kadanoff, A. Libchaber, S. Thomae, X. Z. Wu, S. Zaleski, and G. Zanetti, *Scaling of hard thermal turbulence in Rayleigh-Benard convection*, J. Fluid Mech., **204**, 1–30, 1989.
8. J. Aurnou, *Planetary core dynamics and convective heat transfer scaling*, Geophys. Astrophys. Fluid Dyn., **101**, 327–345, 2007.
9. U. Christensen and J. Aubert, *Scaling properties of convection-driven dynamos in rotating spherical shells and applications to planetary magnetic fields*, Geophys. J. Int., **116**, 97–114, 2006.
10. E. M. King, S. Stellmach, J. Noir, U. Hansen, and J. M. Aurnou, *Boundary layer control of rotating convection systems*, Nature, **457**, no. 7227, 301–304, 2009.
11. R. Kunnen, H. Clercx, and B. Geurts, *Heat flux intensification by vortical flow localization in rotating convection.*, Phys Rev E, **74**, 056306, 2006.

Adaptive Optimization of GOCE Gravity Field Modeling

Wolf-Dieter Schuh, Jan Martin Brockmann, Boris Kargoll, and Ina Krasbutter

Institute of Geodesy and Geoinformation
Department of Theoretical Geodesy, University of Bonn, 53127 Bonn, Germany
E-mail: {*schuh, brockmann, kargoll, krasbutter*}*@geod.uni-bonn.de*

Long years of highly intensive research allowed for the realization of the satellite mission GOCE (Gravity Field and Steady-state Ocean Circulation Explorer), which in 1999 became the first mission to be adopted within the new Living Planet Programme by the European Space Agency (ESA). The primary goal of this mission is the recording of the static component of the Earth's gravity field with an unprecedented global accuracy and resolution of at least $2\,cm$ for the geoid at a global scale of at least $100\,km$. With these results other geoscientific core goals will be realized. On the one hand, the Earth system with all its interacting geophysical and oceanographic processes may be modeled with much higher reliability by means of a high-precision GOCE gravity field. On the other hand, a high-precision geoid will finally enable geodesists to unify and connect the heterogeneous national height reference systems. The goal of this article is to describe an *in situ* approach to determine a global Earth gravity model and its variance/covariance information on the basis of calibrated measurements from the GOCE mission. As the main characteristics of this procedure, the GOCE data are processed *in situ* via development of the functionals at the actual location and orientation of the gradiometer. This high-dimensional data fitting problem (spherical harmonic analysis) with irregularly distributed data can only be managed on a massive parallel computer system using tailored algorithms. The PCGMA (preconditioned conjugate gradient multiple adjustment) algorithm consists of an iterative variance component estimation to handle heterogeneous data groups, a conjugate gradient solver supported by an efficient preconditioning procedure to find the optimal solution of the huge, dense, overdetermined linear system, and a tailored pre-whitening procedure based on digital filters to decorrelate the densely sampled measurements.

1 Introduction

1.1 Overview over the Satellite Mission GOCE

The GOCE satellite (see Figure 1) was launched on 17. March 2009 after decades of scientific research[11] and technological development as the first core mission of ESA's new Living Planet Programme[6]. The main goal of this mission is the determination of the static part of the gravitational field of the Earth, which is inhomogeneous due to mass anomalies in the Earth's interior, with the very high accuracy of at least $2\,cm$ for geoid heights and $1\,mgal$ ($\hat{=}10^{-5}\frac{m}{s^2}$) for gravity anomalies at a global resolution of $100\,km$[6]. The GOCE spacecraft is orbiting planet Earth at a very low altitude of about $255\,km$ and with an inclination of 96.5° (Figure 2). The orbit is sun-synchronous to minimize thermoelastic disturbances at shadow crossings and to guarantee sufficient power supply throughout the at least one year long mission.

The main idea for reaching the unprecedented level of accuracy in global gravity field determination is the fusion of different sensors. The GOCE mission combines in particular two distinct measurement concepts, which are sensitive to the low-frequency part of the gravity field, on the one hand, and the high-frequency part, on the other hand. The low-frequency part is captured via the tracking of the low-orbit satellite using the Global Positioning System (GPS) by precisely measuring the satellite's orbit disturbances[15], known

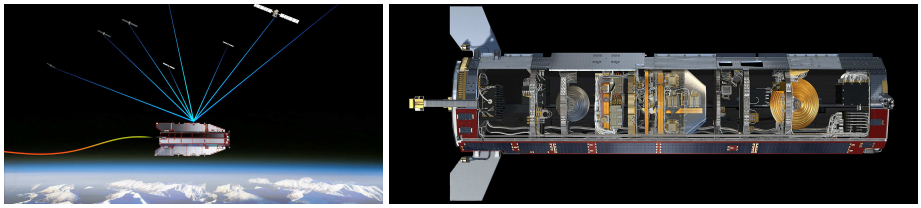


Figure 1. Two illustrations of the GOCE Satellite. Left: GPS tracking of GOCE. Right: A look at the various instruments inside the satellite, the gradiometer as the main instrument can can be seen in the centre. ©ESA

Figure 2. Ground track of the GOCE satellite after the first day, first week and first month in orbit.

as Satellite-to-Satellite Tracking (SST), see Figure 1 left). The high-frequency part of the gravity field signal is determined by means of the innovative measurement principle of Satellite Gravity Gradiometry[11] (SGG). The main instrument is the gradiometer (see Figure 1 right), which measures second derivatives $\mathbf{V}_{ij}$ of the Earth's gravitational potential $\mathbf{V}$ *in situ*. These measurements are derived from accelerations (representing the first derivatives of the gravitational potential), which are measured by six accelerometers, placed some distance ($25\,cm$) apart the centre (of mass) of the satellite, aligned along three orthogonal axes (along-track, across-track and Earth-pointing). Each accelerometer has two extremely sensitive axes. The accelerometer measurements along each axial direction are differentiated, resulting in the desired along-track, across-track and earth pointing component of the gravity tensor $\mathbf{V}_{ij}$. In addition, due to alignment of the sensitive axes of the accelerometers, also the mixed along-track and Earth-pointing component of the gravity tensor can be determined with high accuracy. It is essential to avoid non-gravitational nuisance accelerations such as atmospheric pressure, the satellite is maintained in a compensated drag-free fall via a feedback loop between the accelerometers and the ion thrusters[6].

Now, in order to compute a gravity field solution from SST and SGG measurements, these two independent data groups must be merged in a common statistical model. We will, however, not focus on the SST data in this report, as their quantity and numerical characteristics are not a serious computational challenge in contrast to the SGG data; we obtained the SST data in an appropriately preprocessed model by our project partner. On the other hand, the modeling of SGG data is highly challenging due to various reasons: (1) The operational mission period of 12 months, where we use only the three most accurate gradient components ($\mathbf{V}_{xx}$, $\mathbf{V}_{yy}$ and $\mathbf{V}_{zz}$), sampled at 1 Hz, will result in a huge number of as many as 100 million observations. (2) The SGG data with respect to each of these three components are highly auto-correlated[13], the accommodation of which would be based on a full variance/covariance matrix with required memory of approximately $8\,000$ Terabyte.

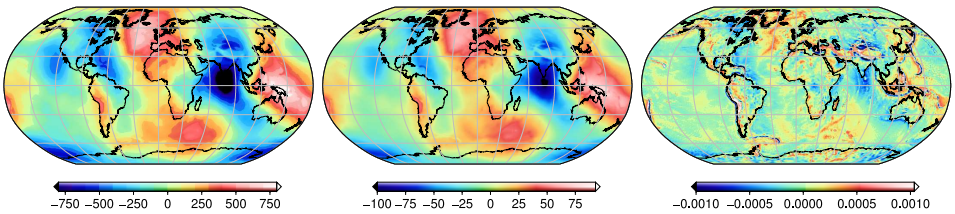


Figure 3. Functionals of the gravity field with maximum achievable resolution of GOCE (sh d/o 270). Left: potential (m^2/s^2). Centre: geoid heights (m). Left: gravity anomalies (m/s^2).

A further challenge arises from the fact that the SST/SGG data combination problem is ill-posed due to data gaps over the polar regions[14] (cf. Figure 2) and the attenuation of the gravity field signal at the satellites altitude. This problem necessitates the regularization of the statistical model by introducing stochastic prior information, which may be viewed as a third independent data group (REG). Thus, the determination of the gravity field requires the solution of a highly over-determined, dense equation system with a huge dimension. For this task we developed a tailored iterative solution strategy based on the method of conjugate gradients (CG)[12].

1.2 Modeling of the Earth's Gravity Field

Before we explain the solution method for determining the gravity field from GOCE observations, we will outline the mathematical parameterization of global gravity field models used in our approach. The basic idea is to describe the Earth's gravitational potential as a spherical harmonic expansion up to a certain maximal degree and order (d/o) $l_{\max}$. Then such a potential may be written, for some evaluation point (r, θ, λ) in an Earth fixed and centred coordinate system, as[8]

$$V(r,\theta,\lambda) = \frac{GM}{a}\sum_{l=0}^{l_{\max}}\left(\frac{a}{r}\right)^{l+1}\sum_{m=0}^{l}\left(c_{lm}\cos(m\lambda) + s_{lm}\sin(m\lambda)\right)P_{lm}(cos\theta)\,, \quad (1)$$

where l and m denote the spherical harmonic degree and order, c_{lm} and s_{lm} the unknown coefficients of the spherical harmonic expansion, a the equatorial radius of the Earth reference ellipsoid, $P_{lm}(\cdot)$ the fully normalized associated Legendre functions, and GM the geocentric gravitational constant. Thus, this model comprises $m = (l_{\max}+1)^2$ unknown coefficients, which may in turn be used to express related geometrical and physical functionals of the Earth's gravity field such as geoid heights (i.e. the metric distances between the geoid as the zero equipotential surface and an approximating ellipsoid) and gravity anomalies (i.e. the local variation of the gravity acceleration with respect to an idealized reference ellipsoid)[8] (see Figure 3). Using this spherical harmonics approach, all coefficients belonging to a particular degree l refer to a unique spatial resolution or frequency.

To determine the unknown gravity field parameters c_{lm} and s_{lm} from the GOCE observations (SST, SGG and REG) they are estimated by applying a linear Gauss-Markov model[9]. In this model, each observation ℓ_i, affected by an observation error v_i, is written as a (linear) function $\mathbf{A}_{i,:}\mathbf{x}$ of the gravity field parameters (denoted by $\mathbf{x}$). This results

in observation equations $\ell_i + v_i = \mathbf{A}_{i,:}\mathbf{x}$, where $\mathbf{A}_{i,:}$ denotes the ith row of the design matrix $\mathbf{A}$. The stochastic behaviour of the measurement errors is usually modeled by a variance/covariance matrix $\mathbf{\Sigma}$, which contains the variances of the observations as well as the covariances between them. Minimizing the weighted squared sum of residuals $\mathbf{v}^T\mathbf{\Sigma}^{-1}\mathbf{v}$ leads to the linear normal equation system $\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{\Sigma}^{-1}\boldsymbol{\ell}$ or $\mathbf{N}\mathbf{x} = \mathbf{n}$. Here $\boldsymbol{\ell}$ comprises the three independently distributed observation groups SST, SGG and REG, thus their corresponding normal equations may be computed separately and added up subsequently. This basic property establishes the mathematical foundation for the data combination in our approach. The weighted least squares solution is then obtained as the solution of the normal equations $\mathbf{N}\mathbf{x} = \mathbf{n}$.

The SST data group is made available to us by our project partners directly in the form of preprocessed normal equations $\mathbf{N}_{\text{sst}}$ and $\mathbf{n}_{\text{sst}}$. As these SST data have a relatively low resolution of d/o 90, they are related to only a small subset ($\approx 8\,000$) of the total parameter vector. Thus, they are easily handled by our solution algorithm. The processing of the regularization data (REG) does not pose a great computational challenge either as the underlying functional model is quite simple. This can be explained by the fact that the parameters are treated as directly observed unknowns with values equal to zero, the corresponding design matrix is given by the unity matrix. Furthermore, these pseudo-observations are uncorrelated and have variances according to Kaula's rule of thumb, represented by the diagonal weight matrix $\mathbf{P}_{\text{reg}}$.

The processing of the SGG data is, however, more intricate. To begin with, the functional model in Eq. (1) is written in the Earth-fixed frame, which is transformed into the gradiometer reference frame via the application of a rotation taking into account the Earth's rotation, the satellite's orientation in space and the gradiometer's orientation within the satellite. This rotated functional model is then differentiated twice, yielding the desired SGG observation equations $\ell_{i,\text{sgg}} + v_{i,\text{sgg}} = \mathbf{A}_{1,:,\text{sgg}}\mathbf{x}$. From simulations we know that these SGG data will be highly autocorrelated. Due to the huge number of observations, these correlations cannot be stored within a variance/covariance matrix. The only feasible way to take these correlations into account is to remove them from the SGG data by means of pre-whitening filters. For this purpose, we use a sequence of high-order autoregressive moving-average (ARMA) filters. The thus decorrelated SGG observation equations read

$$\bar{\boldsymbol{\ell}}_{\text{sgg}} + \bar{\mathbf{v}}_{\text{sgg}} = \bar{\mathbf{A}}_{\text{sgg}}\mathbf{x}. \tag{2}$$

Now, the combined normal equations are established by computing the sum of the three individual normal equations, each weighted by unknown factors ω_i:

$$\left(\omega_{\text{sgg}}\bar{\mathbf{A}}_{\text{sgg}}^T\bar{\mathbf{A}}_{\text{sgg}} + \omega_{\text{sst}}\mathbf{N}_{\text{sst}} + \omega_{\text{reg}}\mathbf{P}_{\text{reg}}\right)\mathbf{x} = \omega_{\text{sgg}}\bar{\mathbf{A}}_{\text{sgg}}^T\bar{\boldsymbol{\ell}}_{\text{sgg}} + \omega_{\text{sst}}\mathbf{n}_{\text{sst}} + \omega_{\text{reg}}\mathbf{n}_{\text{reg}}. \tag{3}$$

2 PCGMA Algorithm

To solve the equation system (3) for the unknown spherical harmonic coefficients, we make use of a tailored version of the conjugate gradient (CG) algorithm denoted as PCGMA[12] (preconditioned conjugate gradient multiple adjustment). This algorithm extends the standard CG algorithm with (a) a data-adaptive preconditioning step, (b) the combination of the various (independent) GOCE data types (given as either observation equations or normal equations respectively) and (c) determination of the weights $\omega_i = \frac{1}{\sigma_0^2}$ via variance component estimation (VCE)[10] using Monte Carlo methods[1,5]. To give some insight

into this computational scheme[4], recall that the CG algorithm is based on the residuals $\mathbf{r}^{(\nu)} = \mathbf{N}\mathbf{x}^{(\nu)} - \mathbf{n}$ of a symmetric equation system $\mathbf{N}\mathbf{x}^{(\nu)} = \mathbf{n}$ which are used as a search direction for finding the minimum L_2-norm solution.

Substituting the combined GOCE normal equation system and some initial parameter values $\mathbf{x}^{(0)}$ leads to initial residuals

$$\begin{aligned}
\mathbf{r}^{(0)} &= \mathbf{N}\mathbf{x}^{(0)} - \mathbf{n} \\
&= \left(\omega_{\text{sgg}}\bar{\mathbf{A}}_{\text{sgg}}^T\bar{\mathbf{A}}_{\text{sgg}} + \omega_{\text{sst}}\mathbf{N}_{\text{sst}} + \omega_{\text{reg}}\mathbf{P}_{\text{reg}}\right)\mathbf{x}^{(0)} - \left(\omega_{\text{sgg}}\bar{\mathbf{A}}_{\text{sgg}}^T\bar{\boldsymbol{\ell}}_{\text{sgg}} + \omega_{\text{sst}}\mathbf{n}_{\text{sst}} + \omega_{\text{reg}}\mathbf{n}_{\text{reg}}\right) \\
&= \omega_{\text{sgg}}\left(\bar{\mathbf{A}}_{\text{sgg}}^T\left(\bar{\mathbf{A}}_{\text{sgg}}\mathbf{x}^{(0)} - \bar{\boldsymbol{\ell}}_{\text{sgg}}\right)\right) + \omega_{\text{sst}}\left(\mathbf{N}_{\text{sst}}\mathbf{x}^{(0)} - \mathbf{n}_{\text{sst}}\right) + \omega_{\text{reg}}\left(\mathbf{P}_{\text{reg}}\mathbf{x}^{(0)} - \mathbf{n}_{\text{reg}}\right) \\
&= \omega_{\text{sgg}}\left(\bar{\mathbf{A}}_{\text{sgg}}^T\mathbf{v}_{\text{sgg}}^{(0)}\right) + \omega_{\text{sst}}\left(\mathbf{N}_{\text{sst}}\mathbf{x}^{(0)} - \mathbf{n}_{\text{sst}}\right) + \omega_{\text{reg}}\left(\mathbf{P}_{\text{reg}}\mathbf{x}^{(0)} - \mathbf{n}_{\text{reg}}\right) \\
&= \omega_{\text{sgg}}\mathbf{r}_{\text{sgg}}^{(0)} + \omega_{\text{sst}}\mathbf{r}_{\text{sst}}^{(0)} + \omega_{\text{reg}}\mathbf{r}_{\text{reg}}^{(0)}.
\end{aligned}$$

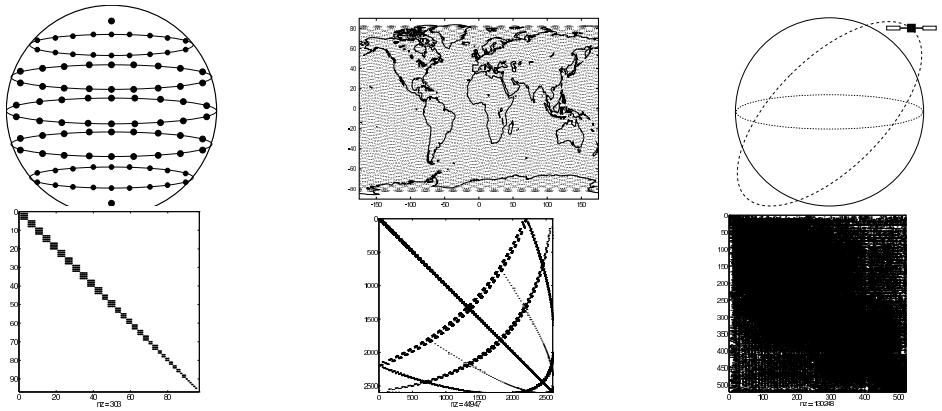
This iterative CG approach has the great benefit that it avoids computation of the joint normal equation matrix $\mathbf{N}$, and especially of the high-dimensional product $\mathbf{N}_{\text{sgg}} = \bar{\mathbf{A}}_{\text{sgg}}^T\bar{\mathbf{A}}_{\text{sgg}}$ in view of the facts that (a) $\bar{\mathbf{A}}_{\text{sgg}}$ could be of dimension $70\,000 \times 100\,000\,000$ or larger and (b) this matrix-matrix product is replaced by two far less expensive matrix-vector multiplications $\mathbf{v}_{\text{sgg}}^{(0)} = \bar{\mathbf{A}}_{\text{sgg}}\mathbf{x}^{(0)} - \bar{\boldsymbol{\ell}}_{\text{sgg}}$ and $\mathbf{r}_{\text{sgg}}^{(0)} = \bar{\mathbf{A}}_{\text{sgg}}^T\mathbf{v}_{\text{sgg}}^{(0)}$. Based on these initial residuals Eq. (3) is solved using the standard preconditioned CG algorithm[7] using $\mathbf{N}_{\oplus}$ as the data-adaptive preconditioning matrix (cf. Sect. 3)[2,3].

PCGMA has been implemented on a computer cluster to distribute the computational workload. As already mentioned above, it is crucial to parallelize the processing of the huge number of SGG data. This is done by distributed row-wise assembling of the design matrix $\mathbf{A}_{\text{sgg}}$. Then the individual parts of the design matrix and the observation vector are transformed by means of digital ARMA filters in order to remove the autocorrelations present in each SGG component. Subsequently, the SGG residuals $\mathbf{r}_{\text{sgg}}$ are computed in parallel, whereas the remaining CG operations are computed in serial mode on the master node, after an append-receive operation of the distributed residuals. The residuals with respect to the REG and SST groups are computed in serial mode, as $\mathbf{P}_{\text{reg}}$ is only a diagonal matrix and $\mathbf{N}_{\text{sst}}$ contains only a relatively small block of non-zero elements (of size less than $10\,000 \times 10\,000$). The serial CG operations have been implemented based on fast level 3 BLAS routines.

3 Computational Aspects

The two parts of the whole PCGMA algorithm of main computational interest are (a) the design of the preconditioner and (b) the assembling and filtering of the SGG observation equations for the purpose of their decorrelation. In this contribution we will concentrate on the choice of a tailored data-adaptive preconditioner for the GOCE observations. The preconditioning matrix serves as an approximation of the joint normal equation matrix for the purpose of accelerating the convergence of the CG algorithm and of stabilizing the numerical procedure.

To obtain the dominant structure of the combined normal equations $\mathbf{N}$ (3) as a sparse, and thus easy-to-handle, approximation, one exploits the properties of the spherical harmonic base functions. These functions satisfy certain orthogonality relations[8]. Conse-



gridded data | regularly distributed data | irregularly distributed data

Figure 4. Resulting structure of normal equations for different scenarios of data coverage.

quently, if the data are globally distributed on the sphere with equidistance along the parallels, only coefficients of the same order are correlated and lead to non-zero elements in the normal equations. If such parallels are even symmetric with respect to the equator, then also the coefficients with odd degrees within each order are independent from the corresponding even coefficients. If the coefficients are then arranged or numbered order-wise within the parameter vector $\mathbf{x}$, the normal equation matrix takes a block-diagonal structure as shown in Figure 4 (left). If some sufficiently dense and global data distribution on symmetric parallels is disturbed by slightly shifting the locations, then the resulting normal equation matrix remains nearly block-diagonal (Figure 4 centre). This matrix loses its block-diagonal structure and becomes increasingly fully populated the more the data coverage differs from a regular distribution (Figure 4 right).

In GOCE data processing, the task is to choose a numbering scheme concerning the parameters such that it allows for a sparse approximation for the combined normal equation matrix. Analyzing the structures of the individual normal equations with respect to the three observation groups, we find three key properties: (a) The SGG data are regularly distributed over the sphere except for the polar regions. With an order-wise numbering, the resulting SGG normal equations are then nearly block-diagonal; therefore, a sparse preconditioning matrix should preserve this property. (b) The REG normal equation matrix is a diagonal matrix, which is naturally sparse and may thus be simply added to the preconditioning matrix. (c) The SST normal equation matrix should be added to the preconditioner as a whole and without using an approximation. Its numbering must be chosen such that the entire populated block remains as coherent as possible. These three desired properties hold when the free kite numbering scheme[2,3] is applied. To understand this kite scheme it is useful to represent the spherical harmonic coefficients c_{lm} and s_{lm} as a triangle in the following way (Figure 5, left): The ordinate is defined by the degree l, which increases from top to bottom, and the abscissa by the order m, with the cosine coefficients c_{lm} to the left and the sine coefficients s_{lm} to the right.

We may now classify the parameters according to the three zones FULL (modeling the full correlations in the sparse preconditioner), INDEPENDENT (parameters where only

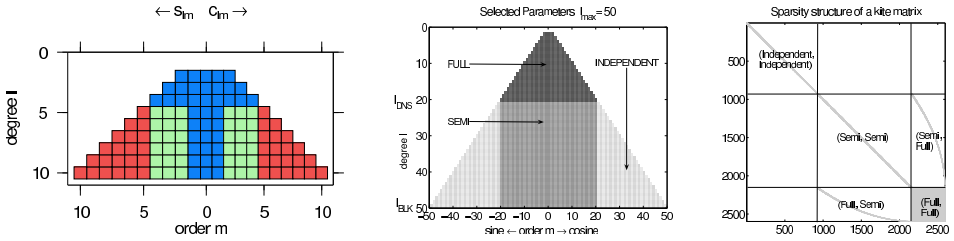


Figure 5. Coefficient triangle with the correlation zones for the free kite numbering scheme and the resulting structure of the sparse preconditioner.

the order-wise correlations are modeled) and SEMI (all the parameters of order m for which some degrees are present in the FULL zone). Then, all of the parameters first within the INDEPENDENT, then within the SEMI, and finally within the FULL zone, are numbered order-wise. From this numbering scheme follows directly the population scheme of the desired preconditioning matrix (Figure 5 right). Such a matrix has the great advantage, that no fill-in values arise when it is reduced via Cholesky's algorithm[2], used in the preconditioning step within PCGMA for solving $\mathbf{N}_{\oplus}\boldsymbol{\rho} = \mathbf{r}$. As far as the SGG data are concerned, all the elements $\mathbf{N}_{\oplus,i,j} = \bar{\mathbf{A}}^{T}_{\text{sgg},i,:}\bar{\mathbf{A}}_{\text{sgg},:,j}$ are computed rigorously according to determined populations scheme. The normal equation matrices concerning the SST and REG groups are then simply added to the matrix $\mathbf{N}_{\oplus}$ using weights ω_i.

The free kite numbering scheme has some interesting additional properties. The more parameters are assigned to the FULL zone, the faster the PCGMA algorithm converges but the larger the size of the preconditioning matrix becomes. The size of the preconditioner used in a typical GOCE simulation is about $1.0 - 2.0GB$, which constitutes a good compromise between a fast convergence rate and memory requirements. The fact that the spectral radius is ≈ 0.1 demonstrates the efficiency of this preconditioner. Thus the PCGMA algorithm converges within $20 - 30$ iterations, allowing for a very efficient solution of the dense normal equation system with $70\,000$ unknowns. In addition, the SGG part of the preconditioner may be computed in parallel, with each node processing parts of the SGG data (in which case the preconditioning matrix must be stored in memory of each of these nodes). The individual parts are collected via a global reduce operation on the master node.

4 Summary and Outlook

The spherical harmonic analysis is a standard procedure to determine the gravitational potential from gravity data. In the context of the GOCE application, the processing of a high-resolution gravity field from the huge number of highly correlated, heterogeneous GOCE measurements, requires both a huge computational effort and large amount of memory. For this purpose, a tailored solution strategy was developed, which can handle heterogeneous data, avoid the expensive computation of the normal equations, allow for a sequential processing of highly correlated measurements, and which uses stochastic trace estimators to overcome the computational burden of VCE. After two decades of research PCGMA has finally been implemented in fully operational mode and used in numerous simulations on

the JUMP cluster (128–256 processors) at Forschungszentrum Jülich. Currently, the research community awaits the GOCE real data, expected to be available at the beginning of October 2009. The results derived from the GOCE real data analysis are expected to revolutionize the understanding of dynamic processes within the Earth system.

Acknowledgements

Parts of this work were financially supported by the BMBF Geotechnologien program GOCE-GRAND II and the ESA contract No. 18308/04/NL/MM. The computations were performed on the JUMP supercomputer in Jülich. The computing time was granted by the John von Neumann Institute for Computing (project 1827).

References

1. H. Alkhatib, *On Monte Carlo methods with applications to the current satellite gravity missions* (PhD thesis, University of Bonn, 2007).
2. C. Boxhammer, *Effiziente numerische Verfahren zur sphärischen harmonischen Analyse von Satellitendaten* (PhD thesis, University of Bonn, 2006).
3. C. Boxhammer and W.-D. Schuh, *GOCE gravity field modeling: computational aspects – free kite numbering scheme* in J. Flury et al. (eds.), *Observation of the Earth System from Space*, Springer, 207–224.
4. J. M. Brockmann, B. Kargoll, I. Krasbutter, W.-D. Schuh and M. Wermuth, *GOCE data analysis: from calibrated measurements to the global Earth gravity field*, (Springer, in press, 2009).
5. J. M. Brockmann and W.-D. Schuh, *Fast variance component estimation in GOCE data processing* in *Gravity, Geoid and Earth Observation*, IAG Symposia (in press, Springer, 2009).
6. ESA, *The four candidate Earth explorer core missions – gravity field and steady-state ocean circulation mission* (ESA Report SP-1233(1), Granada, 1999).
7. M. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems* Journal of Research of the National Bureau of Standards **49**, 2379, 1952.
8. B. Hofmann-Wellenhof and H. Moritz, *Physical geodesy* (Springer, Vienna, 2005).
9. K.-R. Koch, *Parameter estimation and hypothesis testing in linear models* (2nd edition, Springer Berlin Heidelberg, 1999).
10. K.-R. Koch and J. Kusche, *Regularization of geopotential determination from satellite data by variance components* Journal of Geodesy **76**, 641–652, Springer, 2001.
11. R. Rummel, *Satellitengradiometrie* ZfV **6**, 242–257, 1985.
12. W.-D. Schuh, *Tailored numerical solution strategies for the global determination of the Earth's gravity field* Mitteilungen der Universität Graz **81**, TU Graz, 1996.
13. C. Siemes, *Digital filtering algorithms for decorrelation within large least square problems* (PhD thesis, University of Bonn, 2008).
14. N. Sneeuw and M. van Gelderen, *The polar gap* Lecture Notes in Earth Science **65**, 559–568, Springer, 1997.
15. M. Wermuth, *Gravity field analysis from the satellite missions CHAMP and GOCE* (PhD thesis, TU Munich, 2008).

Computer Science and Numerical Mathematics

Michael Griebel

Institute for Numerical Simulation, University of Bonn, Wegelerstr. 6, 53115 Bonn, Germany
E-mail: griebel@ins.uni.bonn.de

Computer science aspects and numerical simulation on modern large scale parallel hardware platforms are discussed in the following section. Here, we put the emphasis on the numerical simulation of certain aspects of quantum computing, i.e. thresholds for the errors in quantum error correction, on the performance-analysis tool Scalasca and its properties to diagnose problems in large-scale parallel applications on e.g. up to 294.912 cores of the full IBM Blue Gene/P system JUGENE, and on an heuristic optimization algorithm for a variant of a multidisperse packing problem, namely the benchmark task of packing circles into a circle.

In their contribution *Threshold Determination for Reliable Quantum Error Correction*, B. Trieu, R. Kennedy and M. Richter deal with a fundamental question in quantum computing: Quantum operations suffer in praxis from operational imprecisions and real quantum computation devices are subject to decoherence. Thus, quantum error correction (QEC) schemes must be employed to overcome these problems. But as the correction circuits themselves are prone to errors, it must be ensured that their error probability is below a specific threshold in order not to spoil the correction process. For this reason, the knowledge of this threshold value is absolutely essential for any reliable QEC. However, it can not be derived theoretically but must be determined via numerical simulations. The authors evaluated the threshold for reliable quantum error correction both in the case of decoherence and in the case of operational errors by large-scale computer simulations. They found that the threshold is $5.2 \cdot 10^{-6}$ for decoherence errors and $4.3 \cdot 10^{-2}$ for operational errors. At present, such thresholds can clearly not be reached by existing quantum computation devices. Nevertheless, since there is in principle no physical limitation that prohibits to achive such thresholds, these results show the way for the developers of future quantum computation devices. Furthermore, on the JUGENE system (Jülich Blue Gene), up to 40 qubits could be simulated by the authors using 49.152 processors and 24 TB of memory. This is the world-wide largest system size that has been directly simulated so far.

In their contribution *Performance Tuning in the Petascale Era*, F. Wolf, D. Böhme, M. Geimer, M. Hermanns, B. Mohr, Z. Szebenyi and B. Wylie describe the research activities of their Helmholtz University Young Investigators Group *Performance Analysis of Parallel Programs* at the Jülich Supercomputing Centre (JSC). Since modern supercomputing applications must exploit a high degree of parallelism to satisfy the growing demand for computing power, the writing of code which runs efficiently on large numbers of processors remains a significant challenge. Furthermore, the rising number of cores in modern and future computer architectures imposes scalability demands not only on applications but also on the software tools needed for their development. To this end, the group developed, in cooperation with the JSC Division *Application Support*, the performance-analysis tool Scalasca. It has been designed for large-scale systems and allows the automatic identifi-

cation of harmful wait states in applications running on tens of thousands of processors. The authors give an overview of Scalasca and highlight the research accomplishments of their group during the past two years with a focus on the analysis of wait states and of time-dependent behaviour. These two examples address the scalability of Scalasca with respect to both the number of processes and the length of execution times.

In his contribution *Optimization of Packing Problems with Heuristic Algorithms Suited for Parallel Enablement*, J. Schneider reports on his recent approach for packing problems which combines global and local optimization methods. It involves the well-known simulated annealing technique and its deterministic threshold accepting variant where an optimum is obtained by an artificial cooling process from an initial configuration created randomly at high temperature. There, a new configuration is successively accepted or rejected with a certain Metropolis probability. The resulting solution is then further improved by a local optimization method. The overall method is parallelized using OpenMP and MPI. The author considers an academic variant of a multidisperse packing problem, i.e. the benchmark problem of packing circles into a circle. There, the task is to find the densest closed packing of N disks with radii $r_i = i, i = 1, \ldots, N$ in a circle with radius R such that R is minimized. In an international contest on that problem, the proposed approach showed excellent results.

Threshold Determination for Reliable Quantum Error Correction

Binh Trieu[1,2]**, Ricardo Kennedy**[1]**, and Marcus Richter**[1]

[1] Institute for Advanced Simulation, Jülich Supercomputing Centre, Research Centre Jülich
D-52425 Jülich, Germany
E-mail: {b.trieu, m.richter}@fz-juelich.de, ricardo_kennedy@web.de

[2] German Research School for Simulation Sciences
D-52425 Jülich, Germany
E-mail: b.trieu@grs-sim.de

We study the threshold for reliable quantum error correction both in the case of decoherence and operational errors. By large-scale computer simulations, we find that the threshold is $5.2 \cdot 10^{-6}$ for decoherence errors and $4.3 \cdot 10^{-2}$ for operational errors.

1 Introduction

During the last two decades it has been demonstrated that quantum computers can solve certain types of problems much more efficiently than classical computers under the assumption that all quantum operations are faultless and the qubits are protected against uncontrolled interaction with their environment. In reality, quantum operations suffer from operational imprecisions and quantum computation devices are subject to decoherence. Therefore, the question arises if quantum error correction (QEC) schemes[1,2] can be employed to overcome these inherent problems. Since the correction circuits themselves are prone to errors one has to ensure that the error probability is sufficiently small in order not to spoil the correction process. But what does sufficiently small mean? One could guess that there might be a threshold for the error probability beyond which QEC works well. The *existence* of such a threshold has mathematically been proven[a]. If one operates below threshold, fault tolerant error correction works well and everything is under control. Above threshold more and more errors are introduced by the error correction scheme itself – in this case QEC is no help at all. For this reason, the knowledge of the exact numerical threshold value is absolutely essential for reliable QEC. Unfortunately, a precise theoretical prediction as well as an experimental determination of the numerical value for the threshold are still far out of reach. We used large-scale computer simulations to determine a precise prediction for the threshold value in case of both operational errors and decoherence for a realistic number of qubits.

2 The Error Model

Based on a simulation package, that simulates an ideal error-free quantum computer at the gate level[4], we have included an error model that considers decoherence as well as

[a] The combination of QEC with fault-tolerant state recovery, fault-tolerant encoding of quantum logic operations and the principle of concatenation of QEC codes culminates in the accuracy threshold theorem[3]. It states that arbitrarily long quantum calculations are possible if the single qubit error rate is below a certain threshold.

imperfect quantum operations and stays within the gate level simulation framework. The new *JUMPIQCS* package[b] incorporates various QEC schemes, i.e. the 5-qubit code, where one logical qubit is encoded into five physical ones[5], Steane's 7-qubit code[6] and Shor's 9-qubit code[1]. Details are given in a previous publication[7]. In this article we focus on a brief summary of the error model.

2.1 Decoherence

The loss of quantum coherence, i.e. the transition from quantum states to classical states, is called decoherence. It occurs whenever a system interacts with additional external degrees of freedom, and this is the main obstacle in the construction of large-scale quantum computers. Therefore, realistic simulations that involve decoherence are essential for the verification and optimization of real quantum computation devices.

We implemented the so-called depolarizing channel[9] as a suitable error model for the simulation of decoherence. In this model general errors are described as linear combinations of basic errors, given by the Pauli matrices σ_x, σ_y and σ_z. A σ_x-error, $\sigma_x = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, can be identified as a bit flip, a σ_z-error, $\sigma_z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, as a phase flip and $-i\sigma_y = \sigma_x\sigma_z$ can be characterized as a combined bit and phase flip. The depolarizing channel assumes an equal probability of $p/3$ for the occurrence of each error and an error-free evolution with probability $1 - p$. The occurring errors are locally and sequentially independent. In order to implement the depolarizing channel error model we introduce a timescale defined by the single qubit operation time δt and let the decoherence operation affect each qubit after each timestep δt.

Since decoherence results in mixed states, and our simulator deals with pure states only, we have to sample many runs using statistically independent random number sequences. The random numbers are drawn from a uniform distribution using the GNU Scientific Library with the Mersenne Twister MT19937 algorithm. The result is an ensemble of pure states weighted by their corresponding probabilities of occurrence. The use of stochastically independent error locations is justified by the observation that the behaviour of a quantum network can be represented as a mixed sum of networks with linear error operators placed at each error location. This approach is also known as error expansion[8].

2.2 Operational Imprecisions

Every experiment in the lab will suffer from operational imprecisions, as quantum superpositions deal with continuous variables and one cannot control quantum operations with infinite accuracy and precision. We implemented a model for operational imprecisions that includes unitary over-rotations[c] to resemble the errors occurring in the lab. It is not specific to a certain realization of a quantum computation device, but its parameters can be adapted to a range of experimental setups.

For our simulations we chose a model that considers plane rotations around the y-axis and the z-axis only, i.e. each single-qubit quantum operation can be constructed from plane

[b] *Juelich Massively Parallel Ideal Quantum Computer Simulator*
[c] We use the term over-rotation in the sense of imperfect rotation for both over- and under-rotations.

rotations in the x-z-plane,

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \tag{1}$$

and phase shifts

$$P(\phi) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix}. \tag{2}$$

This is a restriction of generality, but it is still sufficient to generate every possible rotation on the Bloch sphere. The Hadamard operation for example can be decomposed into

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = R\left(\frac{\pi}{4}\right) P(\pi). \tag{3}$$

This decomposition allows to introduce errors ϵ in the rotation angles:

$$\theta' = \theta + \epsilon_\theta, \tag{4}$$

$$\phi' = \phi + \epsilon_\phi. \tag{5}$$

These angle and phase errors ϵ are generated from a Gaussian distribution ρ with mean μ and standard deviation σ:

$$\rho(\epsilon) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\epsilon-\mu)^2/(2\sigma^2)}. \tag{6}$$

The Gaussian distribution is generated by the GNU Scientific Library, which uses the Box-Muller algorithm, with the previously mentioned random number generator as an underlying source of randomness. This gives uncorrelated stochastic errors.

2.3 Stochastic Error Model

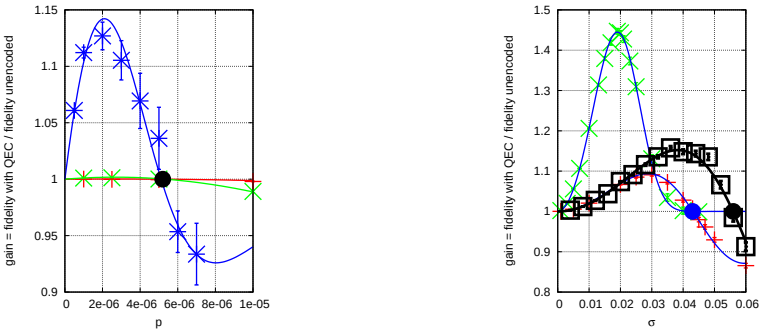
Our simulations including the noise model rely on statistical distributions of single gate errors, so due to the statistical description our readout is not a single shot readout, but averaged over many experiments. All calculations were made by starting with a well-defined pure state vector. The behaviour to noise is then determined by running the algorithm many times with a sample of statistically independent random error events and evaluating the averaged final state. Thus, the fidelity we are measuring is given by

$$F = \frac{1}{m} \sum_{i=1}^{m} |\langle \psi_C | \psi_i \rangle|^2, \tag{7}$$

where m is the number of experiments, each starting with a different initial seed for the random number generator, $|\psi_C\rangle$ is the correct output state vector in the error-free case and $|\psi_i\rangle$ are the single output state vectors of each statistical run.

3 Threshold for Fault-Tolerant Quantum Error Correction

JUMPIQCS has been used to assess the power of QEC schemes, that actively correct errors. We focus on the determination of the threshold value for fault-tolerant QEC using Steane's 7-qubit code[6].



(a) Correction of decoherence errors.

(b) Correction of operational errors.

Figure 1. (a) The ratio between the fidelities with and without QEC is plotted against the single-qubit error rate p. Each point corresponds to a numerical simulation run of the H^{2k}-experiment with $k = 100$ (red), $k = 1000$ (green), and $k = 10000$ (blue). The threshold for fault-tolerant quantum computation is $(5.2 \pm 0.2) \cdot 10^{-6}$ (black bullet). Apparently, the benefit from QEC grows with increasing algorithm length. (b) *Blue curves: single-qubit operations.* Gain by using QEC to correct for operational imprecisions for a sequence of 100 (red) and a sequence of 1000 (green) H^2-iterations depending on the error parameter σ. *Black curve: two-qubit operations.* Instead of Hadamard gates a sequence of 100 $CNOT^2$ operations has been used. The threshold value $(5.5 \pm 0.2) \cdot 10^{-2}$ (black bullet) is of the same order of magnitude as in the single-qubit case (blue bullet).

3.1 H^{2k}- and $CNOT^{2k}$-Algorithm

The algorithms that we are examining are sequences of Hadamard operations for the single-qubit case and controlled-NOT ($CNOT$) operations for the two-qubit case. These operations are involutions, i.e. the repeated application of the operations gives the identity for an even number of repetitions: $H \circ H = \mathbb{1}$ and $CNOT \circ CNOT = \mathbb{1}$. In the error-free case the repeated application of H^2 or $CNOT^2$ does not change the state vector at all. Under the influence of errors, however, the fidelity will decrease.

3.2 Correction of Decoherence Errors

Figure 1(a) summarizes the result of the threshold determination for decoherence errors. The gain g, defined as the ratio between the fidelities with and without QEC, is plotted against the single-qubit decoherence rate p. This plot contains the results of three different lengths of the H^{2k}-algorithm, $k \in \{100, 1000, 100000\}$. The gain is a direct indicator for the range of parameters, where there is benefit from using QEC. The intersection of the curves with the $g = 1$ axis locates the threshold for fault-tolerant quantum computation. In this plot, the interesting part is contained on the left side of the plot, where the gain becomes larger than 1. If p is large, the use of QEC is useless, because a non-encoded system will always perform better. In this case, QEC introduces more errors than it can correct, even if done fault-tolerantly. The lines in Figure 1(a) are fits through all points of each fixed k H^{2k}-iteration run.

Please note that each point is the result of many statistical iterations and especially for low values of p the computational effort can increase dramatically to gather sufficiently high statistics. The computational effort for increasing algorithm length rises linearly for

this simple algorithm. That is why the error bars for the runs with $k = 100000$ iterations are large, because the computational costs prohibit the accumulation of very high statistics.

A conservative estimate for the threshold of fault-tolerant QEC is

$$p_{\text{thr}} = (5.2 \pm 0.2) \cdot 10^{-6}. \tag{8}$$

If we want to give a safe lower bound for the threshold we can state with confidence that the threshold is $5.0 \cdot 10^{-6}$, so that for any single-qubit error rate less than this value we can assure that fault-tolerant QEC will improve the reliability of the quantum computer.

From the results of the H^{2k}-experiment, we see that the gain is marginal for algorithm lengths of $k = 1000$ and increases for longer algorithms.[d] This shows that long algorithms will especially profit from QEC, while for rather short algorithms QEC does work, but the gain is rather marginal. QEC is especially well suited for the protection of quantum memory over large periods of time.

3.3 Correction of Operational Errors

For the analysis of the single-qubit case, we use a sequence of $2k$ Hadamard operations as test algorithm. These operations involve unitary over-rotations as described in section 2.2. With the assumption that fault-tolerant QEC will also lead to an effective decrease of operational errors according to $\sigma_{\text{eff}} = c\sigma^2$, we can again look at the gain g, i.e. the ratio of the fidelity of the encoded case to the fidelity of the unencoded case, with σ being the standard deviation of the Gaussian distributed over-rotations. Figure 1(b) shows the result of this evaluation. The threshold for operational errors $\sigma_{\text{thr}} = 1/c$ can be derived from the fit parameter c. The numerical simulations give the result

$$\sigma_{\text{thr}} = 0.0431 \pm 0.0002. \tag{9}$$

If σ is larger than σ_{thr} there is no possibility to get an improvement by using QEC. A Gaussian width of $\sigma_{\text{thr}} = 0.0431 \pm 0.0002$ corresponds to an over-rotation by $2.5°$, which is a rather large value. For comparison, ion trap quantum computation has to deal with pulse area errors (and therefore operational over-rotations) of about 0.007 ± 0.003[10].

Figure 1(b) clearly shows the benefit from QEC for the correction of Gaussian distributed unitary operational imprecisions. For all points above one, running the algorithm with QEC performs better than using unencoded qubits. Even for relatively short algorithm lengths the gain that comes from using QEC is high. In the case of two-qubit $CNOT$ operations, the resulting threshold is similar to that of the single-qubit Hadamard operations.

3.4 Correction of Decoherence and Operational Errors

For algorithms of the same length the gain from the correction of operational errors is much higher than the gain from the correction of decoherence errors. That means that QEC is especially well suited for the correction of operational errors. In other words: If operations on qubits have to be done with a previously unattainable precision, it is possible to use logical operations on encoded states to achieve effectively more precise quantum operations than the physical apparatus is able to deliver for a single qubit. QEC is also

[d] The maximum gain is 2 for a single qubit, since the fidelity of a maximally noisy state is 0.5.

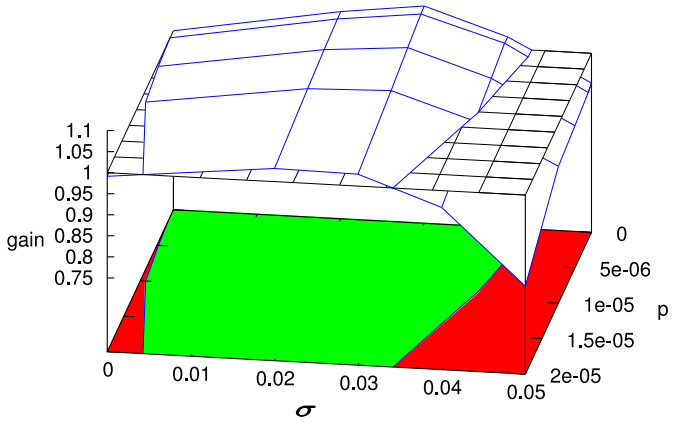


Figure 2. QEC of decoherence errors *and* operational errors in the case of Steane's 7-qubit QEC-code. The gain for 100 H^2-operations is plotted against the single-qubit decoherence probability p and the decoherence parameter σ. The green region marks the parameter space in which fault tolerant QEC gives a benefit.

well suited for the protection of quantum memory, but the benefit is only large for the protection over long periods of time.

With the JUMPIQCS simulation package we are able to quantify the effect of fault-tolerant QEC for any parameter range. Thus, we can for example quantify optimal working points in the presence of both decoherence and operational errors. In this case, one can benefit from using QEC, although the single-qubit decoherence error rate might be above threshold. The improvement from the protection against operational errors can outweigh the loss caused by decoherence errors for a certain range of parameters. With numerical simulations we can determine a combined threshold and optimal working ranges.

4 Computational Resources

Using a complete description of the quantum mechanical system requires exponential resources. Thus a parallel version for large-scale supercomputers has been developed and optimized for performance and scalability. In this work, we deal with pure states, so that the state vector contains the complete information about the quantum state. Since the state vector grows exponentially with the number of qubits, the memory requirements pose a hard limit on possible simulations. A system with n qubits has a state vector with $N = 2^n$ complex-valued amplitudes. Using double precision numbers of 8 bytes length for each the real and the imaginary part, the memory requirement for storing the state vector is 2^{n+4} bytes.

JUMPIQCS has been tested on different computer architectures, such as the IBM Regatta P690+, the IBM Blue Gene/L, and the IBM Blue Gene/P. It has been run on the JUMP supercomputer (Jülich Multi Processor) with a maximum system size of 37 simulated qubits using 1024 processors and 3 TB of memory. On the JUGENE system (Jülich Blue Gene), the simulation included up to 40 qubits using 49152 processors and 24 TB of memory. To our knowledge, this is currently the largest system size that has been simulated without using approximate methods.

On computer systems with distributed memory, the state vector is distributed over different memory locations and each processor has access to the local part of the memory only. Message passing is required to gain access to specific parts of the state vector. In this case the 2^n complex-valued amplitudes of the state vector are distributed to local memory parts, each containing 2^M amplitudes, so that we deal with $L = 2^n/2^M$ MPI processes.

As described in section 2 our simulations are based on statistically independent error locations. We have to make sure to gather sufficiently high statistics. "Sufficiently high" depends on the single qubit error probability p, the length of the algorithm, the number of qubits, the complexity of the algorithm, i.e. the phase space of the system, and of course on the required accuracy of the result. As an example, running the H^{2k}-algorithm with the parameter set $k = 1000$, $p = 10^{-7}$ and fully fault-tolerant Steane encoding requires at least $m = 10000$ statistical iterations to yield a result with a relative error[e] of less than 10^{-3} in the final fidelity. Running this system on a single processor of the JUMP (Power4) system requires about two days of computation time for $m = 10000$ iterations.

5 Conclusion

We have determined a threshold for fault-tolerant quantum computation which is $p_{\text{thr}} = (5.2 \pm 0.2) \cdot 10^{-6}$. At present, an error rate of $5 \cdot 10^{-6}$ is beyond the scope of existing quantum computation devices and for the near future it will be an ambitious goal to endeavor, but in principle there is no physical limitation that prohibits to achieve this threshold. Assuming that one can reach the threshold, fault-tolerant QEC is especially well suited for the protection of quantum bits over long periods of time, e.g. in quantum memory. For the protection over a short timeframe ($\lesssim 1000$ gate operations) we cannot recommend using QEC without reservation, because the benefit does not justify the cost, although using fault-tolerant QEC does not degrade the fidelity of the protected qubit. In this particular case, the improvement by using QEC is only marginal. The simulation results suggest that QEC is especially well suited for the correction of operational imprecisions. There is also a fault-tolerant threshold for this kind of error, that we can quantify as $\sigma_{\text{thr}} = (4.31 \pm 0.02) \cdot 10^{-2}$. Although we have only analyzed a limited set of algorithms using JUMPIQCS, we have developed a simulation package that can, in principle, be used to make statements about requirements of future quantum computer architectures, to calculate limits and thresholds and that can be applied to arbitrary quantum algorithms to determine optimal working points for QEC. Since future quantum computing devices will most probably incorporate some sort of QEC, JUMPIQCS can also be used as a tool for the assessment of future quantum computing architectures.

Acknowledgements

We would like to thank Kristel Michielsen and Hans De Raedt for stimulating and fruitful discussions. Numerical simulations were done on the supercomputers of the Jülich Supercomputing Centre.

[e] We take the standard deviation of the mean as a measure of statistical error.

References

1. P. W. Shor, *Scheme For Reducing Decoherence In Quantum Computer Memory*, Physical Review A, **52**, R2493–R2496, 1995.
2. A. M. Steane, *Error Correcting Codes In Quantum Theory*, Physical Review Letters, **77**, 793–797, 1996.
3. D. Aharonov and M. Ben-Or, *Fault Tolerant Quantum Computation With Constant Error*, quant-ph/9611025v2, 1996.
4. K. De Raedt, K. Michielsen, H. De Raedt, B. Trieu, G. Arnold, M. Richter, T. Lippert, H. Watanabe, and N. Ito, *Massively Parallel Quantum Computer Simulator*, Computer Physics Communications, **176**, no. 2, 121–136, 2007.
5. R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, *Perfect Quantum Error Correcting Code*, Physical Review Letters, **77**, 198–201, 1996.
6. A. Steane, *Multiple-Particle Interference And Quantum Error Correction*, Proceedings of the Royal Society of London, Series A: Mathematical, Physical and Engineering Sciences, **452**, 2551–2577, 1996.
7. D. B. Trieu, *Large-Scale Simulations of Error-Prone Quantum Computation Devices*, PhD thesis, Bergische Universität Wuppertal, 8 2009.
8. E. Knill, R. Laflamme, and W.H. Zurek, *Resilient Quantum Computation*, Science, **279**, no. 5349, 342–345, 1998.
9. M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
10. E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Randomized Benchmarking Of Quantum Gates*, Physical Review A, **77**, no. 1, 012307, 2008.

Optimization of Packing Problems with Heuristic Algorithms Suited for Parallel Enablement

Johannes Josef Schneider

Centre for Computational Research Methods in Natural Sciences
Johannes Gutenberg University of Mainz
Staudinger Weg 7, 55099 Mainz, Germany
E-mail: schneidj@uni-mainz.de

Our aim is to develop a universal high performance general purpose optimization algorithm for packing problems. We focus on established benchmark problems for which no exact algorithms exist, such that only best-so-far solutions or upper and lower bounds are known for these problems. For a specific multidisperse packing problem, we were already able to match and to beat all world records established in an international contest.

1 Introduction

In our daily lives, we deal with packing problems of various kinds, especially when we have to fit large amounts of goods, which we just bought at some supermarkets, into the rear trunks of our cars. These goods exhibit various sizes, shapes, and weights. Often such packing problems inherit constraints, e.g. fruit must not be squashed while being packed into the trunk.

Packing problems also occur in various industrial sectors, like in transport business, where various goods have to be packed optimally in trucks, in textile, wood, and metal-working industry, where the amount of offcuts has to be minimized. Also in physics, the packing of hard particles (usually disks or spheres) is studied, when considering granular matter, colloidal systems, or molecular crystals. Often also here additional constraints have to be considered, like a confined geometry or external forces. Mostly, multidisperse packing problems with many items differing in their sizes, shapes, volumes, and weights occur in logistics and in industrial applications. Often these problems are simplified by using containers with normalized sizes and shapes, in which these items are packed, thus reducing the complexity of the optimization problem, but also giving up the aim of finding the densest packing. Also in physics and mathematics, usually only systems with one type of particles or two types of particles are considered, with A- and B-particles differing in their sizes and/or interactions. Some of these problems have been studied very intensively, e.g. the problem of *packing circles in a circle*: here the task is to find the densest closed packing of N disks with radii $r_i = i$, $i = 1, \ldots, N$ in a circular environment within a circumcircle, such that the radius R of this circumcircle is minimal. There are related benchmark problems, like *packing circles in a square*, *packing circles in a triangle*, *packing squares in a circle*, and so on, in which also N items of the same size and shape have to be packed in a container of a given shape, the size of which has to be minimized. Mathematicians put a great effort in proving the optimality of a solution for some system size N. For most packing problems, exact computer algorithms can only determine the optimum solution for very small values of N, as the computing time increases mostly exponentially

N	contest record	my value	record	N	contest record	my value	record
5	9.0013977	9.0013977	matched	28	94.5499865	94.5265365	beaten
6	11.0570404	11.0570404	matched	29	99.5123179	99.4831116	beaten
7	13.4621107	13.4621107	matched	30	104.5785550	104.5411690	beaten
8	16.2217467	16.2217467	matched	31	109.7719469	109.6824564	beaten
9	19.2331939	19.2331939	matched	32	114.8654383	114.8409490	beaten
10	22.0001930	22.0001930	matched	33	120.2169571	120.0658465	beaten
11	24.9606343	24.9606343	matched	34	125.4335018	125.3669392	beaten
12	28.3713894	28.3713894	matched	35	131.1563546	130.9176279	beaten
13	31.5458670	31.5458670	matched	36	136.5349008	136.4922446	beaten
14	35.0956471	35.0956471	matched	37	142.1749805	142.0513754	beaten
15	38.8379955	38.8379955	matched	38	147.8576914	147.4568453	beaten
16	42.4581164	42.4581164	matched	39	153.5553012	153.3800827	beaten
17	46.2913421	46.2913421	matched	40	159.4890249	159.1824078	beaten
18	50.1197626	50.1197626	matched	41	165.2919097	165.0369006	beaten
19	54.2402936	54.2402936	matched	42	170.9257616	170.8953274	beaten
20	58.4005675	58.4005675	matched	43	177.0743401	177.0513747	beaten
21	62.5588771	62.5588771	matched	44	183.1760616	183.0992248	beaten
22	66.7602862	66.7602862	matched	45	189.6354391	189.2029320	beaten
23	71.1994616	71.1994616	matched	46	195.9107634	195.5264351	beaten
24	75.7527041	75.7491426	beaten	47	202.1856117	201.7279256	beaten
25	80.2858644	80.2858644	matched	48	208.6359467	208.0901593	beaten
26	85.0764012	84.9899391	beaten	49	214.6619520	214.2954475	beaten
27	89.7921816	89.7509627	beaten	50	221.0897526	220.6004187	beaten

Table 1. Comparison of the best results found for $5 \leq N \leq 50$ during the benchmark contest, in which 155 groups from 32 countries took part, with my results.

with N. Sometimes it is possible to determine the optimum solution for some specific larger values of N. But mostly heuristic algorithms have to be used which provide good but not necessarily optimum solutions. The qualities of these heuristics are often compared by their results for well-defined benchmark problems, for which sometimes even contests are organized.

Recently, a multidisperse variant of *packing circles in a circle* was introduced in an international contest, which was organized by Sam Byrd, Jean-Charles Meyrignac, and Al Zimmermann[1]. The task was to find the densest closed packing of N disks with radii $r_i = i$, $i = 1, \ldots, N$ in a circular environment in the way that the radius R of the circumcircle is minimized. This problem can be trivially solved for $N \leq 4$. In the contest, the system sizes $N = 5, 6, \ldots, 50$ were considered. 155 groups from 32 countries took part in the contest, some of which have worked on packing problems for several years already. As shown in Table 1, I was able to match and to beat all world records established during the contest with our new packing algorithm, which is suited for parallel enablement. Figure 1 shows the best results I found for the problem instances with $N = 30$ and $N = 40$ disks. This result gained a lot of attention by various journalists. The packing algorithm was rated to be one of the 50 most important inventions of the year 2009 by the Time Magazine. Reports about this work were shown on TV (German TV channels ARD, ZDF, and SWR), published in online journals (Spiegel Online, New Scientist Online), and printed in various newspapers (e.g. Frankfurter Rundschau, Frankfurter Allgemeine Zeitung).

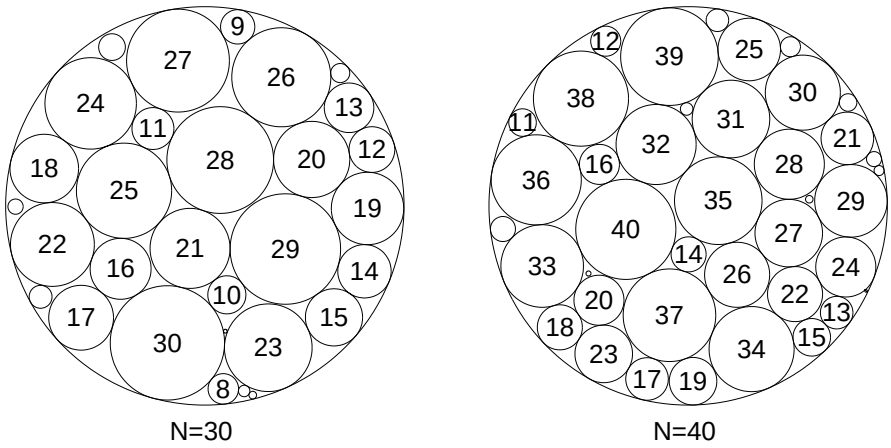


Figure 1. Best solutions found for the multidisperse packing problems with $N = 30$ disks and $N = 40$ disks: the numbers within the larger disks denote their radii.

2 The Packing Algorithm

The packing algorithm we developed is based on a combination of global optimization algorithms from statistical physics and local optimization algorithms from computer science.

2.1 Global Optimization

In a first step, we use *simulated annealing* (SA)[2] and its deterministic variant *threshold accepting* (TA)[3]. Here the problem to be optimized is mapped on a physical system. The cost function or objective function of the optimization problem is interpreted as an energy function $\mathcal{H}$ of the physical system, which is to be minimized. This goal is achieved by an artificial cooling process during a computer simulation, resembling cooling processes in nature: SA and TA start off with a randomly created configuration at a high temperature. Then they subject the system to a cooling schedule, in which the temperature is gradually decreased. This decrease can be performed e.g. in the way $T^{\mathrm{new}} = fT^{\mathrm{old}}$, with $f < 1$, e.g, $f = 0.99$. In each temperature step, the system is altered several times by so-called moves, changing the configuration slightly. Such a small change from the current configuration σ_i to a tentative new configuration σ_{i+1} can be performed by shifting the midpoint (x_i, y_i) of some randomly chosen disk i slightly, by letting the disk jump to a completely new position, and by exchanging two disks with different radii. We found that the quality of the results strongly depends on the parameters of these moves, e.g. the maximum offset in x- and y-direction, by which a disk is shifted[4]. Let $\Delta\mathcal{H} = \mathcal{H}(\sigma_{i+1}) - \mathcal{H}(\sigma_i)$ be the energy or quality difference between the current configuration σ_i and the tentative new configuration σ_{i+1}. Then each move is accepted according to the Metropolis acceptance probability[5]

$$W(\sigma \to \tau) = \begin{cases} 1 & \text{if} \Delta\mathcal{H} \leq 0 \\ \exp(-\Delta\mathcal{H}/T) & \text{otherwise} \end{cases} \tag{1}$$

in case of SA (i.e., the move is always accepted if it leads to an improvement and only with some probability if it leads to a deterioration, the probability depending on the amount of deterioration)[2] or with

$$W(\sigma \to \tau) = \begin{cases} 1 & \text{if} \Delta\mathcal{H} \leq T \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

in case of TA[3] (i.e., the move is always accepted if it leads to an improvement and also always for small deteriorations up to some size T). If the acceptance criterion is fulfilled, the system performs the move $\sigma_i \to \sigma_{i+1}$, in case of rejection, one sets $\sigma_{i+1} := \sigma_i$. Thus, accepting a move can be seen as making a step forward in some landscape, while rejecting it means remaining on the current spot. After the acceptance or rejection of a move, a new move is performed, in which again some randomly chosen disk is shifted to some randomly chosen new location or two randomly chosen disks are exchanged.

For the multidisperse packing problem, the objective function to be minimized is given by the radius R of the circumcircle. If we consider the midpoint of the circumcircle to be located at the origin, R can be written as

$$R = \max_i \left(r_i + \sqrt{x_i^2 + y_i^2} \right) \tag{3}$$

with the maximum taken over all disks i with radii r_i, whose midpoints are located at the coordinates x_i and y_i and whose outermost points have the distance $r_i + \sqrt{x_i^2 + y_i^2}$ to the origin. Furthermore, there is the constraint that the various disks must not overlap. Thus, for each pair (i, j) of disks, the condition

$$P(i,j) = r_i + r_j - \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq 0, \tag{4}$$

i.e., the condition that the sum of their radii is not larger than the distance between their midpoints, must be fulfilled. There are various ways of dealing with such a constraint, e.g. by generating only feasible solutions[6]. However, for this problem, it is advantageous to make use of penalty functions, which are added to the energy function if the corresponding constraints are violated:

$$\mathcal{H} = R + \lambda P(i,j)\Theta(P(i,j)) \tag{5}$$

λ is a Lagrange multiplier weighing the constraints with respect to the radius. It has to be chosen large enough, such that no overlaps remain at the end of the optimization run, and small enough, in order to allow the system to walk in the search space from an island comprised of feasible solutions to another such island, on a path consisting of configurations exhibiting at least one overlap. We found that for our problem $\lambda = 1$ seems to be optimal[4]. $\Theta(x)$ denotes the Heaviside function

$$\Theta(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

ensuring that the penalty is only added if the corresponding constraint is violated.

Summarizing, this global optimization approach subjects the system to be optimized to a cooling process, thus transferring it from a high-energy regime at high temperatures to a low-energy configuration at low temperatures. When using SA, the mean value of the

energy values occurring at temperature T decreases linearly with T at high temperatures[4]. The prefactor of this decrease depends only on the number of disks, not on their sizes.

2.2 Local Optimization

The resulting solutions we get sometimes still contain some small overlaps of the order $10^{-2} - 10^{-5}$. In order to resolve these overlaps and furthermore to improve our solutions, we use a local optimization technique as afterburner. Let (i, j) be a pair of overlapping disks. Then an overlap vector $\vec{v}_{ij}$ can be defined by

$$\vec{v}_{ij} = \frac{r_i + r_j - \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \left(\begin{pmatrix} x_j \\ y_j \end{pmatrix} - \begin{pmatrix} x_i \\ y_i \end{pmatrix} \right). \tag{7}$$

If there is no overlap, then $\vec{v}_{ij} = 0$. For each disk, the sum of these overlap vectors is added to the location of the midpoint of the disk:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix}^{\text{new}} = \begin{pmatrix} x_i \\ y_i \end{pmatrix}^{\text{old}} + \frac{1}{2} \sum_j \vec{v}_{ji} \tag{8}$$

The update of the midpoints is performed in parallel. This routine ensures that if there is only one overlap then it is resolved. But by resolving it, i.e., by shifting two disks away from each other until they only touch each other, new overlaps might occur with other disks. Furthermore, the various overlap vectors might intend to shift a disk in opposite directions. Thus, this resolving routine has to be repeated until no significant overlap remains. (Due to the finite numerical precision, overlaps of the order 10^{-8} might remain.)

This process of resolving remaining small overlaps can be combined with a local approach for improving the solution. This is simply done by shifting the outermost disk slightly towards the center of the circumcircle. By doing so, one might create overlaps with neighbouring disks which have to be resolved as described above. This approach of shifting the outermost disk slightly towards the origin and of resolving the occurring overlaps afterwards is repeated for decreasing values for the length of the shift vector until no further improvement can be found.

3 Parallelization of the Algorithm

Both our global and our local packing algorithm, which were implemented in Fortran 77 with Fortran 8.x extensions, are suited for parallel enablement. For the global optimization approach, the largest amount of computing time has to be used for the determination of the overlaps of the shifted disk or the two exchanged disks with all other disks. This determination can be performed in parallel. Here it is advantageous to make use of OpenMP. Furthermore, the radius of the tentative new configuration has to be determined. We always store the number of the outermost disk, which determines the radius of the circumcircle. If this disk is not selected in the move for shifting it, moving it, or exchanging it, we have only to check whether the shifted/moved/exchanged disks lie partially or in whole outside the circumcircle. However, if this disk is selected, then we also have to go through all disks

for determining the outermost one. Analogously, this loop over the disks can be performed in parallel with OpenMP.

In order to determine the behaviour of physical observables in the cooling process, it is necessary to perform a huge number of optimization runs starting with different seeds for the random number generator in parallel. In order to generate the picture of the specific heat in Ref. 4, we had to perform hundreds of optimization runs, over which we had to take the average, as the data points achieved in the various runs by averaging over a large number of measurements still fluctuated strongly. For this parallelization, we used MPI. Additionally, in thousands of optimization runs, we produced a huge number of quasi optimum solutions, which allowed us to determine their common properties and the differences between them, which led e.g. to the result that the subspace of quasi optimum configurations exhibits the property of ultrametricity[7].

Partially, we also used parallel tempering instead of simulated annealing. This method works like simulated annealing with a unique constant temperature value T_i on a single node i of a parallel computer, but from time to time the configurations are exchanged between the various nodes, in order to allow for treating them at other temperature values.

Also the local optimization method is suited for parallel enablement. In each step, a loop over all disks has to be performed for the determination of the new outermost disk. Additionally, the overlaps between all pairs of disks have to be calculated. Again these loops can be performed in parallel by making use of OpenMP.

Furthermore, by randomizing the strength with which the outermost disk is hammered towards the center of the circumcircle, which is the most crucial parameter of the local optimization method, the local optimization method can be improved by performing it in parallel with MPI and making use of an evolutionary algorithm. We start off with the best configurations achieved by the global optimization methods, which are subjected to the iterative method described above. After some iterations, the best configurations are cloned, while the worst configurations are discarded. This approach is repeated until no significant improvement can be found for some iterations.

Summarizing, these descriptions show that we made various uses of the supercomputing resources at the John von Neumann Institute for Computing.

4 Results for the Multidisperse Packing Problem

4.1 Optimum and Average Results

When plotting the best radius values R we achieved vs. the corresponding system sizes N, as shown in the left graphic of Figure 2, we find that the radius R scales as $R \propto N^{1.5}$. This result can be easily explained by reconsidering that the size of the circumdisk must be at least as large as the sum of the sizes of the various disks, i.e., $R^2\pi \geq \sum_{i=1}^{N} r_i^2\pi = \frac{\pi}{6}N(N+1)(2N+1) > \frac{\pi}{3}N^3$. Analogously, upper bounds for the radius R can be constructed, which also depict a $N^{1.5}$ power law scaling function.

While this result is certainly interesting, we have to state that it does not help us to prove whether our results are optimal or at least close to optimal. However, there exists an interesting hypothesis for a different type of complex optimization problems called spin glass models, which was also found to hold also for other problems, like the traveling salesman problem[6]. This Grest hypothesis allows to estimate the exact optimum value

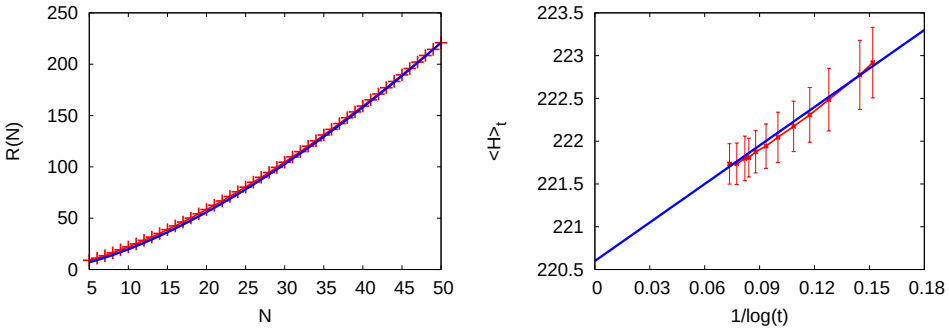


Figure 2. Left: radius R of the best configurations found for various system sizes N. Right: average quality $\langle \mathcal{H} \rangle_t$ of results achieved after calculation time t for the instance with $N = 50$ disks vs. inverse of the logarithm of that calculation time.

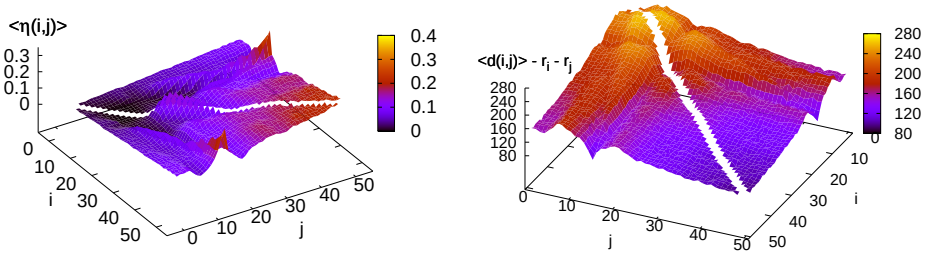


Figure 3. Left: probability with which a disk with radius $r_i = i$ touches a disk with radius $r_j = j$. Right: average distance between disks with radii i and j.

from average results achieved with SA for various calculation times: the deviation of the average quality $\langle \mathcal{H} \rangle_t$ of results achieved after an amount of calculation time t from the ground state energy $\mathcal{H}_0$ decreases slowly with increasing calculation time t according to $\langle \mathcal{H} \rangle_t - \mathcal{H}_0 \sim 1/\log(t)^\zeta$ with $\zeta \approx 1$. The right diagram in Figure 2 shows for the problem with $N = 50$ disks how the ground state energy can be very well estimated with the Grest hypothesis, if comparing the result of ≈ 220.6 to the value in Table 1.

4.2 Neighbourhood Relations of Disks in Quasi Optimum Solutions

Besides the new world record configurations I achieved, we got a huge number of quasi optimum solutions, several of which were better than the world records established during the benchmark contest. We had a closer look at roughly 10,000 of the best solutions we achieved for the problem comprised of $N = 50$ disks and compared them for common properties. Contrarily to our intuition, disks being adjacent to each other in one solution are not necessarily adjacent to each other in other solutions, as shown in the left diagram of Figure 3. There is only a tendency that the larger disks are adjacent to each other more

often. Furthermore, disks with radii $r_i \approx 15$ seem to like to be close to the largest disks. One often finds that two of the largest disks, which lie adjacent to each other and touch the circumcircle, form a hole between them and the circumcircle which can be ideally filled by such a smaller disk.

That the largest disks tend to stay close to each other can also be seen in the right diagram of Figure 3, which shows the average distance between the midpoints of the disks with radii $r_i = i$ and $r_j = j$, from which the sum of these radii is subtracted.

5 Concluding Remarks

From the taxpayers' point of view, one might argue whether it was really worth to spend a lot of computing time for the investigation of an academic problem, which does not seem to have any practical impact at first sight. However, when trying to develop a new computer algorithm, which is intended to be superior to other algorithms, which is intended to be a general purpose algorithm for the solution of any packing problem, and which is indeed intended to be applied not only to academic but also to real-life problems, it is always advantageous to start out with some widely studied benchmark problems, to which the algorithm can first be applied and for which it can afterwards be tuned. Then one is not only able to solve this specific benchmark problem but also to get a feeling for packing problems in general, thus enabling oneself to improve one's own algorithm even more. The benefit of this work is thus not only glory, but also a deep understanding of packing problems, which finally helps to develop a packing algorithm for real-world problems. We are very grateful to the John von Neumann Institute at the Research Centre Jülich for the large grant of computing time we got. Without this grant, we would not have been able to achieve any of the results presented here.

References

1. `http://www.azspcs.net`.
2. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, *Optimization by Simulated Annealing*, Science **220**, 671–680, 1983.
3. G. Dueck and T. Scheuer, *Threshold Accepting: A General Purpose Optimization Algorithm Appearing Superior to Simulated Annealing*, J. Comp. Phys. **90**, 161–175, 1990.
4. A. Müller, J. J. Schneider, and E. Schömer, *Packing a multidisperse system of hard discs in a circular environment*, Phys. Rev. E **79**, 021102, 2009.
5. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations for fast Computing Machines*, J. Chem. Phys. **21**, 1087–1092, 1953.
6. J. J. Schneider and S. Kirkpatrick, *Stochastic Optimization* (Springer, Berlin, Heidelberg, New York, 2006).
7. J. J. Schneider, A. Müller, and E. Schömer, *Ultrametricity property of energy landscapes of multidisperse packing problems*, Phys. Rev. E **79**, 031122, 2009.

Performance Tuning in the Petascale Era

Felix Wolf[1,2], **David Böhme**[1,2], **Markus Geimer**[1], **Marc-André Hermanns**[1], **Bernd Mohr**[1], **Zoltan Szebenyi**[1,2], **and Brian J. N. Wylie**[1]

[1] Forschungszentrum Jülich, Jülich Supercomputing Centre
52425 Jülich, Germany
E-mail: {f.wolf, d.boehme, m.geimer, m.a.hermanns, b.mohr, z.szebenyi, b.wylie}@fz-juelich.de

[2] RWTH Aachen University, Graduate School AICES
52056 Aachen, Germany

Driven by application requirements and accelerated by current trends in microprocessor design, the number of processor cores on modern supercomputers grows from generation to generation. As a consequence, supercomputing applications are required to harness much higher degrees of parallelism in order to satisfy their growing demand for computing power. However, writing code that runs efficiently on large numbers of processors remains a significant challenge. The situation is exacerbated by the fact that the rising number of cores imposes scalability demands not only on applications but also on the software tools needed for their development.

To address this challenge, the Helmholtz University Young Investigators Group *Performance Analysis of Parallel Programs* at Jülich Supercomputing Centre (JSC) in cooperation with the JSC Division *Application Support* creates software technologies aimed at improving the performance of applications running on leadership-class systems. At the centre of our activities lies the development of Scalasca, a performance-analysis tool that has been specifically designed for large-scale systems and that allows the automatic identification of harmful wait states in applications running on tens of thousands of processors. In this article, we highlight the research activities of our group during the past two years and give an outlook on future work.

1 Introduction

Supercomputing is a key technology pillar of modern science and engineering, indispensable to solve critical problems of high complexity. The extension of the ESFRI road map to include a European supercomputer infrastructure in combination with the creation of the PRACE consortium acknowledges that the requirements of many critical applications can only be met by the most advanced custom-built large-scale computer systems. However, as a prerequisite for their productive use, the HPC community needs powerful and robust development tools. These would not only help improve the scalability characteristics of scientific codes and thus expand their potential, but also allow domain scientists to concentrate on the underlying models rather than to spend a major fraction of their time tuning their application for a particular machine.

As the current trend in microprocessor development continues, this need will become even stronger in the future. Facing increasing power dissipation and little instruction-level parallelism left to exploit, computer architects are realizing further performance gains by using larger numbers of moderately fast processor cores rather than by further increasing the speed of uni-processors. As a consequence, supercomputer applications are being required to harness much higher degrees of parallelism in order to satisfy their growing demand for computing power. With an exponentially rising number of cores, the often substantial gap between peak performance and the performance level actually sustained by

production codes is expected to widen even further. Finally, increased concurrency levels place higher scalability demands not only on applications but also on parallel programming tools. When applied to larger numbers of cores, familiar tools often cease to work in a satisfactory manner (e.g. due to escalating memory requirements, failing displays, or limited I/O performance).

To overcome this challenge, the Helmholtz University Young Investigators Group *Performance Analysis of Parallel Programs* at Jülich Supercomputing Centre (JSC) in cooperation with the JSC Division *Application Support* creates software technologies aimed at improving the performance of applications running on leadership-class systems with tens of thousands of cores. At the centre of our activities lies the development of Scalasca[1], a performance-analysis tool that has been specifically designed for large-scale systems and that allows the automatic identification of harmful wait states in applications running on very large processor configurations.

In this article, we give an overview of Scalasca and highlight research accomplishments of our group during the past two years, focusing on the analysis of wait states and of time-dependent behaviour, as these two examples address the scalability of Scalasca regarding both the number of processes and the length of execution, respectively.

2 Scalasca Overview

The current version of Scalasca supports measurement and analysis of MPI applications written in C, C++ and Fortran on a wide range of current HPC platforms. Hybrid codes making use of basic OpenMP features in addition to passing messages are supported as well. Figure 1 shows the basic analysis workflow supported by Scalasca. Before any performance data can be collected, the target application must be instrumented and linked to the measurement library. When running the instrumented code on the parallel machine, the user can choose between generating a summary report ("profile") with aggregate performance metrics for individual function call paths and/or generating event traces recording individual runtime events from which a profile or time-line visualization can later be produced. Summarization is particularly useful to obtain an overview of the performance behaviour and for local metrics such as those derived from hardware counters. Since traces tend to rapidly become very large[2], optimizing the instrumentation and measurement based on the summary report is usually recommended. When tracing is enabled, each process generates a trace file containing records for its process-local events. After program termination, Scalasca loads the trace files into main memory and analyzes them in parallel using as many processors as have been used for the target application itself. During the analysis, Scalasca searches for wait states and related performance properties, classifies detected instances by category, and quantifies their significance. The result is a wait-state report similar in structure to the summary report but enriched with higher-level communication and synchronization inefficiency metrics. Both summary and wait-state reports contain performance metrics for every function call path and process/thread which can be interactively examined in the provided analysis report explorer.

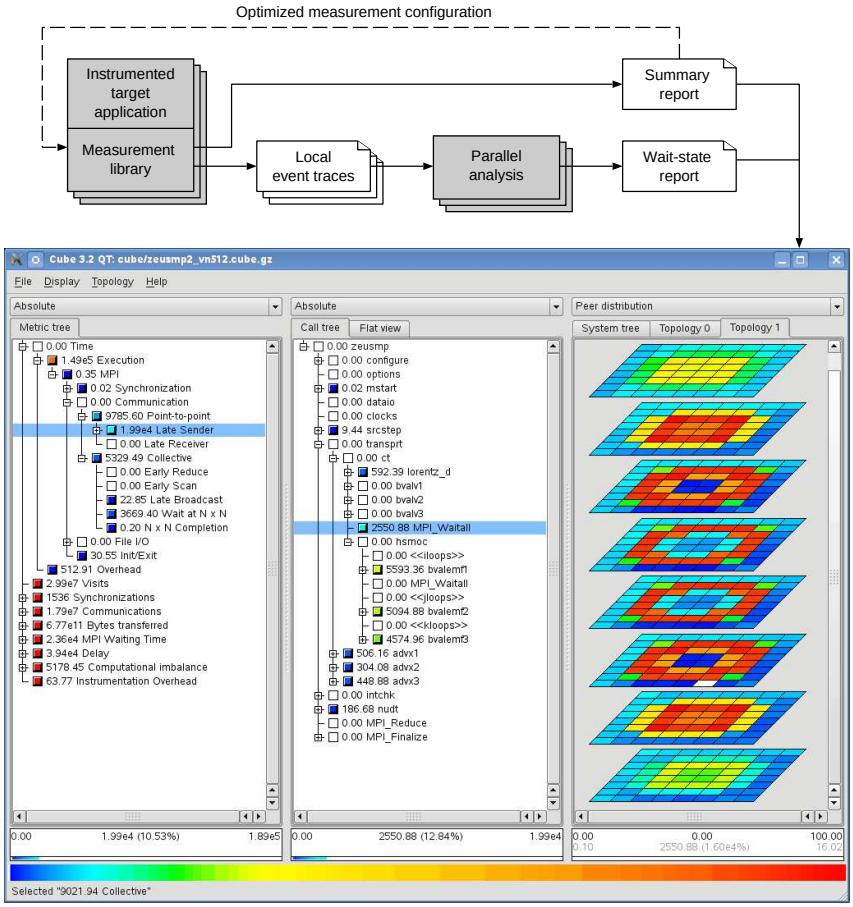


Figure 1. Schematic overview of the performance data flow in Scalasca. Grey rectangles denote programs and white rectangles with the upper right corner turned down denote files. Stacked symbols denote multiple instances of programs or files running or being processed in parallel. The GUI shows the distribution of performance metrics (left pane) across the call tree (middle pane) and the process topology (right pane).

3 Scalable Wait-State Analysis

In message-passing applications, processes often require access to data provided by remote processes, making the progress of a receiving process dependent upon the progress of a sending process. Collective synchronization is similar in that its completion requires each participating process to have reached a certain point. As a consequence, a significant fraction of the communication and synchronization time can often be attributed to wait states, for example, as a result of an unevenly distributed workload. Especially when trying to scale communication-intensive applications to large process counts, such wait states can present severe challenges to achieving good performance.

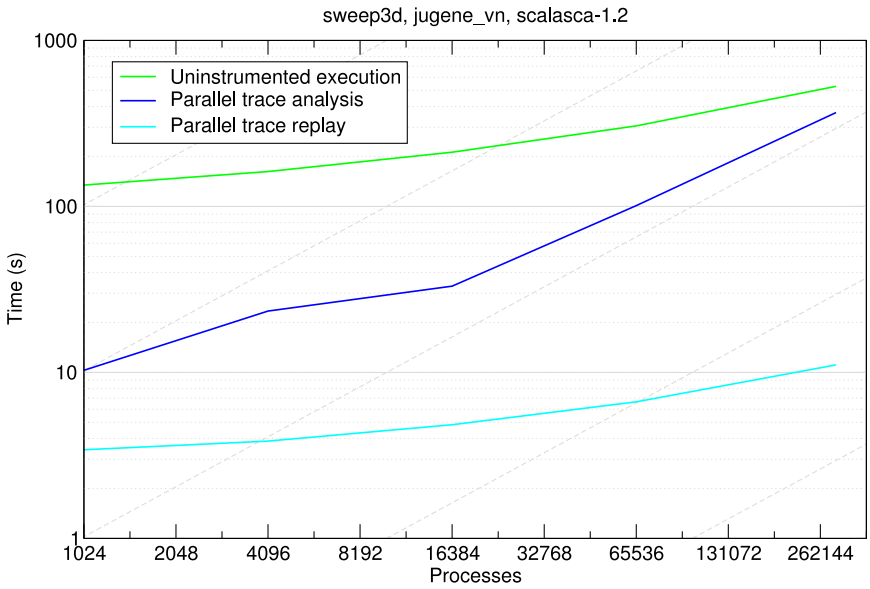


Figure 2. Scalability of wait-state search for the benchmark application SWEEP3D on the full JUGENE system. The graph charts wall-clock execution times for the uninstrumented application and the analyses of trace files generated by instrumented versions with a range of process numbers. The figure shows the total time needed for the parallel analysis including loading the traces and collating the results and the time needed for the replay in isolation. It can be seen that the total analysis takes less than 400 seconds and – for this particular example – is still faster than the application itself.

3.1 Scalability

After the target application has terminated and the trace data have been flushed to disk, the trace analyzer is launched with one analysis process per (target) application process and loads the entire trace data into its distributed memory address space. Future versions of Scalasca may exploit persistent memory segments available on system such as Blue Gene/P to pass the trace data to the analysis stage without involving any file I/O. While traversing the traces in parallel, the analyzer performs a replay of the application's original communication behaviour[3]. During the replay, the analyzer identifies wait states in communication operations by measuring temporal differences between local and remote events after their timestamps have been exchanged using an operation of similar type. Since trace processing capabilities (i.e. processors and memory) grow proportionally with the number of application processes, we can achieve good scalability at previously intractable scales. Recent scalability improvements allowed us to complete trace analyses of runs with up to 294,912 cores on the full IBM Blue Gene/P system JUGENE (Figure 2).

3.2 Delay Analysis

In general, the temporal or spatial distance between cause and symptom of a performance problem constitutes a major difficulty in deriving helpful conclusions from performance data. So just knowing the locations of wait states in the program is often insufficient to

understand the reason for their occurrence. Building on earlier work by Meira, Jr. et al.[4], we are currently extending our replay-based wait-state analysis in such a way that it attributes the waiting times to their root causes. The root cause, which we call a *delay*, is an interval during which a process performs some additional activity not performed by its peers, for example as a result of insufficiently balancing the load.

3.3 Evaluation of Optimization Hypotheses

Excess workload identified as root cause of wait states usually cannot simply be removed. To achieve a better balance, optimization hypotheses drawn from a delay analysis typically propose the redistribution of the excess load to other processes instead. However, redistributing workloads in complex message-passing applications can have intricate side-effects that may compromise the expected reduction of waiting times. Given that balancing the load statically or even introducing a dynamic load-balancing scheme constitute major code changes, they should ideally be performed only if the prospective performance gain is likely to materialize. Our goal is therefore to automatically predict the effects of redistributing a given delay without altering the application itself and to determine the savings we can realistically hope for. Since the effects of such changes are hard to quantify analytically, we simulate these changes via a real-time replay of event traces after they have been modified to reflect the redistributed load[5,6].

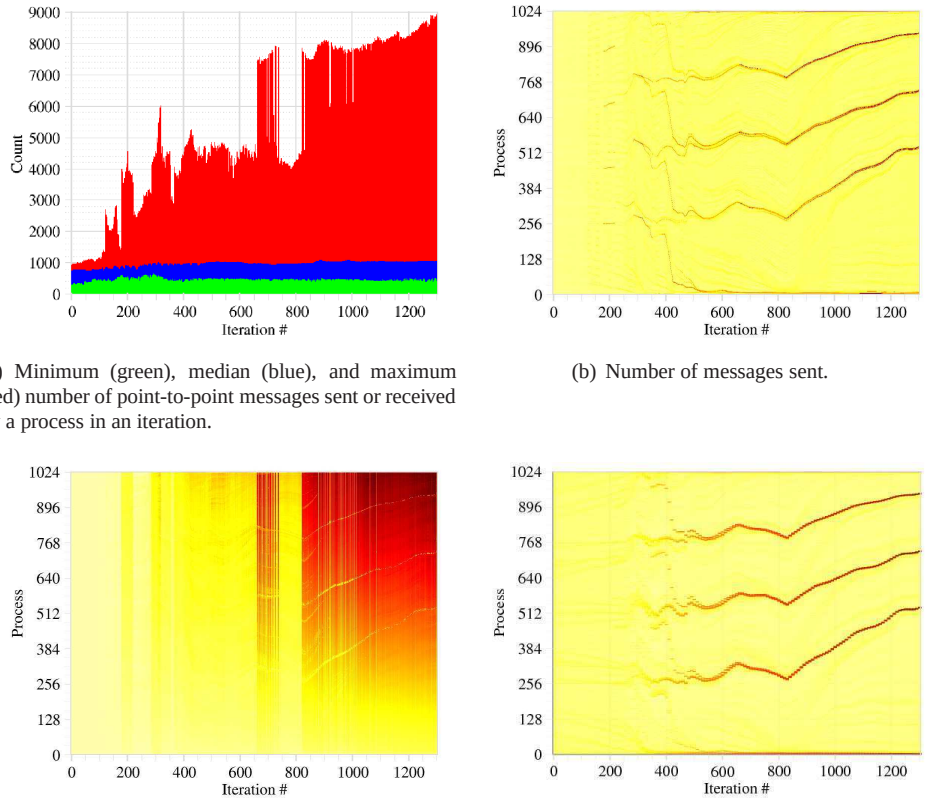
4 Analysis of Time-Dependent Behaviour

As scientific parallel applications simulate the temporal evolution of a system, their progress occurs via discrete points in time. Accordingly, the core of such an application is typically a loop that advances the simulated time step by step. However, the performance behaviour may vary between individual iterations, for example, due to periodically re-occurring extra activities[7] or when the state of the computation adjusts to new conditions in so-called adaptive codes[8].

4.1 Observing Individual Iterations

To study the time-dependent behaviour, Scalasca was equipped with iteration instrumentation capabilities corresponding to dynamic timers in TAU[9] that allow the distinction of individual iterations both in runtime summaries and in event traces. Moreover, to simplify the understanding of the resulting temporal data, we implemented several display tools including iteration graphs with minimum, median, and maximum representation (Figure 3(a)) as well as heat maps to cover the full (process, iteration) space for a given performance metric (Figure 3(b)).

With this new toolbox at our disposal, we evaluated the performance behaviour of the SPEC MPI2007 benchmark suite on the IBM SP p690 cluster JUMP, observing a large variety of complex temporal characteristics ranging from gradual changes and sudden transitions of the base-line behaviour to both periodically and irregularly occurring peaks, including possible noise[10]. Moreover, problems with several benchmarks that limited their scalability (sometimes to only 128 processes) were identified, such as distributing initialization data via broadcasts in 113.GemsFDTD and insufficiently large data sets for several



(a) Minimum (green), median (blue), and maximum (red) number of point-to-point messages sent or received by a process in an iteration.

(b) Number of messages sent.

(c) Late Sender waiting time.

(d) Number of particles owned by a process.

Figure 3. Gradual development of a performance problem along 1,300 timesteps of PEPC on 1,024 processors.

others. Even those codes that apparently scaled well contained considerable quantities of waiting time, indicating possible opportunities for performance and scalability improvement through more effective work distributions or bindings of processes to processors.

Another real-world code with a substantially time-varying execution profile is the PEPC[11] particle simulation code developed at Jülich Supercomputing Centre and subject to an application liaison between our group and the PEPC developer team. The MPI code employs a parallel tree algorithm to efficiently calculate the forces the particles exert on each other and also includes a load-balancing mechanism that redistributes the computational load by shifting particles between processes. However, our analysis[12] revealed a severe and gradually increasing communication imbalance (Figure 3(a)). We found evidence that the imbalance was caused by a small group of processes with time-dependent constituency that sent large numbers of messages to all remaining processes (Figure 3(b)) in rank order, introducing Late Sender waiting times at processes with higher ranks (Figure 3(c)). Interestingly, the communication imbalance correlated very well with the number of particles "owned" by a process (Figure 3(d)), suggesting that the load-balancing scheme smoothes

the computational load at the expense of communication disparities. Since the number of particles also influences the memory requirements of a process, we further conclude that the current behaviour of concentrating particles at a small subset of processes may adversely affect scalability under different configurations. Work with the application developers to revise the load-balancing scheme and improve the communication efficiency is in progress.

4.2 Space-Efficient Time-Series Call-Path Profiling

While call-path profiling is an established method of linking a performance problem to the context in which it occurs, generating call-path profiles separately for thousands of iterations may exceed the available buffer space — especially when the call tree is large and more than one metric is collected. We therefore developed a runtime approach for the semantic compression of call-path profiles[13] based on incremental clustering of a series of single-iteration profiles that scales in terms of the number of iterations without sacrificing important performance details. Our approach offers low runtime overhead by using only a condensed version of the profile data when calculating distances and accounts for process-dependent variations by making all clustering decisions locally.

5 Outlook

Besides further scalability improvements in view of upcoming systems in the range of several petaflops, we plan to extend Scalasca towards emerging programming models such as partitioned global address space languages and general-purpose GPU programming, which we expect to play a bigger role in the future. Moreover, to offer enhanced functionality and share development costs, we will integrate Scalasca closer with related tools including Periscope[14], TAU[15], and Vampir[16].

Acknowledgements

Financial support from the Helmholtz Association of German Research Centres through Grants VH-NG-118 and VH-VI-228 and the Deutsche Forschungsgemeinschaft (German Research Association) through Grant GSC 111 is gratefully acknowledged. Moreover, the authors would like to thank the John von Neumann Institute for Computing for the opportunity to use their high-end computer resources.

References

1. Jülich Supercomputing Centre, *Scalasca*, `http://www.scalasca.org/`.
2. Felix Wolf, Felix Freitag, Bernd Mohr, Shirley Moore, and Brian J. N. Wylie, *Large Event Traces in Parallel Performance Analysis*, in: Proc. of the 8th Workshop on Parallel Systems and Algorithms (PASA, Frankfurt/Main, Germany), Lecture Notes in Informatics, pp. 264–273, Gesellschaft für Informatik. March 2006.

3. Markus Geimer, Felix Wolf, Brian J. N. Wylie, and Bernd Mohr, *A scalable tool architecture for diagnosing wait states in massively-parallel applications*, Parallel Computing, **35**, no. 7, 375–388, 2009.
4. Wagner Meira, Jr., Thomas J. LeBlanc, and Alexandros Poulos, *Waiting time analysis and performance visualization in Carnival*, in: Proc. of the SIGMETRICS Symposium on Parallel and Distributed Tools (SPDT'96), pp. 1–10, Philadelphia, PA, 1996.
5. Marc-André Hermanns, Markus Geimer, Felix Wolf, and Brian J. N. Wylie, *Verifying Causality Between Distant Performance Phenomena in Large-Scale MPI Applications*, in: Proc. of the 17th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP), pp. 78–84, IEEE Computer Society, Weimar, Germany, February 2009.
6. David Böhme, Marc-Andre Hermanns, Markus Geimer, and Felix Wolf, *Performance Simulation of Non-Blocking Communication in Message-Passing Applications*, in: Proc. of the 2nd Workshop on Productivity and Performance (PROPER 2009), August 2009, (to appear).
7. Darren J. Kerbyson, Kevin J. Barker, and Kei Davis, *Analysis of the Weather Research and Forecasting (WRF) Model on Large-Scale Systems*, in: Proc. of the Conference on Parallel Computing (ParCo, Aachen/Jülich, Germany), vol. 15 of *Advances in Parallel Computing*, pp. 89–98, IOS Press. September 2007.
8. Sameer S. Shende, Allen D. Malony, Alan Morris, Steven Parker, and J. Davison de St. Germain, *Performance Evaluation of Adaptive Scientific Applications using TAU*, in: Proc. of the International Conference on Parallel Computational Fluid Dynamics (Washington DC, USA), May 2005.
9. Allen D. Malony, Sameer S. Shende, and Alan Morris, *Phase-Based Parallel Performance Profiling*, in: Proc. of the Conference on Parallel Computing (ParCo, Malaga, Spain), vol. 33 of *NIC Series*, pp. 203–210, John von Neumann Institute for Computing. September 2005.
10. Zoltán Szebenyi, Brian J. N. Wylie, and Felix Wolf, *SCALASCA Parallel Performance Analyses of SPEC MPI2007 Applications*, in: Proc. of the 1st SPEC Int'l Performance Evaluation Workshop (SIPEW, Darmstadt, Germany), vol. 5119 of *Lecture Notes in Computer Science*, pp. 99–123, Springer. June 2008.
11. Paul Gibbon, Wolfgang Frings, Sonja Dominiczak, and Bernd Mohr, *Performance Analysis and Visualization of the N-Body Tree Code PEPC on Massively Parallel Computers*, in: Proc. of the Conference on Parallel Computing (ParCo, Málaga, Spain), vol. 33 of *NIC Series*, pp. 367–374, October 2005.
12. Zoltán Szebenyi, Brian J. N. Wylie, and Felix Wolf, *Scalasca Parallel Performance Analyses of PEPC*, in: Proc. of the EuroPar Workshop on Productivity and Performance (PROPER 2008, Las Palmas de Gran Canaria, Spain), vol. 5415 of *Lecture Notes in Computer Science*, pp. 305–314, Springer. August 2008.
13. Zoltán Szebenyi, Felix Wolf, and Brian J. N. Wylie, *Space-Efficient Time-Series Call-Path Profiling of Parallel Applications*, in: Proc. of the ACM/IEEE conference on Supercomputing (SC09, Portland, OR), November 2009.
14. Technical University of Munich, *Periscope*, `http://www.lrr.in.tum.de/~gerndt/home/Research/PERISCOPE/Periscope.htm`.
15. University of Oregon, *TAU*, `http://www.cs.uoregon.edu/research/tau/`.
16. Technische Universität Dresden, *Vampir*, `http://www.vampir.eu/`.

Hydrodynamics and Turbulence

Dietmar Kröner

Department of Applied Mathematics, University of Freiburg
Hermann-Herder-Str. 10, D-79104 Freiburg, Germany
E-mail: dietmar.kroener@mathematik.uni-freiburg.de

Due to the power of ongoing computer systems there are many promising attempts for the numerical simulations of turbulent flows with additionally difficult and challenging features, e.g. flows through scramjets in aerodynamics, flows in meteorology, global circulation and climate models, flows in astrophysics, formation of planets, flows with phase transitions, rain formation in clouds, flows in combustion engines and turbo-machinery and the transport of impurities. The main difficulties for these problems consist in the modeling and resolution of turbulence, complex geometries, 3D, highly nonlinear effects, reactive flows, particle transport in turbulent flows and the demand for high accuracy and high efficiency. The main tools for treating these problems are highly resolved grids, higher order discretizations in space and time (DNS), adaptive grids, the parallelization on supercomputers, i.e. on a considerable amount of processors and dynamical load-balancing. Some of the above mentioned problems are studied in the following contributions.

The main subject of the first contribution (by A. Bosco and B. Reinartz) is the simulation of the flow through the first part of a scramjet, i.e. a hypersonic air-breathing engine, which compresses the incoming flow without a rotary compressor and achieves a supersonic combustion. The underlying mathematical model for these computations consists of the compressible Navier-Stokes system and different turbulence modeling procedures, based on the Reynolds averaged Navier-Stokes equations. The discretization has been done within the QUADFLOW software tool. The grid consists of up to 5.5 million grid points. The convective terms are dicretized by AUSMDV Riemann solvers and for the time integration Runge-Kutta fifth-order explicit schemes and a first order implicit Euler schemes have been used. Grid adaptation is based on a multiscale analysis and data compression, which is the main features of the QUADFLOW package. The code is parallelized on 64 and 128 processors, including dynamical load-balancing. In a preliminary study a two dimensional hypersonic double ramp has been conducted and in a second step a three dimensional situation with the same conditions as applied for an experimental campaign at the DLR. The main features of the flow fields are presented.

In a second part of this paper, the unsteady, quasi incompressible, inviscid flow in the wake of an airplane has been studied. In this case the fluid flow is described by the Euler equations, which are solved again with QUAFLOW. Because of small Mach numbers a preconditioning, based on a dual-time framework, is applied.

In the second contribution (by H. Homann, J. Bec and R. Grauer) the simulations of particles with diameters larger than the Kolmogorov scale , suspended in a turbulent flow, are considered. The applications range from the formation of planets in astrophysics to the rain formation in clouds in geophysics. The focus lies on a direct numerical simulation, in particular on the calculation of single particle trajectories and statistical quantities like

the variance of acceleration or the rotation rate. The density of the particles is assumed to be the same as that one of the surrounding fluid. The underlying mathematical model consists of the incompressible Navier-Stokes equations, in particular the vorticity transport equation in a cubic volume. Periodic boundary conditions are imposed on the boundary of the cubic volume and suitable boundary conditions on the surface of the sphere. The latter ones are realized by a pseudo penalization method. Due to periodic boundary conditions a standard pseudo-spectral method is used. The grid spacing is much smaller than the particles. The time-discretization is given by a strongly stable, low storage Runge-Kutta third order method and the code is parallelized by MPI. One aim of this work is the comparison of the statistical properties of particles, realized by real spheres with those of simple point-particle methods. It is shown that the observed acceleration variance of the particles follows a scaling law in agreement with theoretical predictions.

The paper of N. Peters, L. Wang, J.-P. Mellado, J. H. Göbbert, M. Gauding, P. Schäfer, M. Gampert concerns a new approach to understand the small scale structures of turbulence by decomposing the entire flow field into relatively simple geometrical units, so called dissipation elements. Starting from any point in the computational domain, its gradient trajectory can be determined by tracing along the descending and ascending directions which are normal to the iso-surfaces of the given scalar, until extremal points have been reached. The ensembles of material points, with which trajectories share the same pair of minimal and maximal points, define the spatial regions, named dissipation elements. This decomposition is deterministic, non-arbitrary and space-filling. It turns out that this method can be used to determine the not well-found semi-empirical coefficients of the widely used $k - \varepsilon$ equations in CFD codes for engineering simulations. The underlying mathematical model consists of the incompressible Navier-Stokes equations in 3D space. They are solved by direct numerical simulations on a grid with up to eight billion nodes by FFT on a domain with periodic boundary conditions. The corresponding code is implemented on the basis of an optimized MPI-parallelized FFT library on a system with up to 16000 CPUs. In order to reconstruct the statistics of the flow field a corresponding probability density function is analyzed.

The impact of phase changes and latent heat release on the turbulence and vice versa for moist thermal convection are studied in the paper of T. Weidauer, C. Junghans, O. Pauluis, M. Pütz, J. Schumacher. This is a challenging problem due to the complex nonlinear thermodynamics of phase changes in addition to the turbulent motion. The mathematical model of this flow consists of the incompressible Navier-Stokes equations and two transport-diffusion equations for the variations of the dry and moist buoyancy fields. The momentum equation is completed by a source term which is a nonlinear function of the variations of the dry and moist buoyancy fields. The equations of motion are solved by a parallelized pseudospectral scheme with volumetric fast Fourier transforms (FFT). Time-stepping is done by a second-order Runge-Kutta scheme. Finally the code is used to study the stability properties for different choices of the dry and the moist Rayleigh numbers in the case of equilibrium initial data with a small perturbation, which evolves into a fully developed turbulent state. This is applied to the evolution of clouds.

Study of Compressible and Quasi-Incompressible Flows with QUADFLOW Flow Solver

Arianna Bosco[1] **and Birgit Reinartz**[2]

[1] AICES Graduate School
Schinkelstraße 2, 52062 Aachen, Germany
E-mail: bosco@aices.rwth-aachen.de

[2] CATS
Schinkelstraße 2, 52062 Aachen, Germany
E-mail: reinartz@cats.rwth-aachen.de

The development and improvement of the in-house code QUADFLOW for the study of compressible and quasi-incompressible flows is the main aim of our research group. In the past two years our work has mainly focused on the investigation of a hypersonic scramjet intake, on the study of wake vortices behind the wing of a commercial aircraft and on the general improvement of numerical methods. The complex geometry required for the simulations and the desire to investigate three-dimensional configurations with adequate grid resolution makes it necessary to parallelize the code on a considerable number of processors.

1 Introduction

The QUADFLOW solver was developed at RWTH Aachen University as part of the Collaborative Research Centre SFB 401 and is continuously updated and improved. Currently two main aspects of the code are of central importance: the turbulence models and the adaptive module. The presence of different types of turbulence models allows the study of a wide range of flow fields varying from subsonic flow up to hypersonic flow as illustrated in the following sections. In addition, an adaptive procedure can be applied in order to better capture the flow field structures thus minimizing the total amount of cells needed. As illustrated below, we see that this gives good results, particularly in configurations with a large far field such as aircraft wings.

2 QUADFLOW Solver

QUADFLOW solves the Euler and Navier-Stokes equations around complex aerodynamic configurations. It uses a cell-centred finite volume method on locally refined grids. The grid adaptation is based on a multiscale analysis and data compression similar to techniques used for image compression. This is a new strategy independent of error indicators and error estimators. The fundamental idea is to rewrite an array of cell averages by which the flow field at hand is characterized into a new data format that allows for data compression. This data set consist of cell averages and details. The details can be interpreted as the difference in the solution between two time steps and they become negligible in the regions where the solution is sufficiently smooth. If, on the contrary, the details are larger than a user-defined threshold, the corresponding cell is refined. Further information on this strategy and its analysis as well as computational aspects can be found in Bramkamp et al.[1].

The computational grids are represented by block-structured parametric B-spline patches. In addition, simulation on structured grids can also be performed. For the computation of the convective flux, the flux-difference splitting AUSMDV Riemann Solver[2] is chosen.

As far as the computation of the viscous fluxes is concerned, the gradients of the variables at the cell interfaces are determined using the divergence theorem. The numerical methods employed for the time integration are a Runge-Kutta fifth-order explicit scheme and a first order implicit Euler. Different turbulence models can be chosen in the QUADFLOW solver. These are all RANS (Reynolds-averaged Navier Stokes equations) models. The well-known 1- and 2-equation turbulence models such as Spalart Allmaras[9] or SST k-ω Menter[7] are available as well as a newly implemented 7-equation Reynolds stress model.

3 Hypersonic Scramjet Intake

A scramjet (supersonic combustion ramjet) is a ramjet characterized by supersonic combustion. A ramjet is a hypersonic air-breathing engine which uses forward motion through the air to compress the incoming flow, without a rotary compressor. In order to achieve a supersonic combustion the scramjet has a minimal functional speed, which is within the range of hypersonic velocities.

A scramjet is an highly integrated system: the forebody and the afterbody are integrated with the combustion chamber. No moving parts are necessary for a scramjet and thus the mechanical complexity of the system is low.

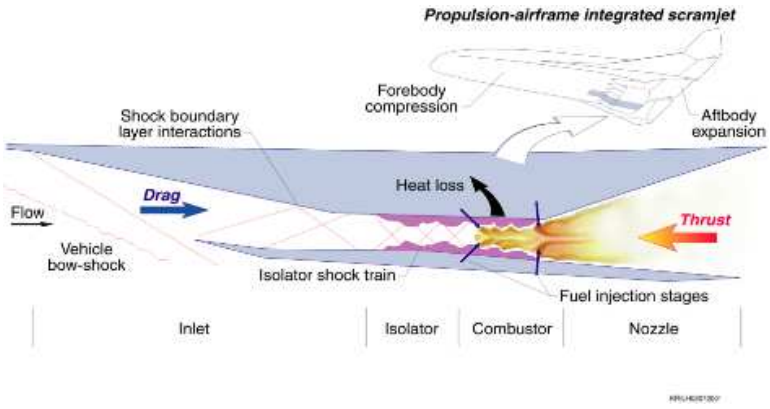


Figure 1. Main features of a scramjet (C.Rogers NASA).

A scramjet consists of four main components as illustrated in Figure 1:

- **inlet**: This consists of a (double) ramp that provides homogeneous high-pressure flow to the engine through a series of oblique shock waves.
- **isolator**: This isolates the pressure-sensitive inlet and allows it to be adapted to the back pressure of the combustor using a shock train.

- **combustor**: This is the area where supersonic combustion takes place. It should provide efficient mixing.
- **nozzle**: This is the area where the gas expands and exits the vehicle producing positive thrust.

In our project we focus our attention on the computation of the inlet and isolator, which together form the so-called scramjet intake. A preliminary two-dimensional study was conducted on a hypersonic double ramp[6]. In a second step, a three-dimensional configuration was studied using the same conditions applied for an experimental campaign at DLR in Cologne[4]. The inflow conditions are shown in Table 1.

free stream Mach number M_∞ [-]	7
free stream pressure p_∞ [Pa]	163
free stream temperature T_∞ [K]	46
free stream density ρ_∞ [km/m^3]	0.012
unit Reynolds number Re_∞ [-]	$4.0 \cdot 10^5$
wall temperature (simulation) $T_w[K]$	300

Table 1. Inflow conditions for scramjet intake.

The geometry under study is constituted by a double ramp intake with a ramp inclination of 9 and 11 degrees, respectively, a rounded leading edge and cowl and side walls to increase the compression of the incoming flow.

A fine grid is required near the solid walls to correctly detect the boundary layer features. For this reason, the computational grid used for this test case containis 5.5 million points. The numerical simulation was run using a Reynolds stress model. This model is known to give better results in flows characterized by high anisotropy, such as the flow under study, due to the 7 turbulent equations that are solved during the computation. The size of the computational grid combined with the high number of equations to be solved (5 for the Navier-Stokes equations and 7 for the turbulence model) make it necessary to parallelize the computation on 64 or 128 processors and to run the computation for at least 100000 iterations in order to obtain good convergence.

The results of the numerical simulation are shown in Figure 2 where the value of the Mach number is presented along the different planes.

In Figure 2 the main features of our flow field can be identified. We see the first ramp shock revealed by a jump in the Mach number along the horizontal planes. At the kink between the two ramps a blue region, characterized by a low Mach number, indicates a boundary layer separation with a recirculation bubble. Further downstream, the flow enters the isolator and a second boundary layer separation is provoked by the impingement of the cowl shock on the lower isolator wall.

4 Wake Vortices

A first assessment is presented to validate the ability of QUADFLOW to simulate the behaviour of the wake of an aircraft. The present study considers, non-stationary, quasi-

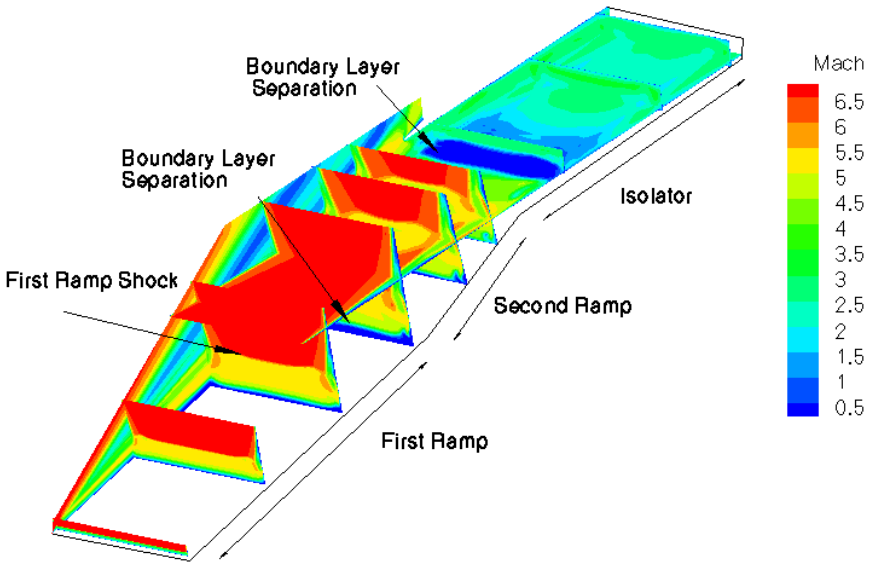


Figure 2. Mach number distribution along a 3-dimensional intake configuration.

incompressible, inviscid fluid flow described by the Euler equations. A velocity field based on the experimental measurements is prescribed as a boundary condition in the inflow plane. The circumferential part of the velocity distribution $v_\Theta(r)$ is described by a Lamb-Oseen vortex according to Eq. 1:

$$v_\Theta(r) = \frac{\Gamma}{2\pi r}\left(1 - e^{-\left(\frac{r}{d_0}\right)^2}\right) \tag{1}$$

The axial velocity component in the inflow direction is set to the constant inflow velocity of the water tunnel. The two parameters of the Lamb-Oseen vortex, circulation Γ and core radius d_0 are chosen in such a way that the model fits the measured velocity field of the wing tip vortex as closely as possible. The radius r is the distance from the centre of a boundary face in the inflow plane to the vortex core.

Instead of water, which is used as the fluid in the experiment, the computation relies on air as fluid. The inflow velocity in the x-direction u_∞ is computed to fulfil the condition that the Reynolds number in the computational test case is the same as in the experiment. The experimental conditions are a flow velocity $u_w = 1.1\frac{m}{s}$ and a Reynolds number $Re_w = 1.9 \cdot 10^5$. With the condition $Re_{air} = Re_w$ the inflow velocity in the x-direction

was computed as $u_\infty = 16.21\,\frac{m}{s}$. For reasons of consistency, the circumferential velocity v_θ was also multiplied by the factor $\frac{u_\infty}{u_w}$. The velocity of the initial solution is set to parallel, uniform flow $u_0 = u_\infty$, $v_0 = w_0 = 0.0$.

The computational domain matches the experimental setup, which extends $l = 6\,m$ in the x-direction, $b = 1.5\,m$ in the y-direction and $h = 1.1\,m$ in the z-direction. The boundaries parallel to the x-direction were modelled as symmetry walls. This domain is discretized by a coarse grid with 40 cells in the flow-direction, and 14 cells in the y-direction and 10 cells in the z-direction, respectively. The number of refinement levels was set to $L = 6$. With this setting, the vortex can be resolved on the finest level by about 80 points in the y-z-plane.

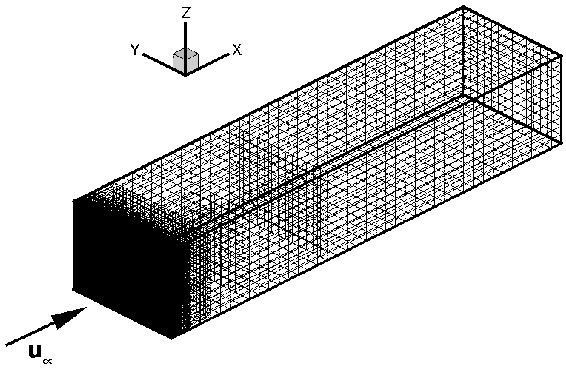


Figure 3. Initial computational grid.

Since QUADFLOW solves the compressible Euler equations, a preconditioning has to be applied for low Mach numbers. This preconditioning is applied in a dual-time framework where only the dual time-derivatives are preconditioned and used for the purposes of numerical discretization and iterative solution, cfr. Ref. 3. The spatial discretization of the convective fluxes is based on the AUSMDV(P) flux vector splitting method[2]. The implicit midpoint rule is applied for time integration. The unsteady residual of the Newton iterations is reduced by about three orders of magnitude in each time step. The physical time step is set to $\Delta t = 5 \cdot 10^{-5}\,s$, which corresponds to a maximum CFL-number of about $CFL_{max} = 28.0$ in the domain. The grid is adapted after each time step. After every 100th time step load balancing is repeated.

To guarantee a sufficiently fine grid to resolve the vortex properly at the start of the computation, the grid on the in-flow plane is refined to the maximum level, see Figure 3. Due to this procedure the first grid contains 384000 cells. When the information at the inlet has travelled through the first cell layer, the forced adaptation of the cells at the inlet is no longer necessary. The grid is subsequently only adapted due to the adaptation criterion based on the multiscale analysis. After 5466 time steps, which corresponds to a computed real time of $t = 0.27\,s$, the grid contains 787000 cells in total. Figure 4 shows six cross sections of the mesh, which are equally spaced in the x-direction with distances $\Delta x = 1.0\,m$. Also presented is the isosurface of the λ_2-criterion with the value $\lambda_2 = -3$. The λ_2 criterion was proposed in Ref. 8 to detect vortices. A negative value of λ_2 identifies

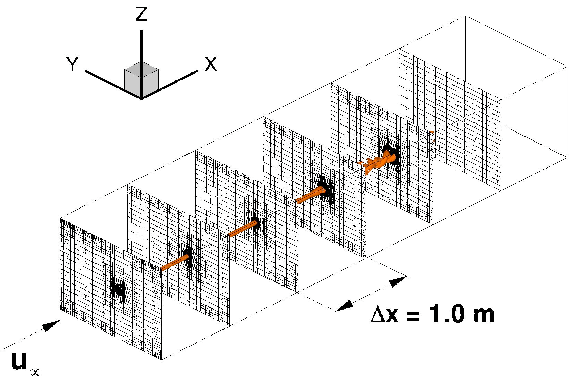


Figure 4. Slices of the computational grid after 5466 time steps at six different positions and the distribution of $\lambda_2 = -3$.

a vortex, whereas the smallest of these negative values marks the core of the vortex. As can be seen from Figure 4, the vortex is transported through the computational domain. The locally adapted grid only exhibits a high level of refinement in the vicinity of the vortex.

5 Conclusion

The computation of flows characterized by a high number of grid points and complex dynamics was achieved by using the QUADFLOW solver. In particular, we focused on hypersonic flows where the thick boundary layer requires a high grid resolution and the study of wake vortices, which needs a large computational domain combined with grid adaptation. It follows that the use of a considerable number of processors on high performing machines is an indispensable requirement for our work.

Acknowledgements

Financial support from the Deutsche Forschungsgemeinschaft (German Research Foundation) through grant GSC 111 and Research Training Group GRK 1095/1 is gratefully acknowledged.

References

1. F. Bramkamp, Ph. Lamby, S. Müller, *An adaptive multiscale finite volume solver for unsteady and steady state flow computations*, Computational Physics **197**, 460-490, 2004.
2. J. Edwards and M. S. Liou, *Low-Diffusion Flux-Splitting Methods for Flows at All Speeds*, AIAA **36**, 1610-1617, 1998.
3. S. A. Pandya, S. Venkataswaran and T. H. Pulliam, *Implementation of Preconditioned Dual-Time Procedures in OVER* , AIAA Paper Vol. **2003-0072**, 2003.

4. J. Häberle and A. Gülhan, *Experimental Investigation of a Two-Dimensional and Three-Dimensional Inlet at Mach 7*, Propulsion and Power **24**, 1023-1034, 2008.
5. B. Eisfeld and O. Brodersen, *Advanced Turbulence Modelling and Stress Analysis for the DLR-F5 Configuration*, AIAA Applied Aerodynamics Conference, Toronto, 6-9 June 2005.
6. A. Bosco, B. Reinartz and S. Müller, *Computation of hypersonic shock boundary layer interaction on a double wedge using a differential Reynolds Stress Model*, International Symposium on Shock Waves 27, St Petersburg, 2009.
7. F. R. Menter *Two-Equation Eddy-Viscosity Turbulence Models for Engineering Applications*, AIAA **32**, 1598-1605, 1994.
8. J. Jeong and F. Hussain, *On the identification of a vortex*, Journal of Fluid Mechanics **285**, 69-94, 1995.
9. P. R. Spalart and S. R. Allmaras, *A One-Equation Turbulence Model for Aerodynamic Flows*, AIAA Paper Vol. 92-04931992.

DNS of Finite-Size Particles in Turbulent Flows

Holger Homann[1,2]**, Jérémie Bec**[2]**, and Rainer Grauer**[1]

[1] Institute for Theoretical Physics I, Ruhr-University Bochum, Germany
E-mail: grauer@tp1.rub.de

[2] Université de Nice-Sophia Antipolis, CNRS
Observatoire de la Côte d'Azur, Laboratoire Cassiopée, Nice, France

We report on direct numerical simulations (DNS) of neutrally buoyant particles suspended in a fully developed turbulent flow. The particles are considered as spheres with a diameter larger than the small scale fluctuations that is the Kolmogorov scale. We adapted a pseudo-penalization technique to this case of moving, non-trivial boundary conditions. This method on the one hand allows for controlling the parameter of the turbulent flow and on the other hand for imposing accurately complex boundary conditions on a regular grid. The underlying method will be presented in detail as well as first results on finite-size effects and comparisons will be drawn to solutions from existing point-particle models.

1 Introduction

Particles suspended in a turbulent flow appear in many environmental and engineering situations. These range from astrophysical problems such as the formation of planets to geophysical questions such as rain formation in clouds. In addition, industrial settings such as material processing involve the transport of impurities by turbulent flows. Inertial particles, having for example a density different from the ambient fluid do not follow exactly the flow but lead to concentrations of particles, called preferential concentrations. If the diameter is much smaller than the Kolmogorov scale these particles can be considered as points and the flow around a sphere is assumed to be a Stokes flow (see Figure 1).

Most of the models rely on this property and a lot of work has been done to explore the dynamical features of small but heavy spheres in turbulence[1–3]. In many cases these

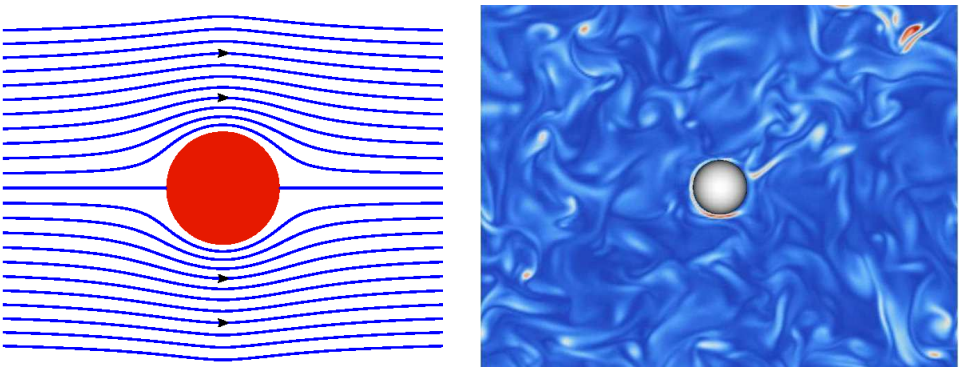


Figure 1. Left: Stokes flow around a particle with diameter much smaller than the Kolmogorov scale. Right: particle that is larger ($r = 20\eta$) than the small scale fluctuations in a fully developed turbulent flow ($R_\lambda = 73$).

particles are assumed to be passive thus do not react back onto the carrier flow. Models for situations where the diameter is of the order of the Kolmogorov scale or larger are often limited to empirical laws for particles in a laminar flow generating a turbulent wake but do not consider the interaction with a fully developed turbulent flow. For the latter case the flow structure is much more complex and cannot anymore assumed to be of Stokes type (see Figure 1).

In this paper we will present DNS of finite-size spheres in hydrodynamic turbulent flows. The focus hereby lies on single particle trajectories as well as statistical quantities such as the variance of acceleration or the rotation rate.

This paper is organized as follows: In the next section the numerical scheme will be presented. Afterwards results will be discussed which deal with time-series of single particles as well as statistical quantities such as probability density functions (PDFs). The last section presents conclusions and perspectives for future works.

2 The Numerical Scheme

In this section we present in detail the strategy we apply to solve the dynamical motion of a finite-size sphere in a turbulent flow. The flow is considered to be incompressible and described by the Navier-Stokes equations with appropriate boundary conditions in a cubic volume with a physical extension of 2π in each direction. These hydrodynamic equations read

$$\partial_t \vec{\omega} = \nabla \times \vec{u} \times \vec{\omega} + \mu \Delta \vec{\omega} + \vec{F}, \qquad \nabla \cdot \vec{u} = 0, \tag{1}$$

$\vec{u}$ denoting the velocity field, related to the vorticity by $\vec{\omega} = \nabla \times \vec{u}$, $\vec{F}$ is an external force and μ is the kinematic viscosity. The right equation of Eq. (1) establishes the condition of incompressibility. The complete boundary of the flow consists of the domain boundaries which are set periodic and of the surface of the sphere.

The periodicity of the domain allows for a standard pseudo-spectral method to solve the Navier-Stokes equations: The underlying equations are treated in Fourier space, while convolutions arising from non-linear terms are computed in real space. In the used code (called LaTu) a Fast-Fourier-Transformation (FFT) is used to switch between these two spaces. The aliasing errors are limited by using a high-exponential cut-off[4,5] which sets less modes to zero than the classic 2/3-rule. The time scheme is a strongly stable, low storage Runge-Kutta of third order. All simulations were done with single floating-point precision.

In order to maintain a statistically steady state we force the flow by prescribing the energy content of the Fourier vectors with moduli 1 and 2. The energy content of each of these two shells is kept constant while the individual amplitudes and phases are evolved piecewise linearly in time between several random configurations separated by a time $10\,T_L$. The advantages of such a forcing are two-fold: it allows one to achieve a statistically isotropic large-scale flow and limits the fluctuations to only approximatively 10% of the mean.

The interprocess communication uses the Message Passing Interface (MPI). More than half of the computational time is spend on the FFTs. Therefore, an efficient implementation deserves attention. The LATU-code is able to use different implementations for the FFT-algorithm. First, it can use the MPI-parallel, portable FFTW library (version 2.1.5

(`http://www.fftw.org`)) which decomposes the entire domain via slab-geometry, depicted in the left part of Figure 2; every CPU processes an ensemble of 2D-slices. This FFT parallelizes very efficiently on mid-size computers, with up to several hundred CPUs. However, the number of used CPUs is limited to the number of grid points in each direction. To overcome this restriction, the LATU-code employs the San-Diego P3DFFT library (`http://www.sdsc.edu/us/resources/p3dfft`) which is a parallel interface to 1D-FFT routines. For these one can use for example the ESSL library or the FFTW3 library. The P3DFFT permits a 2D-decomposition of the computational domain, depicted in the right part of Figure 2. Hence, the limiting restriction concerning the number of CPUs is weakened.

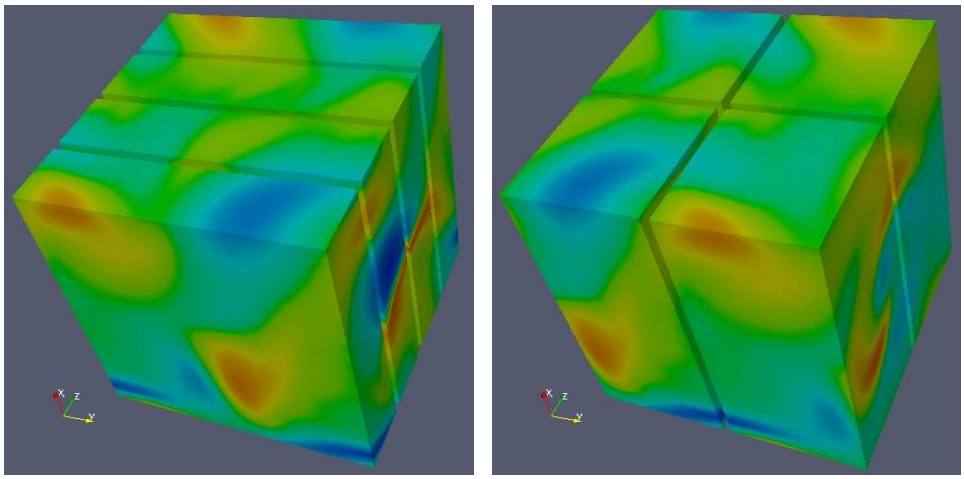


Figure 2. Partitioning of the domain for four processes. Left: slab-geometry (FFTW). Right: 2D-decomposition (P3DFFT).

Now we turn to the numerical scheme for implementing the boundary conditions of a sphere inside the simulation volume. The sphere is supposed to be much larger than the grid-spacing and to react back on the flow thus posing a boundary condition to the Navier-Stokes equations (1). For this we make use of a pseudo-penalization method[6]

$$\frac{\alpha_0}{\tau}\vec{u}^{i+1} - \nu\Delta\vec{u}^{i} + \nabla p^{i+1} = (1-\chi)\left((\vec{u}\cdot\nabla\vec{u})^{i} + \frac{1}{\tau}\sum_{q=1}^{Q}\alpha_q\vec{u}^{i+1-q}\right), \qquad (2)$$

where the α_k depend on the time-scheme. The function χ describes the position of the sphere via $\chi = 0$ for $|\vec{x}| > r$ else $\chi = 1$. Here r denotes the radius of the sphere.

As the considered spheres are allowed to move in the turbulent flow it is appropriate to smooth the function χ. For this we use a volume-fraction scheme which in 1D averages χ over two grid cells thus

$$\bar{\chi} = \frac{1}{2h}\int_{(i-1)h}^{(i+1)h}\chi dx. \qquad (3)$$

The equation for the particle acceleration is given by Newton's equations

$$m_p \ddot{\vec{X}} = \int_S \nabla \cdot T dV = \int_{\partial S} T dS, \qquad T = -pI_3 + \frac{v}{2}\left(\nabla \vec{u} + \nabla \vec{u}^T\right), \tag{4}$$

T denoting the stress-tensor. The force

$$\vec{f_T} = \int_{\partial S} T = \vec{f_p} + \vec{f_\nu} \tag{5}$$

from the stress-tensor can be split up into a force

$$\vec{f_p} = -\int_{\partial S} p\vec{n} dS \tag{6}$$

due to pressure differences over the surface of the sphere and a force

$$\vec{f_\nu} = \nu R^2 \partial_r \int (\vec{u} + u_r \vec{n}) d\Omega \tag{7}$$

due to viscosity and shear at the surface of the sphere. We solve the appearing surface integrals by numerical summation over a homogeneous grid of points on the surface of the sphere (see Figure 3). The grid-spacing is chosen to resolve the Kolmogorov scale. As in general the points of the surface grid do not lie on the cubic grid their value has to be interpolated from the latter. For this we use a tri-cubic scheme which on the one hand provides a high degree of accuracy[7] and on the other hand parallelizes efficiently.

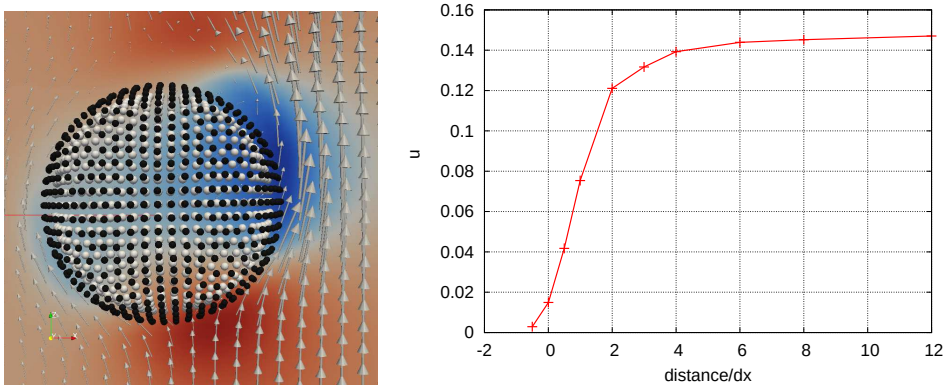


Figure 3. Left: homogeneous grids of points at the surface (white) and slightly above (black) together with the velocity field (arrows) and pressure (colour) in the equatorial plane of the sphere. Right: Rms-velocity in the vicinity of the surface of the sphere.

The spheres are allowed to rotate and the rotation rate satisfies

$$\frac{2}{5} m R \dot{\vec{\Omega}} = \int_{\partial S} \vec{n} \times (T \cdot dS) = \vec{t_\nu} \tag{8}$$

where the torque $\vec{t_\nu}$ is given by

$$\vec{t_\nu} = \nu R^2 \partial_r \int (\vec{n} \times \vec{u}) d\Omega. \tag{9}$$

R_λ	u_{rms}	ε	ν	δt	η	τ_η	L	T_L	N^3
32	0.17	$4.5 \cdot 10^{-3}$	$3 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$5 \cdot 10^{-2}$	0.8	1.2	6.5	512^3
50	0.4	$5 \cdot 10^{-2}$	$3 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$2.8 \cdot 10^{-2}$	0.25	1.3	3.3	128^3
73	0.13	$1 \cdot 10^{-3}$	$8 \cdot 10^{-4}$	$1.5 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$	0.9	2.2	17	256^3

Table 1. Parameters of the numerical simulations. $R_\lambda = \sqrt{15 u_{\mathrm{rms}} L/\nu}$: Taylor-Reynolds number, u_{rms}: root-mean-square velocity, ε: mean kinetic energy dissipation rate, ν: kinematic viscosity, δt: time step, $\eta = (\nu^3/\varepsilon)^{1/4}$: Kolmogorov dissipation length scale, $\tau_\eta = (\nu/\varepsilon)^{1/2}$: Kolmogorov time scale, $L = (2/3E)^{3/2}/\varepsilon$: integral scale, $T_L = L/u_{\mathrm{rms}}$: large-eddy turnover time, N^3: number of collocation points.

The derivatives in Eq. (7) and (9) are computed by means of a second mesh located at one grid-spacing distance of the surface (see left part of Figure 3) which is inside the linear viscous sub-layer (see right part of Figure 3).

3 Results

In this section we are going to present results from different simulations. The aim is a better understanding of the dynamics of spheres with diameter larger than the Kolmogorov scale in fully developed turbulence. As we want to avoid interactions between different particles each simulation contains only one sphere of a given size. All particles are neutrally buoyant, i.e. the mass density of the particle ρ_p equals the mass density of the fluid ρ_f. The main parameters of the simulations are given in Table 1.

We begin with time-series of dynamical quantities of the particles. In the left part of Figure 4 the rms-acceleration is given for two different sizes. Both functions show a chaotic behaviour as expected in a turbulent flow. The smaller sphere has a larger acceleration than the bigger sphere, which is on the one hand due to the differing masses and on the other hand due to the finite correlation length of pressure and velocity. The right part of Figure 4 shows time-series of the rotation rate from the same simulations. The rotation rate of the smaller particles is larger than that of the bigger particle. The explanation of these findings is analogous to that of the acceleration.

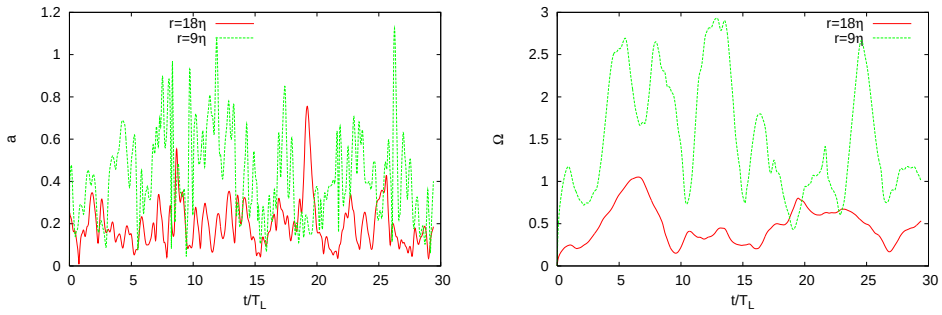


Figure 4. Time-series of particles with different sizes ($R_\lambda = 50$). Left: rms-acceleration. Right: rms-rotation-rate.

The next quantities we are going to consider are of statistical nature. It is clear that one needs long runs in order to gather enough data to extract reliable statistical informations. Depending on the resolution and size of the particle we performed simulations up to 400 large-eddy-turnover times (LET) per simulation. Here we begin with acceleration probability density functions (PDFs) corresponding to the time-series discussed in the previous paragraph. The left part of Figure 5 shows these PDFs for four different particle sizes. The smaller the size the larger are the characteristic tails. Once normalized to unit-variance we obtain a collapse of the PDFs as observed in experiments[8].

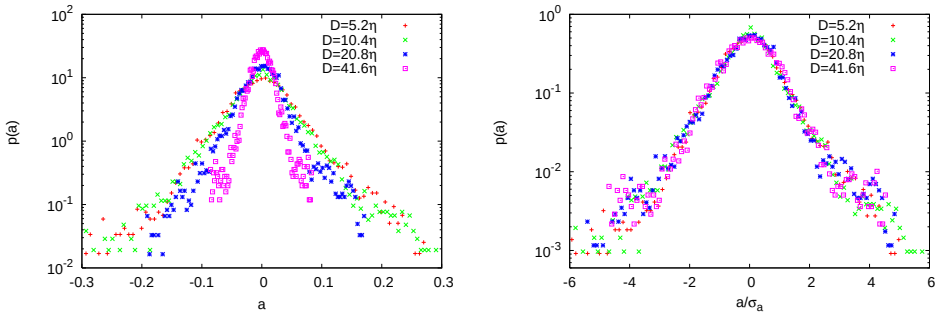


Figure 5. PDF of acceleration of particles with different diameter ($R_\lambda = 73$). Left: not-normalized. Right: normalized to unit variance.

Because it is computationally demanding to directly solve the boundary layer and therefore the dynamics of a finite-size sphere, it is worth looking for models. Simple models neglect the back-reaction of the particles onto the flow and try to integrate the particle trajectory from local values of the flow (usually at the centre of the particle).

One of the aims of this work is to compare the statistical properties of real spheres with those of a simple point-particle model. Newton's equation for the most simple model reads

$$\frac{d\vec{v}}{dt} = \frac{D\vec{u}}{Dt} + \frac{3\nu}{r^2}(\vec{u} - \vec{v}), \tag{10}$$

v denoting the particle velocity and r its radius. It only takes into account the fluid-acceleration (first term right hand side) and the viscous-drag (second term on right hand side)[9]. We performed simulations where a statically stationary flow is seeded with particles described by this equation. Furthermore simulations with the same flow-parameters were done for spheres computed with the pseudo-penalization method.

The left part of Figure 6 shows the normalized acceleration variance $a_0 = \langle(\mathrm{d}v_\mathrm{p}^i/\mathrm{d}t)^2\rangle\epsilon^{-4/3}\eta^{2/3}$ for both the model as well as for the real spheres. The accuracy of the presented pseudo-penalization method is confirmed by the approach of the latter to the tracer value. The acceleration decreases with size as was already discussed in previous paragraphs. For diameters $d > 4\eta$ one observes an inertial-range behaviour with $a_0 \propto d_\mathrm{p}^{-4/3}$. It is in agreement with the observation that pressure scaling is dominated by sweeping at moderate Reynolds numbers[3].

In contrast to the curve of the spheres the acceleration variance of the model-particles is approximately constant with a slight bump after $D = 12\eta$. Therefore important features

of the dynamics of finite-size spheres are not correctly represented by the point-particles. Further terms in Eq. (10) such as the Faxén-corrections have to be taken into account. It is questionable whether one can in general model the motion of a finite size particle by a point-model. The fluid velocity at the centre is generally a key-ingredient of such models, which is not well defined for spheres larger than the Kolmogorov scale.

The right part of Figure 6 shows the rotation rate as a function of the particle diameter. One observes a decrease of the variance of the rotation rate in the non-smooth scales of the flow. The straight line indicates a quadratic decrease.

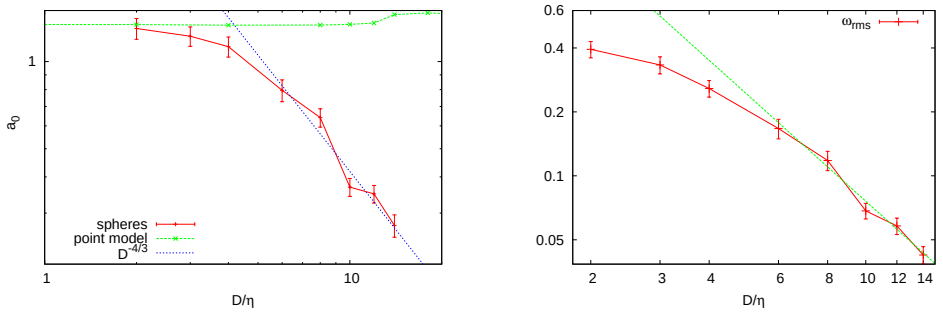


Figure 6. ($R_\lambda = 32$) Left: Acceleration variance for spheres computed with the pseudo-penalization method and with a point model. Right: Variance of the rotation rate for the penalty-spheres.

4 Conclusions and Perspectives

In this paper we report on an application of a dynamical pseudo-penalization technique on the dynamics of finite-size spheres in fully developed turbulent flows. This method provides the opportunity to control the turbulent flow and impose complex boundary conditions given by moving particles. It shows good scaling properties on massive parallel computers and has been used on the Blue Gene/P system JUGENE at the NIC, Jülich, Germany. The penalization technique is shown to accurately reproduce the tracer dynamics in the limit of small diameter. By means of well resolved direct numerical simulations it is observed that the acceleration variance of neutrally buoyant spheres with radii larger than four Kolmogorov scales follows a scaling law in agreement with dimensional predictions.

By means of this method a lot of important, unresolved problems can be tackled. We mention the dynamics of particles with densities different from the surrounding flow, the influence of non-spherical shapes and particle-particle interactions in turbulence.

Acknowledgements

This study benefited from fruitful discussions with M. Bourgoin, E. Calzavarini, R. Pasquetti, and Y. Ponty. This research was supported by the Agence Nationale de la Recherche under grant No. BLAN07-1_192604. During his stay in Nice, H. Homann benefited from a

grant of the DAAD. During his stay in Bochum he benefited from partial support through DFG-FOR 1048. Access to the IBM Blue Gene/P computer JUGENE at the FZ Jülich was made available through project HBO22. Part of the computations were performed on the "mésocentre de calcul SIGAMM" and using HPC resources from GENCI-IDRIS (Grant 2009-i2009026174).

References

1. K. D. Squires and J. K. Eaton, *Preferential concentration of particles by turbulence*, Phys. Fluids A, **3**, 1169–1178, 1991.
2. E. Balkovsky, G. Falkovich, and A. Fouxon, *Intermittent distribution of inertial particles in turbulent flows*, Phys. Rev. Lett., **86**, 2790–2793, 2001.
3. J. Bec, L. Biferale, M. Cencini, A. S. Lanotte, S. Musacchio, and F. Toschi, *Heavy particle concentration in turbulence at dissipative and inertial scales*, Phys. Rev. Lett., **98**, 84502, 2007.
4. T. Y. Hou and R. Li, *Computing Nearly Singular Solutions Using Pseudo-Spectral Methods*, J. Comp. Phys., **226**, 379 – 397, 2007.
5. T. Grafke, H. Homann, J. Dreher, and R. Grauer, *Numerical simulations of possible finite time singularities in the incompressible Euler equations: comparison of numerical methods*, Physica D, **237**, 1932–1936, 2008.
6. R. Pasquetti, R. Bwemba, and L. Cousin, *A pseudo-penalization method for high Reynolds number unsteady flows*, Appl. Numer. Math., **58**, no. 7, 946–954, 2008.
7. H. Homann, J. Dreher, and R. Grauer, *Impact of the floating-point precision and interpolation scheme on the results of DNS of turbulence by pseudo-spectral codes*, Comput. Phys. Comm., **177**, 560–565, 2007.
8. Nauman M. Qureshi, Mickaël Bourgoin, Christophe Baudet, Alain Cartellier, and Yves Gagne, *Turbulent transport of material particles: an experimental study of finite size effects*, Physical Review Letters, **99**, no. 18, 184502, Nov. 2007, PMID: 17995412.
9. M. R. Maxey and J. Riley, *Equation of motion for a small rigid sphere in a nonuniform flow*, Phys. Fluids, **26**, no. 4, 883, 1983.

Geometrical Properties of Small Scale Turbulence

Norbert Peters, Lipo Wang, Juan-Pedro Mellado, Jens Henrik Göbbert, Michael Gauding, Philip Schäfer, and Markus Gampert

Institute for Combustion Technology, RWTH Aachen University
52056 Aachen, Germany
E-mail: {n.peters, l.wang, j.p.mellado, j.h.goebbert, p.schaefer, m.gauding}@itv.rwth-aachen.de

In order to extract small-scale statistical information on turblulent flows from passive scalar fields obtained by direct numerical simulation (DNS), a new method of analysis is introduced. It consists of determining local minimum and maximum points of the fluctuating scalar field via gradient trajectories starting from every grid point in the directions of ascending and descending scalar gradients. The ensemble of grid cells from which the same pair of extremal points is reached determines a spatial region which is called dissipation element. Dissipation elements are highly convoluted and have on average an elongated shape. Their diameter is nearly constant and of the order of a few Kolmogorov scales while the length is highly variable and has a mean of the Taylor length. Since dissipation elements are spacefilling, it is possible to reproduce the overall statistics of the turbulent fields based on the comparably simple statistics of the elements.

1 Introduction

Due to their nonlinear and nonlocal nature turbulent flows are highly complex systems. In contrast to other subjects, there is a strong need to study fundamental turbulence, of which our understanding is still at an infancy stage. Although an unresolved classical problem in physics, turbulent flows occur in many applications not only in aerodynamics, meteorology, geo- and astrophysics, but also in internal flows in combustion engines and turbo-machinery. Extensive research has been carried out for more than one century, nonetheless the understanding of turbulence is still far from being satisfactory and many aspects remain to be explained.

Especially the behaviour of small scale structures of turbulence is still not well understood. Galilei realized in the 17th century, “I had less difficulty in the discovery of the motion of heavenly bodies in spite of their astonishing distances, than in the investigation of the movement of flowing water before our very eyes.”

A new approach can help understand these small scale structures of turbulence by decomposing the entire flow field into relatively simple geometrical units, so called dissipation elements (Figure 1). Understanding the behaviour and properties of dissipation elements leads to a better understanding of turbulence itself.

One of the most important improvements for the research of turbulence comes from the use of Direct Numerical Simulation (DNS) on supercomputers like JUGENE, an IBM Blue Gene/P system. Only the use of these supercomputers made it possible in recent years to resolve the small scale structures of turbulence.

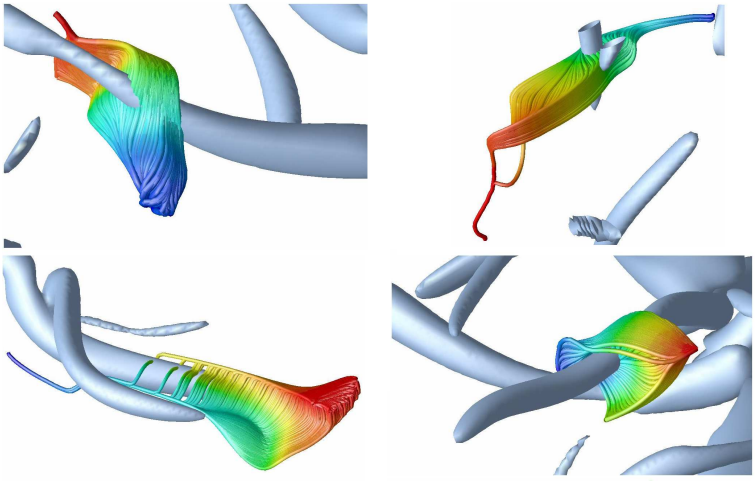


Figure 1. Examples of dissipation elements interacting with vortex tubes.

2 Direct Numerical Simulations

Direct Numerical Simulation (DNS) is of indispensable importance in the study of turbulence. Compared with experiments, DNS has the advantage of providing a complete set of three-dimensional data with high spatial resolution.

Incompressible flows are described by the Navier-Stokes equations:
Continuity

$$\nabla \cdot U = 0, \tag{1}$$

Momentum

$$\frac{\partial U}{\partial t} + U \cdot \nabla U = -\frac{1}{\rho}\nabla p + \nu \nabla^2 U, \tag{2}$$

where U, ρ, p and ν denote the velocity field, the fluid density, the pressure field and the kinematic viscosity, respectively. By nondimensionalizing the Navier-Stokes equations with a reference velocity U_0 and a reference length L, one obtains the Reynolds number Re_L, defined as $Re_L = \frac{U_0 \cdot L}{\nu}$, as the only remaining parameter in Eq. 2. Thus the Reynolds number Re_L, which indicates the ratio of the inertial to the viscous forces on L, is the most fundamental nondimensional parameter to characterize flows. Mathematically, the Navier-Stokes equations with a given set of initial and boundary conditions are deterministic in describing flows. However, there exists a critical value of the Reynolds number, above which the flow may transit from laminar to turbulent. Consequently, completely new types of motion and flow structures will appear in the whole flow field. Once the turbulent motion is introduced, further descriptions such as statistics are needed to investigate the dynamics of the flow.

Turbulence is a multiscale phenomina[10]. The integral scale L describes the size of the largest structures and in contrast the Kolmogorov scale η, which is of the order of the size

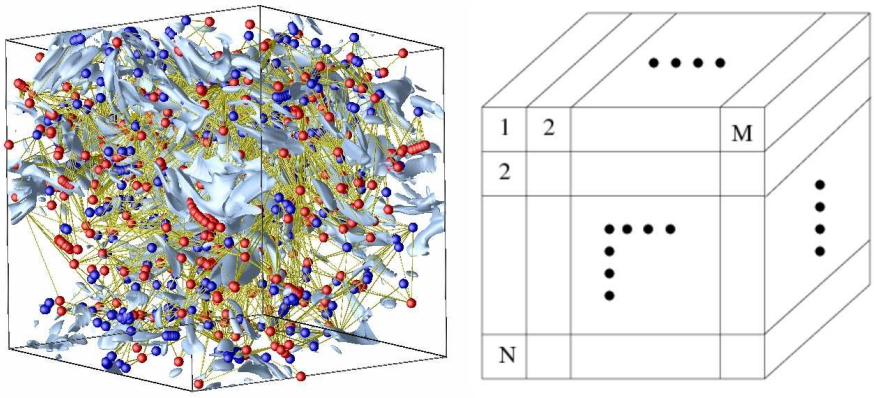


Figure 2. Direct Numerical Simulation (DNS) of a turbulent flow showing strain sheets and extremal points and its data decomposition on a supercomputer with N*M processors.

of the smallest structures of the flow. To fully resolve all properties of a turbulent flow, η must be proportional to the grid size Δx in the calculation domain. Therefore $\frac{L}{\eta} \propto Re_L^{3/4}$ leads to the total number of grid points needed, which is $N^3 \propto Re_L^{9/4}$.

Generally it is believed that the spatial resolution for DNS is about $\Delta x \propto \mathcal{O}(\eta)$. Practically depending on different calculations, this criterion can vary to some extent. To study the overall flow parameters, like the mean kinetic energy and energy dissipation, 2η is a reasonable choice to determine Δx and therefore the grid resolution[9]. These resolutions are adequate to obtain first and second-order statistics. Therefore it is claimed that demanding η to be resolved as the smallest scale is probably too stringent[8]. But to investigate higher order derivatives or special properties, the above criteria are too loose. More strictly, Sreenivasan[6] pointed out that due to intermittency, scales which are much smaller than η can locally occur in turbulence. To also resolve these local small scales, or spots of violent events, in a strict sense, the grid size must be determined in a more stringent way.

To analyze the structures of dissipation elements, the resolution criterion is even more stringent. As shown in Figure 1, the geometrical structure of dissipation elements is very complicated in 3D space. The number of dissipation elements in a flow field is determined by the number of extremal points. It has been found that quite frequently extremal points will cluster together with a distance of the order of η. To resolve extremal points in clusters in a sufficient way, Δx must be smaller than η.

3 Dissipation Element Analysis

"Is geometry 'naturally' identifiable in turbulent flows[4]?" In the past years many efforts have been devoted to study flow geometries (i.e. vortex tubes or critical points). However, due to the difficulty of clearly defining shapes and scales in turbulence, there was no success to obtain a determined and complete partition of the entire flow fields. One needs to construct a suitable method, which can identify specific geometrical elements in turbulent flows. Which method to choose is by no means evident. Dissipation element analysis is

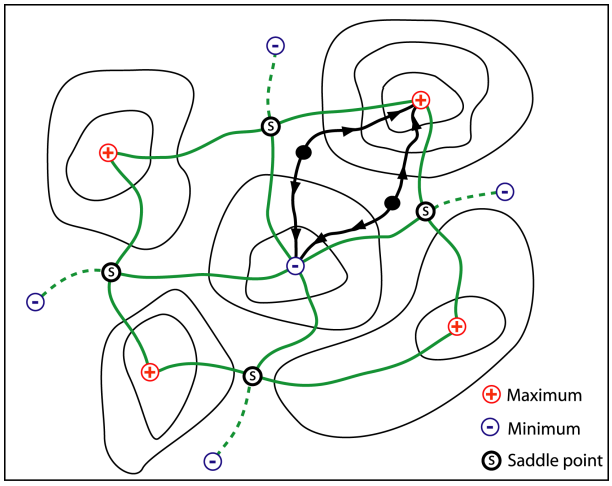


Figure 3. Schematic sketch of a two-dimensional scalar field including the trajectories from an initial point to the minimum and maximum points. The solid lines are geodetic lines bounding four repeating units corresponding to dissipation elements, while the dashed lines are geodetic beyond these units.

a new idea to decompose the entire flow fields into relatively simple geometrical units according to the distribution of some predefined scalars[1]. For any turbulent scalar variable which is diffusion-controlled at small scales, in most parts of the flow field locally this scalar varies monotonously, except for the regions around zero gradient points. Starting from any grid point, its gradient trajectory can be determined by tracing along the descending and ascending directions which are normal to the iso-surfaces of the given scalar, until extremal points have been reached (Figure 3). The ensembles of material points, with which trajectories share the same pair of minimal and maximal points, define the spatial regions named dissipation elements. This decomposition is deterministic, non-arbitrary and space-filling. Some examples of dissipation elements from numerical simulations are shown in Figure 1. In this figure the colouring shows the scalar value along trajectories, red at the maximum point and blue at the minumum point. Dissipation elements can be characterized by two parameters: the linear length l between maximum and minimum point and the scalar difference $\Delta\phi$. It has been shown[1,2] that the mean length of dissipation elements l_m is of the order of the Taylor scale λ. More details of the physical properties of dissipation elements have been discussed in Ref. 1, 2. For example, the trajectories of the gradient field are intrinsic coordinates which are imposed by the flow field in contrast to an arbitrary cartesian coordinate system. It is possible to calculate statistics of the turbulent field from dissipation element analysis, which exhibits linear scaling over a large range of scales. These statistics can be found and confirmed by classical turbulence theory.

An important result of dissipation element analysis is the joint probability density function (PDF) of the linear lenght l and the scalar difference $\Delta\phi$, depicted in Figure 4. This function contains most of the information required to predicatively reconstruct the statistics of the scalar field. Based on highly resolved direct numerical simulations an equation was derived and validated[1] in order to model the joint probability density function.

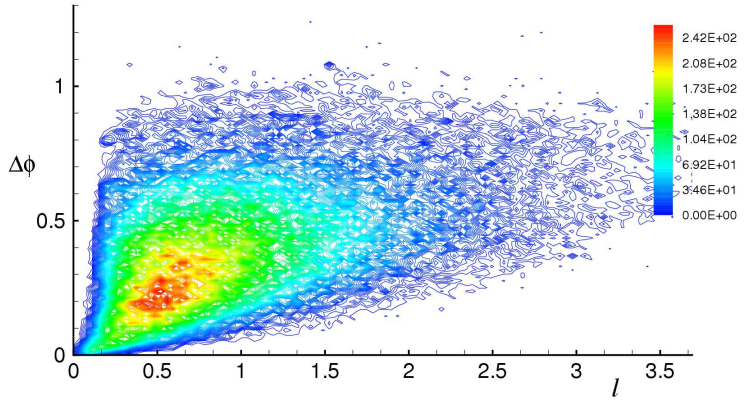


Figure 4. A representative joint propability density function (PDF) from direct numerical simulation (DNS).

4 Numerics

In order to reduce the numerical diffusion and resolve structures η, temporal and spatial difference schemes with high-order accuracy are required. Spectral methods using fast fourier transformation (FFT) are applied due to their numerical accuracy. To apply FFT, the transformed Navier-Stokes equations need to be solved in a moving coordinate with periodic boundary conditions. For better numerical stability under the control of mass conservation, the skew-symmetric formulation of the convection terms[7] and a third-order Runge-Kutta time discretization are used.

Most of the CPU time of our DNS code is consumed by the frequently invoked FFT subroutines. An optimised MPI-parallelized FFT library is very important to utilize the hardware resource effectively. We have adopted the dubbed P3DFFT[11], a library for large-scale computer simulations in a wide range of sciences. The scalability of P3DFFT with the number of CPUs plays a key role in the overall performance.

The real running of the DNS code, with a grid of 2048^3=8 billion nodes, does show that the iteration time is almost inversely proportional to the CPU number up to 16384 CPUs. The calculation consumes 1000 GByte memory and creates an output file for a single timestep with the size of 400 GByte. The iterative solver takes three minutes for one cycle, which gives a total calculation wall-clock time of at least 10 days for the first reasonable results.

5 Applications

In recent work[3] a new approach to partition shear layer flows has been developed, which is an important progess in understanding the geometrical structures in an unbiased way. Based on our method of gradient trajectories, the turbulent/non-turbulent interface can be defined non-arbitrarily. Figure 5 shows the interaction of scalar gradient trajectories with a scalar isosurface in the interface zone.

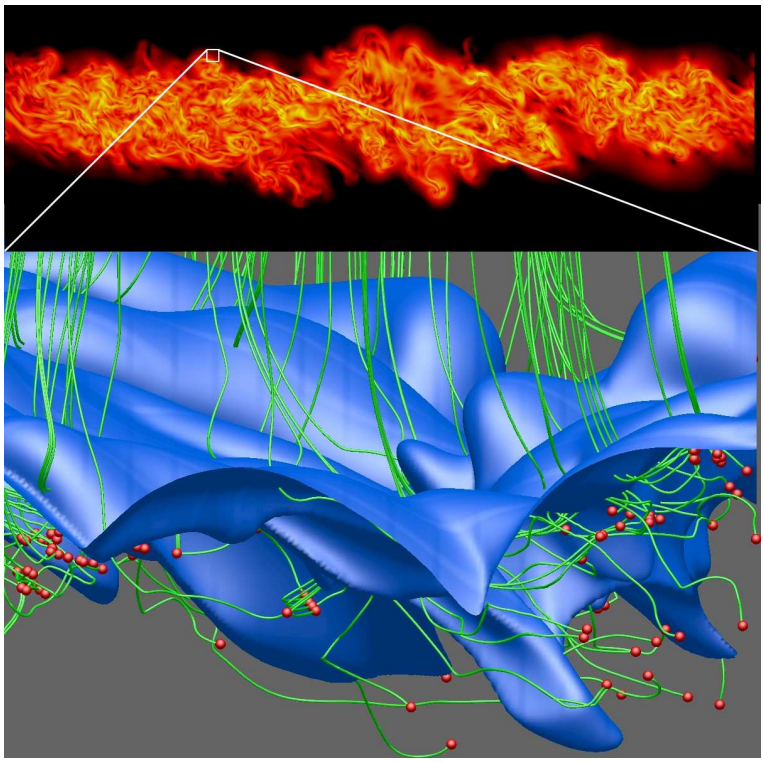


Figure 5. The interaction of a scalar isosurface with gradient trajectories in shear layer flows.

Furthermore, the not well-founded semi-empirical coefficients of the widely used k-ϵ-equation in CFD codes for engineering simulations could be reconstructed[2] using dissipation element analysis.

As this model is widly used for industrial simulations it serves as an important link between fundamental research and engineering applications. The results of the currently employed models are far from being satisfactory. Therefore a combination of theory and its testing in highly resolved DNS can help to improve modeling accuracy[5].

6 Concluding Remarks

Fundamental theoretical research in turbulence helps us to understand the physical basis of this complex phenomenon. From an engineering view point, turbulence modeling serves as the most important linkage between fundamental research and applications. The available turbulent models, for instance the k-ϵ-model and subgrid models of large eddy simulation (LES), may work for specific cases, but generally are far from being satisfactory. However, to construct a more accurate and wider valid model, both theoretical studies as well as numerical simulations which are only realizable via supercomputing are of vital importance for the investigation of the Reynolds number dependency.

Acknowledgements

This project has been continuously sponsored by NIC at the Research Centre Jülich. The authors greatly appreciate the necessary support. We also want to thank Dmitry Pekurovsky for the P3DFFT library[11], which is an important part of our DNS code. Furthermore, the funding from Deutsche Forschungsgemeinschaft under Grant Pe 241/38-1 and Cluster of Excellence "Tailor-Made Fuels from Biomass" are greatly acknowledged.

References

1. L. Wang, N. Peters, *The length-scale distribution function of the distance between extremal points in passive scalar turbulence*, J. Fluid Mech. **554**, 457-475, 2006.
2. L. Wang, N. Peters, *Length scale distribution functions and conditional means for various fields in turbulence*, J. Fluid Mech. **608**, 113-138, 2008.
3. J. Mellado, L. Wang, and N. Peters, *Gradient trajectory analysis of a scalar field with external intermittency*, J. Fluid Mech. **626**, 333-365, 2009.
4. S. Corrsin, *Random geometric problems suggested by turbulence*, Lecture Notes in Physics **12**, 300-316, 1972.
5. P. Schaefer, M. Gampert, J. H. Goebbert, L. Wang and N. Peters, *A-priori and a-posteriori testing of different model equations for the mean dissipation using Direct Numerical Simulations of Kolmogorov flows*, submitted to Flow, Turbulence and Combustion.
6. K. R. Sreenivasan, J. Schumacher, V. Yakhot, *Intermittency and Direct Numerical Simulations*, (58th Annual Meeting of the Division of Fluid Dynamics, APS, Chicago, USA, 2005).
7. W. J. Feiereisen, J. H. Ferziger et al., *Numerical simulation of a compressible, homogeneous, turbulent shear flow*, (Department of Mechanical Engineering, Stanford Univ. TF-13, 1981).
8. P. Moin, K. Mahesh, *Direct Numerical Simulation: A Tool in Turbulence Research*, Annual Review of Fluid Mechanics **30**, 539-578, 1998.
9. S. B. Pope, *Turbulent flows*, (Cambridge University Press, 2000).
10. U. Frisch, *Turbulence*, (Cambridge University Press, 1995).
11. D. Pekurovsky, *Parallel Three-Dimensional Fast Fourier Transforms*, `http://www.sdsc.edu/us/resources/p3dfft`.

Shallow Moist Convection

T. Weidauer[1], C. Junghans[2], O. Pauluis[3], M. Pütz[2], and J. Schumacher[1]

[1] Institute of Thermodynamics and Fluid Mechanics, Technische Universität Ilmenau
D-98684 Ilmenau, Germany
E-mail: joerg.schumacher@tu-ilmenau.de

[2] Deep Computing – Strategic Growth Business, IBM Deutschland GmbH
D-55131 Mainz, Germany

[3] Center for Atmosphere Ocean Science, New York University
New York City, NY 10012-1185, USA

Convective turbulence with phase changes and latent release is an important dynamical process in the atmosphere of the Earth which causes, e.g. the formation of clouds. Here we study moist convection in a simplified setting – shallow and nonprecipitating moist Rayleigh-Bénard convection with a piecewise linear thermodynamics on both sides of the phase boundary. The presented model is a first nontrivial extension of the classical dry Rayleigh-Bénard convection. The equations of motion and linear stability studies of equilibria are discussed and supercomputations of the fully developed turbulent dynamics in very flat Cartesian cells are presented.

1 Introduction

Moist thermal convection combines turbulent convection with phase changes and latent heat release. It is ubiquitous throughout the atmosphere of the Earth[8]. When a parcel of air rises in convective motion, it expands adiabatically. As a consequence, its temperature and pressure drop and at some point during its ascent the air parcel becomes saturated. The water condensation sets in and clouds are formed. The large range of temporal and spatial scales in the convective turbulent motion causes big variations in the lifetime and shape of clouds, such as shallow stratocumulus or isolated cumulus clouds, and thus in the cloud cover and radiation budget. Cloud parametrizations remain therefore one of the big uncertainty factors for global circulation and climate models[8].

Despite its enormous importance, the small-scale structure and statistics of moist convective turbulence has been studied relatively little compared to its dry convection counterpart. The reason for this gap is that turbulent convection in moist air includes the complex nonlinear thermodynamics of phase changes in addition to the turbulent motion[2,14]. Phase changes lead to discontinuities of the partial derivatives in the equation of state at the saturation point[6]. The associated latent heat release in the bulk provides a rapidly changing local source of buoyant motion, in addition to the buoyancy (or heat) flux from the bottom to the top planes as already present in dry convection. Significant progress in understanding the global and local mechanisms of turbulent heat transfer in dry convection has been made in the last decade (for a comprehensive review see Ref. 1).

In this work, we aim at transfering the numerical analysis concepts from the well-investigated dry convection case[1,7,11,12] to the less-explored moist convection case. We make a first step by considering moist Rayleigh-Bénard convection with a linearized thermodynamics of phase changes[9]. The model is a straightforward extension of the well-known dry Rayleigh-Bénard convection case in the Boussinesq approximation[11,12]. It is

also an extension of a moist convection model which was discussed originally by Bretherton[4,5] for the onset of linear instability and weakly nonlinear regime. Here, we conduct direct numerical simulations of the turbulent nonlinear stage of moist convection and discuss the computational challenges for studies in flat cells[10].

2 Moist Boussinesq Model

2.1 Thermodynamic Equilibrium Approximation for a Shallow Layer

The buoyancy B in atmospheric convection is given by[6]

$$B = -g\frac{\rho(S, q_v, q_l, q_i, p) - \overline{\rho}}{\overline{\rho}}\,, \tag{1}$$

with g being the gravity acceleration, $\overline{\rho}$ a mean density, p the pressure, S the entropy and q_v, q_l, q_i the mixing ratios of water vapor, liquid water and ice. So-called warm clouds result in $q_i = 0$. If we assume local thermodynamic equilibrium, which means for example that precipitation is absent, the two remaining mixing ratios are combined to the total water mixing ratio, $q_T = q_v + q_l$. Thus dependencies are reduced to $B(p, S, q_T)$. The Boussinesq approximation considers pressure variations about a mean hydrostatic profile and thus one is left with a function $B(S, q_T, z)$. However, this functional dependence is still highly nonlinear and contains the full thermodynamics of phase changes. The next step is to approximate B as a piecewise linear function of S and q_T for both phases (vapor and liquid) close to the phase boundary. This step preserves the main new physical ingredient, the discontinuity of partial derivatives (e.g. the specific heat) at the phase boundary, but allows for an explicit determination of whether the air is locally saturated or not. Since B becomes a linear function of S and q_T, we can introduce two new prognostic buoyancy fields, a *dry buoyancy field* D and a *moist buoyancy field* M, which substitute S and q_T. Consequently, the buoyancy $B(M, D, z)$ is a linear function of M and D and is given at each space-time-point by

$$B = \max\left(M, D - N_s^2 z\right)\,, \tag{2}$$

where N_s is the Brunt-Vaisala frequency. An air parcel at position $\mathbf{x}$ at time t is unsaturated if $M(\mathbf{x}, t) < D(\mathbf{x}, t) - N_s^2 z$; it is saturated if $M(\mathbf{x}, t) > D(\mathbf{x}, t) - N_s^2 z$.

2.2 Equations of Motion and Dimensionless Parameters of the Model

The dry and moist buoyancy fields can be decomposed in a linear mean and variations about the mean

$$D(\mathbf{x}, t) = \overline{D}(z) + D'(\mathbf{x}, t) = D_0 + \frac{D_H - D_0}{H} z + D'(\mathbf{x}, t) \tag{3}$$

$$M(\mathbf{x}, t) = \overline{M}(z) + M'(\mathbf{x}, t) = M_0 + \frac{M_H - M_0}{H} z + M'(\mathbf{x}, t)\,. \tag{4}$$

The variations about the mean profiles of both fields have to vanish at $z = 0$ and H, which imposes the boundary conditions $D' = 0$ and $M' = 0$. Eq. (2) can now be transformed into $B = \overline{M}(z) + \max\left(M', D' + \overline{D}(z) - \overline{M}(z) - N_s^2 z\right)$. Note that the first term on the right-hand side is horizontally uniform. This implies that it can be balanced by a horizontally

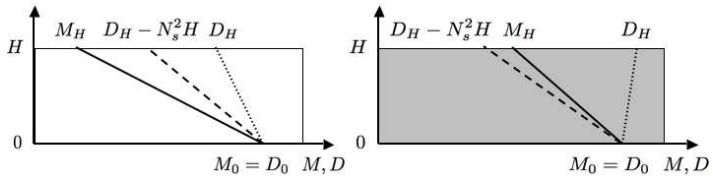


Figure 1. Left: Completely unsaturated intital equilibrium condition with $\overline{M}(z) < \overline{D}(z) - N_s^2 z$. Both buoyancy fields are linearly unstable for sufficiently high Rayleigh numbers. Right: Conditionally unstable equilibrium case. The dry buoyancy field is linearly stable. The whole slab is saturated and $\overline{M}(z) > \overline{D}(z) - N_s^2 z$. D_0, M_0, D_H and M_H are prescribed values of both fields at $z = 0$ and H.

uniform pressure field given by $p(z) = -M_0 z - [(M_H - M_0)/(2H)]z^2$. We can thus remove the mean contribution from the buoyancy field without any loss of generality.

A dimensionless version of the equations of motion is obtained by defining the characteristic quantities. These are the height of the layer H, the free-fall velocity $U_f = \sqrt{H(M_0 - M_H)}$, the time $T_f = H/U_f$, the characteristic (kinematic) pressure U_f^2, and the characteristic buoyancy difference $M_0 - M_H$. The Boussinesq equations are given by

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\nabla p + \sqrt{\frac{Pr}{Ra_M}} \nabla^2 \mathbf{u} + B(M, D, z)\mathbf{e}_z \tag{5}$$

$$\nabla \cdot \mathbf{u} = 0 \tag{6}$$

$$\frac{\partial D'}{\partial t} + (\mathbf{u} \cdot \nabla)D' = \frac{1}{\sqrt{PrRa_M}} \nabla^2 D' + \frac{Ra_D}{Ra_M} u_z \tag{7}$$

$$\frac{\partial M'}{\partial t} + (\mathbf{u} \cdot \nabla)M' = \frac{1}{\sqrt{PrRa_M}} \nabla^2 M' + u_z \tag{8}$$

We identify three non-dimensional parameters: the Prandtl number, Pr, the dry and the moist Rayleigh numbers, Ra_D and Ra_M, which are defined as

$$Pr = \frac{\nu}{\kappa}, \qquad Ra_D = \frac{H^3(D_0 - D_H)}{\nu\kappa}, \qquad Ra_M = \frac{H^3(M_0 - M_H)}{\nu\kappa}\,. \tag{9}$$

The kinematic viscosity is ν and the diffusivity of the buoyancy fields is κ. Two more parameters are hidden implicitly within the definition (2) of the buoyancy B which is given in dimensionless form by

$$B = \max\left(M', D' + SSD + \left(1 - \frac{Ra_D}{Ra_M}\right) z - CSAz\right)\,. \tag{10}$$

The so-called *Surface Saturation Deficit* SSD and the *Condensation in Saturated Ascent* CSA are defined as

$$SSD = \frac{D_0 - M_0}{M_0 - M_H}\,, \qquad CSA = \frac{N_s^2 H}{M_0 - M_H}\,. \tag{11}$$

These two new non-dimensional parameters respectively measure how close the lower boundary is to be saturated, and how much water can condense within the shallow layer. When $D_0 - M_0$ is positive, the air at the lower boundary is unsaturated, and $D_0 - M_0$ is

Run	iproc	jproc	T_{P3DFFT}	$T_{PB3DFFT}$	Speedup
1	64	32	10910 secs	10829 secs	$\sim 1\%$
2	64	64	7301 secs	7127 secs	2%
3	128	64	6934 secs	5585 secs	19%

Table 1. Runtime tests on Blue Gene/P for the full moist convection code in production mode on a $2048 \times 2048 \times 513$ grid. The number of MPI tasks is iproc×jproc. Integration was done for 300 time steps.

the amount of water vapor that must be added to an air parcel to become saturated. For convection over the ocean, the lower boundary can just be saturated, i.e. $SSD = 0$. In the following we will consider a three-dimensional subspace of the five-dimensional parameter space only. We will restrict the present study to $Pr = 0.7$ and $SSD = 0$.

2.3 Numerical Scheme and Computational Challenges

We solve the model equations in a flat Cartesian slab with $L_x = L_y = \Gamma H$ with $\Gamma \gg 1$ being the aspect ratio. In the lateral direction we apply periodic boundary conditions, in the vertical direction free-slip boundary conditions. They are given by

$$u_z = D' = M' = 0 \quad \text{and} \quad \frac{\partial u_x}{\partial z} = \frac{\partial u_y}{\partial z} = 0 \tag{12}$$

at $z = 0, H$. The equations of motion are solved by a pseudospectral scheme with volumetric fast Fourier transformations (FFT) and 2/3 de-aliasing. Time-stepping is done by a second-order Runge-Kutta scheme. The smallest resolved scale is the Kolmogorov dissipation length which limits the Rayleigh numbers of our DNS to values of $\sim 10^8$. We use B' in the momentum Eq. (5) instead of B since the mean contribution is $\overline{B}(z)$ and can be added to the kinematic pressure, i.e. $\partial_z p + B = \partial_z \tilde{p} + B'$.

Convection in flat and highly resolved Cartesian cells with grid point numbers $N_x = N_y \gg N_z$ causes challenges for a parallel implementation of the code. On the one hand, massively parallel computers are necessary to process such problem at hand. On the other hand, the geometry requires a two-dimensional parallelization, i.e. the volume $V = \Gamma H \times \Gamma H \times H$ has to be decomposed into iproc×jproc pencils each associated with a MPI task[10]. The resulting volumetric FFTs have to be highly scalable, i.e. increasing the number of CPUs to solve the problem should also substantially decrease the time-to-answer. Recall that three-dimensional FFTs require several global communication steps across the full processor grid. In our code, the FFTs in total consume 87% of the computing time. We compared here two code versions with three-dimenional FFT packages, the P3DFFT-package by D. Pekurovsky[15] and an improved package PB3DFFT with a cache blocking such that the data portions fit into the fast L2 cache of the processors. This avoids several reloadings into L2 during execution of the FFTs along the three coordinate directions and causes a slight speed-up in some cases. In Table 1, we report a test series of runs for $\Gamma = 4$. Although the listed results look promising further efforts of improvement are under way, in particular for very flat cells with $\Gamma = 16$ or 32. The MPI task mapping on the 3d Torus in Blue Gene/P becomes then more and more essential. It should be stated that cache blocking works significantly better on POWER-6 systems where 8 times more memory per core is available than on Blue Gene/P.

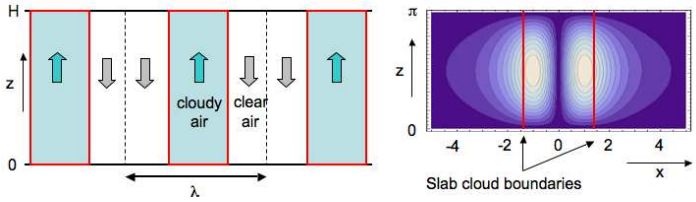


Figure 2. Left: Sketch of the configuration in a conditionally unstable layer that can be solved analytically. Red lines are the cloud boundaries where matching conditions have to be satisfied. The arrows indicate up- and downwelling air. Right: Contours of the velocity stream function which correspond with streamlines of the 2D velocity field of a linearly unstable mode. The parameters are $Ra_M = 3409$, $Ra_D = -955$ and $Pr = 0.7$. The periodicity length is $\lambda \to \infty$. The growth rate is $\Omega = 0.2485$, $t = 0.1$ and $\gamma = 2.8$.

3 Linear Stability of a Shallow Moist Convection Layer

Case of $Ra_D > 0$ and $Ra_M > 0$: Starting point of our dynamical studies is an equilibrium configuration which is infinitesimally perturbed and evolves into a fully developed turbulent state. The equilibrium configuration for a convection layer is a fluid at rest ($\mathbf{u} = 0$) and linear profiles of both buoyancy fields, $\overline{D}(z)$ and $\overline{M}(z)$ as defined in Eq. (3) and (4). In case that both Rayleigh numbers are positive, the linear instability sets in at a critical value[6], i.e. Ra_D and $Ra_M > Ra_c = 27\pi^4/4 \approx 657.5$. Under most circumstances, the amount of water in the atmosphere decreases with height. This implies that the moist Rayleigh number should be larger than the dry Rayleigh number, $Ra_M \geq Ra_D$ (see Figure 1, left).

Case of $Ra_D < 0$ and $Ra_M > 0$: In the atmosphere, another situation is often present. An air parcel is unstable if it is saturated, but stable if it is unsaturated. In case of saturation condensation sets in which is connected with a local release of latent heat and an upward motion of the parcel. We call this regime *conditional instability*[3,6]. In this case, the dry Rayleigh number $Ra_D < 0$, which is sketched in Figure 1 (right). The particular case of $\overline{M}(z) = \overline{D}(z) - N_s^2 z$ was discussed in a closely related model by Bretherton[4,5]. He was able to progress for the two-dimensional case (no y-dependence of all fields) along the same lines as a classical linear stability analysis of dry convection. For the onset of linear instability a periodic slab cloud configuration, as sketched in Figure 2 (left), will be obtained. Moist cloudy air is rising up inside the slab clouds and clear air is falling down outside. Since $\overline{M}(z) = \overline{D}(z) - N_s^2 z$, one needs to discuss one buoyancy field only, M'. The partial derivatives are discontinuous at the cloud boundary and thus additional matching conditions in x direction have to be satisfied. This prohibits a normal mode ansatz in x. With the moist potential, $L(x,z,t) := \int M'(x,z,t)\mathrm{d}x$, Bretherton ended up at a 6th-order ordinary differential equation

$$\left[Ra_M \frac{\partial^2}{\partial t^2}\nabla^2 - \left(\sqrt{\frac{Ra_M}{Pr}} + \sqrt{PrRa_M} \right) \nabla^4 \frac{\partial}{\partial t} + \nabla^6 \right] L = R \frac{\partial^2}{\partial x^2} L \,, \qquad (13)$$

with $R = Ra_M$ if $M'(x,z,t) > 0$ and $R = Ra_D$ otherwise. The free-slip boundary conditions are given by $L = \partial^2 L/\partial z^2 = \partial^4 L/\partial z^4 = 0$ for $z = 0, H$. The matching conditions in x result in six relations $\partial^n L/\partial x^n = 0$ for $x = \pm\gamma/2$ where $n = 0, ..., 5$. The separation ansatz $L(x,z,t) = \exp(\Omega t) G(x) \sin(\pi z/H)$ is now used where $G(x + n\lambda) = G(x)$, as

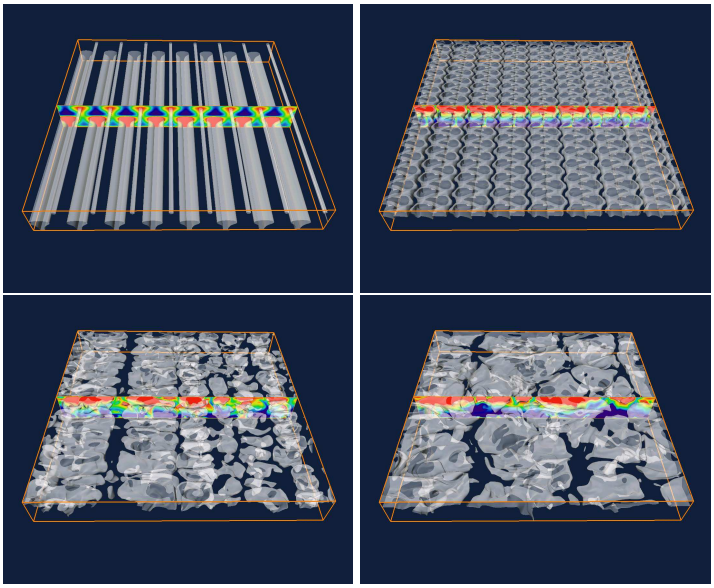


Figure 3. Evolution of clouds from the initial equilibrium configuration to the fully developed turbulent state. Times are $t/T_f = 11.9$ (top left), $t/T_f = 41.6$ (top right), $t/T_f = 64.0$ (bottom left) and $t/T_f = 96.7$ (bottom right). Isosurfaces of the cloud boundary and a two-dimensional contour slice of q_l are shown. Here, $CSA = 0.35$, $Ra_D = 7.0 \times 10^5$, $Ra_M = 1.4 \times 10^6$ and $\Gamma = 8$.

shown in Figure 2 (left). Solutions G in dry and moist areas have to be matched at yet not known cloud boundary positions $x = \pm\gamma/2$. Figure 2 (right) shows an example of the flow field triggered by the linear instability for an isolated cloud with $\lambda \to \infty$. The slab of cloudy air tends to become narrower than the dry air regions surrounding it, as already shown in a simple model by Bjerknes[3].

4 Numerical Simulations of the Full Turbulent Regime

Case of $Ra_D > 0$ *and* $Ra_M > 0$: Our numerical computations of the fully developed turbulent regime were mostly focused so far on the case where both buoyancy fields are linearly unstable. An initial equilibrium configuration is perturbed infinitesimally. After a transient phase of $t/T_f \sim 10^2$, the convective flow is relaxed into a statistically stationary state. This evolution is illustrated by the four snapshots in Figure 3. An important aspect of the simulations is the evolution of the cloud patterns. Clouds are defined in our model as the space-time regions with a liquid water mixing ratio $q_l(\mathbf{x}, t) = M(\mathbf{x}, t) - (D(\mathbf{x}, t) - N_s^2 z) \geq 0$. Our parametric studies found that the size of the cloud patterns depends on both, the parameter CSA and the two Rayleigh numbers Ra_D and Ra_M. The covering of the layer by clouds is decreased with increasing Rayleigh numbers. Furthermore, the larger CSA the more water can be condensed during an adiabatic ascent. A decrease of this parameter allows to study a transition from a closed cloud layer to a broken one. Figure 4 shows an example with a small cloud cover[13]. Isolated clouds are found in regions where air is rising up from the ground (red isosurfaces). The

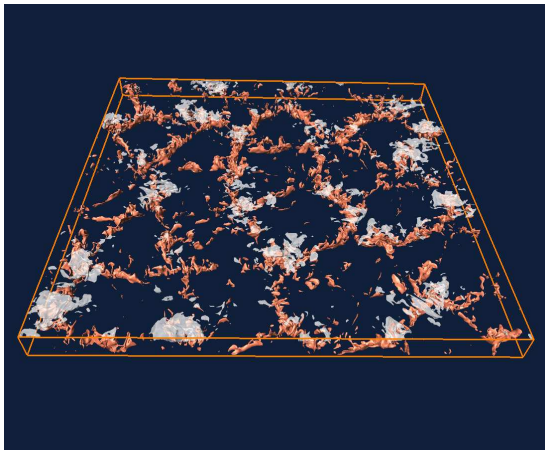


Figure 4. Snapshot of a turbulent shallow convection layer. The cell with an aspect ratio of 16 is resolved by $2048 \times 2048 \times 129$ grid points. The simulation was run on one Blue Gene/P rack (4096 MPI tasks) and took 10 days. The red isocontours show upwelling motion with $u_z \geq 0.23 U_f$. Right above this skeleton of upwelling fluid, clouds are formed which are displayed as grey transparent isosurfaces.

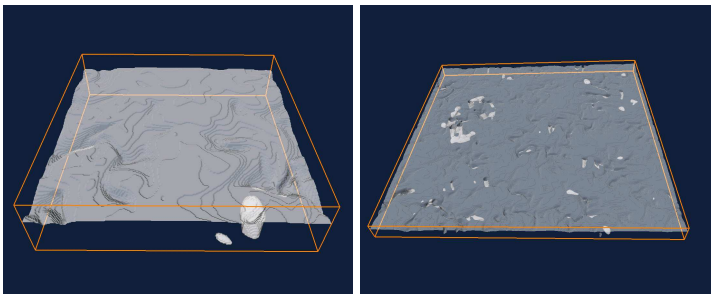


Figure 5. Cloud base isosurface for moist convection in a conditionally unstable layer. The numbers here, $Ra_D = -1.0 \times 10^5$ and $Ra_M = 2.5 \times 10^5$. Left: $\Gamma = 4$. Right: $\Gamma = 16$. Clouds fill the whole slab above the isosurface.

perimeter-area analysis of vertically averaged cloud distribution gave a scaling dimension of 1.27 which is close to 1.32 from large eddy simulations of cumulus convection.

Case of $Ra_D < 0$ and $Ra_M > 0$: In the case of conditional instability, we applied a finite perturbation to a configuration sketched in Figure 1 (right) in order to overcome the stabilizing effect of the dry buoyancy field. Note that this initial configuration differs from the slab cloud ansatz (see Figure 2 (right)) in the linear stability analysis of Bretherton. Our first studies indicate that this regime is very sensitive to the aspect ratio Γ and the amplitude of the perturbation (not shown here). This is illustrated in Figure 5 for $\Gamma = 4$ and $\Gamma = 16$. The cloud base, $q_l = 0$, is shown for snapshots of the turbulent states. Scattered across the nearly unbroken cloud layer are cloud pieces that reach down to the bottom layer. They reproduce the feature which was illustrated in Figure 2. We see also that their number grows with Γ since the system gains a bigger variability to compensate upwelling moist air by downwelling dry air.

In summary, we have shown that the present model can display several features of moist convection. It is thus an appropriate base for the investigation of the impact of phase changes and latent heat release on the turbulence and vice versa. The model allows to include further effects such as rotation of the plane or more importantly radiative cooling at the top.

Acknowledgements

The computer time on the Blue Gene/P system JUGENE at the Jülich Supercomputing Centre (Germany) was provided by the DEISA grant "Cloud09" and by grant HIL02. JS wishes to thank Florian Janetzko for his help and acknowledges support by the DFG. OP is supported by the NSF.

References

1. G. Ahlers, S. Grossmann and D. Lohse, *Heat transfer & large-scale dynamics in turbulent Rayleigh-Bénard convection*, Rev. Mod. Phys. **81**, 503-537, 2009 .
2. P. R. Bannon, *Theoretical foundations for models of moist convection*, J. Atmos. Sci. **59**, 1967-1982, 2002.
3. J. Bjerknes, *Saturated-adiabatic ascent of air through dry adiabatically descending environment*, Quat. J. R. Met. Soc. **64**, 325-330, 1938.
4. C. S. Bretherton, *A theory for nonprecipitating moist convection between two parallel plates. Part I: Thermodynamics and "linear" solutions*, J. Atmos. Sci. **44**, 1809-1827, 1987.
5. C. S. Bretherton, *A theory for nonprecipitating moist convection between two parallel plates. Part II: Nonlinear theory and cloud field organization*, J. Atmos. Sci. **45**, 2391-2415, 1988.
6. K. A. Emanuel, *Atmospheric convection*. Oxford University Press, Oxford, 1994.
7. M. S. Emran and J. Schumacher, *Fine-scale statistics of temperature and its derivatives in convective turbulence*, J. Fluid Mech. **611**, 13-34, 2008.
8. J. Heintzenberg and R. J. Charlson, *Clouds in the Perturbed Climate System*. Strüngmann Forum Reports, MIT Press, Cambridge, Massachusetts, 2009.
9. O. Pauluis and J. Schumacher, *Idealized moist Rayleigh-Bénard convection with piecewise linear equation of state*, Comm. Math. Sci. **8**, 295-319, 2010.
10. J. Schumacher and M. Pütz, *Turbulence in laterally extended systems*, in Proc. of the International Conference ParCo 2007, Editors: C. Bischof, M. Bücker, P. Gibbon, G. Joubert, T. Lippert, B. Mohr und F. Peters, NIC Series 38, Jülich 2007, 585-592 (2007).
11. J. Schumacher, *Lagrangian dispersion and heat transport in convective turbulence*, Phys. Rev. Lett. **100**, 134502 (4 pages), 2008.
12. J. Schumacher, *Lagrangian studies in convective turbulence*, Phys. Rev. E **79**, 056301 (13 pages), 2009.
13. J. Schumacher and O. Pauluis, *Buoyancy statistics in moist turbulent Rayleigh-Bénard convection*, J. Fluid Mech., in press, 2010.
    `http://arxiv.org/abs/0912.5195`.
14. B. Stevens, *Moist convection*, Annu. Rev. Earth Planet. Sci. **33**, 605-643 (2005).
15. `http://www.sdsc.edu/us/ressources/p3dfft.php`.

Computational Plasma Physics: Simulations on Extreme Space and Time Scales

Karl-Heinz Spatschek

Institut für Theoretische Physik, Heinrich-Heine-Universität Düsseldorf
D-40225 Düsseldorf, Germany
E-mail: khs@tp1.uni-duesseldorf.de

The notation plasma was introduced by Langmuir, Tonks, and their collaborators in the 1920's when they studied processes in electronic lamps filled with ionized gases, i.e. low-pressure discharges. Parallel to the developments in gaseous electronics, basic plasma physics evolved from astrophysics during the last century; Hannes Alfvén, Lyman Spitzer, Jr., and Subrahmanyan Chandrasekhar may be named as representatives of the originators of basic plasma physics. Modern plasma physics emerged in 1950's, when the idea of thermonuclear reactor was put forward. The tremendous achievements of magnetic fusion research during the last decades substantiated some years ago the decision to build ITER as an international fusion test reactor which is going to produce on earth reactions similar to those powering the sun. On the other hand, breakthroughs about 20 years ago in laser technology also affected the development of plasma physics considerably. Laser pulses up to powers of several Petawatt and pulse durations of femtoseconds can now routinely be produced. The combination of ultra-short pulse duration and the enormous peak power achieved has opened up unique opportunities for the study of fundamental physics processes and for the development of major applications in completely new regimes of laser-matter-interaction. Laser-produced plasmas deliver unique conditions for studies of relativistic physics, sometimes called relativistic nonlinear optics. As mentioned already, the third area with persistent interest in fundamentals of plasma physics is space physics and astrophysics. Most of the matter in the universe occurs in the plasma state. During the last decades the improvements in observational techniques led to snowballing amounts of observed phenomena and collected data which call for interpretations from plasma physics.

All three areas, magnetic fusion, laser-matter-interaction, and plasma astrophysics, necessitate computational support at highest level to achieve their goals. In magnetic fusion, the "numerical tokamak" is a great dream to better understand the complex processes in bulky machines like tokamaks and stellarators. Short-pulse laser plasmas are inherently nonlinear, difficult to diagnose experimentally, and need extensive numerical simulations. Practically most of the new regimes in electron and ion acceleration, X-ray sources, etc. were proposed and supported by numerical modeling. Finally, knowledge in space physics and plasma astrophysics lives on modeling, which is nowadays, because of the complexity of the systems, in most cases purely numerical.

According to the rich variety of plasmas and their different time and space scales, the theoretical models are multifaceted, e.g. classical, quantummechanical, non-relativistic, or relativistic, respectively. They apply to high-temperature or low-temperature, weakly or strongly coupled plasmas. Nonlinearities play the central role. Numerical approaches

encompass single particle dynamics, particle-in cell codes, kinetic, gyro-kinetic, hydrodynamic (plasmadynamic), and magnetohydrodynamic (MHD) treatments, just to name the most popular ones.

Considering the abundance of models and methods, the following two articles do exemplary represent results from high-performance computing (HPC) in important areas of plasma physics. The summary of Gibbon et al. gives an overview over the state-of-art kinetic particle codes and their applications in laser-plasmas and plasma-wall interaction, i.e. for systems changing on small time scales. The interesting aspect here is, besides the physical results which can be found in the cited papers, the discussion of mesh-free approaches in comparison to, e.g., particle-in-cell codes. It seems that mesh-free techniques offer new possibilities on present-day and future HPC systems. However, the parallel scalability still has to be improved; this is a challenge for the next period. The second contribution gives a taste how mesh-adaptive MHD-computations contribute to both: the understanding of the statistical properties of the interstellar medium and the development of (macroscopic) turbulence theories. Grauer et al. developed the raccoon II code, implemented it on JUGENE, and now publish results for supersonic, compressible MHD turbulence. The obtained velocity power spectra on large space scales will be of interest for the space physics community. They are also important to discriminate between, at least two, theoretical models for compressible MHD turbulence.

The two contributions being presented in this section on plasma physics are just examples of projects in plasma physics that are carried out in HPC. The impact of supercomputing on plasma physics in general is enormous and still growing. For example, in magnetic fusion research, recently the European fusion community put into operation in Jülich a powerful supercomputer, the European “High Performance Computer For Fusion” (HPC-FF), entirely dedicated to analyze the behaviour of burning plasmas and of the materials surrounding them in a fusion reactor. This is a crucial step to ensure the success of ITER. Similar examples can be found in other sections of plasma physics. The recently established Plasma Physics Simulation Laboratory at JSC is a place of national and international importance to bring together scientists of different specializations to exchange ideas, develop new methods, and produce original results in the prosperous area of computational plasma physics.

Plasma Simulation with Parallel Kinetic Particle Codes

Paul Gibbon[1], Robert Speck[1], Benjamin Berberich[1], Anupam Karmakar[1,2], Lukas Arnold[1], and Martin Mašek[3]

[1] Institute for Advanced Simulation, Jülich Supercomputing Centre
Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: {p.gibbon, r.speck, b.berberich, a.karmakar, l.arnold}@fz-juelich.de

[2] ExtreMe Matter Institute EMMI, GSI Helmholtzzentrum für Schwerionenforschung
Planckstraße 1, 64291 Darmstadt, Germany

[3] Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
E-mail: masekm@fzu.cz

Recent work on plasma simulation with kinetic particle codes at JSC is described. Special attention is given to developments in mesh-free plasma modelling using parallel tree codes, covering algorithmic and performance issues, simulations of ion acceleration by high intensity lasers, and new applications of this technique to vortex-fluids and edge-plasma modelling in tokamaks.

1 Introduction

Kinetic simulation methods are an indispensible tool for plasma modelling over large ranges of density and temperature. These techniques, which treat the plasma as a 6-dimensional (3 space, 3 momenta) phase-fluid, are more complex and expensive than hydrodynamic or magnetohydrodynamic models. In the project JZAM04, several complementary kinetic codes are being applied to model a variety of plasma physical problems. These include BOPS, a versatile, 1-dimensional particle-in-cell (PIC) code capable of modelling laser-solid interaction experiments; PSC, an open-source 3D PIC code from LMU Munich; ILLUMINATION, a similar 3D code (with different wave-solvers) developed by Queen's University Belfast; and PEPC, a parallel, mesh-free tree code developed by the authors. The development and performance optimisation work with these codes forms a major activity of the recently established Plasma Physics Simulation Laboratory at JSC.

In the present article we review recent applications and development work involving these codes, which have played a central role in two separate 3rd-party projects, and also form the basis of several Diploma and Ph. D. theses.

2 Hard x-Ray Sources with Thin Foils

Recent experiments by the U. Duisburg-Essen group to optimize hard x-ray line emission from thin coating targets were analysed with the help of the 1D PIC code BOPS. Hard x-rays are generated when a short-pulse, high-intensity laser is used to ionize a solid target, generating high-velocity electrons which proceed to create inner-shell holes in nearby atoms. The resulting hard x-ray pulse can be used for real-time imaging of rapid structural changes in solids or large molecules. It is found that the x-ray pulse efficiency for

poor-contrast laser systems in which a large preplasma is suspected can be enhanced by using a near-normal incidence geometry even at high laser intensities. It can be shown that independently of the laser contrast ratio, i.e. for a steep density gradient or even in the presence of an extended density shelf, it is possible to control and optimise the K_α emission from a femtosecond laser plasma by adjusting the prepulse delay and/or angle of incidence appropriately. Laser systems with poor contrast can be better utilised by using target configurations with small incidence angles to optimise (resonant) absorption. The use of large incidence angles – such as the standard 45-degree geometry – may be contributing to poor K_α conversion efficiencies recorded in many recent experiments. A high-contrast system can be optimised either by controlled prepulse or by going to larger incidence angles, exploiting the tendency for the density profile to be remoulded by the main pulse, allowing better absorption[1,2].

3 Heating and Ion Acceleration in Nanostructured Foils

Nanostructure surfaces are especially promising as highly absorbing targets for high-peak-power sub-picosecond laser-matter interaction. Efficient hot electron, fast ion, and thermonuclear neutron production with moderate laser intensity have already been reported, but theoretical investigations on the use of porous targets for these purposes are still scarce. In a recent study using PEPC, a new phenomenon of Coulomb *implosion* has been identified[3]. The implosion effect is caused by hot electrons circulating inside the shells, drawing ions inwards, where they eventually meet in the centre – Figure 1.

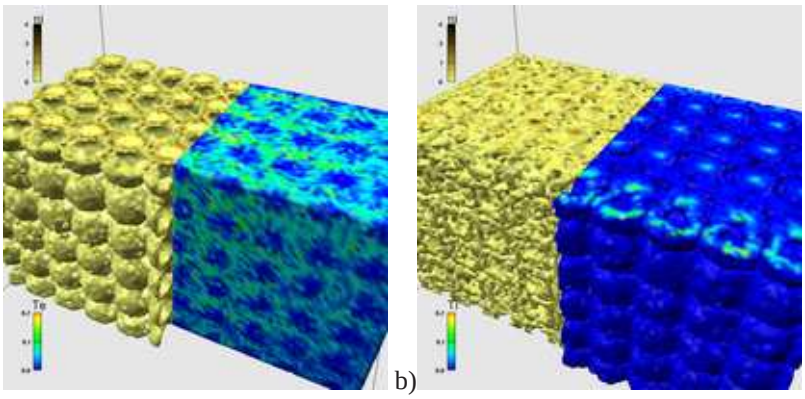
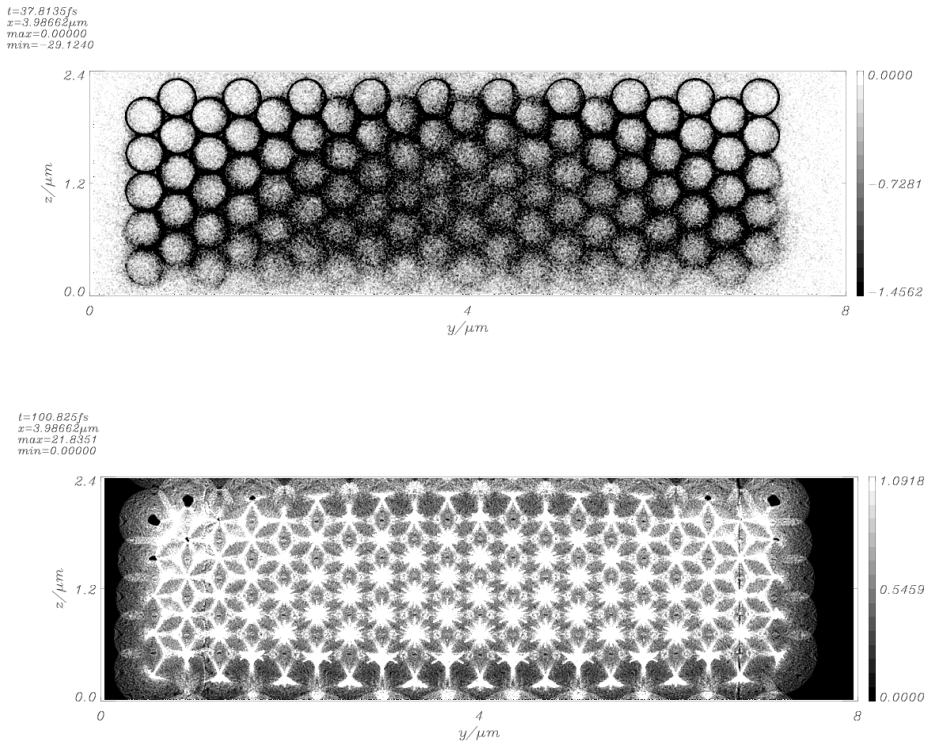


Figure 1. Coulomb implosion of thin shells initially arranged in a bcc foam matrix with ionized electrons confined to the inner shell surfaces – a). Acceleration of ions off the inner surfaces leads to uniform convergence and peaked ion temperatures (hot spots) at the shell centres – b).

Under the same irradiation conditions, a single shell simply blows apart, and does not exhibit the symmetric collapse observed in the foam lattice simulation here. In this case, some hot electrons circulate inside the shell, but most are dragged outside, leading to a net force on the ions directed radially outwards. (A fully stripped ion shell will, by Gauss' Law, have zero electric field inside). This implies that the laser heating is strongly modified by

the regular lattice structure, which in turn radically alters the ion dynamics. These findings have recently been corroborated by 2D PIC simulations in which the electron heating and laser propagation is treated self-consistently with a fully electromagnetic field solver – Figure 2.

a)



b)

Figure 2. 2D PIC simulation of foam array irradiated by high intensity pulse (incident from the bottom) with similar parameters to those in the tree-code simulations. a) Electron density at a few fs, b) ion density after 100 fs.

Should this nano-implosion phenomenon scale to higher, relativistic intensities, it might also have potential as a compact neutron source. A recent comparison between atomic clusters and aerogels[4] suggests that the latter are capable of yielding 10 times as many neutrons for the same laser energy because of the higher kinetic energy imparted to deuterons contained within the aerogel lattice. The present study indicates that density enhancements created by heating *regular* porous lattice structures should result in even higher neutron yields.

4 Mass Limited Ion Acceleration

One of the hot topics of laser-matter interactions is ion acceleration. Current experiments have made dramatic progress in producing multi-MeV beams of light ions and protons which may eventually be used in tumor therapy. Before this goal can be realised however, a number of challenges have to be overcome regarding the beam quality. Simulations are essential here for exploring novel target configurations in order to provide experimental guidance. Microstructured targets have been proposed to reduce the ion beam energy bandwidth and emittance. Our simulations show that a pure proton microdot target does not by itself result in a quasimonoenergetic proton beam: in fact, such a beam can only be produced with a very lightly doped target, in qualitative agreement with one-dimensional theory[5] – Figure 3. The simulations suggest that beam quality in current experiments[6]

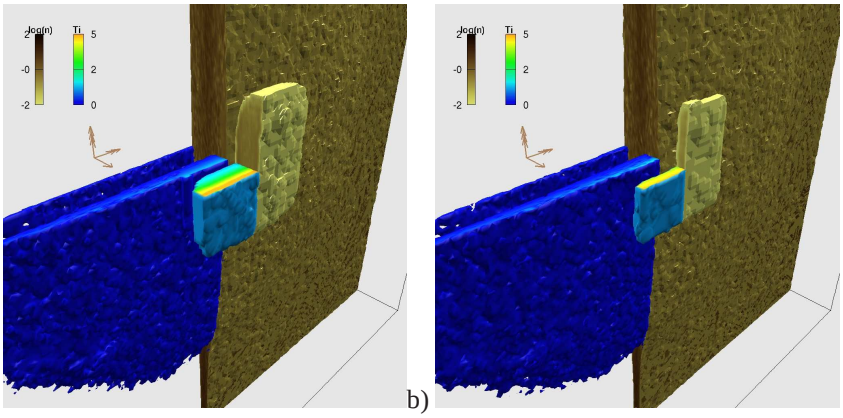


Figure 3. Laser acceleration of proton microdot (central feature). Energy spread of the protons can be reduced by decreasing the relative proton density – here 50% (a) and 5% (b) respectively.

could be dramatically improved by choosing microdot compositions with a 510 times lower proton fraction. Further investigations have further quantified these findings, setting lower limits on the useful proton fraction, beyond which an energy filtering scheme becomes more effective[7].

5 Mesh-Free Modelling of Tokamak Edge Physics

The group “plasma edge simulations for fusion plasmas” in IEF-4 (Prof. Reiter) develops and applies 2- and 3-dimensional computer simulation codes for interpretative and predictive studies of physics close to the plasma container wall (vacuum chamber). This domain is characterised by a complex interaction of plasma-chemical and turbulent processes. The models contain both empirical and first-principles based modules. The long term goal of model development is step by step replacement of the *ad hoc* by *ab initio* models, aided by increasing HPC resources based on the parallel computing paradigm. The self-consistent electrical field, obtained from the position of charged test particles, or,

in fluid approximations, of charged fluid parcels, is one such component.
The mesh-free method offered by the tree code PEPC-B developed at JSC is a promising candidate either to provide self-consistent fields in tandem with existing modelling tools, or as a stand-alone (radiation-free) code for fusion plasma simulations. Currently this is being jointly developed by JSC and IEF-4 to model plasma edge phenomena such as the potentially catastrophic edge-localised modes (ELMs).

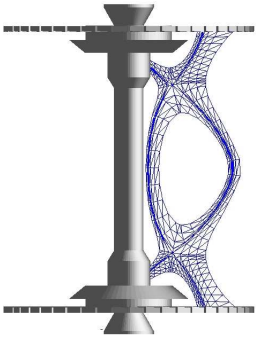


Figure 4. *Example of a triangular mesh fitted in the vessel of the MAST-tokamak.*

In a first step interfaces have been developed to provide PEPC-B with fusion relevant magnetic fields. For this purpose equilibrium fields similar to those used in the transport code EIRENE[8] have been introduced. To define the external magnetic field PEPC-B assigns vector data according to a 2d triangular mesh (5), assuming toroidal symmetry in the tokamak. This feature in principle enables PEPC-B to run full tokamak core simulations. Next, the field data provided from external sources have to be smoothed and adjusted to guarantee $div(\vec{B}) = 0$.
A further addition is the inclusion of a collision module based on the Monte-Carlo model of Takizuka and Abe[9]. This permits collisions between the injected tracer particles and the plasma backround to be taken into account. An example application is the modelling of impurities such as C^+ ions occurring in fusion plasmas after sputtering or gas puff scenarios in a realistic plasma environment. A current priority is to simulate gas puff experiments recently performed on TEXTOR and to see how these findings scale to larger machines such as ITER.

6 Darwin Model

A strategic goal with this project is to extend the fast summation algorithm deployed in PEPC to include self-generated magnetic fields. This “Darwin” or magnetoinductive approach has been pursued within the particle-in-cell paradigm for some time[10]. A mesh-free Darwin model would open up a large range of new applications of mesh-free plasma simulation, not least in the tokamak modelling scenarios described above, and also in particle beam transport in dense plasmas. The model incorporates slowly varying magnetic fields

by neglecting the transversal part of displacement current in Ampères law. This transforms the hyperbolic equation system into a set of elliptic equations, which can again by solved by a fast summation algorithm.

$$\phi_i(\boldsymbol{r}_i) = \sum_{j \neq i} \frac{q_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1}$$

$$\boldsymbol{E}_i^l(\boldsymbol{r}_i) = \sum \frac{q_j \boldsymbol{r}_{ij}}{r_{ij}^3} \tag{2}$$

$$\boldsymbol{A}_i(\boldsymbol{r}_i) = \frac{1}{2c} \sum \frac{q_j \boldsymbol{v}_j}{r_{ij}} + \frac{1}{2c} \sum \frac{(q_j \boldsymbol{v}_j \cdot \boldsymbol{r}_{ij}) \boldsymbol{r}_{ij}}{r_{ij}^3} \tag{3}$$

$$\boldsymbol{B}_i(\boldsymbol{r}_i) = \nabla \times \boldsymbol{A}_i = \frac{1}{c} \sum \frac{q_j \boldsymbol{v}_j \times \boldsymbol{r}_{ij}}{r_{ij}^3} \tag{4}$$

$$\boldsymbol{E}_i^t(\boldsymbol{r}_i) = -\frac{1}{c} \frac{\partial \boldsymbol{A}_i}{\partial t} = -\frac{1}{2c^2} \sum \frac{q_j \dot{\boldsymbol{v}}_j}{r_{ij}} - \frac{1}{2c^2} \sum \frac{(q_j \dot{\boldsymbol{v}}_j \cdot \boldsymbol{r}_{ij}) \boldsymbol{r}_{ij}}{r_{ij}^3} \tag{5}$$

On the other hand, numerical difficulties arising from this approximation have already been reported by previous authors who implemented the Darwin model within PIC or Vlasov codes. The standard scheme known from the fully electromagnetic codes used for the calculation of time derivative of the vector potential causes a violent numerical instability destroying the whole run in a few time-steps. One of the possible solutions to this problem is to express the quantities in terms of Hamiltonian generalized variables, which avoids the time derivative of the vector potential in the the equation of motion. Our magnetoinductive model employs a multipole expansion of the Darwin field equation, modified to account for finite-sized particles and evaluated within the PEPC tree algorithm framework. First tests of this model have been made on charged particle beam evolution in vacuum.

7 Tree Code Scalability/Performance

The PEPC code is written in a generic fashion without the usage of external libraries. This results in excellent portability. In the PRACE benchmarking framework PEPC was run on four different computer architectures, namely: IBM Blue Gene/P (jugene), IBM Power6 (huygens), Cray XT5 (louhi), Intel Nehalem (juropa). As a test case we used $5 \cdot 10^7$ particles and a cubic, homogeneous initial distribution. The scaling behaviour is shown in Figure 5, where the time needed for one inner loop step is shown as a function of the peak performance of the used partition. As shown, PEPC is able to utilize the given partition performance independently of the architecture. Only the Intel Nehalem architecture shows a significantly better performance. PEPC's overall scaling behaviour is also impressive for a tree code, although the current version is only able to use up to approximately 8192 cores. But with these the code is capable of simulating more than 10^8 particles.

Three different areas have been identified as challenges for further development of the code: The inevitable need for global communication in Barnes-Hut tree codes, a strong memory dependency of the number of cores and a reliable load balancing scheme. These problems have to be well analysed in order to make resonable use of PEPC on larger partitions. While the second topic is a question of data structure and organisation the two others focus on communication patterns and their impact on the performance. It is clear that although the tree code is an intrinsically adaptive algorithm which performs very well even with very inhomogeneuous particle distributions, its performance relies strongly on a sophisticated load balancing scheme, which is closely linked to the domain decomposition method. But even with a wisely chosen load balancing approach the need for non-local communication is an inevitable consequence of the elliptic nature of the Poisson equation. On the other hand, recent analyses of the communcation pattern reveal that nearly every communication instance is actually required at some point in the force calculation. For covering the span of necessary non-local communication and demanding memory requirements a physical separation of communication and computation process may provide a way to scale beyond 10^4 cores.

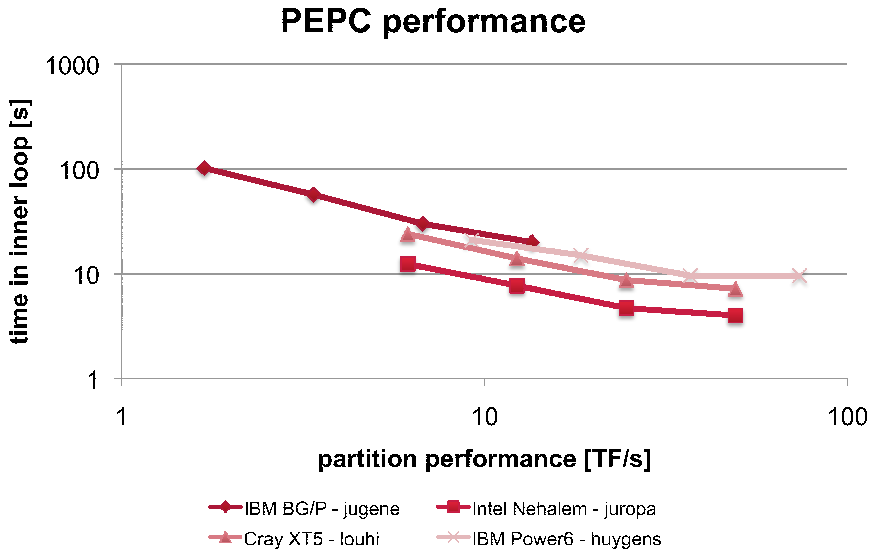


Figure 5. Performance of PEPC on various HPC architectures, normalised to the theoretical peak performance of a given multi-core partition.

8 Concluding Remarks

To summarize, we have reported on various activities concerning kinetic plasma modelling at JSC. Most of this work involves application and further development of the mesh-free

code PEPC, but recently we have also started using fully electromagnetic PIC codes to model laser-plasma interaction processes. Future work will include rigorous comparisons between the PIC and mesh-free approaches, in order to more clearly identify the advantages of the latter. The mesh-free Darwin or magnetoinductive approach offers completely new possibilities in many areas of plama physics from magnetic fusion to space physics. A major challenge for the usage of the mesh-free technique on contemporary HPC systems is to improve its parallel scalability. Efforts are underway to improve the current limit of 8192 cores on BG/P to at least 64k cores.

Acknowledgements

This work was supported by the Alliance Program of the Helmholtz Association (HA216/EMMI). Simulations were performed with computing resources granted by the VSR of the Research Centre Jülich unter project JZAM04.

References

1. P. Gibbon, M. Mašek, U. Teubner, W. Lu, M. Nicoul, U. Shymanovich, A. Tarasevitch, P. Zhou, K. Sokolowski-Tinten, and D. von der Linde, *Modelling and optimisation of fs laser-produced K_α sources*, Appl. Phys. A, 2009.
2. W. Lu, M. Nicoul, U. Shymanovich, A. Tarasevitsch, P. Zhou, K. Sokolowski-Tinten, D. von der Linde, M. Mašek, P. Gibbon, and U. Teubner, *Optimized K_α x-ray flashes from femtosecond-laser irradiated foils*, Phys. Rev. E, **80**, 026404, 2009.
3. P. Gibbon and O. N. Rosmej, *Stability of nanostructure targets irradiated by high intensity laser pulses*, Plasma Phys. Contr. Fus., **49**, 1873–1883, 2007.
4. V. P. Krainov and M. B. Smirnov, *Nuclear fusion induced by a super-intense ultrashort laser pulse in a deuterated glass aerogel*, JETP, **93**, 485–490, 2001.
5. A. P. L Robinson and P. Gibbon, *Production of proton beams with narrow-band energy spectra from laser-irradiated ultrathin foils*, Phys. Rev. E, **75**, no. 1, 015401, 2007.
6. S. M. Pfotenhauer, O. Jäckel, A. Sachtleben, J. Polz, W. Ziegler, H.-P. Schlenvoigt, K.-U. Amthor, M. C. Kaluza, K. W. D. Ledingham, R. Sauerbrey, P. Gibbon, A. P. L. Robinson, and H. Schwoerer, *Spectral shaping of laser generated proton beams*, New J. Phys., **10**, 033034, 2008.
7. A. P. L. Robinson, P. Gibbon, M. Zepf, S. Kar, R. G. Evans, and C. Bellei, *Relativistically correct hole-boring and ion acceleration by circularly polarized laser pulses*, Plasma Phys. and Cont. Fus., **51**, 024004, 2009.
8. D. Reiter, M. Baelsmans, and P. Börner, *The Eirene and B2-Eirene Codes*, Fusion Science and Technology, **47**, 172, 2005.
9. T. Takizuka and H. Abe, *A binary collision model for plasma simulation with a particle code*, J. Comp. Phys., 1977.
10. D. W. Hewett, *Elimination of electromagnetic radiation in plasma simulation: the Darwin or magnetoinductive approximation*, Space. Sci. Rev., **42**, 29–40, 1985.

Statistics, Structures and Alignment in Compressible MHD Turbulence

R. Grauer, J. Dreher, C. Beetz, and C. Schwarz

Institute for Theoretical Physics I, Ruhr-University Bochum, Germany
E-mail: grauer@tp1.rub.de

Compressible, supersonic, magnetohydrodynamic (MHD) turbulence, as found in the interstellar medium (ISM), is not very well understood. In a set of simulations we want to address the following questions: What is the correct scaling behaviour for compressible, supersonic MHD turbulence? What are the dominant structures? What is the mechanism, from which a supersonic, turbulent MHD-flow becomes quasi-incompressible but still shows compressive qualities in the density? How do those properties change in sub- and superalfvenic flows? Are Lagrangian statistics able to shed a brighter light on the process of star formation in the ISM?

1 Introduction

The interstellar medium (ISM) shows turbulent behaviour[1] and is characterised by high Mach number flow, fractal density distribution and a magnetic field. The strength of the magnetic field varies widely. The large scale galactic field is rather weak but in molecular clouds equipartition of magnetic and kinetic energy has been observed[2]. Knowledge of basic statistical properties of this turbulent, compressible and supersonic flow would ease the understanding of many astrophysical processes in the galaxy like star formation or transport phenomena (e.g. cosmic rays).

1941 Kolmogorov postulated the assumption of universality for incompressible hydrodynamical flows with high Reynolds numbers which lead to the famous $k^{-5/3}$ scaling relation for the velocity power spectrum of fully developed, isotropic and homogeneous turbulence[3]. Since this scaling behaviour was experimentally confirmed even under conditions violating the basic assumptions, Kolmogorovs phenomenology became a working hypothesis for all sorts of turbulent flows.

But unlike in incompressible Navier-Stokes turbulence, there are at least two models for MHD flows: the Iroshnikov-Kraichnan (IK)[4,5] model for homogeneous, isotropic turbulence and the Goldreich-Shridar (GS)[6] model which assumes that MHD turbulence is anisotropic. The former predicting a velocity power spectrum proportional to $k^{-3/2}$, the latter a Kolmogorov scaling perpendicular to the mean field ($k_\perp^{-5/3}$). Numerical findings[7,8] revealed that, depending on the strength of the magnetic field, the spectrum of the perpendicular velocity fluctuations changed from GS to IK as the field increased. A possible solution to this incongruity could be the phenomenological approach of Boldyrev[9] in which he takes into account that by dynamically alignment of the velocity and the magnetic field the eddies become locally anisotropic in the mean-field perpendicular direction. Thus, the turbulence is getting even more anistotropic than in the GS model and the scaling formally coincides with the IK model. A hint towards the explanation may rise from the geometry of the dissipative structures in the flow where the $k^{-3/2}$-scaling would reveal current sheets while the GS model would correspond to filament-like structures.

Figure 1. Volume rendering of the density (MHD-simulation with a resolution of 1024^3 grid-points).

We want to take those considerations as a guideline to investigate compressible turbulence, for which the scientific community has just recently begun to discuss which statistical measures are appropriate. Kritsuk et. al.[10] examined supersonic hydrodynamical (HD) turbulence and suggested, based on the model of Fleck[11], the density weighted velocity $\boldsymbol{u} = \rho^{1/3}\boldsymbol{v}$ as the measure of a natural extension of Kolmogorov phenomenology. While compressibility and density statistics depend strongly on the Mach number of the flow, the spectra and structure functions of the new combined measure appear to be invariant. While the universality of turbulence seems to be violated in compressible supersonic flows[12] the scaling of the mass-weighted velocity remains independent of the large scale properties. And, similar to the incompressible case, in compressible MHD the velocity $\boldsymbol{v}$ exhibits different scaling perpendicular and parallel to the mean magnetic field, while the scaling of $\boldsymbol{u}$ looks similar in all directions[13].

In our own simulation of supersonic compressible MHD turbulence we came across another peculiarity of this kind of flow. While the density distribution follows a log-normal distribution as in highly compressible HD cases[10,14] the mean compressibility $\mathcal{C} =< (\mathrm{div}\ \boldsymbol{v})^2 > / < \sum_{i,j=1..3} (\partial_i v_j)^2 >$ remains, after a short transient phase, on a very low i.e. incompressible level (Figure 2, bottom left). The velocity power spectra (Figure 2, top) and the fraction of the dissipation rate that depends solely on the solenoidal velocity components $\varepsilon_s/\varepsilon = < |\nabla \times \boldsymbol{v}|^2 > / (< |\nabla \times \boldsymbol{v}|^2 > + 4/3 < |\nabla \cdot \boldsymbol{v}|^2 >)$ (Figure 2, bottom right) give clear evidence that the flow is basically incompressible. And yet we get an Kaplan-Yorke fractal dimension of the mass distribution of $D_L = 2.77$[15], not far away from the value of 2.6 given by Sanchez et. al.[16] for the ISM.

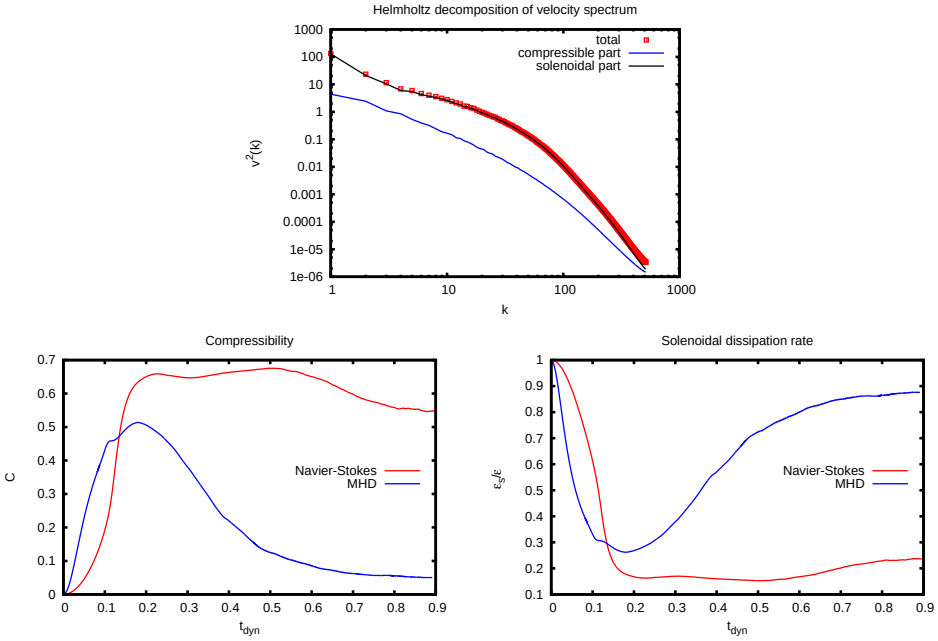


Figure 2. Compressible MHD at high Mach number (M=6) show incompressible flow features.

2 The Numerical Scheme

racoon II[17] is a framework for time-dependant mesh-adaptive computations on structured grids, developed especially for the use on massively parallel computers. We use it here to solve the conservative MHD-equations in direct numerical simulations on a Cartesian grid by a finite-volume discretisation. The numerical difficulties we have to deal with in supersonic, compressible turbulence are, for instance, steep shocks and intervalls where the density is almost zero. To handle these conditions we use an overall third order semidiscrete central upwind scheme[18] with CWENO reconstruction. The solenoidal constraint of

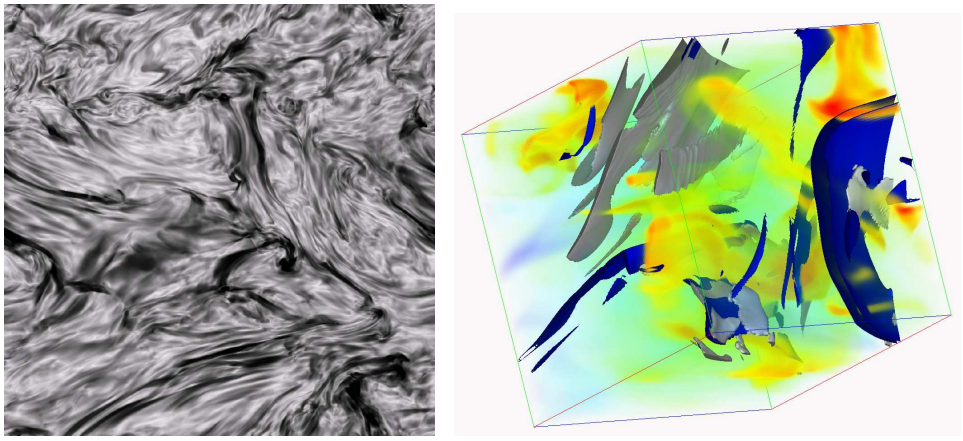


Figure 3. The volume rendering of a thin slice of the vorticity reveals turbulent and intermittent properties (left). Volume rendering of the denisty (yellow to red), isosurfaces of vorticity (grey) and current density (blue) show sheet-like structures (right).

the magnetic field is ensured due a constrained transport method[19]. Time integration uses a strongly stable and memory saving third order Runge-Kutta scheme[20].

The grid is divided into cubic blocks which are distributed over the processors. Such a block, along with its ghost cells, can then be calculated autarchically. Data exchange between blocks is done by the Message Passing Interface (MPI) library. To minimize the communication between processors the blocks are distributed via a 3-dimensional Hilbert-curve-partitioning[21], which is also used for the dynamical load balancing in the octree adpative mesh refinement technique (AMR) we will apply in the upcoming simulations (Figure 4, left).

Figure 4, right, summarizes the scaling behaviour of *racoon II* on the JUGENE super-

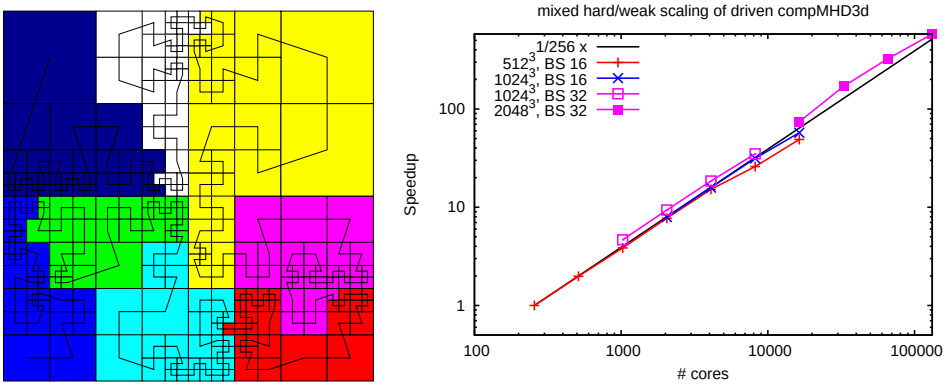


Figure 4. Hilbert-curve-partitioning of the workload (left) and the resulting scaling of *racoon II* on JUGENE (right).

computer. In this graph we took data from scaling tests as well as from production runs. Hence, we used different basic blocksizes. Enlarging the blocksize means fewer but larger messages. Thus larger blocksizes are bandwidth limited whereas smaller blocksizes are latency limited.

Weak scaling (constant load per processor) is perfect linear up to **131072 processors**.

References

1. J. W. Armstrong, B. J. Ricket and S. R. Spangler, ApJ **443** (1995) 209.
2. E. Falgarone, T. H. Troland, R. M. Crutcher and G. Paubert, A& A **487** (2008) 247.
3. U. Frisch, Turbulence, Cambridge University Press, 2004.
4. P. S. Iroshnikov, AZh **40** (1963) 742.
5. P. S. Iroshnikov, Sov. Astron. **7** (1964) 566.
6. P. Goldreich, S. Sridar, ApJ **438** (1995) 763.
7. J. Maron and P. Goldreich, ApJ **554** (2001) 1175.
8. W.-C. Mueller and R. Grappin, Phys. Rev. Lett. **95** (2005) 114502.
9. S. Boldyrev, ApJ **626** (2005) L37.
10. A. G. Kritsuk, M. L. Norman, P. Padoan and R. Wagner, ApJ, **665** (2007) 416.
11. R. C. Fleck Jr., ApJ **458** (1996) 739.
12. W. Schmidt, C. Federrath and R. Klessen, Phys. Rev. Lett. **101** (2008) 194505.
13. G. Kowal and A. Lazarian, ApJ **666** (2007) L69.
14. C. Beetz, C. Schwarz, J. Dreher, R. Grauer, Phys. Lett. A **372** (2008) 3037.
15. C. Schwarz, C. Beetz, J. Dreher, R. Grauer, submitted to Phys. Lett. A (2009).
16. N. Sanchez, E. J. Alfaro and E. Perez, ApJ **641** (2006) 347.
17. J. Dreher and R. Grauer, Parallel Comp. **31** (2005) 913.
18. A. Kurganov and D. Levy, SIAM J. Sci. Comput. **22** (2000) 1461.
19. U. Ziegler, J. Comput. Phys. **196** (2004) 393.
20. C.-W. Shu and S. Osher, J. Comput. Phys. **77** (1988) 439.
21. G. Zumbusch, Angew. Math. Mech. **81** (2001) 25.

Schriften des Forschungszentrums Jülich
IAS Series

1. **Three-dimensional modelling of soil-plant interactions: Consistent coupling of soil and plant root systems**
   by T. Schröder (2009), VIII, 72 pages
   ISBN: 978-3-89336-576-0
   URN: urn.nbn:de:0001-00505

2. **Large-Scale Simulations of Error-Prone Quantum Computation Devices**
   by D. B. Trieu (2009), VI, 173 pages
   ISBN: 978-3-89336-601-9
   URN: urn.nbn:de:0001-00552

3. **NIC Symposium 2010**
   Proceedings, 24 – 25 February 2010 | Jülich, Germany
   edited by G. Münster, D. Wolf, M. Kremer (2010), V, 395 pages
   ISBN: 978-3-89336-606-4
   URN: urn:nbn:de:0001-2010020108

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 amd 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 fifth NIC Symposium are from projects that have been supported by the IBM super-computers JUMP and Blue Gene/P in Jülich and the APE topical computer at DESY in Zeuthen. 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, Hydrodynamics and Turbulence, and Plasma Physics.

IAS Series
Volume 3
ISBN 978-3-89336-606-4

JÜLICH
FORSCHUNGSZENTRUM