UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Annual Report 2006/2007

PADERBORN
CENTER FOR
PARALLEL
COMPUTING

University of Paderborn
Paderborn Center for Parallel Computing
Fürstenallee 11, D-33102 Paderborn

www.upb.de/pc2

Table of Contents

1 Preface

Das PC^2 ist eine zentrale wissenschaftliche Einrichtung der Universität Paderborn und arbeitet als Forschungs- und Dienstleistungszentrum im Bereich des parallelen und verteilten Rechnens. Die wesentlichen Aufgaben des PC^2 liegen in der Entwicklung und Bereitstellung innovativer paralleler Rechnersysteme für die Hochschulen des Landes Nordrhein-Westfalen sowie in der Entwicklung von Methoden und Verfahren zur effizienten Nutzung paralleler und verteilter Systeme.

Die Jahre 2006 und 2007 standen im Zeichen des Wandels und der Veränderung – insbesondere in Anbetracht der am PC² beteiligten Personen. Anfang 2006 trat zunächst mit Prof. Gero Schmidt, Theoretische Physik, ein sehr aktiver und engagierter Nutzer unserer Dienstleistungen in den Vorstand des PC² ein und erhöhte damit die fachliche Breite. Im August 2006 nahm allerdings Prof. Odej Kao, seit April 2003 geschäftsführender Leiter des PC², einen Ruf an die Technische Universität Berlin auf den Lehrstuhl für komplexe und verteilte IT-Systeme, verbunden mit der Leitung des IT-Service-Centers der TU Berlin, an und schied damit aus dem PC² aus. Dies war der Anlass, die Leitungsstruktur des PC² neu auszurichten, so dass im Jahr 2007 Prof. Burkhard Monien den Vorsitz des Vorstandes an Prof. Holger Karl übergab. Prof. Monien hat als Initiator und langjähriger Vorsitzender des PC² viele Forschungsarbeiten in Paderborn maßgeblich geprägt und die vom PC² verfolgten Forschungsrichtungen an dieser Universität überhaupt erst ermöglicht. Ihm gilt unser aller Dank und Anerkennung für seine unermüdliche und erfolgreiche Arbeit!

Die Führung des PC² soll im Jahr 2008 durch eine gemeinsam mit dem Institut für Informatik zu besetzende Juniorprofessur verstärkt werden, die neben Aufgaben der Geschäftsleitung wesentliche Themen neuer Hochleistungsrechner-Infrastrukturen in Lehre und Forschung vertreten soll. Der Schwerpunkt liegt dabei auf Virtualisierungstechniken in Hochleistungssystemen.

Mit der Ausrichtung dieser Juniorprofessur und dem Wechsel des Vorstandsvorsitzes ist auch eine inhaltliche Neuorientierung des PC² verbunden. Einige der bisher bearbeiteten Themen – als Beispiel sei die numerische Mathematik genannt – werden derzeit nicht mehr am PC² selbst bearbeitet. Dafür sind Themen wie flexible Rechnerstrukturen und rekonfigurierbare Systeme (insbesondere in Zusammenarbeit der mit der Forschungsgruppe von Prof. Marco Platzner, Institut für Informatik), Betriebskonzepte und Ressourcenverwaltung für Hochleistungssysteme, oder auch Vernetzung solcher Systeme samt experimenteller Testbeds verstärkt in den Fokus gerückt.

Diese Neuorientierung wirkt sich auch auf die Personalstruktur des Kernteams aus, bzw. wurde teilweise durch die personellen Veränderungen bedingt. Tatsächlich waren die beiden vergangenen Jahre durch eine sehr hohe Fluktuation bei den wissenschaftlichen Mitarbeitenden geprägt. Erfreulich ist dennoch zu berichten, dass drei Mitarbeiter promoviert das PC² verlassen haben: Dr. Felix Heine (August 2006), Dr. Matthias Hovestadt (Dezember 2006) und Dr. Oliver Marquardt (November 2007). Und erneut konnte eine Ausbildung im PC² erfolgreich absolviert werden: Frau Diana Hunecke verließ das PC² als Fachinformatikerin mit Fachrichtung Systemintegration. Ende 2007 waren damit insgesamt 13 wissenschaftliche Mitarbeiter im PC² tätig; vier mehr als Ende 2005.

Wie in den Jahren zuvor war das PC² im Berichtszeitraum wieder erfolgreich an Projekten auf nahezu allen Förderebenen (EU, BMBF, DFG, MIWFT) beteiligt. Diese breit angelegte Basis unterstreicht die Vielfältigkeit des PC². Daneben wurden die direkten industriellen Kooperationen mit den Firmen Fujitsu Siemens Computers und Intel fortgeführt.

Neben der Projektarbeit wurde die wissenschaftliche Arbeit ebenso nachhaltig und erfolgreich weiter geführt. Insgesamt entstanden 48 internationale Publikationen, die auf einschlägigen Konferenzen präsentiert bzw. durch 3 Artikel in wissenschaftlichen Zeitschriften veröffentlicht wurden.

Das PC² unterstützt die Universität Paderborn auch im Bereich der Lehre und studentischen Ausbildung. Vorlesungen und Projektgruppen werden regelmäßig für Studierende angeboten. Im Berichtszeitraum wurden 35 Studien-/Bachelorarbeiten und 20 Diplom-/Masterarbeiten von Mitarbeitern des PC² betreut.

Alle Projekte wie auch andere in der Universität durchgeführte Forschungsarbeiten profitieren dabei auch von der Dienstleistungsrolle des PC². Diese wird insbesondere durch die Bereitstellung von Rechenleistung im „Arminius“-Cluster wahrgenommen. Dieses System ist das „Arbeitspferd“ des PC² und steht inzwischen verlässlich und einfach nutzbar zur Verfügung. Insgesamt wurde in 2007 eine Auslastung von 90% erzielt; ein für Rechenzentren in der Größenordnung des PC² ausgezeichneter Wert. Die Bedeutung dieses Systems ist dabei nicht auf Paderborn beschränkt: Ca. 57% der Rechenleistung wurde für nicht in Paderborn arbeitende Forscher zur Verfügung gestellt, alleine 38% wurden von der Universität Bielefeld genutzt. Dies zeigt eindrucksvoll die Bedeutung des Hochleistungsrechnens für die gesamte Region Ostwestfalen-Lippe.

Flankierend zum Arminius-System konnte Ende 2007 ein weiteres, wenn auch deutlich kleineres System durch eine Sonderinvestition im Rahmen der Deutschen-Grid-Initiative (D-Grid) für das BIS-Grid-Projekt beschafft werden. Dieser Rechner ist

vornehmlich für Experimente und Arbeiten im Kontext der Virtualisierung von Grid-Anwendungen und der verlässlichen Bereitstellung von Grid-Ressourcen, insbesondere für Anwendungen der Wirtschaft, vorgesehen.

Allerdings ist mit diesen Neuanschaffungen – und den für 2008 absehbaren weiteren Installationen – das Ende der bisherigen Kapazität praktisch erreicht. Für Folgesysteme besteht derzeit keine Möglichkeit der Unterbringung, da sowohl Stellfläche als auch Kühlkapazität ausgeschöpft sind. Hier hat eine im Jahr 2007 mit dem Präsidium der Universität geschlossene Zielvereinbarung die Weichen für die Zukunft gestellt und den Weg zu neuen Möglichkeiten, insbesondere zu einem neuen Rechnerraum, geöffnet. Damit werden das PC² und die Universität Paderborn auch in Zukunft eine bedeutende, regionale und überregionale Position im Hochleistungsrechnen in Deutschland – auch in der sich derzeit formierenden „HPC-Allianz Deutschland“ – einnehmen.

Erstmalig umfasst der PC² Jahresbericht einen Zeitraum von zwei Jahren. Alle Informationen wurden sorgfältig zusammengetragen und thematisch sortiert. Mehr als 30 Beiträge sind dabei zusammengekommen. Ich möchte an dieser Stelle allen Beteiligten hierfür danken. Den Lesern wünsche ich nun eine spannende Lektüre und freue mich auf Ihre Reaktionen, Fragen und Anregungen.

Paderborn, im März 2008

Prof. Dr. Holger Karl, Vorsitzender des Vorstandes

Preface

The Paderborn Center for Parallel Computing (PC²) is one of University of Paderborn's central scientific institutes. It serves as a research and service center for parallel and distributed computing. The main tasks of PC² are the development and operation of innovative parallel computing systems for the universities in the federal state of North Rhine-Westphalia (NRW). Moreover, PC² scientists develop methods for the efficient use and management of parallel and distributed systems.

The years 2006 and 2007 were characterized by renewal and change – in particular concerning the people involved with PC². First, early in 2006, Prof. Gero Schmidt, theoretical physics, a very active and engaged user of the PC² services, joined the directorate extending the technical expertise. In August 2006, Prof. Odej Kao, managing director of the PC² since April 2003, accepted a position at the Technical University of Berlin as the chair of Complex and Distributed IT Systems jointly with the management of the IT-Service Center. His departure initiated a realignment of the PC² management structure. In 2007, Prof. Burkhard Monien turned over the chairmanship to Prof. Holger Karl. Prof. Monien decisively influenced as the initiator and long lasting chairman of the PC² board many research activities in Paderborn; he made the PC² research areas at the University of Paderborn possible in the first place. Our special thanks and respect to Prof. Monien for his persistent and successful work!

In 2008, the PC² management is to be strengthened by appointing, jointly with the department of computer science, a "Juniorprofessur" (assistant professor). Besides partially assuming responsibility for PC² management, this professorship will represent relevant topics of new high-performance infrastructures in research and teaching; it will emphasize virtualization techniques in high-performance systems.

With the "Juniorprofessur" and the change of the chair of the PC² board a reorientation of contents is happening at PC². Some of the so-far covered questions – e.g. numerical mathematics – are currently not actively pursued at the PC² itself. Instead, topics like reconfigurable computer systems (particularly in joint work with the research group of Prof. Marco Platzner, department of computer science), operational concepts and resource management of high-performance systems, or networking of such systems including experimental testbeds, are coming to the fore.

This reorientation affects the personnel structure of the core team and was partly caused by changes in personnel. The last two years are characterized by a high fluctuation of scientific employees. Three members of staff completed their jobs having received a PhD: Dr. Felix Heine (August 2006), Dr. Matthias Hovestadt

(December 2006), and Dr. Oliver Marquardt (November 2007). And, once more an apprenticeship at PC² ended with success: Ms. Diana Hunecke finished as an IT specialist in systems integration.

Similar to previous years, the PC² was successfully involved in projects. During the reporting period, they were supported by nearly all funding organizations (EC, BMBF, DFG, MIWFT), underlining the broad spectrum of topics successfully research at the PC². In addition, the direct cooperation with the companies Fujitsu Siemens Computers and Intel continued.

Besides the project work, the scientific work was continued successfully. In total 48 international publications appeared and were presented at relevant conferences; three publications in leading journals were achieved.

The PC² also offers lectures and project groups to students of the University of Paderborn. In the last two years, 35 bachelor's theses and 20 master's theses were supervised by PC² staff.

All our projects as well as other research work within the university benefit from PC²'s role as a service provider of high performance computer systems. In particular, our "Arminius Cluster" is the workhorse of the PC²; it is available absolutely reliably and easy to use. All in all, a utilization of 90% was achieved in 2007 – an excellent value for a computing center the size of PC². Arminius plays an important role even outside of Paderborn: About 57% of computing capacity was provided to external researchers; 38% were used by the University of Bielefeld alone. This highlights the importance of high-performance computing for the region of Eastern Westphalia-Lippe.

The Arminius system was amended by a further, albeit significantly smaller cluster. It was made possible by a special program in the context of the German grid initiative's (D-Grid) BIS-Grid project at the end of 2007. This system will be used for experiments in the context of virtualisation of grid applications and the dependable provisioning of grid resources, especially for economically relevant applications.

However, with these new acquisitions – and some foreseen installations in 2008 – the limits of housing capacity are almost reached. At the moment, it would not be possible to accommodate any futher systems as both footprint and cooling capacities are exhausted. Here, an agreement completed at the end of 2007 with the University's president offers new prospects, particularly for a new computing area. Therewith, the PC² and the University of Paderborn will assure their significant regional and national position for high-performance computing in Germany – also

within the “HPC Allianz Deutschland” (German Alliance for high-performance computing).

For the first time the PC² annual report covers a period of two years. All information was compiled carefully and sorted according to topic. More than 30 contributions were collected. We thank all contributors and hope that you will enjoy reading. We are looking forward to feedback, questions and suggestions!

Prof. Dr. Holger Karl
Chairman of the PC² board
March 2008

2 Inside PC²

2.1 Board

The PC² is headed by an interdisciplinary board comprising professors from various working groups.

2.2 Members of the Board

Prof. Dr. Holger Karl (Chairman)
Faculty of Electrical Engineering, Computer Science and Mathematics

Prof. Dr. Wilhelm Dangelmaier
Faculty of Business Administration and Economics

Prof. Dr. Michael Dellnitz
Faculty of Electrical Engineering, Computer Science and Mathematics

Prof. Dr. Hans-Ulrich Heiß
Institute for Telecommunication Systems, Technical University Berlin

Prof. Dr. Odej Kao (until August 2006)
Institute for Telecommunication Systems, Technical University Berlin

Prof. Dr. Joachim Lückel
Faculty of Mechanical Engineering

Prof. Dr. Burkhard Monien
Faculty of Electrical Engineering, Computer Science and Mathematics

Prof. Dr. Marco Platzner
Faculty of Electrical Engineering, Computer Science and Mathematics

Prof. Dr. Franz Josef Rammig
Faculty of Electrical Engineering, Computer Science and Mathematics

Prof. Dr. Ulrich Rückert
Faculty of Electrical Engineering, Computer Science and Mathematics

Prof. Dr. Otto Rosenberg
Faculty of Business Administration and Economics

Prof. Dr. Wolf Gero Schmidt
Faculty of Theoretical Physics

Prof. Dr. Hans-Joachim Warnecke
Faculty of Science

Dr. Jens Simon
Paderborn Center for Parallel Computing
Assistant researchers´ representative

Dipl.-Inform. Sabina Rips
Faculty of Electrical Engineering, Computer Science and Mathematics
Assistant researchers´ representative

Dipl.-Inform. Axel Keller
Paderborn Center for Parallel Computing
Non researchers´ representative

Patrick Koch
Student representative

2.3 PC² Staff

The following people were assigned to the PC² for the period of time covered by this report.

Dipl.-Inform. Dominic Battré (until December 2006)
Dipl.-Inform. Bernard Bauer
Dipl.-Inform. Georg Birkenheuer (since April 2006)
Dr. Stephan Blazy (until August 2007)
MSc. Mariusz Grad (associated with since October 2007)
Dr. Felix Heine (until December 2006)
Dipl.-Inform. André Hoeing (until April 2007)
Dr. Matthias Hovestadt (until December 2006)

Diana-Mercedes Hunecke (Trainee until June 2007, staff member July and August 2007)
Dipl.-Inform. Dipl.-Math. Paul Kaufmann
Dipl.-Inform. Axel Keller
Michaela Kemper (Secretary)
Dipl.-Ing. Andreas Krawinkel
Dipl.-Inform. Stefan Lietsch
Dipl.-Inform. Jens Lischka (since June 2007)
Dipl.-Ing. Enno Lübbers (since January 2006)
Tanja Müller (Trainee since September 2006)
Dr. Oliver Marquardt (until September 2007)
Holger Nitsche
Dr. Christian Plessl (since September 2007)
Dipl.-Inform. Tobias Schumacher (since January 2006)
Dr. Jens Simon
Dipl.-Inform. Kerstin Voß

Within the reporting period additional support was provided by students and graduate assistants who were engaged part time (9.5 h/week and 19 h/week) in tasks which included programming, user support, system administration, etc.

Andreas Agne	Tobias Beisel	Tobias Bettmann
Hennrich Blöbaum	Alexander Boschmann	Daniel Breitlauch
Robert Breitrück	Martin Eikermann	Dominic Eschweiler
Christian Fromme	Alexander Gretencord	Stefan Grösbrink
André Höing	Diana Kleine	Christoph Konersmann
Alexander Kujat	Sven Kurras	Jens Lischka
Katrin Mätzel	Robert Meiche	Dirk Meister
Björn Meyer	Roland Mühlenbernd	Benjamin Potyka
Peter Quiel	Christian Todtenbier	Elmar Weber
Jessica Wiegelmann	Jan Henrik Wiesner	Adrian Wilke
Jakob Wisor	Veit Wittenberg	

In the year 2006/2007 the PC² employed two trainees to learn the trade of a “computer specialist” (Fachinformatiker) in the field of system integration. With the source required to employ trainees provided by the North Rhine-Westphalia government, the PC² was able to oversee this priority assignment.

3 Research and Projects

3.1 Research Areas

Research interests of the PC² are parallel and distributed large scale systems.
The focus of research is on

- Distributed & parallel applications
- Middleware & system software
- Computer architecture
- Testbeds & Benchmarking

Actual information are also presented on the web pages of PC²:
(http:/www.upb.de/pc2).

Distributed & parallel applications
Over the last years, computational power available for HPC simulations increased exponentially and so did the data that such simulations produce. Therefore scientists have to use visualization techniques in a comfortable way to analyze, understand, and interact with this large amount of generated data. The PC² is working on a seamless integration of visualization and simulation in HPC systems.
The Arminius cluster, with its tightly connected visualization and computation nodes, enables researcher to immediately see what their simulation calculates, and to plan further steps based on this information. There is no standard solution for visualization itself or the coupling between visualization and simulation. Therefore different approaches to ease the usage of these complex systems are considered. Currently the PC² is developing a computational steering platform as well as an advanced remote rendering framework. For this purpose these techniques are used for example in the driving simulator *Nighdriver* and the Computational Fluid Dynamics framework *padfem².*
Additionally other computational intensive applications are considered. Modern medicine more and more tends to do non-invasive examination before going into surgery. Methods in this field are Computer Tomography (CT), Magnetic Resonance (MR), and Ultrasonography. The medical imaging methods are using digital geometry processing to generate a three-dimensional image. Different kinds of image algorithms are needed to generate the images out of the data provided by the medical scanner systems. Obviously this is a very challenging task which also

demands a lot of computational power. In cooperation with partners the PC² wants to find new ways to accelerate this image processing and to enhance its quality. The main idea is to parallelize existing algorithms to reach the goal. But also new ways like the utilization of multi-core systems, Graphics Processor Units (GPU), or new architectures like the Cell-Processor are taken into account.

Research Topics	***Contact***	***Email***
Computation Steering, Remote rendering, Medical image processing	Stefan Lietsch	slietsch@upb.de
Acceleration of compute-intensive algorithms	Dr. Jens Simon	simon@upb.de

Middleware & system software

Like the World Wide Web the Grid will revolutionize the world of computers by offering on demand worldwide access to computing power.

The PC² has been working in this sector for years and has established itself as a competence center in the field of resource management. Current research focuses on the problem of how to guarantee the use of the resources with Service Level Agreements (SLA). This research includes the ability to perform fault tolerance mechanisms like checkpointing and migration of SLA agreed jobs and research on the assessment about likelihood of SLA violations. The combined instruments, risk assessment and fault tolerance mechanisms, allow a powerful risk aware management of grid jobs. This improves the guaranteed service quality of the resource management. In addition the PC² works on the integration of Web service based Enterprise Application Integration (EAI) into the Grid. Our aim is to combine the strengths of the two areas, loosely coupled services and the secure and easy to deploy grid infrastructures. The result will be the ability of creation and secure performing of business workflows on grid infrastructures. These fields are vital requirements for future commercial use of grid environments.

Resource Management Systems (RMS) are needed for the grid as well as for compute clusters. They allow users and system administrators to access and manage various computing resources like processors, memory, networks, or storage. PC² has developed an expandable and modular RMS, called Computing Center Software, which uses a planning based job scheduler. This OpenCCS is used in several projects and its features are continuously extended.

Research Topics	*Contact*	*Email*
Risk Assessment / Management, Service Level Agreements, Resource Management	Kerstin Voß	kerstinv@upb.de
Grid-Computing, Scheduling, Pricing	Georg Birkenheuer	birke@upb.de
Grid-based integration and orchestration of business information systems	Holger Nitsche	hn@upb.de
Computing Center Software (OpenCCS)	Axel Keller	kel@upb.de

Computer Architecture

Application-specific coprocessors can significantly accelerate many high-performance computing (HPC) applications. Designing fast accelerators and optimizing their performance remains a difficult task requiring significant hardware design expertise.

The PC² has a long term experience with innovative cluster systems based on commodity as well as on specialized hardware components. Different techniques to accelerate compute nodes are considered, like multi-core processors, graphical processing units (GPUs), and acceleration cards equipped with field-programmable gate arrays (FPGAs). Hence, vendors of supercomputers and high-performance workstations are beginning to integrate reconfigurable accelerators in their products, which makes this custom computing technology available to a broader user group. One of our missions is to make the potential of custom computing more accessible to users. To this end, we work on basic infrastructure for reconfigurable computers, specifically we work on flexible and portable communication infrastructures and on runtime systems that support dynamic reconfiguration. In several application projects we are exploring the applicability of these infrastructures by building scalable accelerators for HPC applications that exploit the performance of FPGAs and standard CPUs.

Maximizing the performance of an application consisting of many tasks is challenging, since the hardware accelerator cores affect each other when accessing shared resources. Hence, meticulous care has to be taken to avoid bottlenecks in an implementation. To support the designer with performance optimization, we are working on estimating the application's performance with a model-based approach. By combining a model of the application and a model of the execution architecture, we can study the influence of various system parameters, such as communication bandwidths and latencies, and can use this information for performance optimization.

Research Topics	*Contact*	*Email*
Multi-Objective Intrinsic Evolution of Embedded Systems	Paul Kaufmann	paul.kaufmann@upb.de
Reconfigurable Hardware Operating Systems	Enno Lübbers	enno.luebbers@upb.de
High performance custom computing	Dr. Christian Plessl	christian.plessl@upb.de
Reconfigurable architectures	Tobias Schumacher	tobe@upb.de
Computer system architecture	Dr. Jens Simon	simon@upb.de

Testbeds & Benchmarking

The development of software components for high complex networked systems requires besides analytical and simulation-based evaluation methods more and more experiments in large real live environments. One method to build a new system on top of an existing system is to use virtualization. Virtualization of resources can be found in all areas of computing. Also in the domain of networking, virtualization is used to hide the characteristics of network resources (like routers, switches, etc.) from the way in which other systems interact with them. The PC² is engaged in investigating how virtualization can be utilized as a concept in the context of building new network testbeds.

The PC² benchmarking center is specialized in evaluating the performance of high-speed networks and parallel computer systems. Typically, these are based on cluster technology. Functional parts or complete systems are evaluated with the help of so-called low-, system-, and application-level benchmarks. Derived from this evaluation new system architectures will be developed. In addition, the PC² offers assistance in finding a high performance and cost efficient solution for parallel computers for already existing application programs and those which are under development.

Research Topics	*Contact*	*Email*
Testsbed for network virtualization	Jens Lischka	jeli@mail.upb.de
System Evaluation, Benchmarking, Experimental Cluster Systems	Dr. Jens Simon	simon@upb.de

3.2 Projects

Projects finished in 2006 and 2007

EU	HPC4U	June 2004	May 2007
EU	DELIS	January 2004	December 2007
DFG	SFB 376-A3	January 2004	December 2006
DFG	ReconOS	August 2005	December 2007
DFG	Moves	October 2005	August 2007
MIWFT	VisSim	January 2005	December 2007
Intel Academic Award		December 2005	December 2006

Ongoing Projects

BMBF	D-GRID (DGI)	September 2005	February 2008
Fujitsu Siemens Computer-Cooperation		Since 1999	
Intel Cooperation		Since 2005	

Projects started in 2006 and 2007

EU	AssessGrid	April 2006	December 2008
BMBF	BIS-Grid	May 2007	April 2009
DFG	ReconOS-2	January 2008	December 2009
DFG	Moves-2	September 2007	August 2009
BMBF	Tandem	July 2007	December 2009

For current information about our projects please have a look on our web site.

3.3 Awards / Grants

- Intel Advanced Computing Center University Award, 2006, Advanced Computing Frameworks using Heterogeneous and Reconfigurable Multicore Architectures,
  Prof. Dr. M. Platzner, Dr. U. Lorenz, Dr. J. Simon

- Altera-AMD-Sun-XtremeData University Program donation, XD1000 development system, 2007,
  Prof. Dr. M. Platzner, Dr. J. Simon

3.4 Publications

Papers 2007

Tobias Schumacher, Enno Lübbers, Paul Kaufmann, Marco Platzner
Accelerating the Cube Cut Problem with an FPGA-augmented Compute Cluster
In: Proceedings of the ParaFPGA Symposium, International Conference on Parallel Computing (ParCo), Aachen/Jülich, Germany, September 2007.

Dominic Battré, Matthias Hovestadt, Odej Kao, Axel Keller, Kerstin Voss
Planning-based Scheduling for SLA-awareness and Grid Integration
In: Proceedings of the 26th W n Proceedings of the ASME 27th Computers and Information in Engineering Conference (ASME CIE07)orkshop of the UK PLANNING AND SCHEDULING Special Interest Group (PlanSIG2007), 2007

Dominic Battré, Giovanni Cortese, Felix Heine, André Höing
Accessing XML Documents using Semantic Meta Data in a P2P Environment
In: Siegfried Handschuh, Andreas Abecker, Tudor Groza (Ed.), the 20th International FLAIRS Conference, Special Track on Artificial Intelligence and Social Semantic Collaboration (SEMSOC 2007), 2007

Dominic Battré, Karim Djemame, Odej Kao, Kerstin Voss
Gaining User's Trust by Publishing Failure Probabilities
In: Grid-STP 2007, First International Workshop on Security, Trust and Privacy in Grid Systems, 2007

Dominic Battré, Matthias Hovestadt, Odej Kao, Axel Keller, Kerstin Voss
Transparent Cross-Border Migration of Parallel Multi Node Applications
In: The 7th Cracow Grid Workshop, Academic Computer Center CYFRONET AGH, 2007

Dominic Battré, Odej Kao, Kerstin Voss
Implementing WS-Agreement in a Globus Toolkit 4.0 Environment
In: Usage of Service Level Agreements in Grids Workshop in conjunction with The 8th IEEE International Conference on Grid Computing (Grid 2007), 2007

Stefan Gudenkauf, Wilhelm Hasselbring, Felix Heine, André Höing, Odej Kao, Guido Scherp
BIS-Grid: Business Workflows for the Grid
In: The 7th Cracow Grid Workshop, Academic Computer Center CYFRONET AGH, 2007

Enno Lübbers, Marco Platzner
ReconOS: An RTOS supporting Hard- and Software Threads
In: Proceedings of the 17th International Conference on Field Programmable Logic and Applications (FPL), Amsterdam, Netherlands, August 2007. IEEE.

Paul Kaufmann, Marco Platzner
MOVES: A Modular Framework for Hardware Evolution
In: Proceedings of the NASA/ESA Conference on Adaptive Hardware and Systems (AHS), 2007.

Paul Kaufmann, Marco Platzner
Toward Self-adaptive Embedded Systems: Multiobjective Hardware Evolution
In: Proceedings of the 20th International Conference on Architecture of Computing Systems (ARCS), Zurich, Switzerland, March 2007. Springer, LNCS 4415

Georg Birkenheuer, Holger Nitsche, Kerstin Voß, Elmar Weber, Peter Majlender
Gather and Prepare Monitoring Data for Estimating Resource Stability
In: The 7th Cracow Grid Workshop, Academic Computer Center CYFRONET AGH, 2007

Stefan Lietsch, Jan Berssenbrügge
Parallel, Shader-Based Visualization of Automotive Headlights
In: Proceedings of the 7th Eurographics Symposium on Parallel Graphics and Visualization (EGPGV07) - Short Paper, 2007

Stefan Lietsch, Jan Berssenbrügge
Verteilte Berechnung und Darstellung automobiler Scheinwerfer
In: Proceedings of the 6. Paderborner Workshop Augmented & Virtual Reality in der Produktentstehung (AR/VR07) - in german, 2007

Kerstin Voß, Karim Djemame, Iain Gourlay, James Padgett
AssessGrid, Economic Issues Underlying Risk-Awareness in Grids
In: Proceedings of the 4th International Workshop on Grid Economics and Business Models (GECON), hold in conjunction with EuroPar 2007, Rennes, France, August 2007, Springer Verlag.

Kerstin Voß
Comparing Fault Tolerance Mechanisms for Self-organizing Resource Management in Grids
In: Proceedings of the 3rd International Conference on Semantics, Knowledge and Grid (SKG) 2007, Xi'an, China, October 2007, IEEE Digitial Library.

Kerstin Voß
Enhance self-managing Grid s by Risk Management
In: Proceedings of the IEEE International Conference on Networking and Services (ICNS'07), Athens, Greece, June 2007, IEEE Computer Society

Stefan Lietsch, Jan Berssenbrügge, Henning Zabel
Computational Steering of Interactive and Distributed Virtual Reality Applications
In: Proceedings of the ASME 27th Computers and Information in Engineering Conference (ASME CIE07), 2007

Stefan Lietsch, Oliver Marquardt
A CUDA-Supported Approach to Remote Rendering
In: Proceedings of the 3rd Inernational Symposium on Visual Computing (ISVC2007), 2007

Papers 2006

Paul Kaufmann, Marco Platzner
Multi-objective Intrinsic Hardware Evolution
In: 2006 MAPLD International Conference, Washington D.C., USA, September 2006

Dominic Battré, Felix Heine, André Höing, Odej Kao
On Triple Dissemination, Forward-Chaining, and Load Balancing in DHT based RDF stores
In: Sonia Bergamaschi, Sam Joseph, Jean-Henry Morin, Gianluca Moro (Ed.), Fourth International Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P) in scope of 32st International Conference on Very Large Data Bases, Springer, 2006

Dominic Battré, Felix Heine, André Höing, Odej Kao
Load-balancing in P2P based RDF stores
In: Holger Wache, Heiner Stuckenschmidt, Bijan Parsia, Yuanbo Guo, Tim Finin, Dave Beckett (Ed.), Second International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2006) in scope of the International Semantic Web Conference, pp. 21-34, 2006

Dominic Battré, Felix Heine, Odej Kao
Top k RDF Query Evaluation in Structured P2P Networks
In: Wolfgang E. Nagel, Wolfgang V.Walter, Wolfgang Lehner (Ed.), Euro-Par 2006 Parallel Processing: 12th International Euro-Par Conference, Springer, pp. 995-1004, 2006

Georg Birkenheuer, Sven Döhre, Matthias Hovestadt, Odej Kao, Kerstin Voß
On Similarities of Grid Resources for Identifying Potential Migration Targets
In: Cracow Grid Workshop, 2006

Georg Birkenheuer, Karim Djemame, Iain Gourlay, Matthias Hovestadt, James Padgett, Kerstin Voß
Introducing Risk Management into the Grid
In: Proceedings of the eScience Conference (eScience2006), 2006

Felix Heine
Scalable P2P based RDF Querying
In: Xiaohua Jia (Ed.), First International Conference on Scalable Information Systems (INFOSCALE06), ACM Press, 2006

Kerstin Voß, Christian Schindelhauer
Oblivious Parallel Probabilistic Channel Utilization without Control Channels
In: IEEE International Parallel & Distributed Processing Symposium (IPDPS) 2006, p.121

Matthias Hovestadt, Odej Kao
Sicherung von Dienstgüten in Grid Systemen
In: 20. DFN-Arbeitstagung, Heilbronn, 2006

Stephan Blazy, Oliver Marquardt
Parallel Computation of Three-Dimensional Flows on Unstructured Grids
In: 6th International Heinz Nixdorf Symposium 2006

Felix Heine
Scalable P2P based RDF Querying
In: First International Conference on Scalable Information Systems (INFOSCALE06), 2006

Felix Heine, Dominic Battré, Odej Kao
Top k RDF Query Evaluation in Structured P2P Networks
In: Euro-Par 2006 Parallel Processing: 12th International Euro-Par Conference, LNCS 4128, pages 995-1004

Kerstin Voß
Risk-aware Migrations for Prepossessing SLAs
In: IEEE International Conference on Networking and Services (ICNS'06), p. 68, Silicon Valley, July 2006, IEEE Computer Society.

Matthias Hovestadt, Odej Kao, Kerstin Voß
The First Step of Introducing Risk Management for Prepossessing SLAs
In: IEEE International Conference on Services Computing (SCC) 2006, pp. 36-43, Chicago, September 2006, IEEE Computer Society.

Stefan Lietsch
First Steps Towards a Platform for Distributed, Interactive Visualization and Simulation
In: Proceedings of the 3rd High-End Visualization Workshop (2006)

Stefan Lietsch, Jan Berssenbrügge, Henning Zabel, Martin Eikermann
Light Simulation in a Distributed Driving Simulator
In: Proceedings of the 3rd High-End Visualization Workshop (2006)

Dominic Battré, Felix Heine, André Höing, Odej Kao
On Triple Dissemination, Forward-Chaining, and Load Balancing in DHT based RDF stores
In: Proceedings of Fourth International Workshop on Database, Information Systems and Peer-to-Peer Computing (DBISP2P 2006), 2006

Dominic Battré, Felix Heine, André Höing, Odej Kao
Load-balancing in P2P based RDF stores
In: Proceedings of Second International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2006), 2006, 21-34

Matthias Hovestadt, Lerch, Nicolas, Holger Nitsche, Kerstin Voß
First Steps of a Monitoring Framework to Empower Risk Assessment on Grids
In: Cracow Grid Workshop, 2006, to appear

Georg Birkenheuer, Sven Döhre, Matthias Hovestadt, Odej Kao, Kerstin Voß
On Similarities of Grid Resources for Identifying Potential Migration Targets
In: Cracow Grid Workshop 2006

Georg Birkenheuer, Karim Djemame, Iain Gourlay, Matthias Hovestadt, James Padgett, Kerstin Voß
Introducing Risk Management into the Grid
In: Proceedings of the 2nd IEEE International Conference on e-Science and Grid Computing 2006 (eScience2006), p. 28, Amsterdam, Netherlands, December 2006.

Matthias Hovestadt, Odej Kao, Kerstin Voss
The First Step of Introducing Risk Management for Prepossessing SLAs.
In: IEEE SCC, pp. 36-43, 2006

Matthias Hovestadt, Nicolas Lerch, Holger Nitsche, Kerstin Voß
First Steps of a Monitoring Framework to Empower Risk Assessment on Grids
In: Cracow Grid Workshop, 2006

Lars-Olof Burchard, Felix Heine, Hans-Ulrich Heiss, Matthias Hovestadt, Odej Kao, Axel Keller, Barry Linnert, Jörg Schneider
The Virtual Resource Manager: Local Autonomy versus QoS Guarantees for Grid Applications
In: Vladimir Getov, Domenico Laforenza, Alexander Reinefeld (Ed.), Future Generation Grids, Springer, pp. 83-98, 2006

Journal Contributions

Sven Groppe, Stefan Böttcher, Georg Birkenheuer, André Höing
Reformulating XPath Queries and XSLT Queries on XSLT Views
In: Data and Knowledge Engineering, 57, pp. 64-110, 2006

Matthias Hovestadt
Operation of an SLA-aware Grid Fabric
In: Journal of Computer Science, 2 (6), 2006

Dominic Battré, David S. Angulo
MPI framework for Parallel Searching in Large Biological Databases
In: Journal of Parallel and Distributed Computing
Volume 66, Issue 12, December 2006, Pages 1503-1511 Special Issue: Grids in Bioinformatics and Computational Biology

Dissertations

Felix Heine
P2P based RDF Querying and Reasoningfor Grid Resource Description and Matching, Universität Paderborn, 2006

Matthias Hovestadt
Service Level Agreement aware Resource Management, Universität Paderborn, 2006

Oliver Marquardt
Massiv parallele, adaptive FEM-Simulation auf Tetraedernetzen: Objektmodell, Algorithmen und Datenstrukturen, Universität Paderborn, 2007

3.4 Workshops

PC² Open Day: Schwerpunkt Grid Computing, May 2006
PC² Open Day: Schwerpunkt FPGA-Computing, October 2006
Workshop: Windows Compute Cluster Server 2003, February 2007

4 Services

4.1 Operated Parallel Computing Systems

In 2006 and 2007 the PC² operated 8 high performance computing systems and one parallel file system. The main compute resource of the PC² is the Arminius cluster. This system and the small Itanium2 cluster were available for all users. 6 of the HPC systems were dedicated to specific projects and/or working groups. At the end of 2007 an additional cluster system dedicated to the BIS-Grid project of the German Grid Initiative D-Grid was installed.

4.1.1 Publicly Available Systems

Arminius Cluster

The Arminius cluster was installed at the end of 2004 with the financial support of the state North Rhine-Westphalia and the federal republic of Germany. The system architecture was designed by Fujitsu-Siemens Computers in collaboration with PC². The system consists of 200 compute and 8 visualization nodes. The visualization nodes are tidily connected with a stereoscopic rear projection located in the PC² user center (F0.346).

The Arminius cluster is available as the main compute resource for all users of the PC² since begin of 2005.

Hardware	***Description***
200 nodes, dual socket Intel Xeon DP	400 processors, each node with • two 3.2 GHz, 1 MByte L2-cache Xeon • 4 GByte DDR2 main memory • InfiniBand SDR 4 x PCI-e HCA
8 nodes, dual socket AMD Opteron	16 processors, each node with • two 2.4 GHz, 1 MB L2-cache Opteron • 12 GByte DDR main memory • InfiniBand SDR 4x PCI-e HCA • nVidia Quadro FX 4500G PCI-e graphics card
216 port InfiniBand SDR 4x switch Fabric	central switch fabric with 18 switch modules each with 12 ports
7 TByte parallel file system	Accessible from all nodes
1 front-end	Nodes are used for compiling and starting user applications
Stereoscopic rear projection	1.80m x 2.40m screen 2 D-ILA projectors 3D tracking system

Table 1: Hardware specification of the Arminius cluster

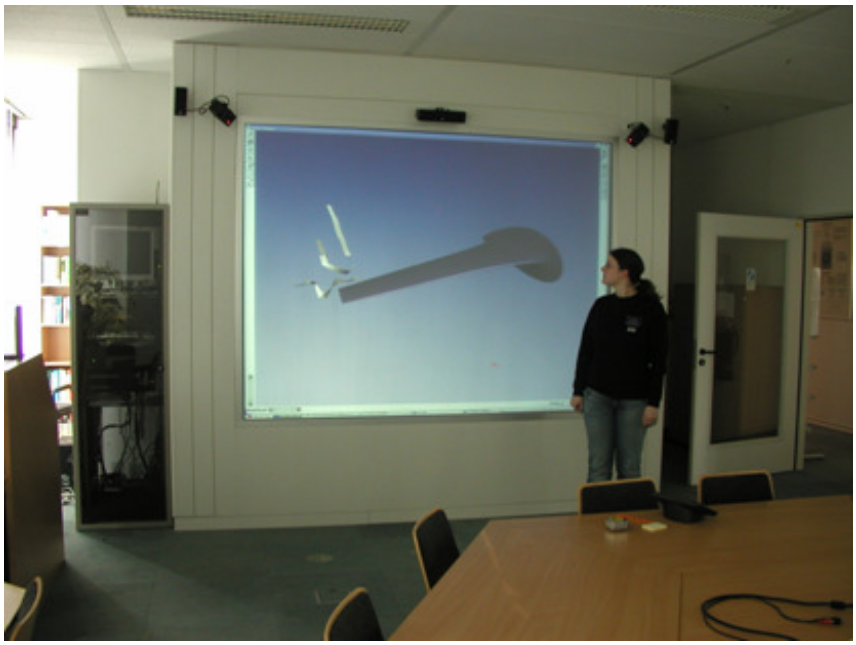


Fig. 1 The stereoscopic rear projection of the Arminius cluster

The Arminius cluster with 208 nodes (416 processors) has a peak performance of 2.6 TFlop/s. The Arminius cluster has a particular cooling system and housing. A special fluid based cooling system is used inside the system. All processors of the compute nodes have special heat sinks which are connected via a heat exchanger to the cooling system of the building. This technique is able to move 50 to 60 percent of the thermal energy directly out of the building. The rest is cooled with the air conditioner. Special cabinets and housings for the nodes are used to adopt the processor cooling technique. We are using housings for the compute nodes which allow maximum flexibility in upgrading the system with additional I/O cards, and/or acceleration cards, e.g. GPUs, FPGAs.

We provide all standard system software for cluster systems. A Linux operating system with its software development tools is installed. Additionally, some MPI message passing libraries thereunder, three MPI versions optimized for InfiniBand are available. Scientific libraries for numerical applications are available and the Intel compiler suite optimized for the Intel Xeon processor can be used. The software environment of the Arminius cluster is shown in table 2.

Software	***Description***
RedHat Advanced Server Release 5	Linux operating system 2.6.9 kernel
GNU Tools	e.g. gcc
Intel compiler	C/C++, Fortran
Scali-MPI-Connect	MPI 1 compliant, fail-over from IB to GbE
MPICH-vmi	MPICH 1.2.6 for VMI 2.1 from NCSA
MvAPICH	MPICH 1.2.6 on VAPI from Ohio State University
Intel MKL	Math Kernel Library

Table 2: System Software of the Arminius Cluster

The Arminius cluster has an excellent performance with user applications. It is able to sustain 1.978 TFlop/s out of 2.6 TFlop/s peak performance. Based on the Linpack benchmark for supercomputers, the Arminius cluster has achieved rank #205 (word-wide) and rank #13 (Germany) in the 25th list of Top-500 supercomputers (6/2006).
Parts of the compute resources of the Arminius cluster were provided for the German D-Grid Initiative. Two worldwide used Grid Computing environments, Globus and UNICORE, are available on the system.

Utilzations

The utilization of the Arminius cluster in 2006 is shown in figure 2. Downtimes of the system are marked in the figure with corresponding event numbers. The outage dates and the reason why the system was not accessible for users are listed in table 4. The average load of the system in 2006 was 79.7%.

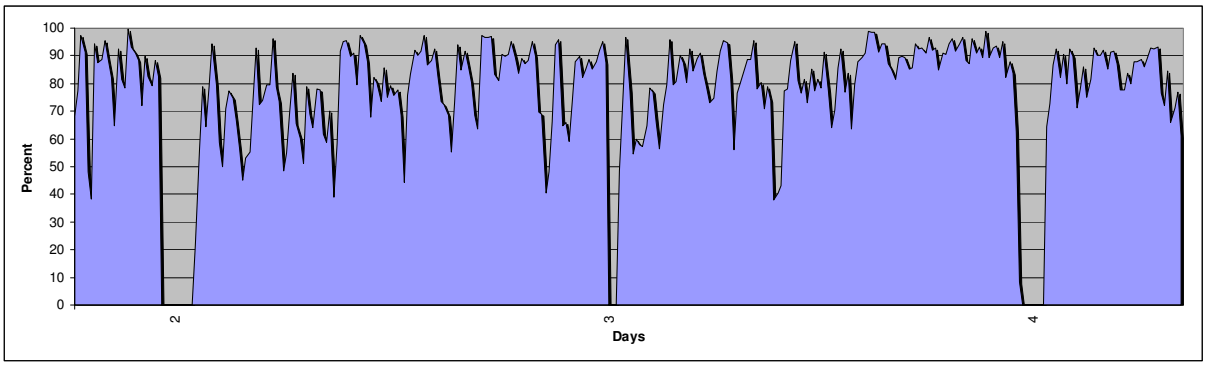


Figure 2 Utilization of the Arminius cluster in 2006 (24 hours per day).

Event	Date	Reason
1	02.01.06 – 10.01.06	Maintenance
2	02.03.06 – 13.03.06	PC² has been hacked
3	13.07.06 – 16.07.06	Maintenance
4	13.11.06 – 21.11.06	Maintenance

Table 4 Dates in 2006 were the system was not available for users.

Figure 3 depicts the utilizations of the Arminius cluster in 2007. The average load for the system in that year was 86.9%. One reason for the high system utilization was an reduction of maintenance dates (see table 5).

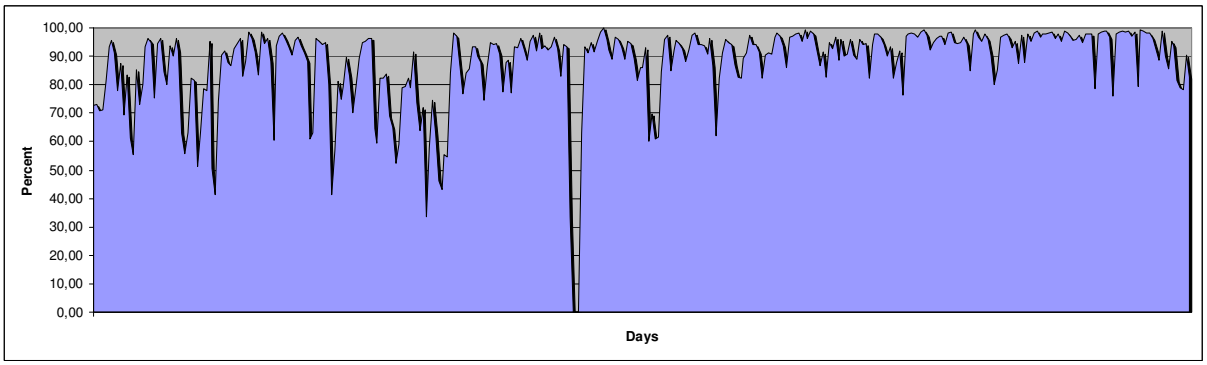


Figure 3 Utilization of theArminius cluster in 2007 (24 hours per day).

Event	Date	Reason
1	11.06.07 – 15.06.07	Maintenance

Table 5 Dates in 2007 the system was not usable.

Itanium2 / InfiniBand Cluster

Installed	2002
Vendor	Hewlett-Packard
Number of nodes / CPUs	4 / 8
Node type	HP ZX6000 (2x Intel Itanium-2, 1GHz)
Node memory	12 MByte
System memory	48 GByte
Node peak performance	8 GFlop/s
System peak performance	32 GFlop/s
High speed network type	Infiniband
High speed network topology	switched
Infiniband performance (MPI)	Bandwidth: 446 MByte/s, Latency: 7.79 µs
Myrinet performance (MPI)	Bandwidth: 270 MByte/s, Latency: 11.81 µs
Operating system	Linux
Compiler	GNU, Intel
Message Passing SW	MPICH, ScaMPI

The Itanium2 cluster is accessible via the Globus Toolkit and is dedicated to the members of the German D-Grid Initiative.

4.1.2 Dedicated Systems

Paderborner Linux Cluster Next Generation (PLING)

Installed	2003
Vendor	Hewlett-Packard
Number of nodes / CPUs	32 / 64
Node type	HP RX-2600 (2x Itanium-2 1.3 GHz)
Node memory	4 GByte
System memory	128 GByte
Node peak performance	10.4 GFlop/s
System peak performance	332 GFlop/s
High speed network type	Infiniband SDR 4x
High speed network topology	switched
Infiniband performance (MPI)	Bandwidth: 751 MByte/s, Latency: 6.51µs
Operating system	Linux
Message Passing SW	MPICH, ScaMPI
Compiler	GNU, Intel

The system is owned by and dedicated to the working group of Prof. Dr. Gero Schmidt.

BisGrid Cluster

Installed	2007
Vendor	Fujitsu-Siemens Computers / hpcLine
Number of nodes / CPUs / cores	8 / 32 / 64
Node type	Supermicro Server 2041M
Processors	Opteron 8220, 2.8 GHz, 2 x 1MB-L2
Node memory	64 GByte
System memory	512 GByte
Node peak performance	44.8 GFlop/s
System peak performance	360 GFlop/s
High speed network type	Infiniband DDR 4x
High speed network topology	switched
Infiniband performance (MPI)	Bandwidth: 2.1 GByte/s, Latency: 1.4 µs
Operating system	Linux (Windows Server)
Message Passing SW	MPICH, MvAPICH
Compiler	GNU, Intel

The system is operated for the BIS-Grid project of the D-Grid Initiative. This system will be available also for member of D-Grid.

WinCCS - Paderborner Windows Compute Cluster

Installed	2006
Vendor	FSC
Number of nodes / CPUs	6/ 12
Node type	4x Intel Xeon 5160, 3.0 GHz
Node type	2x AMD Opteron 270, 2.0 GHz
Node Memory	8 GByte
System memory	48 GByte
Node peak performance	48/16 GFlop/s
System peak performance	224 GFlop/s
High speed network type	Infiniband SDR 4x
High speed network topology	Switched
Operating system	Windows Compute Cluster Server 2003

The WinCCS cluster is dedicated to execute Windows based applications. All nodes are running the operating system Windows Server 2003.

Available Software

- Intel Compiler Suite C/C++ / Fortran
- MS MPI
- ANSYS V11.0
- Altera Quartus II 7.1
- Xilinx ISE 9.2i

4-way Opteron Cluster

Installed	2004
Vendor	AMD, Fujitsu-Siemens
Number of nodes / CPUs	2 / 8
Node type	4x Opteron 848, 2.2 GHz
Node Memory	32 GByte
System memory	64 GByte
Node peak performance	17.6 GFlop/s
System peak performance	35.2 GFlop/s
Operating system	Linux

The system is owned by the SFB 376 “Massive Parallelität, Algorithmen, Entwurfsmethoden, Anwendungen” and is dedicated to the working group of Prof. Dr. Monien.

KAO Cluster

Installed	2003
Vendor	Dell
Number of nodes / CPUs	4 / 8
Node type	2x Intel Xeon, 2.4 GHz
Node Memory	1 GByte
System memory	4 GByte
Node peak performance	4.8 GFlop/s
System peak performance	19.2 GFlop/s
High speed network type	Myrinet
High speed network topology	Switched
Disks	365 GByte SCSI-RAID
Operating system	Linux

The system is dedicated to the EU projects AssessGrid and HPC4U.

Paderborn SCI Cluster-2 (PSC2)

Installed	1999
Vendor	Fujitsu-Siemens
Number of nodes / CPUs	96 / 192
Node type	Primergy Server (2x Pentium III, 850 MHz)
Node memory	512 MByte
System memory	48 GByte
Node peak performance	850 MFlop/s
System peak performance	163 GFlop/s
High speed network type	SCI (Scalable Coherent Interface)
High speed network topology	12x8 torus
SCI performance (MPI)	Bandwidth: 84 MByte/s, Latency: 5 µs
Operating system	Linux
Message Passing SW	PVM, MPICH, ScaMPI
Compiler	GNU, Intel, PGI, Lahey-Fujitsu
Debugger	Totalview
Performance Analyzer	Vampir

The PSC2 system is dedicated to research projects e.g. HPC4U and vDrive.

4.1.3 System Access

The access to the systems at the PC² is free of charge for all users coming from the academic world e.g. universities or schools. Users from commercial sites are also welcome but may have to pay a fee for using the systems. Please contact the PC² administration phone number +49 5251 606291.
Access to systems dedicated to specific user groups may be denied depending on the requirements of the owner.
To apply for an account for the PC² systems one has to fill in small application forms available on our web server. Refer to http://www.upb.de/pc2/services/access.
After processing the application all necessary information will be sent via email within one or two days.
The registration information is kept private and will not be disclosed to third parties. It helps us to survey users of our parallel systems.
To register for a new project one has to provide:

A project description:

- The name and office address of the research manager and the project leader.
- The name and office address of each project member using the system.
- Additionally needed requirements like disk space or special software packages.

Specialist counseling is available for the following fields:

- Compiler
- Debugging
- Grid Computing
- MPI
- Optimization
- Performance Profiling
- System Access and CCS
- System-Benchmarking and -Evaluation

For detailed information about how to use our systems please also refer to this URL: http://www.ubb.de/pc2/services/systems
Please report your problems to:
pc2-gurus@upb.de or call our service number +49 5251 606303.

Available Software

Software	Purpose	Licence	Available on
Abaqus	Finite element analysis	Dedicated	Arminius
Amira	Advanced Visualization, Data Analysis and Geometry Reconstruction	PC²	Arminius
Ansys	3D FEM solvers	Dedicated	Arminius
ATLAS	Automatically Tuned Linear Algebra Software	None	Arminius, SFB
Fluent	Computational fluid dynamics	Dedicated	Arminius, SFB
FFTW	library for computing the discrete Fourier transform (DFT)	None	Arminius, SFB
Gaussian	Electronic Structure	PC²	Arminius
Gromacs	Molecular dynamics	None	Arminius, SFB
MKL	Intel Math Kernel Library	PC²	Arminius, SFB
MOE	Molecular operating environment	Dedicated	Arminius
Matlab	Technical computing	Campus	Arminius
MPICH	MPICH 1.2.6 for Ethernet	None	Arminius, SFB
MPICH-vmi	MPICH 1.2.6 for VMI 2.1 from NCSA	None	Arminius, SFB
MvAPICH	MPICH 1.2.6 on VAPI from Ohio State University	None	Arminius
NWChem	High Performance Computational Chemistry	PC²	Arminius
OpenFoam	Finite element analysis	None	Arminius
Padfem2	Finite element analysis	None	Arminius
Siesta	Electronic Simulations	PC²	Arminius,
ScaMPI	MPI 1 compliant, fail-over from IB to GbE	PC²	Arminius, SFB
StarCD	Finite element analysis	Dedicated	Arminius
VMD	Visualization	Dedicated	Arminius
VASP	Ab-initio quantum-mechanical molecular dynamics	PC²	Arminius
xmgrace	Two-dimensional plots of numerical data	None	Arminius

4.2. Collaborations

4.2.1 Ressourcenverbund – Nordrhein-Westfalen (RV-NRW)

Project coordinator:	Dr. Jens Simon, PC², University of Paderborn
Project Members:	Axel Keller, PC², University of Paderborn

The *Ressourcenverbund – Nordrhein-Westfalen* (RV-NRW) is a network of university computer centers of the state North Rhine-Westphalia which provides a network of excellence and cooperative resource-usage of high performance compute systems 1. Targets of this network are:

- Outsourcing of work besides the main focus of each computer center.
- Providing access to short and expensive resources.

Active member organisations of the RV-NRW are:

- RWTH Aachen
- University Köln
- University Paderborn
- University Münster
- University Siegen
- University Dortmund
- University Duisburg-Essen
- Ruhr-University Bochum
- Open University Hagen

In generally, all systems and services of the Ressourcenverbund are available for all scientists of RV-NRW members. The use of the resources is free of charge for this community.

Problem details and work done

The RV-NRW excellence network provides different kind of services to researchers of universities and institute of the state North Rhine-Westphalia.

Consulting HPC users: The RV-NRW provides a primary point of contact for users for all resources provided within the network. Expert advice will be provided by the appropriate compute center staff responsible for the requested resources. PC² provides all technical services and user support for its systems. Additionally, courses

and material concerning high performance computing are offered to increase the skills and qualifications of the users.

HPC systems and application software:
Several high-performance computer systems are available for the users of the RV-NRW. The Rechen- und Kommunikationszentrum of the RWTH Aachen provides a Sun Sparc processor based cluster system consisting of 24 SMP-Servers, each with up to 144 processors. The Paderborn Center for Parallel Computing of the University Paderborn provides the Arminius cluster system with 200 nodes, each node with two Intel Xeon processors. The University Siegen operates a 128 nodes cluster with two AMD Opteron processors per node.
The following centers are providing resources to RV-NRW, but they are up to now not integrated in the unified user management: The Zentrum für Informationsverarbeitung of the University Münster operates a cluster system with 120 Intel Xeon single processor nodes. The Zentrum für Angewandte Informatik of the University Köln provides a 256 nodes cluster system with AMD Opteron dual processor nodes. The Rechenzentrum of the Ruhr-University Bochum operates an out-dated 28 processor Hewlett-Packard Superdome SMP system. Finally, 224 nodes with 464 processors are provided by the University Dortmund.
Interested scientists apply for access to the RV-NRW compute resource at their local compute center.

Certificate Registration Authority: The Open University Hagen provides a Public Key Infrastructure (PKI) for an automatically issue of X.509v3 certificates. Members of the RV-NRW are free to use the dedicated certificate-server.
The University Paderborn, PC² is a registration authority for Grid certificates. In Germany two Certificate Authorities (CAs), Karlsruhe GridKa and Hamburg DFN, are established for grid services. The standard DFN certificates, used for the encryption of e-mails, can not be used for grid services.

Resource Usage
PC² provides about 30 percent of the compute resources of the Arminius cluster to users of universities and institutes of North Rhine-Westphalia. Researchers from University Münster, RWTH Aachen, and Ruhr-Universität Bochum are currently using RV-NRW accounts to access the PC² cluster system.
Further information about the RV-NRW network of excellence is available on the web-pages of the Ressourcenverbund-NRW [1].

References
[1] Ressourcenverbund Nordrhein-Westfalen (in German), http://www.rv-nrw.de

4.3. Teaching

4.3.1 Thesis and Lectures in PC²

Lectures

- Web Services (WS05/06 - Prof. Dr. Odej Kao)
- Architektur paralleler Rechnersysteme (WS05/06, WS06/07 - Dr. Jens Simon)

Project-Groups

- GOmputer – the GO machine (2006 - Dr. Ulf Lorenz, Prof. Dr. Marco Platzner, Tobias Schumacher, Dr. Jens Simon, Gunnar Steinert)
- CSI:PC2 – Biometric Computing (2007 - Prof. Dr. Marco Platzner,Tobias Schumacher, Dr. Jens Simon)

Bachelor Thesis / Studienarbeiten

- Al Bunni, Maher: Entwurf und Implementierung eines Diagnosetools für vernetzte Komponenten, 2006
- Bergen, David: Präsize Positionsbestimmung mittels wireless LAN, 2006
- Bettmann, Tobias: Integration von Checkpointing-Funktionalität mittels BLCR in das PIRANHA-System, 2006
- Böttcher, Henning: Definition und Verarbeitung von topologischen Informationen innerhalb eines Location Servers
- Brückner, Anne: Ein Framework für die verteilte Simulation regelungstechnischer Modelle
- Dau, Daniel: Entwicklung eines Location Servers für WLAN Positionierung in Java, 2006
- Doehre, Sven: Grid Monitoring als Datenbasis für die Auswahl gleicher und ähnlicher Ressourcen, 2006
- Dohle, Steffen: Globus-Integration des BabelPeer-Systems
- Dyck, Eugen: Ortsbezogene Druckdienste für PDAs
- Eschweiler, Dominic: Benchmarkprogramme zur Leistungsbewertung von Multi-Core Architekturen, 2007
- Ewers, Ralf: Access Control Listen für Router in Voice over IP Netzwerken, 2006

- Felix, Peter: Eine Public-Key-Infastruktur zur sicheren Chipkartengestützen Authentisierung auf Basis des Microsoft Active Directory
- Freitag, Thomas: Vergleich von Verfahren zur Auffindung von Diensten und Ressourcen
- Fromme, Andre: Ressourcenmanagement mit optimierten Service-Levels
- Gretencord, Alexander: Entwicklung eines Frameworks für ortsbezogene Dienste mittels CORBA Trading Service
- Haertel, Christine: Visualisierung von WS-Resource Properties und WS-Service Groups der WSRF-Spezifikation
- Herlich, Matthias: Execution Enviroment for Multi Core systems with Varying core Counts, 2007
- Jaeger, Thomas: Integration des PIRANHA-Systems in das Globus Toolkit 4, 2006
- Koch, Philipp: Entwurf eines Replikationskonzeptes mit SQL Server
- Krieger, Alexander: Videostreaming von OpenGL-Applikationen auf einem HPC-Cluster, 2007
- Lerch, Nicolas: Server Monitoring und Leistungmessung mit Nagios, 2006
- Mense, Holger: Einführung des Network Itrusion Detection Systems „Snort“
- Müller, Stephan: Konzeption und Entwicklung eines alternativen WLAN Zugangsverfahren für die Universität Paderborn, 2006
- Niehörster, Oliver: Ortsabhängige Zugriffssteuerung auf Dokumentendienste 2006
- Oertel, Markus: Entwicklung und Umsetzung eines identitätsbasierten Verfahrens zur Abwehr von unerwünschten E-mails (Teil 2), 2006
- Paul, Oxana: Untersuchung der Effizienz von relationalen Datenbanken zur Speicherung von XML-basierten Service Level Agreements
- Riemann, Tim: MPI-Kommunikation für qualitätsbasierende verteilte Algorithmen
- Rikanovic, Igor: Klassifizierung von Positionierungstechniken für Location-based Services
- Stobbe, Boris: Lastbalancierung im Peer-to-Peer Netzwerk mittels Sampling, 2006
- Stolzenhain, Till: Prognosemodell für Serverleistung am Beispiel PRIMERGY, 2006
- Swoboda, Tobias: Processing Area Network (PAN) Monitoring in heterogenen Umgebungen, 2007
- Twiste, Nicola: Konzeption und Implementierung eines verteilten Agentensystems zur Synchronisation von Verzeichnisdiensten am Beispiel Microsoft Active Directory
- Warneke, Daniel: Entwicklung und Umsetzung eines identitätsbasierten Verfahrens zur Abwehr von unerwünschten E-mails (Teil 1), 2006

- Wasfi, Jaradat: Entwicklung und Realisierung einer verteilten Umgebung für den Mehrgrößenoptimierer MOPO 2006
- Zhi, Yuan: Positionsbestimmung in WLAN-Umgebungen durch Signalausbreitungsmodellierung 2006

Master Thesis / Diplomarbeiten

- Beisel, Tobias: Entwurf und Evaluation eines parallelen Verfahrens zur Bildrekonstruktion in der Positionen-Emissions-Tanographie auf Multi-Core-Architekturen, 2007
- Birkenheuer, Georg: A Framework for Semantic Web Service Development, 2006
- Eikermann, Martin: A flexible Framework for Computational Steering, 2007
- Finke, Stefan: Einigungsprozesse mit Nachweispflicht auf XML-Basis – Ein Konzept zur zertifikatsbasierten Definition, Aushandlung und Archivierung von Vertragsdokumenten, 2006
- Flachsel, Michael: Aufbau und Evaluierung eines Microsoft Windows Compute Cluster Server 2003 basierten HPC Systems, 2006
- Fromme, Christian: Konzeption und Entwicklung eines automatischen Inventarisierungs- und Lizenzsystems mit Ersatzteilverfolgung, 2006
- Gerdemann, Sebastian: Analyse der Backboneauslastung und Bandbreitenplanung, 2006
- Hagemeier, Björn: Automatische, dynamische Abbildung und Ausführung Grid-basierter Workflow mit Unicore, 2006
- Höing, André: Lastbalancierung von Speicherbedarf und Anfragen in BabelPeers, 2006
- Lerch, Nicolas: Konzeption und Entwicklung eines Monitoring Frameworks für Grid-Umgebungen, 2007
- Müller, Stephan: Etablierung lokaler Sicherheitsmechanismen im Kontext einer Grid-Integration von Ressource Management Systemen, 2007
- Schumacher, Tobias: Untersuchung von Kommunikationsmethoden zwischen FPGA-basierten Systems-on-Chip auf Basis von Message-Passing, 2006
- Thiele, Kai: Konzept und Entwicklung von kontextbasierten Diensten in Wireless-LAN Umgebungen, 2006
- Thygs, Manuela: Swapping Mechanismen für Smart Cards, 2006
- Todtenbier, Christian: Kompatibilitätsaspekte bei der Migration von Checkpoint-Datensätze, 2006
- Pieper, Carsten: Kopplung von Simulation und Visualisierung am Beispiel von Chromium, 2006
- Wan, Kan-Sing: Ontology-Based Resource Matching, 2006

- Weking, Michael: Erfassung und Visualisierung der physikalischen Topologie von ethernetbasierenden Netzwerken in der Automatisierungstechnik, 2006
- Wittenberg, Veit: Ein Konzept zur Verteilung shader-basierter Visualisierung auf einem Hochleistungsrechner, 2006
- Wolf, Björn: Approach for evaluating and selecting an efficient storage system for a customized environment, 2007

PhDs Thesis

- Dr. Felix Heine: “P2P based RDF Querying and Reasoning for Grid Resource Description and Matching”, 2006
- Dr. Matthias Hovestadt: “Service Level Agreement aware Resource Management”, 2006
- Dr. Oliver Marquardt: "Massiv parallele, adaptive FEM-Simulation auf Tetraedernetzen: Objektmodell, Algorithmen und Datenstrukturen", 2007

PC² Colloquium

- Bahr, Andreas (Studienarbeit am PC²): Realisierung virtueller Organisationen mit Hilfe des "Virtual Organisation Membership Service" (VOMS)
- Breitrück, Robert (PC²): Exakte Stromlinien auf unstrukturierten Tetraedernetzen
- Härtl, Christine (Studienarbeit am PC²): Entwicklung eines Clients zum WS-Agreement als Anwendung der WSRF-Spezifikation
- Keller, Axel (PC²): PC2-Workshop: CCS: Tipps und Tricks
- Keller, Axel und Simon, Dr. Jens (PC²): Vorstellung des neuen PC² Cluster-Systems
- Klevhag, Jonas und Möhl, Stefan (Mitrionics AB): FPGA-Programming
- Lerch, Nicolas: Server Monitoring und Leistungsmessung mit Naigos
- Lietsch, Stefan (PC²): Grundlagen und Einsatzgebiete eines Hochleistungs-Visualisierungs-Clusters
- Mense, Holger (Universität Paderborn): Network Intrusion Detection am Beispiel von Snort (Studienarbeit)
- Ndong, Steeve (Diplomarbeit am PC2): Entwicklung einer erweiterbaren Architektur für das Tool ccsMon
- Paderborn Center for Parallel Computing: Tag der offenen Tür
- Porta, Dr. Juan José (IBM): Broadband Processor Architecture: power-efficient and cost-effective high-performance processing for a wide range of applications

- Specovius, Prof. Dr. Maria (Universität Kassel): Wie man Randwertprobleme auf PC-Grösse zuschneidet
- Teilnehmer der Projektgruppe: Abschlussvortrag Projektgruppe PeerThing
- Tobias Schumacher (PC²): Untersuchung von Kommunikationsmethoden zwischen FPGA-basierten Systems-on-Chip auf Basis von Message-Passing

4.3.2 PhD at PC²

Dr. Felix Heine: "P2P based RDF Querying and Reasoningfor Grid Resource Description and Matching"

Abstract:

In this thesis, we look at the problem how resources in large, heterogeneous Grids can be discovered. In world-sized Grids, two aspects of the **resource discovery** problem become especially important: **heterogeneity** and **size**. **Heterogeneity** means that the types of resources included in the Grid are highly diverse.Additionally to traditional resources like high performance clusters and storage devices, any kind of service including arbitrary applications and expensive physical instruments aretreated as resources. No single standard can encompass any resource to be described.

As soon as there are multiple standards, additional knowledge is needed to mediate between these standards. **Size** means that a scalable solution to the resource discovery problem is needed. Although there are numerous reasoning systems, they typically assume that all knowledge is collected at a single system, which is infeasible for arbitrary large collections of information.

We present a system that contributes the initial steps to the solution of the described problem. Although Grid computing is the motivating application, the scope of this thesis is larger. Its goal is to provide **a scalable approach to combine and query large-scale collections of machine-readable information**.

The main elements of the thesis are the description of the system architecture based on a structured p2p network, a dissemination algorithm that places information on well-defined nodes, a reasoning mechanism that derives new knowledge from the existing, combining information which originates from different nodes, and two query evaluation strategies.

The first evaluation strategy aims to extract all matches for a given query. We describe various strategies to minimize the network load. However, in case of queries with large result sets, an exhaustive evaluation is infeasible. Thus we present a second strategy targeting queries with a huge number of results that retrieves only a restricted number of results according to some sorting criterion. We use caching and look ahead strategies to make the algorithm efficient.

We have implemented the system prototypically. Using this implementation, we perform various experiments both on a simulation base and using real test runs to show the efficiency of the system.

Keywords
Grids, Resource Description, Resource Matching, Semantic Grid, Information Integration, Semantic Web, RDF, RDFS, Reasoning, SPARQL, Query Processing, Top k Query Processing, Scalability, Data Organization, P2P, DHT, Structured Overlay Networks

Dr. Matthias Hovestadt: "Service Level Agreement awareResource Management"

Abstract:

Next Generation Grids aim at attracting commercial users to employ Grid environments for their business critical compute jobs. These customers demand for contractually fixed service quality levels, ensuring the availability of results in time In this context, a Service Level Agreement (SLA) is a powerful instrument for defining a comprehensive requirement profile.

Numerous research projects worldwide already focus on integrating SLA technology in Grid middleware components like broker services. However, solely focusing on Grid middleware services is not sufficient. Services at Grid middleware may accept compute jobs from customers, but they have to realize them by means of local resource management systems (RMS). Current RMS offer best-effort service only, thus they are also limiting the service quality level the Grid middleware service is able to provide.

In this thesis the architecture and operation of an SLA-aware resource management system is described, which allows Grid middleware components to negotiate on SLAs. The system uses its internal mechanisms of application-transparent fault tolerance to ensure the terms of these SLAs even in case of resource outages. The main parts of this work focus on scheduling aspects and strategies for ensuring SLA compliance, respectively design aspects on implementation. Scheduling strategies significantly determine the level of fault tolerance that the system is able to provide. After presenting requirements of Grid middleware components on service qualities and a description of operation phases of an SLA-aware resource management system, intra-cluster scheduling strategies are described. Here, the system solely uses its own resources and mechanisms for coping with resource outages.

For further increasing the level of fault tolerance, strategies for cross-border migration are presented. Beside a migration to other cluster systems in the same administrative domain, the system uses also Grid resources as migration targets. For ensuring the successful restart, mechanisms for describing the compatibility profile of a checkpointed job are presented.

The concept of the SLA-aware resource management system has been implemented in the scope of the EC-funded project HPC4U. We will describe design aspects of this realization and show results from system deployments at use-case customers.

Keywords

Grid Computing, Grid Middleware, Grid, Resource Management, Service Level Agreement, SLA, Negotiation, Fault Tolerance, Scheduling, Process Checkpointing, Storage Snapshot, Network Fault Tolerance, Quality of Service, QoS

Oliver Marquardt: "Massiv parallele, adaptive FEM-Simulation auf Tetraedernetzen: Objektmodell, Algorithmen und Datenstrukturen"

Abstract:

The research and analysis of partial differential equations is the basis for fundamental understanding of natural phenomena, practical science procedures and industrial applications. The numerical solution of such applications based on partial differential equations on unstructured grids in two or three dimensions is one of the most important problems in mathematical computation and simulation nowadays. To obtain reasonable solutions the approximation usually involves a large number of unknowns. Thus, it can only be solved in a reasonable amount of time by utilizing (massively) parallel computer systems with large memory space in a reasonable amount of time.

One of the main problems of numerical simulation software running on parallel computer systems is scalability and effcient usage of such a system. Therefore, a distributed data and object model especially designed for massively parallel finite element applications is presented in this thesis. The main characteristic of this object and data model is a local namespace usage for all elements within a partition of a distributed mesh. Mesh consistency on partition boundaries is automatically maintained by the distributed object model itself.

The object and data model of a distributed mesh is a key component for a simulation environment, because it connects the three main modules of a numerical simulation software, namely the numerical module, the (geometric) adaptation module and the workload balancing and data migration module. All these modules work on their own data model. Thus, an efficient conversion technique between these data models is required. The distributed object and data model developed in this thesis, offers an efficient and scalable approach for this important requirement in parallel simulation applications. The utilization of this mechanism is presented in detail for all three modules. Mesh modifying modules like the adaptation and the migration module represent bottlenecks for the efficiency of the data structure in the object model. For this reason, two algorithms for these modules, both especially developed for massively parallel usage and working on the distributed data and object model, are introduced in this thesis. The geometric adaptation algorithm is based on irregular refinement and extended with an additional set of rules for quality conservation of element shapes. The migration algorithm works efficiently on large distributed meshes and provides an automatic scalable partition boundary reconstruction, which maintains the local namespace consistency requirement.

To evaluate and verify the object model and the algorithms working on it, a practical implementation in the framework padfem2 has been carried out. Several artificial benchmark sets are used for analysis of the three main modules and the results are presented in this thesis. Finally, a comprehensive numerical simulation benchmark

for computational fluid dynamics is evaluated within the padfem2-environment to proof the efficiency of the developed framework including object model, data structures and algorithms.

4.3.3 Project Group: GOmputer – The GO Machine

Project coordinator	Prof. Dr. Marco Platzner, University of Paderborn
Project members	Dr. Jens Simon, PC², University of Paderborn
	Dr. Ulf Lorenz, University of Paderborn
	Gunnar Steinert, University of Paderborn
	Tobias Schumacher, PC², University of Paderborn

General Problem Description

GO is an extremely popular board game, especially in the Asian world. The rules of this game are rather simple; however until now no computer program exists that is able to beat a strong GO player. Current chess programs normally use a technique called α/β-pruning for searching the best move a player can perform. However, this algorithm is rather time-consuming, especially for wide search trees. While in chess the number of moves possible in a certain situation is rather limited, in GO the player can almost freely place a tile on a board. Therefore, current GO programs usually do not use any search algorithms for finding out the best move but really only on a static evaluation of the current board position.

On the other hand, there are several possibilities to accelerate a search algorithm like α/β-pruning. The related Hydra project implements a chess game that uses a parallel version of the α/β-algorithm for accelerating the search and therefore provides a greater search depth. Additionally, the board evaluation is accelerated by using reconfigurable hardware. This project is very successful and provides one of the strongest chess programs in the world.

The goal of the project group “GOmputer” was to adopt these techniques used in Hydra to the game of GO. Several subgoals were defined:

- implement a program that plays the game of GO
- perform a parallel game tree search using α/β-pruning for finding the best move
- accelerate the final board evaluation in the leaves using FPGAs
- implement a GUI to interact with the GOmputer

Problem Details and Work Done

The project group was split up into three subgroups:

- GO-Base/Cluster (Baster) implemented the core application performing a parallel α/β-search on a cluster and software evaluation of the board in the leaves of the search tree
- GO-FPGA implemented the FPGA-based board evaluation
- GOOI developed the graphical front end for user interaction with the GOmputer

After one year of development, the GOmputer now is in its final stage. The GO-Baster is based on the Message Passing Interface MPI and therefore is able to run on a wide range of parallel computers. Benchmarks were performed using up to 128 processors. A speedup could be observed for partitions of up to 16 processors, depending on the search depth performed. Comparable results can also be observed in similar projects.
The Baster-group also implemented several board evaluation features in software, namely:

- Influence
- Piece-Square-Table
- Dragons
- Holes

Details to these features and their implementation can be found in the project documentation [1]. The impact of these features for the final evaluation result can be adjusted by a wide range of parameters. These parameters are calibrated using the graphical frontend (GOOI). GOOI connects to a database that provides a large number of games played by DAN-level players before. Several random moves are taken out of this database and transmitted to the Baster for evaluation using different parameter sets. A parameter set that performs moves very similar to those the DAN-player performed before gets a good ranking.
The GOOI connects to the Baster by a text based protocol using standard I/O. This provides a great flexibility – Baster can be directly started through GOOI on the same machine as well as on any other machine accessible by ssh or rsh. The command for starting the Baster is configurable so that the user can directly connect to a remote machine, allocate a partition in a reservation system like CCS on that machine and start the Baster in one step.

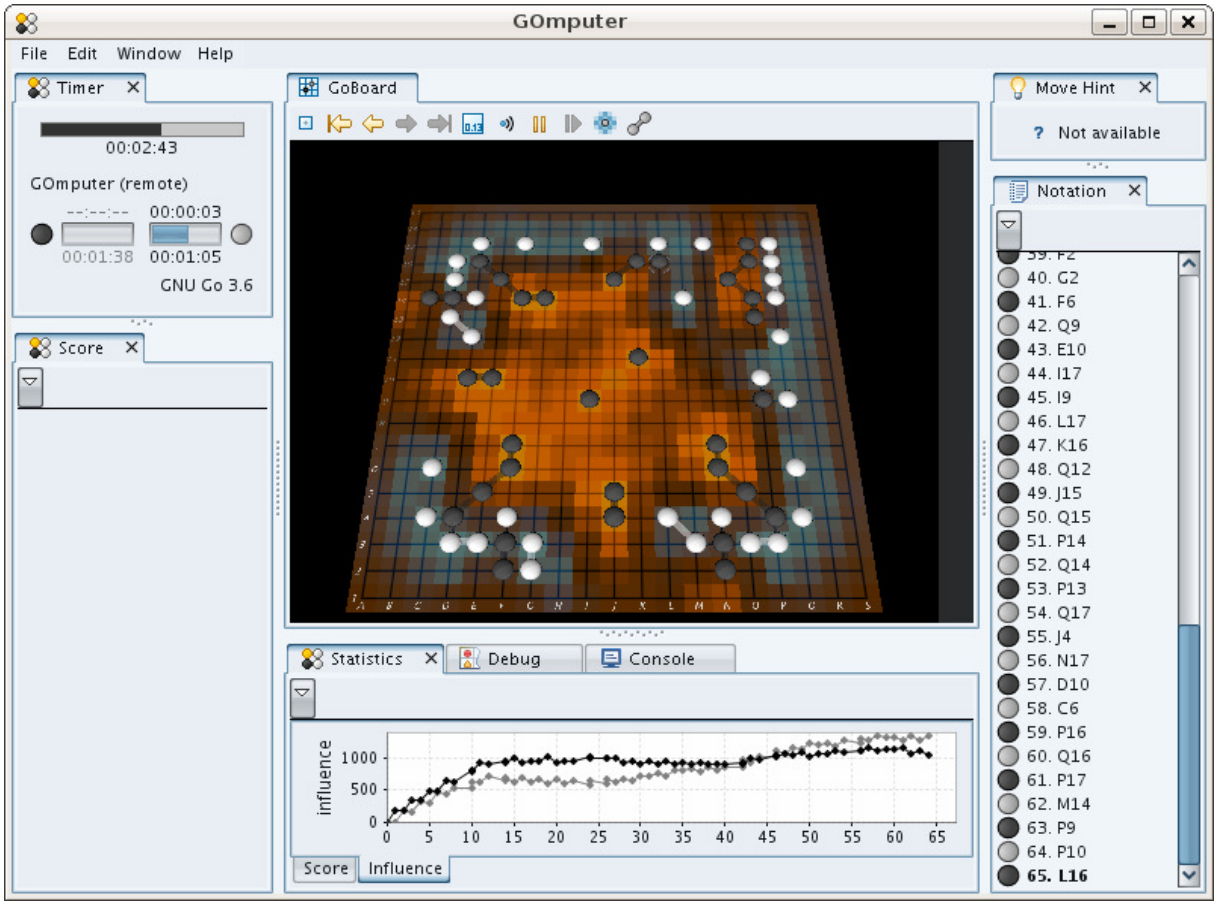


Figure 1: The graphical user interface visualizing the influence feature

Additionally to connecting to the Baster, GOOI provides the possibility to connect to third party GO games like GNU-GO. This way, the GOOI can not only be used for games with a human player playing against the GOmputer Baster, but also for human vs. GNU-GO or even GOmputer vs. GNU-GO games. Lots of other features are provided by the GOOI, especially for analyzing the search tree and evaluation functions of Baster. Details on these features can also be found in the project documentation [1].

The FPGA group successfully implemented the evaluation functions provided by Baster in an FPGA. The target platform for this work was the AlphaData ADM-XP FPGA board available in four nodes of the Arminius cluster. The board is connected to the host computer using a 64bit PCI bus running at 66MHz and provides a Xilinx Virtex II-Pro 2VP70 FPGA.

One major challenge that arised during the implementation was the size of the FPGA. The influence was the only feature that could be implemented in this chip for the maximum board size of 19x19. The logic for the other features was too complex to fit into the chip for a full-sized board. However, the FPGA implementation provides

a good speedup for every feature, so porting this work to other architectures providing more and/or bigger FPGAs seems to be a desirable step.
Additionally, the FPGA group implemented several other evaluation features in hardware that are better suited for the existing FPGAs. Small-holes is a simplified variation of the original holes feature that only takes up 64% of the logic resources available. The other new feature is the Euler value whose implementation only takes 24% of the logic resources. Again, details on these features, their implementation and speedups achieved can be found in the project documentation [1].

Resource Usage

The project was mainly developed using the Arminius cluster at PC². Baster was run on up to 128 Processors of this machine, using the Infiniband network for communication. When using the FPGA accelerated version of the GOmputer, the FPGA island of Arminius was used.. Additionally several development and testing systems equipped with Infiniband and FPGA boards were used.

References

[1] Abschlussbericht der Projektgruppe GOmputer, PC² November 2007

4.3.4 Project Group: CSI PC² - Biometric Computing

Project coordinator	Prof. Dr. Marco Platzner, University of Paderborn
Project members	Dr. Jens Simon, PC², University of Paderborn
	Tobias Schumacher, PC², University of Paderborn
Supported by:	TST Biometrics

General Problem Description

Identification of individuals by biometric properties is a topic of major interest in many different areas today. For example, fingerprints or iris scans can be used by the police for identifying offenders. Another field of application for these techniques is to control access to rooms, buildings or computer systems. All these systems roughly work in the same way. First, a fingerprint or iris scan is taken from a person. Several identification features are extracted from this scan and saved to a database. When someone now wants to enter a room that is controlled by such a biometric system, his fingerprint or iris scan is taken again, the features are extracted and compared to those in the database. Several challenges arise in this step:

- The scans can include a certain noise, so one cannot expect an exact match to the features stored in the database. The algorithms used must take this into account and provide a good hit ratio while minimizing the number of false positives.
- Depending on the area of application, the database can become very large, which directly impacts the performance. In most application areas, it is desirable to get the results fast. For example, if one wants to enter a room or building, he does not want to wait for minutes until the system checked his authorization.

These problems directly lead to the goals of this project. We want to implement a biometric system that maximizes the hit ratio while minimizing the number of false positives returned. Additionally, the system should provide a small response time.
These goals shall be achieved by using a parallel computer for the database matching, where the database will be stored on a parallel file system. An additional speedup for the matching algorithm will be gained by using reconfigurable hardware for the matching. The initial implementation will be done on the Arminius cluster operated at PC². This machine provides a parallel file system, powerful compute nodes, a fast communication network and several nodes equipped with additional

FPGA boards. Therefore, it is an ideal platform for development. However, the system will be designed as portable as possible.
The user frontend to the system will be a separate PC providing a graphical user interface that communicates with the cluster via a socket based protocol. The frontend is equipped with a fingerprint sensor and an iris camera.

Problem Details and Work Done

The project group is still ongoing. Conforming to the project plan, the first milestone was successfully achieved. For the frontend side this means that the fingerprint sensor and the iris camera can be accessed. The data received can be processed, and relevant features are extracted. The GUI is also operational and the system can send the features extracted to the cluster for matching them against the database. A screenshot of the GUI is shown in Figure 1. Work that still needs to be done consists of improving the algorithms for feature extraction as well as parameter and performance tuning.

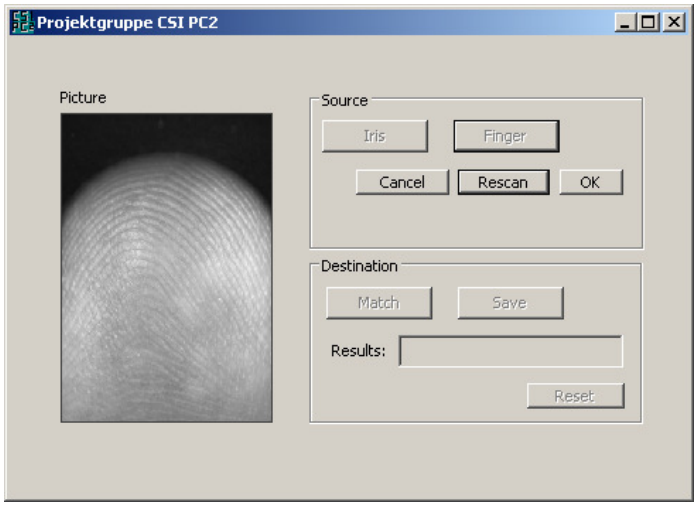


Figure 2: The graphical user interface

The central parts of the parallel database matching are also operational. The database format has been defined; methods exist for efficiently reading from and writing to the database. A cluster control daemon exists receiving connections from the frontend and distributing the work to be done. A first version of the matching algorithms has also been implemented. The next steps will include the FPGA implementation of the matching algorithms. Here we hope to achieve great speedups, because FPGAs can perform lots of feature comparisons in parallel as opposed to the sequential nature of a general purpose CPU. To sum up, the complete system is already in an initial working state.

Resource Usage

The main work of the project group is done on the Arminius cluster. Most of the features this machine provides are utilized, for example the high speed network, the parallel file system and the FPGA boards available in four of the compute nodes. The communication between the compute nodes is implemented using MPI. For testing purposes, two other machines (ic13 and ic14) are also used, which provide an Infiniband network and two FPGAs per node.

The frontend is a dedicated computer running Microsoft Windows. Connected to this machine are a fingerprint sensor and an iris camera. To simplify user interaction, the system provides a touch screen.

Additionally, the project group will likely use the Windows CCS computer cluster for generating artificial fingerprints needed for performance benchmarks. The program used for this task is not a parallel application and therefore cannot directly utilize multiple processors or even multiple compute nodes and a high speed network, but it is possible to start the program with different initial parameters on different processors.

The project was partially supported by TST biometrics who donated a fingerprint sensor and the APIs available.

References

[1] http://www.csipc2.de

5 User Projects

5.1 Parallel Architectures

5.1.1 System Evaluation, Benchmarking and Operation of Experimental Cluster System

Project coordinator	Dr. Jens Simon, PC², University of Paderborn
Project members	Axel Keller, PC², University of Paderborn
	Andreas Krawinkel, PC², University of Paderborn
	Holger Nitsche, PC², University of Paderborn
Work partly supported by	Fujitsu-Siemens Computers, Intel, MyriCom, Voltaire

General Problem Description

In the years 2006 and 2007, PC² has installed two cluster systems. One system consists of 6 compute nodes with 24 processor cores and the other system consists of 8 powerful nodes with 64 processor cores, 512 GByte main memory, and 10 TByte parallel file system. The later system is financed by the BMBF project BIS-Grid and the other system was bought within an industrial cooperation. Emerging technologies, computer systems, interconnects, and software systems are evaluated by the PC² for these installed cluster systems and further next generation systems.
Besides system evaluation and benchmarking of new cluster technologies, different experimental or special purpose cluster systems are operated for research groups of the University of Paderborn.

Problem details and work done

Different computer systems and cluster technologies have been evaluated. The tested systems are up-to date two sockets Intel Xeon systems with single, dual, and quad core processors, two and four sockets AMD Opteron with single and quad core processors, and some special purpose computer systems with reconfigurable hardware. These systems were supported in different configurations of high-speed interconnects (InfiniBand single- and double data rate, Myrinet 2 Gb/s and 10 Gb/s, and Gigabit Ethernet) and different operating systems of Linux and Microsoft Windows Server 2003. All benchmarking results are published on the web sides of the Paderborn Benchmarking Center 0.

Co-operations: The PC² benchmarking center is also doing system evaluation and benchmarking for external companies and organizations. The PC² has a long term co-operation with Fujitsu-Siemens Computers where Paderborn acts as a Competence Center for High Performance Computing. System benchmarking is also done for the company ICT Solutions AG. We presented our performance results on the International Supercomputer Convergence (ISC2006). Also workshops and user meetings have been visited to present the system architecture and the performance benefits of the Arminius cluster system.

Company	Provided Equipment
Ict AG	Loan equipment 6 two sockets dual core Opteron systems 6/06 - 10/06
Voltaire	Loan equipment 6 IB NICs, InfiniBand switch 6/06 – 8/06
	Loan equipment InfiniBand Switch Fabric 24p DDR 4/07 – 10/07
MyriCom	Loan equipment 4 NICs MyriNet 10G and 8 ports switch 1/07 – 3 /07
	2 NICs MyriNet 10G Since 9/07

References

[1] Simon, J., PC² Benchmarking Center, http://wwwcs.uni-paderborn.de/pc2/about-us/staff/jens-simons-pages/benchmarkingcenter.html

5.1.2. Evaluation of Microsoft Windows Compute Cluster Server

Project coordinator	Dr. Jens Simon, PC², University of Paderborn
Work supported by	Fujitsu-Siemens Computers, Microsoft Cooperation

General Problem Description

After first successful tests with Microsoft Compute Cluster Server 2003 in the year 2005/2006 PC² has built a Windows cluster with 24 processor cores and InfiniBand interconnect [1]. Now, different applications are under evaluation.

Problem details and work done

Windows CCS is made up of several tools layered onto the standard Windows Server 2003. The clustering tools are using existing Microsoft technology like Remote Installation Services (RIS), Microsoft Management Console with a cluster management console plug-in, and Active Directory for network-wide authentication.

The Windows CCS Architecture consists of several required and optional components. The optimal deployment can be configured depending on the applications being run on the cluster system. For deployment and management of the cluster a dedicated computer system is needed.

The head node provides user interfaces for administration as well as management services for all compute nodes. Management services include job scheduling and resource management. Automatic compute node deployment can be done by Remote Installation Services (RIS). Internet Connection Sharing (ICS) and Network Address Translation (NAT) can be used to connect private compute nodes with public networks. The head node can also provide the Dynamic Host Configuration Protocol (DHCP) and Domain Name System (DNS) services. For authorization and authentication each node of the cluster must be a member of a special Active Directory domain.

The compute nodes provide the computational resources of the cluster. Heterogeneous configurations of compute nodes are supported as long as each node fulfills the hardware and software requirements. In the Windows cluster system of the PC² six compute nodes with 4 processor cores each are available.

The Job Scheduler is running on the head node and coordinates the execution of jobs on the compute nodes. The Scheduler manages the job queues and all resource allocations. The execution of jobs is done by the local Node Manager Service.

The Message Passing Interface (MPI) is integrated in the Windows CCS to provide an environment for the execution of parallel applications. MPICH-2 from the Argonne

National Laboratory is used which utilizes any Ethernet connection as well as the high-performance communication networks InfiniBand and MyriNet. Winsock Direct driver is used as the network protocol. Therefore in principle all networks with Winsock Direct support can be used as cluster interconnect.
Multiple network interfaces are supported. Especially a private, high-speed interconnect dedicated for message passing purposes can be configured. We have installed the InfiniBand interconnect on all compute nodes and some compute nodes are additionally equipped with a MyriNet interconnect. The Gigabit-Ethernet is for all standard network traffic and the high-speed interconnects InfiiniBand and MyriNet are dedicated for MPI traffic.
In this project, the PC² evaluates the usability and performance of Windows CCS for different commercial and user applications. Together with the University of Applied Science Lemgo the rendering software 3D Studio Max is used in a project where small movies are photo-realistically rendered. Electrical engineers and computer engineers from University of Paderborn are using software from the companies Xilinx and Altera to develop circuits for programmable hardware (FPGAs) and application-specific chips (ASICs). And the commercial software ANSYS is used from researchers of the Heinz-Nixdorf-Institute, Mechantronics and Mechanics for their simulations of structural mechanics. Besides the application tests on Windows CCS several benchmarks of kernel programs and MPI programs are done [2].

Co-operations: PC² takes part in the Microsoft test program for the Windows CCS. Since mid-2006 a stable version of Windows Compute Cluster Server is available. Besides standard Gigabit-Ethernet networks, the high-speed interconnect InfiniBand from QLogic and Voltaire and MyriNet 2G and MyriNet 10G are used [2].

Resource Usage
In this project the dedicated Windows CCS cluster with 6 compute nodes is used.

References
[1] Windows Compute Cluster Server 2003,
http://www.microsoft.com/windowsserver2003/ccs/
[2] Simon, J., PC² Benchmarking Center,
www.upb.de/pc2/about-us/staff/jens-simons-pages/benchmarkingcenter.html

5.1.3 Operating Systems for Reconfigurable Hardware (ReconOS)

Project coordinator	Prof. Dr. Marco Platzner, University of Paderborn
Project members	Enno Lübbers, PC², University of Paderborn
Work partly supported by	DFG SPP 1148 – Rekonfigurierbare Rechensysteme

General Problem Description

The project ReconOS aims at the investigation and development of a programming and execution model for dynamically reconfigurable hardware devices. These devices, such as Field-Programmable Gate Arrays (FPGAs), are being used more and more in embedded systems but also as accelerators in general purpose computing. Being a rather new technology, dynamically reconfigurable systems are not sufficiently supported by current design methodologies and tools. Especially the unique feature of partial reconfiguration is still difficult to exploit except for hand-crafted designs.

Extending the established programming model of existing real-time operating systems to the hardware domain would greatly simplify module reusability and design-space exploration for hybrid hardware/software systems. At the same time, it would provide designers with a well-known model for specifying thread synchronization and communication that can be transparently used for hardware and software design.

The main challenges in adapting this programming model to the hardware domain lie in

- identifying which operating system objects' semantics can be extended to hardware threads
- how these objects and their underlying mechanisms have to be modified to support both hard- and software
- to what extend these implementations can be customized or supported by reconfigurable hardware to meet the requirements of specific embedded applications, such as low memory footprints and real-time performance

Problem Details and Work Done

We have extended the popular real-time operating system eCos with support for hardware threads running on reconfigurable devices, such as FPGAs. eCos is highly configurable and provides a flexible bases for customizing our operating system to application requirements. Building on the existing eCos kernel API and POSIX

compatibility layers allows us to seamlessly extend the programming model used for traditional software-based embedded systems to the hardware domain. We provide transparent mechanisms that permit hardware and software threads to use the same OS services for communication and synchronization. Two communicating threads use only operating system objects and thus do not need to know the execution environment (hard- or software) of their communication partners. This greatly simplifies design space exploration and thread reusability for hybrid hardware/software systems.

A ReconOS hardware thread is usually written in a hardware description language like VHDL and consists of at least two parallel processes, an OS synchronization state machine and one or more user processes. Figure 1 shows the basic structure of a hardware thread. The OS synchronization state machine specifies the OS interactions of the thread and serializes accesses to the operating system. It can be blocked by the OS, thus implementing the semantics of blocking function calls for hardware threads. Any user processes, however, can continue running in parallel. These processes implement the thread's actual functionality and interact with the synchronization state machine through user-defined signals.

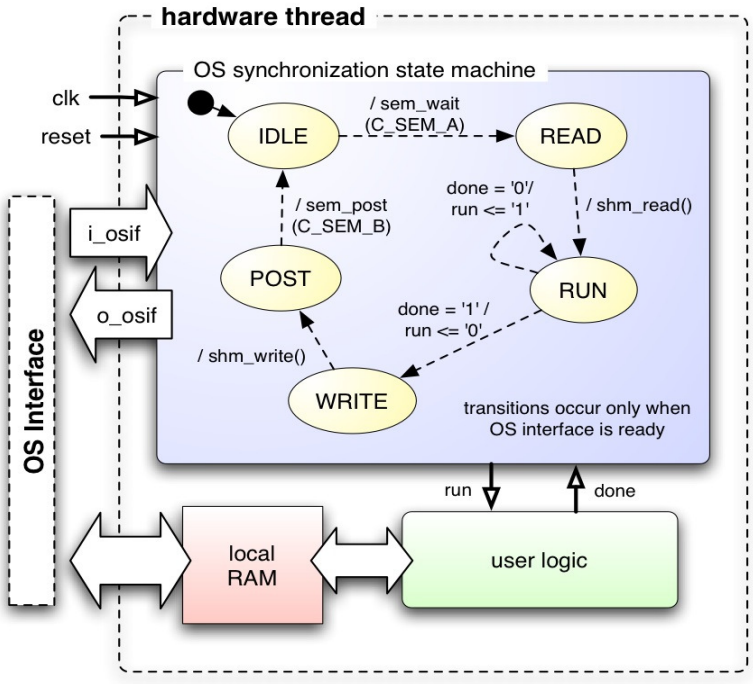


Figure 3. Hardware thread structure

The ReconOS execution model is based on a CoreConnect bus topology implemented in Xilinx FPGAs, an example of which is depicted in Figure 2. The kernel of the operating system and all software threads run on the CPU. Hardware threads are implemented in the FPGA fabric and are connected to an OS interface

(OSIF) block, which is connected to the system bus and provides direct access to the entire system memory space as well as communication registers accessible for OS interaction. Whenever a hardware thread issues a request for an operating system service, its OSIF either executes it directly (if it is implemented in hardware, such as memory accesses) or raises an interrupt to relay the call to the CPU, where the request is executed as a regular software API function call.

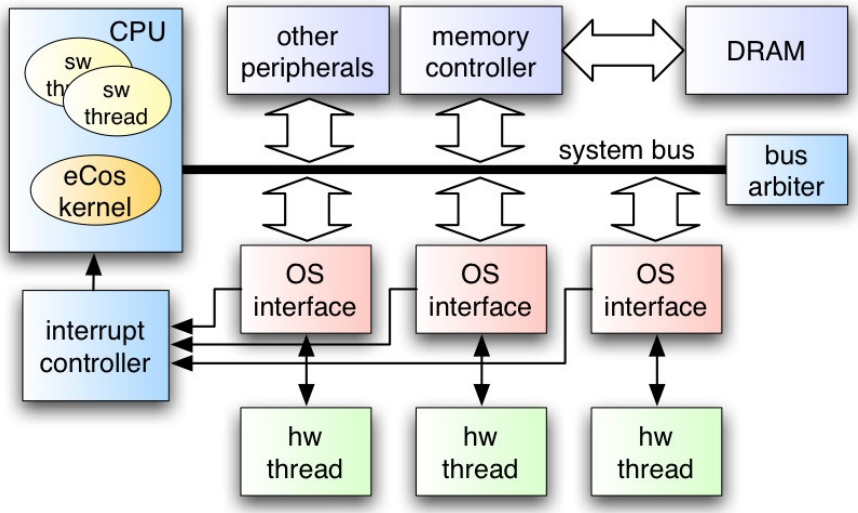


Figure 4. ReconOS run-time system architecture

An intriguing feature of modern FPGAs is the ability to dynamically reconfigure parts of the device while the remaining portions of the system keep operating. ReconOS tries to exploit this feature to enable multithreading of hardware threads, similar to how multithreading is done on the CPU for software threads. This technique would raise the utilization of the reconfigurable logic area, since inactive hardware threads could be dynamically replaced by active ones by partially reconfiguration. However, this technique involves dedicated logic resources for supporting dynamic reconfiguration, a custom tool chain for specifying hardware thread layout and scheduling parameters, mechanisms for saving and restoring thread state as well as novel scheduling techniques that differ from established scheduling algorithms for single- or multiprocessor systems due to the hybrid nature of the system. These are major areas of our current research.

References

[1] Lübbers; E. and Platzner, M.: „ReconOS: A RTOS Supporting Hard- and Software Threads“, In *Proceedings of 2007 IEEE International Conference on Field Programmable Logic and Applications (FPL'07)*, Amsterdam, August 27-29, 2007

5.1.4 Multi-objective Intrinsic Evolution of Embedded Systems

Project coordinator	Prof. Dr. Marco Platzner, PC², University of Paderborn
Project members	Paul Kaufmann, PC², University of Paderborn
Supported by:	DFG, priority program 1183 – Organic Computing

General Problem Description

This project aims at the investigation and development of intrinsically evolvable embedded systems. Simulated evolution will provide such systems with a means to react properly to unforeseen changes in the environment and the system resources. In an intrinsically evolved system the evolutionary process runs together with the function under evolution on the same target platform. This is a necessary precondition for autonomous operation. While evolutionary techniques have already been applied to the design of software and hardware, intrinsic evolution as an adaptation method is a novel approach. We view intrinsic evolution as a promising mechanism to provide autonomous embedded systems with self-adaptive and self-optimization capabilities. We achieve our goals by combining research in bio-inspired computing with modern embedded system architectures. The vision behind this project is that novel bio-inspired algorithms paired with modern system-on-chip architectures will allow us to construct future embedded systems that exhibit intelligent behavior.

Problem details and work done

To investigate intrinsic digital logic design and optimization methods we have implemented a generic modular framework for evolutionary computations. It is capable of visualizing the optimization and adaptation process and creating statistics, and helps users to find the right parameter settings for an experiment (see Figure 1). Due to framework's modular structure, experiments for a specific application using one of the implemented evolutionary algorithms, representation models and operators can be easily set up and executed. The average and asymptotic behavior of a benchmark can be quickly investigated by the framework's interface to the grid software `Condor` which allows running a large number of experiments on a compute cluster. Within the framework we have implemented several hardware representation models, including the classical FPGA structure oriented Cartesian Genetic Programming Model and embedded CGP, an extension to automatically defined functions. Beside the single-objective classical genetic algorithm, the multi-objective

algorithms SPEA2, NSGAII, µGA(2) and IBEA allow to optimize circuits not only regarding functional quality but also for small size and delay.

Figure 1 Graphical user interface of the MOVES-Framework

The performance comparison of optimization methods and representation models for digital logic design is often done by evolving arithmetic functions. This class of functions is widely used to benchmark new approaches and algorithms. However, such functions are not considered as the primary application scenario for an autonomous

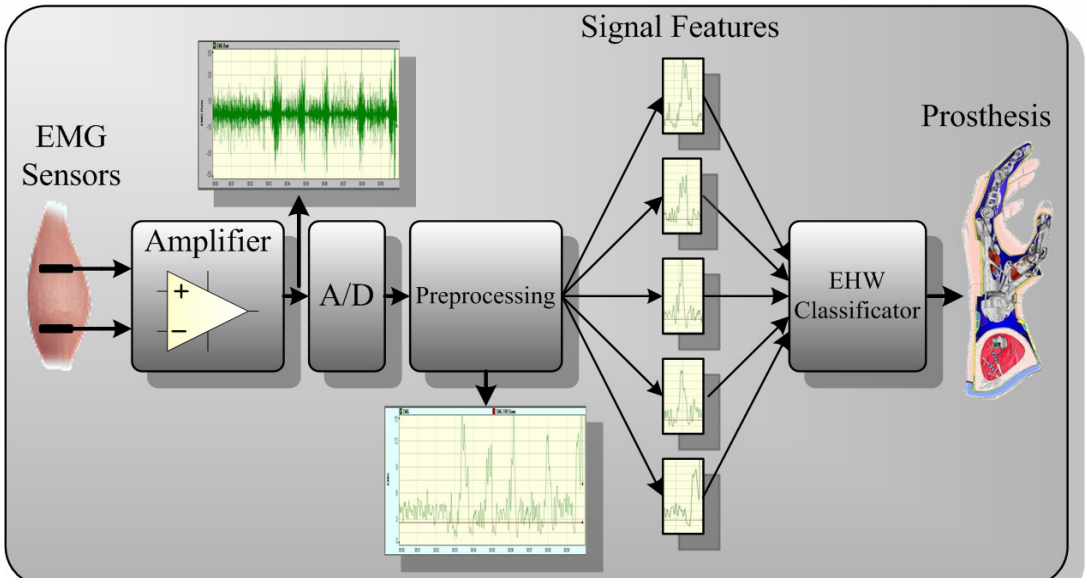


Figure 2 Architecture of muscle signal classifier

embedded system. Here, we are interested in functions that depend on the distribution of the input data. In such cases, the optimal solution is often unknown or would overextend the computational resources of the embedded system. Signal classification is a typical target application domain. To verify our methods we have set up a system for gathering and classifying muscle signals (see Figure 2). The goals for an evolvable classifier are to adapt continuously and autonomously to varying sensor positions and skin conductance, to compensate noise and cross-talk effects from adjacent muscles and to have a considerable recognition rate even after long unmaintained periods. Another application area for testing autonomous adaptation capabilities is robot control and navigation. Here, we are using EyeSim, a simulation environment for the EyeBot robot platform, to simulate agents based on self-adaptive evolvable hardware. Research questions to be solved are: how to gather test data, how to define the fitness and how to validate the evolved circuits? Investigating these questions, we try to understand which modes of adaptation can be handled by an embedded system.

References

[1] Kaufmann, P. and Platzner, M.: MOVES: A Modular Framework for Hardware Evolution. In *Proc. of the 2nd NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pp.447-454, Aug. 2007. Received the Best Paper Award in the Evolvable Hardware Category.

[2] Kaufmann, P. and Platzner, M.: Toward Self-adaptive Embedded Systems: Multiobjective Hardware Evolution. *In Proc. Int'l Conf. on Architecture of Computing Systems (ARCS)*, March 2007. Springer, LNCS 4415.

[3] Kaufmann, P. and Platzner, M.: Multi-objective Intrinsic Hardware Evolution. In *Proc. of the MALD Int'l Conference*, Sept. 2006.

5.1.5 FPGA accelerated high performance computing

Project coordinator:	Prof. Dr. Marco Platzner, University of Paderborn
Project members:	Dipl.-Inform Tobias Schumacher, PC², University of Paderborn
Supported by:	XtremeData Inc.

General Problem Description

Today's high performance computers usually consist of many ordinary processors connected by a fast communication network. This applies to off-the-shelf machines as well as to clusters of workstations that are connected by a high-speed interconnect like Infiniband or Myrinet. The programming models did not change for a long time - depending on the machine's architecture one can use a shared memory programming model or parallelize an application using a message passing library like MPI.

However, things have changed recently. Current trends in high performance computing also take custom accelerators like graphics hardware, special processors or reconfigurable hardware into account. While HPC users usually are in some way experienced in writing parallel applications using message passing libraries like MPI, porting applications to FPGAs is a challenging task for them. First of all, FPGAs are traditionally programmed using a hardware description language like VHDL or Verilog, which are not well known by most of the HPC users. Several C-like dialects exist, but they mostly do not achieve the maximum performance possible. Additionally, a wide range of different FPGA boards exists, each with different hardware parameters and different APIs. Therefore, already deciding which parts of an algorithm are suitable for FPGA acceleration becomes a challenging task. Some of the questions that arise are:

1) How does the CPU communicate with the FPGA?
2) Which bandwidth can be achieved when sending data to the FPGA and read back results?
3) Where should data be stored? Host memory or memory local to the FPGA?
4) Which compute cores need to communicate to each other?

In this project we address these problems by defining a generic model for FPGA-based accelerators. The model helps the user deciding which parts of an algorithm to implement on an FPGA and where to store the data. When moving to a new

architecture (e.g. a new FPGA board), the user can easily modify the model parameters to cover the new features.

Problem Details and Work Done

Many common modeling and analysis techniques for parallel algorithms, e.g., PRAM, describe algorithms by simple operations and express the execution times by counting the number of such operations performed. While these models are of great importance from a theoretical point of view, their usefulness is often also rather limited when it comes to implementations on real parallel machines. The major problem with these models is that they consider only the time spent for performing calculations and do not regard the time spent for data accesses, i.e., memory and I/O.

To overcome this limitation, the LDA (latency of data-access) model was created previously. It consists of an architecture model defining the various properties of the target architecture (memory hierarchy, bandwidths, latencies, ...), an execution model describing the algorithms in terms of LDA operations and an execution time analysis. The execution model operates on LDA instructions which differentiate between accesses to L1-/L2-cache, main memory and remote memory.

We take up on the LDA modeling approach to cover also FPGA accelerators which poses a number of challenges. FPGAs are connected to the host by a bus system like PCI-X or PCI-Express; current trends lend to connecting an FPGA more closely to the CPU using Hypertransport or the Frontside bus. The FPGA can access the host memory using DMA transfers, or the CPU can directly program registers on the FPGA. The parameters for these transfers depend on the kind of Host-to-FPGA connection. Additionally, most FPGA boards provide a certain amount of different types of memory (e.g. DDR-SDRAM, QDR-SSRAM). Unlike on the CPU, on the

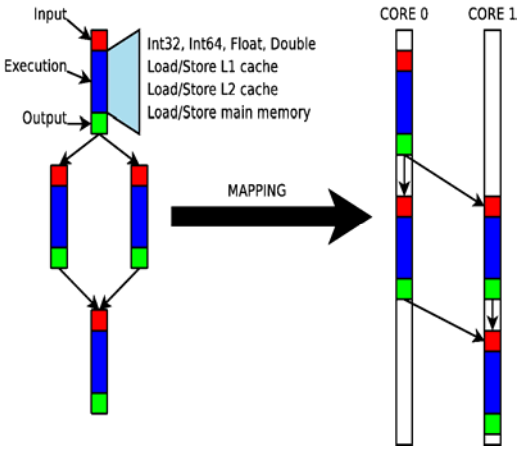


Figure 1: Task graph and mapping to an architecture in the original LDA model

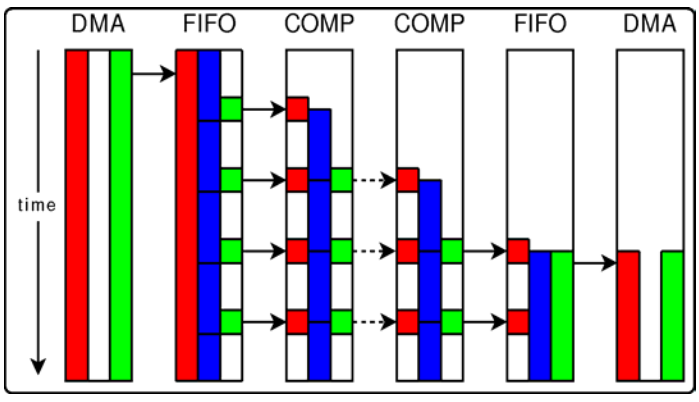


Figure 2: Task graph of an algorithm mapped to a pipelined architecture in an FPGA

FPGA the user has to decide which memory to use for storing a certain type of data. Additionally, no predefined instructions exist, these need to be defined by the user.

Our LDA-based modeling approach is not targeted at modeling the execution of algorithms on existing architectures, but on defining an architecture that efficiently executes the algorithms. There are no predefined compute cores with complete instruction sets. Rather, the user has to define several cores that provide a reduced instruction set suitable for one special task. Even the memory accesses cannot be modeled by simply inserting load/store instructions into the tasks. In the modified LDA model, each memory location is modeled as a separate execution unit that communicates with other compute cores. Every execution unit later gets mapped to an appropriate resource of the target architecture (a CPU, an FPGA, main memory, SRAM/SDRAM on the FPGA-board).

This modeling approach helps the user in deciding which parts of an application are applicable to FPGA acceleration, deciding where different data should be stored and where to place custom compute cores. Additionally it helps the user in predicting the performance achievable by the accelerator. Another central point is the portability of existing FPGA-based applications to new hardware. By mapping a task graph to the new architecture the user directly gets an impression of the performance improvement possible by using new FPGA systems.

An initial version of our model was already successfully used in the development of an FPGA accelerator for the Cube-Cut problem [1]. While the main loop of this algorithm can be implemented in a straight-forward pipeline, we used the modified LDA model to determine optimal parameters like the pipeline depth and width of the accelerator. The first version of the accelerator was generated for the AlphaData ADM-XP based machines. A port to the XtremeData XD1000 was started by changing the model's parameters (Host-FPGA bandwidth, memory bandwidth). The resulting core still uses the same basic instructions but pipeline width and depth changed to maximize the impact of the improved transfer rates the XD1000 provides.

Resource Usage

Many different machines were used in this work. First of all, for testing generated designs the FPGA-equipped machines available in PC² were used, namely:

1) the FPGA island of arminius (4 nodes, AlphaData ADM-XP, Infiniband)
2) the FPGA test cluster (4 nodes, 2xAlphaData ADM-XP each, Myrinet)
3) ic13/ic14 (2 test nodes, 2xAlphaData ADM-XP each, Infiniband)
4) XD1000 (1 node, Altera Stratix II in Hypertransport)

Additionally to these FPGA-equipped machines, several machines were used for generating bitstreams out of the VHDL implementations. These synthesis jobs can only use one processor core efficiently, but need much compute power and main memory. They were run on several testing machines as well as on the Arminius and Windows CCS cluster in batch mode.

The work was supported by a consortium of XtremeData, Altera, AMD and Sun who gratefully provided us with an XtremeData XD1000 workstation.

References

[1] Schumacher, T.; Lübbers, E.; Kaufmann, P. and Platzner, M.: Accelerating the Cube Cut Problem with an FPGA-augmented Compute Cluster, In Proceedings of the ParaFPGA Symposium, International Conference on Parallel Computing (ParCo), Aachen/Jülich, Germany, September 2007.

5.1.6 Support of the ALICE experiment of the European Center for Particle Physics CERN

Project coordinator:	Prof. Dr. Volker Lindenstruth, Kirchhoff-Institut für Physik, Heidelberg
Project members:	Axel Keller, PC², University of Paderborn Andreas Krawinkel, PC², University of Paderborn Holger Nitsche, PC², University of Paderborn KIP Team

General Problem Description

The ALICE heavy-ion particle physics experiment is currently being built at CERN near Geneva. It will use a PC cluster of about 500 machines for the last stages of the data readout process. A group from the University Heidelberg participates in the development of this ALICE High Level Trigger (HLT) system. For a large test of the system's software the group used the 200 node Arminius cluster at the Paderborn Centre for Parallel Computing to test the scalability of the system and evaluate the use and operation of a system of this size.

Problem details and work done

The ALICE experiment shown in Figure 1 is A Large Ion Collider Experiment [1] currently in construction at the European Centre for Particle Physics CERN near Geneva. Together with the three other experiments CMS, ATLAS, and LHCb it will be used for the new Large Hadron Collider (LHC) particle accelerator, also under construction.

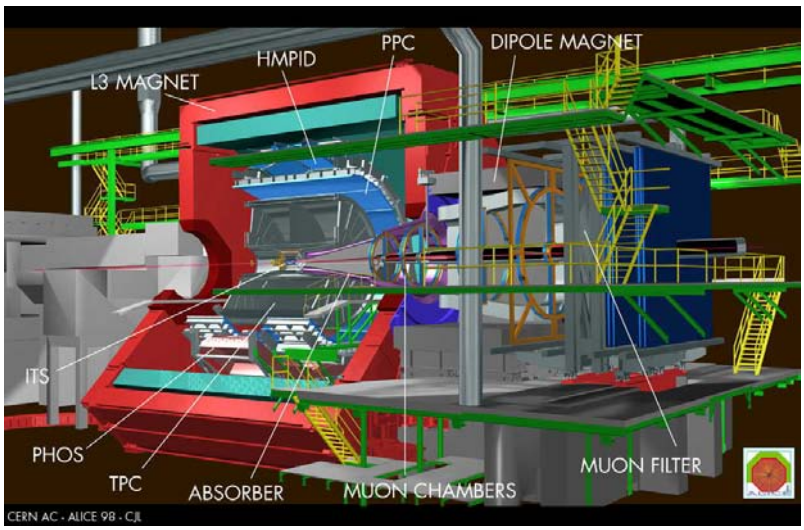


Figure 1: ALICE – A Large Ion Collider Experiment

Like most particle physics experiments, ALICE is a very large scale collaboration, both in the number of participating groups as well as in its physical dimensions:

- more than 90 institutes with more than 1000 people from 29 countries
- 16 m diameter, 25 m long, total weight 10000 t
- inner magnetic field of 0.4 Tesla

The aim of this large experiment is to search for and study the Quark-Gluon-Plasma (QGP). QGP is a state of matter in which quark and gluon elementary particles exist under extremely high temperatures and densities so that they can move freely and are not confined by the forces binding them together in the states of matter seen under more ordinary conditions. In order to produce the required conditions for the existence of QGP, lead nuclei traveling at almost the speed of light will be collided together by the LHC in the centre of the ALICE detector. Due to the large number of particles in the colliding lead ions, the number of particles produced is also quite large. Models predict up to 15000 charged particles passing through the detector for a single collision or event. This also leads to very large measured data size for each event in the detector. Unlike the approximate 2 MB event size of the CMS and ATLAS detectors, ALICE can produce up to 80 MB in a single event. As collisions happen up to 8000 times a second, the resulting maximum theoretical total data rate produced by ALICE could reach up to 640 GB/s. This rate is equivalent to more than 900 CDs per second, filling even then largest currently available hard drives in fractions of a second. The data rate is too high to be handled by any computer system and would also result in high operating costs just for storing the produced data volume. For this reason ALICE, like any other particle physics experiment, employs a multi-level readout architecture which selects only those events from the data which are of most interest to the ALICE physicists.

The ALICE High Level Trigger: The final stage of this selection architecture before the data is written to mass storage is the High Level Trigger (HLT) system. It will consist of a large Linux PC cluster with up to 400 nodes each using multiple CPUs and connected by a fast network. About 250 of these PCs will be equipped with optical links in special PCI add-on cards which will feed the data from the detector into the system. These nodes are called Front-End-Processors (FEPs). At this last stage the input rate is already considerably reduced but can still be up to 25 GB/s, more than 35 CDs each second. At the output, the HLT is allowed to produce a maximum of 1.2 GB/s, so the data stream has to be reduced by a factor of up to 20 inside the system. The communication software for this system is currently being developed in Heidelberg, and consists of multiple components communicating with each other via a defined interface. Components can be arranged in complex hierarchies to form distributed processing chains or pipelines. Some of the components have the task of defining the flow of data in the system by

- transparently connecting other components on separate nodes,
- merging data streams containing parts of the data from different areas of the detector,
- and splitting up and re-merging a data stream into several smaller streams, e.g. for load balancing among multiple nodes.

Other components have the task of actually processing the data received from the detector. This processing will be done in multiple steps, from a full reconstruction of the event from the binary detector data via 3D coordinates of produced charge clouds (cluster finding), to a representation of the full particle trajectories after the collision (tracking). Processing in the lower steps of this hierarchy is highly local and independent for different parts of the detector. It can therefore take place indepedently of other parts directly in the FEPs. In the higher processing levels the dependence on data from other parts of the data increases with each level.
At the top level the data from all sub-detectors of ALICE is used to build a completely reconstructed event. Based on this reconstructed data a decision is then made whether the event or parts of it are to be transferred to permanent storage for later thorough analysis.
The sub detector of the ALICE experiment which produces the largest amount of output data is the Time Projection Chamber (TPC) shown in Figure 2.

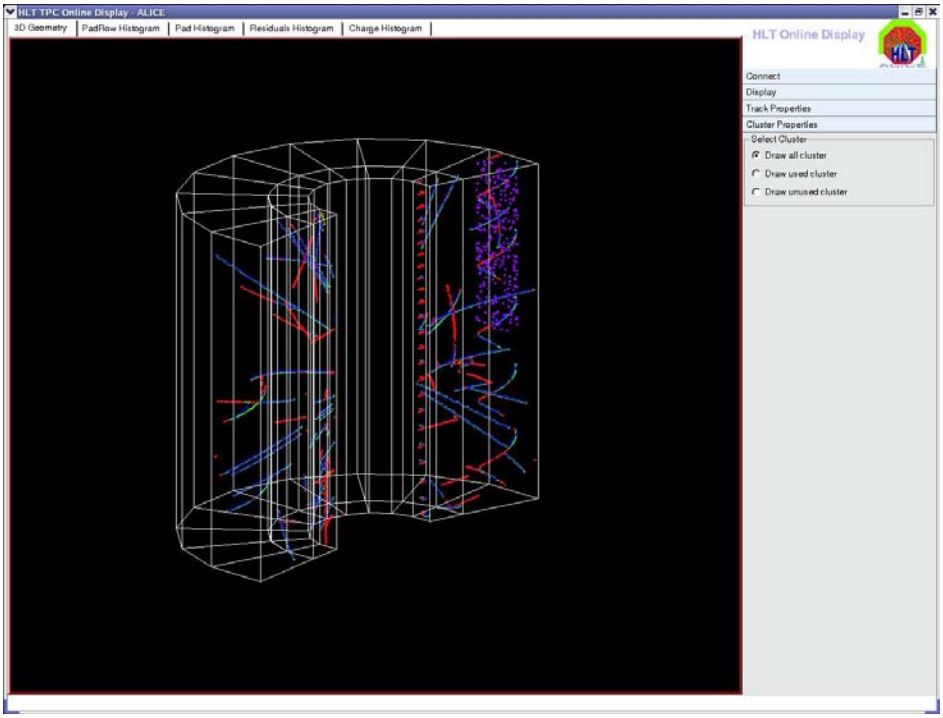


Figure 2: TPC Event Display

It is a so called tracking detector as it detects particle trajectories in a large cylindrical gas volume. The
TPC consists of 36 slices with 6 patches each. As there is one output channel for each patch, the total number of output channels is 216.

Test Setup: The main task during the test was to run the event reconstruction of the whole TPC with 1000 incoming events per second. In order to achieve this high input rate the cluster finding and tracking processes had to be distributed well over the whole cluster. As the PCs had two processors each, they could handle two detector channels simultaneously. Since there was no real detector data available, 4 GB of simulated protonproton collision events were used as input data. Two cluster finding (CF) processes were used for every input channel, meaning a total of 4 cluster finders per node. In order to find all particle trajectories in one slice, two nodes with two tracking processes (TR) each were assigned to every slice. Afterwards the data coming from all the 72 tracking nodes were merged together in 8 global merging processes (GM) nodes to reconstruct trajectories in the whole TPC volume. At the very end of this online analysis chain there was a so called Online Display to visualize the data in 3D and histograms.
About 188 PCs were used for this test. This is approximately half of the final size of the upcoming ALICE HLT cluster.

Cluster Monitoring and System Management: The components running on the nodes will be controlled by a further hierarchy of processes, the TaskManagers. On each node one slave TaskManager process will control all components running locally on that node. Related nodes, e.g. nodes which process data from related sub-detectors, can be grouped together.
Their slave TaskManagers can be controlled by one (or more for fault-tolerance) intermediate TaskManager. At the top-level a group of equal master TaskManagers will provide the top-level control for all components via the tree of intermediate and slave TaskManagers. These top-level TaskManagers will be connected both to the central ALICE Experiment Control System as well as the cluster management and monitoring subsystem.
During the test it was of particular interest to monitor the usage of the different cluster resources, for instance CPU and memory usage, generated network traffic, and disk usage. However, in general it is not trivial to collect and display information about the state of such a large system in an effective way. On the one hand, the information should always be up-to-date, but on the other hand the monitoring system should consume as little additional resources as possible. During the test, the freely available monitoring system Ganglia [2] was used. Ganglia is a scalable, distributed monitoring system for high-performance computing systems that promises very low per-node overheads and high concurrency. Ganglia provides an easy-to-use plug in

architecture which makes it possible to quickly develop extensions to the functionality of the system.
Typically, monitoring information is presented via a web interface, showing a summarized global view of the cluster. However, it is also possible to obtain very detailed reports and extensive statistics for each of the individual cluster nodes. In addition to the standard functionality, we developed a module that monitors the interrupt rate of the nodes and a separate visualization module that presents information about the individual nodes in a very compact, yet easy to comprehend, way (Figure 3).

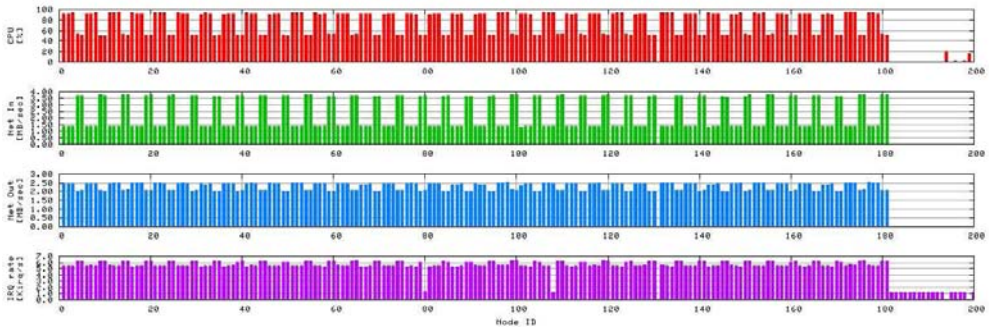


Figure 3: Monitoring View

Parallel to Ganglia we have used our own tool named SysMES for the cluster management. Amongst others, this tool implements functionality for active and dynamic job management. We use these capabilities for configuring, starting and stopping the analysis chain and for a range of management jobs to retrieve the cluster status. SysMES is a scalable, distributed, and decentralized management system capable of performing the complex management tasks required by the (HLT) Cluster. All monitoring information is passed on to the management framework, which runs the cluster based on this data. Management entails, for example, detecting and repairing faults in the analysis chain, rerouting the data flow so as to bypass overloaded computers. To do this, SysMES implements the functionalities for monitoring, messaging, jobs, rules and configuration management. More detailed information about SysMES can be found in [3].

Resource Usage

During the tests on the Arminius cluster with up to 400 processors, a number of important milestones and conclusions were reached. A test which processes simulates processing of event data for half the TPC, including cluster-finding and tracking, ran successfully. During this test a maximum event rate of up to 220 Hz

could be reached. A standalone test with empty event data ran at a rate of 2.9 kHz, demonstrating that the main bottleneck is the processing of the data itself and that the system is not limited by the framework used for the transport of the data. Beyond these system 4 tests, implementations of new network parts for the high-speed InfiniBand interconnect could be tested and benchmarked. These modules performed very well, showing a significant improvement over the standard TCP over Gigabit Ethernet implementation, both in throughput as well as in CPU utilization. Furthermore, the tests also revealed some issues with using the software at large scale. Overall the test was performed successfully and the lessons learned from working on a system of that scale will be taken into account in the further development of the HLT software.

We wish to thank the Paderborn Center for Parallel Computing for all their help and support, without which the tests would not have been possible. We particularly extend our thanks to Axel Keller and Holger Nitsche who gave us a lot of excellent technical support in their facility.

References

[1] The Alice Experiment, http://aliceinfo.cern.ch/
[2] Ganglia Monitoring System, http://ganglia.info
[3] The SysMES project, http://www.kip.uni-heidelberg.de/wiki/SysMes/index.php
[4] Paderborn Center for Parallel Computing, www.upb.de/pc2

5.2. Grid Technologies

5.2.1 HPC4U – Highly Predictable Clusters for Internet Grids

Project coordinator	Prof. Dr. Odej Kao, PC², UPB / TU Berlin
Project members	Dr. Felix Heine, PC², UPB / TU Berlin
	Dr. Matthias Hovestadt, PC², UPB / TU Berlin
	Axel Keller, PC², University of Paderborn

General Problem Description

During the last years, Grid computing has become an established instrument for academic users worldwide. Research has led to Grid middleware systems like “Unicore” or the “Globus Toolkit”. Now the focus turns on attracting commercial users for using Grid infrastructures. However, this new user community has new challenging requirements on the Grid. In this context, already in 2002 the European Commission convened a group of experts. Their task was to identify demands of future Grid systems, which should be commercially successful. By this, they clarified properties and capabilities missing in existing Grid infrastructures.

The work of this group resulted in the idea of the Next Generation Grid (NGG) [2]. Among the core requirements on such an NGG are issues on reliability, transparency, and assurance of Quality of Service (QoS) parameters. In fact, a commercial user will not use Grid infrastructures for computing his deadline bounded or business critical jobs, if the Grid is only able to follow the best-effort approach. In contrast, the commercial user demands for contractually fixed service quality levels, on which the user is able to rely on.

In this context, a Service Level Agreement (SLA) is a powerful instrument, since it allows the description of all expectations and obligations in the business relationship between service provider and the (commercial) service customer [1]. This way, the customer is able to unambiguously specify the requirement profile of his job. In turn, the provider is able to define the service to be provided. Hence, the SLA is only accepted and enforced, if both parties agree on its content.

Numerous research projects already focus on the integration of SLA negotiation and management procedures within the Grid middleware. However, solely focusing on Grid middleware is not sufficient. Since Grid middleware is mediating between incoming user jobs to available local resource management systems (RMS), which

are providing their resources to Grid infrastructures, also these RMS have to be aware of the requirements of the job defined in the SLA. Moreover, RMS have to realize the agreed service quality level in their local domain [4]. However, current RMS are operating on best-effort approach only, not allowing Grid middleware to substantially guarantee SLA aspects.

A major research focus at PC² is on resource management systems providing an increased level of quality of service as a software-only solution. At this, application transparency is crucial. Arbitrary applications should benefit without recompilation and linkage against special libraries from an increased level of QoS, like fault tolerant job execution. This is of particular interest for commercial Grid environments, since source code of commercial applications is normally not available.

Within the EU-funded project “Highly Predictable Cluster for Internet-Grids” (HPC4U) [3] the PC² is working on an SLA-aware resource management system, utilizing the mechanisms of process-, storage- and network-subsystem for realizing application-transparent fault tolerance. The HPC4U project started in June 2004 and will end in May 2007. The partners are IBM France (FR), Fujitsu Systems Europe Ltd (UK), Seanodes SA (FR), Dolphin Interconnect Solutions AS (NO), Scali AS (NO), Paderborn Center for Parallel Computing, CETIC (BE), and National Supercomputer Centre (SE).

The goal of the HPC4U project (Highly Predictable Cluster for Internet Grids) is to provide an application-transparent and software-only solution of a reliable RMS. It will allow the Grid to negotiate on Service Level Agreements, and it will also feature mechanisms like process and storage checkpointing to realize Fault Tolerance and to assure the adherence with given SLAs. The HPC4U solution will act as an active Grid component, using available Grid resources for further improving its level of Fault Tolerance.

Problem Details and Work Done in the Reporting Period

Typically, a resource management system (RMS) is responsible for the runtime management of a compute cluster. Beside its scheduling and system management procedures, it also has mechanisms for detecting of and reacting on failures, e.g. outages of compute nodes. In case of such an event, the means of an RMS depend on the capabilities of its mechanisms. In the case of HPC4U, the RMS is able to generate transparent checkpoints of running applications. Hence, neither the user has to know about the checkpointing capabilities, nor does he have to change the application itself or how it is actually invoked. The RMS can checkpoint the

application in the background and use the checkpoints it in case of resource outages for restart.

Even if this restart mechanism allows the RMS to restart the application, so that it does not have to restart from scratch, the mechanism depends on the availability of spare resources, i.e. resources that are idle and can be used for restarting the job checkpoint. In case of spare resource unavailability, the resource management system would either have to cancel one of the running jobs, or the checkpointed job has to wait until spare resources are available again. Since the resource operator strives for utilizing his resources as best as possible, preferably having valuable SLA-bound jobs, this reduces the number of free spare resources. Hence, the more the operator reaches his goal of increasing system utilization with valuable SLA-bound jobs, the more he reduces the system's options on reacting on resource outages. Thus, by reaching his goal, he increases the general risk of violating SLAs due to the general inability of applying fault tolerance mechanisms. This results in the necessity of trading off between the two goals: system utilization on the one hand, and fault tolerance respectively guaranteed service provision on the other. In this context, the system administrator can introduce general buffer nodes that must not be used for SLA-bound jobs, but for best-effort jobs only. This way, node outages effecting SLA-bound jobs can always compensated by suspending best-effort jobs running on spare resources.

However, the approach of introducing spare nodes is static, generally reducing the overall system utilization, so that resource operators may strive to setting the number of spare nodes as low as possible. Another approach in handling resource outage situations in full-load scenarios is the introduction of cross-border job migration. Here, jobs may not only be restarted on resources of the same cluster system, i.e. the cluster system where the job has been originally executed and that is now affected by the resource outage, but also on other cluster systems. In the scope of this workpackage, the restart of jobs on remote cluster systems within the same administrative domain is to be realized. Hence, the migration process can be denoted as "inter-cluster, inner-domain".

The main difference between inner-domain cross-border migration (i.e. the migration between two cluster systems belonging to the same administrative domain) and inter-domain cross-border migration (i.e. the migration to cluster systems outside the same administrative domain, e.g. other cluster resources within the Grid) is the resource selection process. While there is no knowledge about cluster resources outside the same domain, the administrator has full knowledge about the cluster resources within his own domain. This means that he can statically configure dependencies between the cluster systems, e.g. specifying that clusters A, B, and C are compatible to each other, therefore serving as backup resource and target for cross-border migration

processes. As a matter of fact, also questions on security, authorization, authentication, and accounting come into play when moving to the inter-domain cross-border migration case. However, Grid systems like the Globus Toolkit offer mechanisms for these questions, so that this is basically transparent for the central RMS services.

Resource Usage at PC²

For the runtime of the HPC4U project, a cluster system has been installed as reference platform for the HPC4U project. This cluster consists of five Dell PowerEdge 2650 nodes. Each of these nodes is equipped with two Intel Xeon 2.4Ghz processors, 1GB of RAM, 73GB harddisk and a Myrinet M3F-PCIXD-2 card. Furthermore a Dolphin Interconnect SCI card (AS PSB66) has been integrated into these nodes, since the SCI interconnect is used as standard in HPC4U.
One of these five nodes is used as a frontend node for the cluster system, where users are submit their jobs. This frontend node also represents the interface for Grid middleware. All other nodes of the cluster are used as dedicated compute nodes.

References

[1] Sahai et al. Specifying and Monitoring Guarantees in Commercial Grids through SLA. Technical Report HPL-2002-324, Internet Systems and Storage Laboratory, HP Laboratories Palo Alto, November 2002.

[2] H. Bal et al. Next Generation Grids 2: Requirements and Options for European Grids Research 2005-2010 and Beyond. ftp://ftp.cordis.lu/pub/ist/docs/ngg2_eg_final.pdf, 2004.

[3] Highly Predictable Cluster for Internet-Grids (HPC4U), EU-funded project IST-511531. http://www.hpc4u.org.

[4] Burchard, L.O. et al. The Virtual Resource Manager: An Architecture for SLA-aware Resource Management. In *4th Intl. IEEE/ACM Intl. Symposium on Cluster Computing and the Grid (CCGrid) Chicago, USA*, 2004.

5.2.2 AssessGrid – Advanced Risk Assessment and Management for Trustable Grids

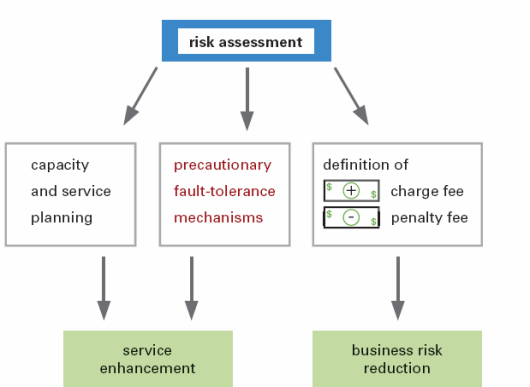
Project coordinator	Prof. Dr. Odej Kao, PC², UPB / TU Berlin
Project members	Georg Birkenheuer, PC², University of Paderborn Kerstin Voß, PC², University of Paderborn Dominic Battré, PC², UPB / TU Berlin Christer Carlsson, Abo Akademie University, Finnland Franck Tetard, Abo Akademie University, Finnland Irina Georgescu, Abo Akademie University, Finnland Melanie Biette, Atos Origin, Spain Josep Martrat, Atos Origin, Spain Igor Rosenberg, Atos Origin, Spain Manuel Quijada, Atos Origin, Spain Simon Alexandre, CETIC asbl, Belgien Stephane Mouton, CETIC asbl, Belgien Christoph Ponsard, CETIC asbl, Belgien Karim Djemame, School of Computing, University of Leeds Iain Gourlay, School of Computing, Universtity of Leeds Jamed Padgett, School of Computing, Universtity of Leeds Matthias Hajari, Wincor Nixdorf, Paderborn Jörg Stümke, Wincor Nixdorf, Paderborn
Supported by:	European Commission (IST – 031772)

General Problem Description

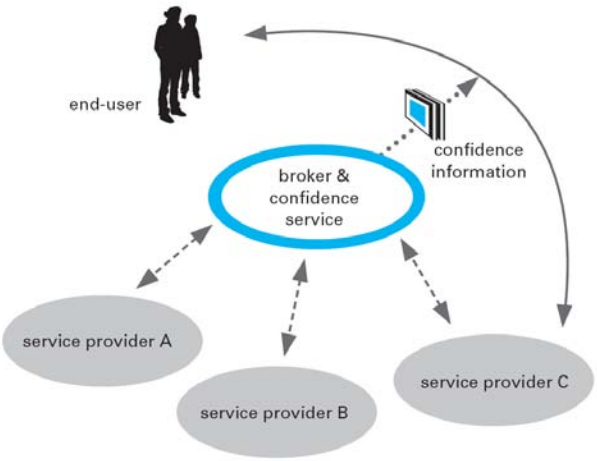
In our everyday life we book various services and benefit from a broad public and commercial infrastructure making our life more convenient. Usually there are several providers offering identical or similar services and products. The selection of an appropriate service is based on various parameters: price and availability are major criteria, but the reputation and thus the assumed quality of service also play a major role. The reputation expresses a long-term tradition and customer trustworthiness. New companies without this type of reputation publish test and evaluation reports to prove their quality and to attract customers. Similar mechanisms exist in other fields, e.g. hotel guests study the number of stars, put these in relation to the national

standard and try to get as much quality of service (QoS) as possible within their budget. Furthermore, they visit web sites where former guests comment their stay and give their subjective impression in different categories. Auction buyers for instance take a close look at the past evaluation of the seller and have a preference for well-ranked persons following the locality principle: A good performance in the past is an indicator for a good performance in the near future. We send our children to schools with an excellent reputation, give money to research institutions with outstanding records and hope, the job/product/service will be delivered with the highest possible quality.

In case of Grids a significant need for information on reputation exists. Should Grid users send a job/mission into a cloud of resources and should not care, who is going to execute the work? In particular should they do it if holding deadlines is crucial for their project progress or research work? The complete virtualization of resources is a powerful technological development, but Grid services are only used in the academic environment. Customers need some information about the provider reputation and quality on one hand and a legal agreement (implicit or explicit) on the other hand, before they are ready to assign a mission to a certain provider. While the Grid community undertook significant efforts by developing Service Level Agreements (SLAs), the need for reputation for gaining trustworthiness in the Grid technology is uncovered. This is one reason, why adopters underline–despite the high level of Grid technology development–core shortcomings related to security, trustworthiness, and dependability of the Grid for commercial applications and services. Since the SLA concept is a risk for both sides, a commercial SLA provisioning and usage is not available,. Users want to be sure that their job will be treated according to the SLA. But providers are cautious on adoption, as agreeing on SLAs including penalty fees is a business risk: system failure, operator unavailability etc. can lead to an SLA violation. Providers need risk assessment methods as decision support for accepting/rejecting SLAs, for price/penalty negotiation, for activating fault-tolerance actions, and for capacity and service planning. Customers need the estimation and aggregated confidence information for provider selection and fault-tolerance/penalty negotiations.



AssessGrid addresses the issues of reputation and trustworthiness for all groups of Grid participants and provides methods and tools to assess and to manage a possible probability of failure in all Grid layers. The investigated risk management scenarios reflect the perspective of Grid end-users, brokers, and providers. The results will support all Grid actors by increasing the transparency, reliability, and trustworthiness as well as providing an objective foundation for planning and management of Grid activities. End-users will have a connection to a confidence service and thus access to reputation indicators suitable for the specific job. The providers will receive objective indicators about the quality of their own infrastructure, methods for risk estimation for different situations (low/high loaded resources, vacation time, overloaded network, etc.), and a decision-support for system development, management, and planning in order to detect and remove bottlenecks. Self-organizing fault tolerance mechanisms use risk indicators as thresholds to increase the reliability. In case of failures and thus risk above the threshold, the business policy will be adapted, e.g. longer slack-times will be negotiated, the penalty fee will be reduced or even SLAs will be rejected. Furthermore, spare resources or a redundant processing will be activated. Providers will also need a platform for an objective comparison and competition, which is likely to be located at the Grid broker. They have a large number of submitted jobs, so over time they can learn a lot about the quality of a certain provider and thus create and publish reputation indicators. These indicators express the risk of failure or the risk of job execution with lower quality and offer a platform for provider rankings and competition.
The demonstration of the development will be based on three sample scenarios, related to the three different groups of Grid participants – end-users, broker, and providers.



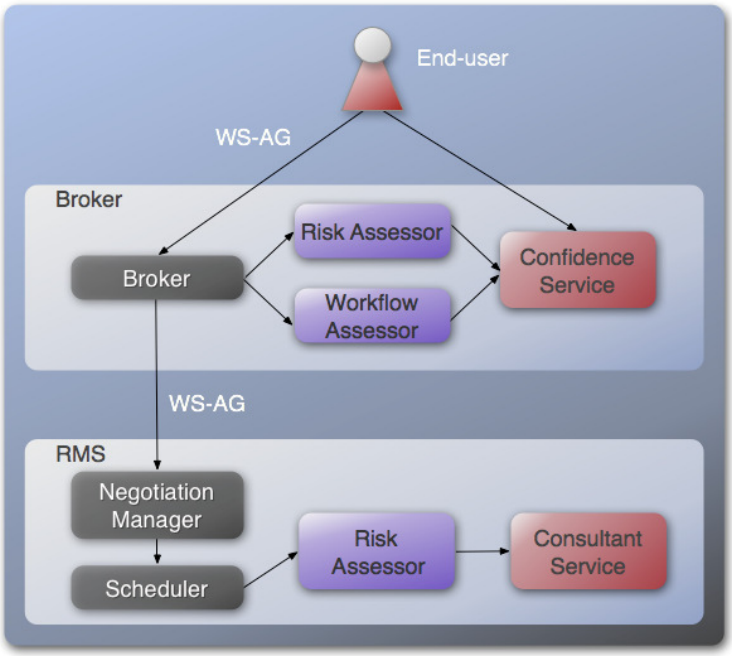
AssessGrid is funded the European Commission as a STREP in the scope of the FP6 Specific Programme for “integrating and strengthening the European Research Area” (SP1). After one negotiation meeting, the project started in April 2006. It has a duration of 33 months and will end in December 2008.

Problem Details and Work Done

The AssessGrid project started on 1st April 2006 with a kick-off meeting held in Paderborn. The main issues within the first six months were the analysis and collection of requirements as well as the evaluation of existing methods for risk management and risk assessment. The requirement analysis was lead by CETIC. The PC² was strongly involved in the requirement analysis since it developed the overall project idea and PC² performed a detailed view of the provider's developments.. A structured approach for identification of Grid participants and pooling requirements was successfully applied and implemented by elaboration and distribution of questionnaires and face-to-face interviews. The resulting deliverable D1.1 "Requirement Analysis" encompasses all relevant requirements and offers an extensive and high-quality summary of the needs for a risk-aware Grid middleware. In parallel, methods for risk assessment and risk management, which are already developed and applied e.g. in financial or insurance sector, were summarized and evaluated regarding their usability and adaptability for the Grid scenarios under investigation. The result in deliverable D1.2 allows the re-use of knowledge from other areas and the creation of a roadmap towards development and integration of risk methods into the Grid. Based on the requirements, promising methods for risk assessment and management as well as current developments of the Grid middleware, the system architecture for AssessGrid prototypes was developed in a strong collaboration between the University of Leeds and PC².

At the time of the proposal set-up there were several groups world-wide working on the development and standardization of a WS-Agreement [3] negotiation manager. However, they failed and a crucial component of AssessGrid could not be re-used as planned. Since as a consequence the WS-AG implementation had been performed by AssessGrid [4] it is the first world-wide working prototype, which increased the visibility of the project significantly and allows us to strengthen the exploitation, dissemination, and standardization activities.

WP 2 started in October 2006 and implements the first of the three scenarios named "customer scenario" by including the basic set of risk management functionality, primarily focusing on the integration within the Grid end-user interface and to build up the specific AssessGrid architecture (see figure below). The main goal is to introduce the risk assessment as an additional quality parameter in the Grid service, middleware, and fabric so the user is aware of possible probability of failure prior to the selection of a provider for the current job/data. The applied risk management methods are limited in the sense that they use static data only, which is provided by monitoring components. The result is the first AssessGrid outcome with integrated risk functionality on all Grid layers, which have been presented as a deliverable D2.1 "Risk-aware End-user Client" on 10th October 2007. This deliverable was accompanied by extended verification and validation cases, which have been published as a report in deliverable D2.2.



Within the first prototype the PC² integrated risk awareness in the Grid fabric by introducing a risk assessor module and appropriate services. To provide the risk assessment the required input data, a consultant service had been developed which extracts monitoring information from the UMF. Furthermore the resource management system OpenCCS had been enhanced in order to support SLA negotiations and the communication with the risk assessor. After selecting a suitable slot for the job execution, OpenCCS asks the risk assessor for the probability of an SLA violation if the resources reserved would be used. The Grid service consumer can determine a maximum probability of failure (PoF) they are willing to accept. If the maximum PoF is higher than the feasible PoF estimated for the reservation, OpenCCS can reserve dedicated spare resources for the job which lowers the PoF. Hence, by reserving additional resources the provider will be able to fulfill the maximum PoF.

The risk assessment and management functionalities will be enhanced within the WP3 started in July 2007. The risk assessment model will consider additional input data and reflect the execution circumstances in order to estimate accurate PoFs. The Abo Akademi started to enhance the model based on the requirements defined by PC². To integrate the new model in the Grid fabric, the consultant service had been extended in order to provide the input data required. In November 2007 the implementation of OpenCCS has started and will be completeled in March 2008.

The PC² has been very active in establishing and starting collaboration activities. Thanks to the European Grid Technology Days in September 2006, AssessGrid intensified the contact with selected EC-funded projects such as BeInGrid, GridTrust,

GridEcon, and HPC4U. These projects have a similar focus: GridTrust helps the consortium to collect and verify requirements, BeInGrid provides a platform for Grid technologies applicable for the Grid commercialization like the self-organizing fault tolerance mechanisms developed in scope of HPC4U.
Furthermore a collaboration with CERN had started in which Grid sites of EGEE should provide the AssessGrid project with monitoring data. Several monitoring statistics from Ganglia have been used in AssessGrid for the initial estimations of average node uptimes since Unified Monitoring Framework (UMF) [5] was still under development.

References

[1] Highly Predictable Cluster for Internet-Grids (HPC4U), EU-funded project IST-511531. http://www.hpc4u.org.

[2] Djemame, K.; Haji,M. and Padgett, J.: SLA Management in a Service Oriented Architecture, In Proceedings of ICCSA'2005, Singapore, May 2005, Lecture Notes in Computer Science 3483, pp.1282-1291

[3] Andrieux , A.: Web Services Agreement Specification (WS-Agreement), et. Al, OGF, (GRAAP) WG, http://www.ogf.org/Public_Comment_Docs/Documents/Oct-2006/WS-AgreementSpecificationDraftFinal_sp_tn_jpver_v2.pdf

[4] Battré, D.; Kao, O. and Voss,K.: Implementing WS-Agreement in a Globus Toolkit 4.0 Environment, In Usage of Service Level Agreements in Grids Workshop in conjunction with The 8th IEEE International Conference on Grid Computing (Grid 2007), September 2007

[5] Hovestadt, M.; Lerch, N.; Nitsche, H. and Voss, K.: First Steps of a Monitoring Framework to Empower Risk Assessment on Grids, Cracow Grid Workshop, 2006

5.2.3 Computing Center Software (CCS)

Project coordinator	Prof. Dr. Holger Karl, PC², University of Paderborn
Project members	Axel Keller, PC², University of Paderborn
	Lars Schäfers, University of Paderborn

General Problem Description

The availability of commodity high performance components for workstations and networks made it possible to build up large, PC based compute clusters at modest costs. Large clusters with hundreds of processors operated as multi purpose systems demand for multi user management and efficient administration tools.

A resource management system is a portal to the underlying computing resources. It allows users and administrators to access and manage various computing resources like processors, memory, networks, or permanent storage.

Workstation clusters are often not only used for high-throughput computing in time-sharing mode (i.e. multiple jobs are sharing the same resources at the same time) but also for running complex parallel jobs in space-sharing mode (only one job is using the resources). This poses several difficulties to the resource management system, which must be able to reserve computing resources for exclusive use and also to determine an optimal process mapping for a given system topology.

The Computing Center Software is such a resource management system. It provides a homogeneous access interface to a pool of different High Performance Computer (HPC) systems, while for system administrators it provides a means for describing, organizing and managing HPC systems that are operated in a computing center. Hence the name “Computing Center Software”, CCS for short.

CCS is in itself a distributed software system possibly consisting of hundreds of entities. Its functional units have been kept modular to allow adaptation to new environments. A CCS island is responsible for one underlying system. It consists of six components. Each part contains several modules and/or daemons which may run asynchronously on different hosts to improve the performance.

Figure 1 depicts the principal architecture of a CCS island.

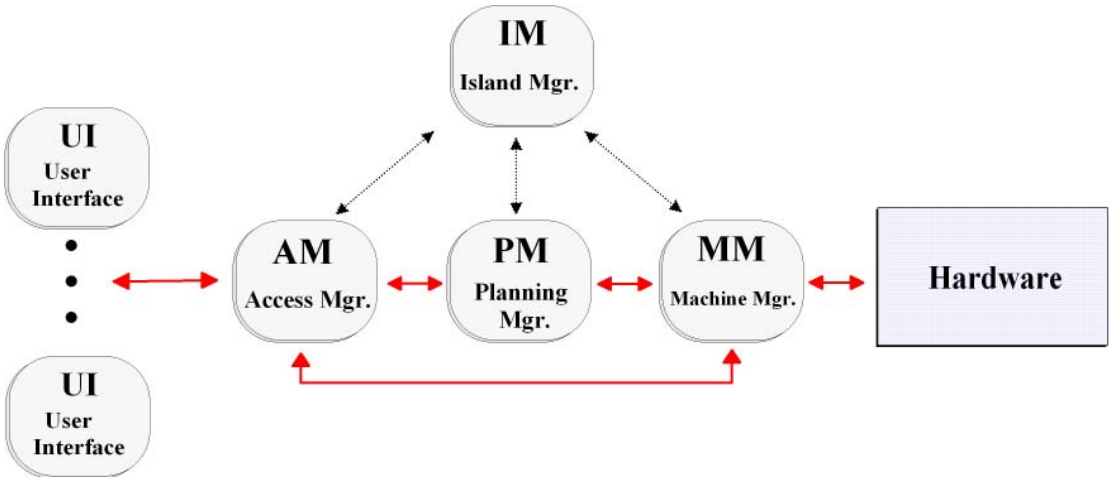


Figure 1 Architecture of a CCS island

1) The *User Interface (UI)* offers X-window or ASCII access to the machine.
2) The *Access Manager (AM)* manages the user interfaces and is responsible for authentication, authorization, and accounting.
3) The Planning *Manager (PM)* schedules the user requests onto the machine.
4) The Machine Manager (MM) provides an interface to the machine specific features like partitioning, job controlling, etc.
5) The *Island Manager (IM)* provides name services and watchdog functions to keep the island in a stable condition.
6) The *Operator Shell (OS)* is the X-window based interface for system administrators to control CCS, e.g. by connecting to the other components.

Problem Details and Work Done in the Reporting Period

Unlike well known resource management systems like Condor [4], LSF [5], PBS [6], or SGE [7] CCS is targeted to support of space-sharing parallel computers. Its resource description facility qualifies CCS to compute an efficient mapping of partitions onto the nodes. Another important difference to the mentioned resource management systems is the way how CCS schedules incoming requests. The criterion for the differentiation of resource management systems concerning the aspect of scheduling is the planned time frame.

Queuing systems try to utilize currently free resources with waiting resource requests. Future resource planning for all waiting requests is not done. Hence waiting resource requests have no proposed start time.
Due to their design, queuing systems provide no information that answers questions like “Is tomorrow's load high or low?” or “When will my request be started?”. Hence reservations in advance are troublesome to implement which in turn makes it difficult to participate in a multi-site grid workflow run.

Planning systems in contrast plan for the present and future. Planned start times are assigned to all requests and a complete schedule about the future resource usage is computed and made available to the users.
Hence planning systems are well suited to participate in grid environments and multi-site application runs. There are no queues in planning systems. Every incoming request is planned immediately. However, planning systems also have drawbacks. The cost of scheduling is higher than in queuing systems.

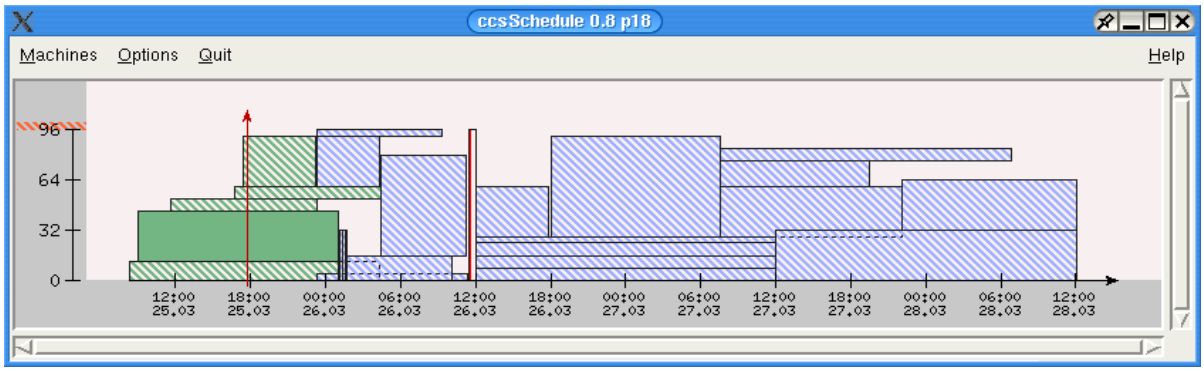


Figure 2 A CCS Schedule

Figure 2 shows an example of a schedule planned by CCS. The X-axis represents the time and the Y-axis the used number of nodes.

Based on this scheduling paradigm CCS provides the following features:

- *Show Planned Start Times:* The CCS user interface shows the estimated start time of interactive requests directly after the submitted request has been planned. This output will be updated whenever the schedule changes.
- *Advance Reservations:* CCS can be used to reserve system resources for a given time. Once CCS has accepted a request, the user is guaranteed access to the requested resources. During the reserved time frame a user can start an arbitrary number of interactive or batch jobs

- *Deadline Scheduling:* Batch jobs can be submitted with a deadline notification. Once a job has been accepted, CCS guarantees the job to be completed at (or before) the specified time.
- *Limit Based Scheduling:* In CCS authorization is project based. One has to specify a project at submit time. CCS knows two different time slots: weekdays and weekend. In each slot CCS distinguishes between day and night. All policies consider the project specific node limits (given in percent of the number of available nodes of the machine). This means that the scheduler will sum up the already used resources of a project in a given time slot. If the time dependent limit is reached, the request in question is planned to a later or earlier slot (depending on the request type: interactive, reservation, deadline etc.).
- *System Wide Node Limit:* The administrator may establish a system wide node limit. It consists of a time slot [start, stop], a number of nodes (N), a threshold (T), and duration (D). The limit defines that during the interval [start, stop] N nodes are kept free for requests which consists of less than T nodes and have duration not longer than D. This ensures that small partitions are not blocked by large ones during the given interval.
- *Admin Reservations:* The administrator may reserve parts or the whole system for a given period of time for one or more projects. Only the specified projects are then able to allocate and release an arbitrary number of requests during this interval on the reserved nodes. Requests of other projects are planned to an earlier or later time. An admin reservation overrides a project limit and the current system wide node limit. This enables the administrator to establish virtual machines with restricted access for a given period of time and a restricted set of users.
- *Duration Change at Runtime:* It is possible to manually change the duration of already running or waiting requests.

Resource Usage at the PC²

CCS is used for the management of the Arminius cluster [8]: The system consists of 400 Intel-Xeon processors.

References

[1] Battré, D.; Hovestadt, M.; Kao, O.; Keller, A. and Voß, K.: Planning Based Scheduling for SLA-awareness and Grid Interaction Proceedings of the 26th Workshop of the UK Planning and Scheduling Special Interest Group (PlanSIG2007)

[2] Battré, D.; Hovestadt, M.; Kao, O.; Keller, A. and Voß, K.: Transparent Cross-Border Migration of Parallel Multi Node Applications Proceedings of the 7th Cracow Grid Workshop, Academic Computer Center CYFRONET AGH 2007

[3] Burchard, L.-O.; Heine, F.; Heiss, H.-U.; Hovestadt, M.; Kao, O.; Keller, A.; Linnert, B. and Schneider, J.: The Virtual Resource Manager: Local Autonomy versus QoS Guarantees for Grid Applications Future Generation Grids, Springer, pp 83-98, 2006
[4] Condor. http://www.cs.wisc.edu/condor
[5] LSF - Load Sharing Facility. http://www.platform.com/products/wm/LSF/index.asp
[6] PBS - Portable Batch System. http://www.openpbs.org
[7] SGE - Sun Grid Engine. http://www.sun.com/software/gridware
[8] ARMINIUS-Cluster. http://www.upb.de/services/systems

5.2.4 D-Grid: German Grid Initiative

Project coordinator	Prof. Dr. Odej Kao, PC², UPB / TU Berlin
	Prof. Dr. Holger Karl, PC², University of Paderborn
Project members	Dominic Battré, PC², UPB / TU Berlin
	Georg Birkenheuer, PC², University of Paderborn
	Bernard Bauer, PC², University of Paderborn
	Dr. Felix Heine, PC², UPB / TU Berlin
	André Höing, PC², UPB / TU Berlin
	Dr. Matthias Hovestadt, PC², UPB / TU Berlin
	Holger Nitsche, PC², University of Paderborn

General Problem Description

The D-Grid (www.d-grid.de) is a German Grid computing initiative started in September 2005 and has ended in 2007. It is supported by the BMBF (Bundesministerium für Bildung und Forschung) and encompasses over 100 institutions all over Germany. The initiative consists of several Grid[1] application projects from scientific communities like High Energy Physics, Life Sciences, Climate Research, or business communities like the BIS-Grid [2] project. The technological basis for the project is provided by the so-called D-Grid *integration project* (DGI). This project will develop and package a middleware for the community projects and provide a Grid generic infrastructure consisting of high performance compute resources, storage space, and other specialized devices. The PC^2 is part of the D-Grid initiative in the context of the integration project and has several tasks. First of all, PC^2 is responsible to develop and run the D-Grid portal in close cooperation with FZK (Forschungszentrum Karlsruhe).

Furthermore, PC^2 integrates its ARMINIUS cluster with the D-Grid using commonly used middleware packages like Globus [3] and UNICORE [4], [5]. Thus, each user of D-Grid may use the IC-Cluster or parts of the ARMINIUS [6] cluster in a transparent way running its local Grid client. An important part of the security of Grid infrastructures are digital certificates. Each user of D-Grid is required to have a certificate which is used during authentication. The certificates used in D-Grid follow the European policy for Grid certificates, which permits only one tier of Certificate Authorities (CAs) in each nation. Thus, in Germany there are only two authorized CAs. To ease the application and the issuing process for new certificates, the PC^2 has established and runs a local Registration Authority (RA) for grid certificates. This allows users to apply for certificates locally on site.

Problem details and work done

Registration Authority

As stated in the introduction, a registration authority is an important entity improving the usability of Grid systems. Without a local RA, Grid users from Paderborn would have to travel either to the DFN office in Hamburg or to the GridKA center in Karlsruhe for authentication as these are the places where the German Grid CA's are located. Thus, PC^2 established a local RA. The RA has customized Web-Pages where users can apply for a personal certificate. From these pages, a form is generated and printed, which has to be signed by the user. Users then have to visit PC^2 personally, carrying an id-card or passport, in order to verify their identity. Employees of PC^2 will sign their request and forward it to the mentioned CA. The certificate is then issued to the user via e-mail. The procedure is described in detail on the PC^2 web pages, together with all relevant links, and addresses to get support.

User accounts can be requested at dgrid-usermngt@upb.de. General support requests for either of the systems can be issued to dgrid-support@upb.de.

UNICORE

UNICORE is a vertically integrated Grid middleware. It has a three layer architecture shown in Figure 1.

Users define and submit their jobs locally using the graphical UNICORE client. The client provides various plug-ins to enter application-specific parameters during the job definition in a convenient way. The job is submitted to the UNICORE gateway, which serves as a dispatcher forwarding the job to the target system. The various systems running in the UNICORE framework are virtualized behind a so-called network job supervisor (NJS), which gets the responsibility for the jobs running on the system. The incarnation database (IDB) describes both the hardware aspects and the installed software of the underlying system. The user database (UUDB) maps the grid certificates to local usernames and defines who is permitted to use the system. Finally, the target system interface translates high-level UNICORE commands into commands of the local resource management system, which is the computing center software (CCS) and the Portable Batch System PBS.

At PC2, we have currently 2 clusters running in our UNICORE infrastructure, the Arminius Cluster [6] and the Itanium-based SFB (IC-)Cluster [7]. They are integrated in our UNICORE infrastructure in the past already and in the scope of the D-Grid initiative. The installed UNICORE has been completed by a new UUDB version, which eases the user-authorization setup, as users in different VO's can know choose between different accounts which have the access-rights for their projects.

Various UNICORE systems are now integrated the D-Grid using a central gateway, thus forming a powerful Grid system. The UNICORE Vsite names are upb-arminius for the ARMINIUS cluster and upb-ic-master for the SFB cluster.

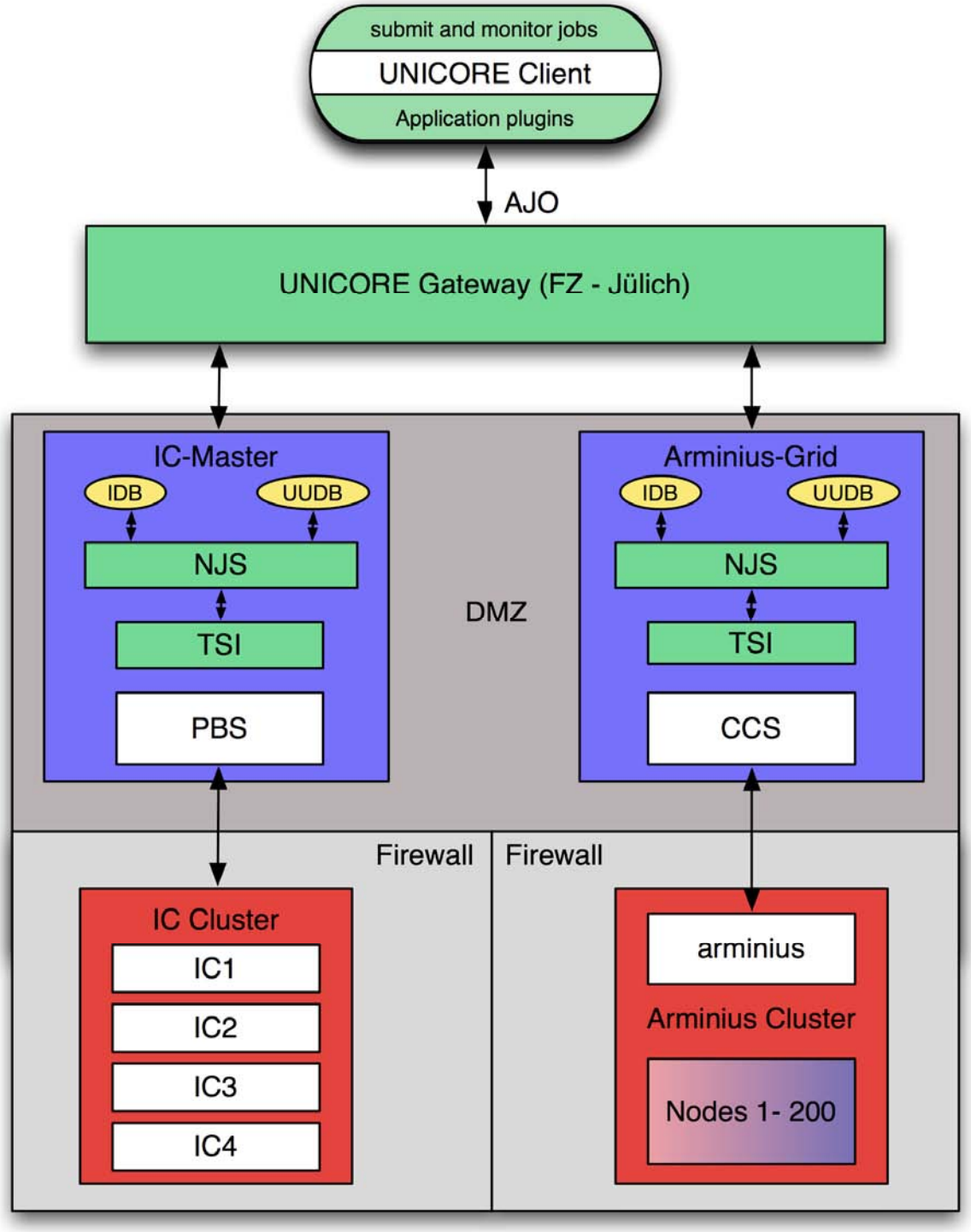


Figure 1: UNICORE at PC²

Globus Toolkit

The work on Globus Toolkit [8] started in 1998 at Argonne National Laboratory and University of Chicago. The goal was to establish an open platform for Grid computing, following open standards, e.g. defined in the Open Grid Forum (OGF) [9]. Meanwhile version 4 of the Globus Toolkit is available, which is based on open standard Grid services. Since Globus Toolkit has rapidly evolved and is used in a multitude of projects, it has become the "de-facto standard" in Grid computing. The Globus Toolkit is in use at the PC^2 since 2001, allowing users to access compute resources at PC^2 by submitting compute jobs via the Globus toolkit infrastructure.

For the D-Grid project, the Globus Toolkit installation at the PC^2 has been updated to the most recent version 4.0.4. Having the new interface available, users from D-Grid are now able to access resources from the ARMINIUS Cluster and the IA-64 architecture SFB cluster using their local Globus Toolkit 4 system as shown in Figure 2.

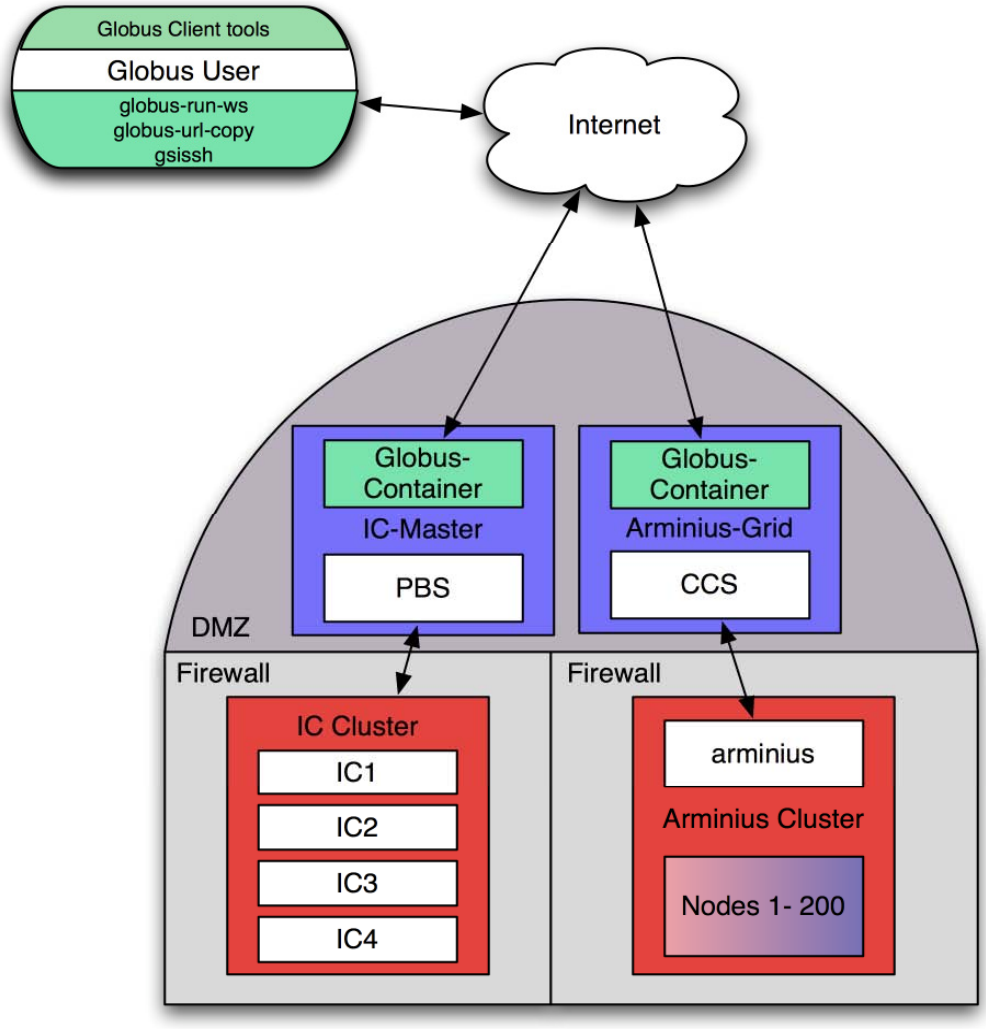


Figure 2: Globus Toolkit at PC²

The gateways are accessible under the URL arminius-grid.uni-paderborn.de:8443 and ic-master.uni-paderborn.de:8443. The Paderborn Globus-Toolkit resources are attached to the Core D-Grid Web-MDS Monitoring system of the LRZ in München.
This way, users have been able to easily access resources at the PC^2 using standard Globus mechanisms - without knowing about the usage of the resource management systems CCS and PBS, which is used for operating cluster systems at the PC^2.

Portal

The D-Grid web portal [10] serves the purposes of external presentation and internal collaboration.

Figure 3: D-grid Web Portal

The Typo3 system, a content management system used for the web sites of many enterprise level companies, has been installed to allow members of the D-Grid initiative and its community projects an easy access to their individual web appearances while preserving a consistent layout throughout the whole web site. Apart from an online rich text editor, Typo3 provides necessary authentication functionalities, automatic menu generation, and support for bilingual content. Within the scope of the portal project, we have implement a Typo3 plug-in to support a certificate based single-sign-on solution throughout various web sites of the D-Grid Initiative and the community projects. This sign-on grants access to internal areas, which provide online collaboration tools, such as document sharing facilities, access to an event calendar, and detailed information about resources such as mailing lists, and the trouble ticket system. The start page of the web site support the user with randomly chosen information about the various D-Grid Communities. Figure 3 shows the start page of the D-Grid Web Portal. The PC^2 is in charge of creating and maintaining a central point of information and supporting other members of the D-Grid initiative at questions related to the common web appearance.

Further Development Tasks

Within the work in the DGI the PC2 committed itself to some other tasks which were outside the original description of work.

Firstly, PC2 has lead a taskforce, which developed recommendations for the nomenclature in the D-Grid. This work addresses the labels and names used to identify and access services and machines within the D-Gird infrastructure. The providers within the D-Grid are requested to abide to the rules defined. A common use of the nomenclature rules will provide an easy to use infrastructure, as the URLs of services will be intuitional guessable and thus easy to access. The results of the taskforce were presented at a Core-Grid meeting and the document with the specification is accessible under [11].

Secondly, PC2 participated at the work for the D-Grid Betriebskonzept [12]. In this document the operational concepts for the whole infrastructure are defined. This includes beneath a description of the hardware and software operation policies for providers, the VO management policies as well as user policies. The Betriebskonzept is an ongoing work. When new requirements will occur in the future, for, instance from the new community projects, they will be integrated into the Betriebskonzept.

Resource Usage

The ARMINIUS and IC cluster are part of the D-Grid infrastructure. For this to work, also a large number of other servers at PC^2 are used.

References

[1] Foster and C. Kesselman (Eds.). The Grid 2: Blueprint for a New Computing Infrastructure. Morgan Kaufmann Publishers Inc. San Francisco, 2004
[2] BIS-Grid, https://bi.offis.de/bisgrid/tiki-index.php
[3] The Globus Alliance, http://www.globus.org
[4] UNICORE Forum e.V., http://www.unicore.org
[5] UNICORE at SourceForge, http://unicore.sourceforge.net
[6] The ARMINIUS Cluster at PC2: http://www.upb.de/pc2/services/systems/arminius
[7] The SFB Cluster at PC2. http://www.upb.de/pc2/services/systems/ic
[8] The Globus Toolkit, http://www.globus.org
[9] The Open Grid Forum, http://www.ogf.org
[10] The D-Grid web portal, http://www.d-grid.de
[11] Empfehlungen zur Nomenklatur von Ressourcen und Diensten im D-Grid, http://www.d-grid.de/fileadmin/dgi_document/FG2/Kern-D-Grid/Dokumente/D-Grid-URI-NomenklaturV1.1.pdf
[12] Betriebskonzept für die D-GridInfrastruktur: http://www.d-grid.de/fileadmin/user_upload/documents/Kern-D-Grid/Betriebskonzept/D-Grid-Betriebskonzept.pdf

5.2.5 DELIS: Large-Scale P2P Data Management

Project coordinator	Prof. Dr. Friedhelm Meyer auf der Heide, HNI, University of Paderborn Prof. Dr. Odej Kao, PC², UPB / TU Berlin
Project members	Dr. Felix Heine, PC², UPB / TU Berlin André Höing, PC², UPB / TU Berlin Giovanni Cortese, Telekom Italia Learning Services
Supported by:	6th Framework Programme of the European Commission

General Problem Description

Large scale data management is a challenging task. Through modern information technologies, more and more data and information is available online. However, it is often difficult to find the relevant information, and to efficiently combine the pieces found in various data sources. The problems lie both in the sheer amount of data and in syntactic and semantic heterogeneities.

Even within single organizations, sometimes large numbers of different data sources are available. The need to process this data collection as a whole and to draw conclusions from the aggregated information residing in the sources has lead to the field of data warehousing. The typical approach is to copy every relevant piece of information into a large, centrally managed data warehouse (DWH), which is then used for evaluating queries, e.g. in so called decision support systems [1].

However, in many cases the relevant data sources span multiple organizations. In the collaboration of multiple companies, or when mining the information contained in the World Wide Web, it is necessary to combine multiple heterogeneous information sources which are decentrally controlled. The Semantic Web initiative [2] aims at this goal.

To face these challenges, a system needs to be capable of integrating a huge number of information sources which are syntactically and semantically heterogeneous, decentrally managed, and which may be highly dynamic. Two important aspects have to be regarded. First, the underlying infrastructure must be scalable and flexible. Second, the system must handle the heterogeneities of the data sources.

With respect to infrastructure, p2p systems [3] have gained much attention of the research community in recent years. They provide a good basis for large-scale data management systems. Compared to traditional approaches, p2p systems offer good scalability features combined with decentralized control and flexibility.

In the context of the EU project DELIS we are working on the BabelPeers p2p data management system which fulfills the above mentioned demands. We utilize a structured p2p network called Pastry [4] and use the W3C standard Resource Description Framework (RDF) [5] for knowledge description. We are already developing this system since 2004. Within the reporting period, we focused on top k query evaluation, load balancing issues, and we have integrated the system with an XML based system called XGR developed by a partner in DELIS.

Problem details and work done

In general, we assume to have a large p2p network. Each node in the network has some local knowledge stored using RDF. Knowledge in RDF is described as a set of triples, which can be regarded as short sentences of the form subject, predicate, object. A set of triples forms a labeled, directed graph.

The nodes also have local schema knowledge stored as RDF Schema [6] triples. The schema knowledge does not need to be the same for every node. In fact, we are convinced that it is impossible to ensure synchronization of schema knowledge in large world-wide distributed environments or to restrict the schema to a single common standard. Moreover, it is desirable to allow each node to add locally needed schema information on the fly. If new entities need to be described, new classifications may become necessary. Waiting for a new version of some standard schema does not solve this problem.

However, we assume that there is an ontology which serves as a common schema, at least for some subsets of the nodes. This ontology will be the basis which can be extended locally. Additional schema knowledge may be stored to allow translation from one ontology to the other. Without such common understanding, no interoperability would be possible.

Our desired result is to put all this knowledge from all the nodes virtually in one pool, apply RDFS entailment rules to this pool and evaluate queries with respect to the union of the knowledge, see Figure 1. This approach is very beneficial, as overlaps in the schema knowledge are used to build bridges between different schemas used by different nodes.

A query is formulated as a pattern consisting of multiple triples where parts of the labels are replaced by variables. Thus query answering is essentially sub-graph matching, see [7]. However, in our scenario, the model graph is distributed over the nodes of the p2p network. Our first stragegy, developed in the previous year, works in two phases. In the first phase, a sub-graph is fetched from the network which has to be large enough to contain all query results, but should be as small as possible to save network transmission time. In the second phase, the matching sub-graphs for the query are searched locally within the retrieved sub-graph.

However, this kind of query processing is only suitable for small result sets. For very generic queries with millions of hits, the generated network traffic is too large. Furthermore, users are normally not interested in every hit. Just like a Google search, the desired match is typically within the first 10 or 20 matches.

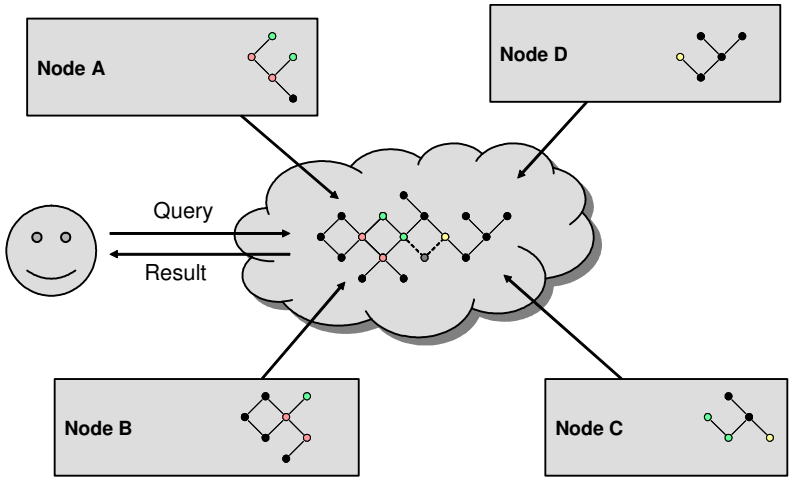


Figure 1: Virtual pool of knowledge

For this situation, we developed a top *k* distributed search algorithm that finds the *k* best matches for a given query. “Best” in this scenario means those hits that have the highest value for a given attribute. The algorithm uses a single integrated phase to fetch the candidates and to generate the results, in contrast to our previous strategy. Thus the candidates are fetched from the network as needed. As a nave implementation would result in large amounts of very small packets transferred over the network, improvement was necessary.
For this, we used a combined caching and lookahead strategy. Caching means that past candidates are cached, because it is likely that they will be reused during the backtracking algorithm that does the sub-graph matching. Lookahead means that when fetching candidates for a given query, the current query situation is analyzed to guess which candidates might be needed in the near future from the same peer that is being asked. This guess is then used to fetch more than a single candidate.
Both methods together lead to an efficient algorithm. We have evaluated this algorithm and compared it to our previous method.

The evaluation has been performed on the Arminius cluster of PC². We have generated an artificial model graph together with a set of 1000 test queries. We have started the network using various node counts and have measured the overall throughput of the system, by starting a query client on each node which looped over the set of test queries. The result is depicted in Figure 2.

Both the exhaustive evaluation method (named SG1/B) and the top k method scale well. Although the top k version has smaller speedup values, the overall performance is much better. We can see that on average about 450 queries per second are answered in the 128 node network using top k evaluation, while the exhaustive version can only process around 120 queries. The smaller speedup values in the top *k* case are mainly caused due to the effect with networks with a few nodes.

The knowledge distribution in BabelPeers suffers from a load balancing problem. The RDF triples are distributed three times, by hash values of its subject, predicate, and object. So the query evaluator can find all triples, even if it knows only one part and the other are undefined. The hash value additionally determines the node where the triple is stored. Because of the unequal appearances of URIs, there are some nodes that suffer from load imbalance. This load imbalance has effects on the query evaluation performance and on storage load distribution. Existing load balancing strategies like virtual nodes [17] do not work in this case because they cannot split the amount of data mapped to a single hash id.

Therefore we examined other load balancing possibilities regarding to two different kinds of load: storage load and query load. Without global knowledge it is hard to determine which nodes are overloaded because the load can only be regarded with respect to the load of the other nodes in the network. Therefore we developed a simple, sample based load detection mechanism which compares the own load with samples collected from various other nodes from the peer-2-peer network. Load indicators can be either the storage load or the network load caused by query answering traffic. On this basis, we implemented a load balancing strategy which relocates triples from an overloaded node to a less loaded node. So a tree of helper nodes is created by the system and each node forwards the query to its children. Each child sends its own result set to the requesting node which combines all answers to one common answer.

Load balancing flattens the amount of network traffic over the peers so that the overall query evaluation performance grows and the system becomes faster.

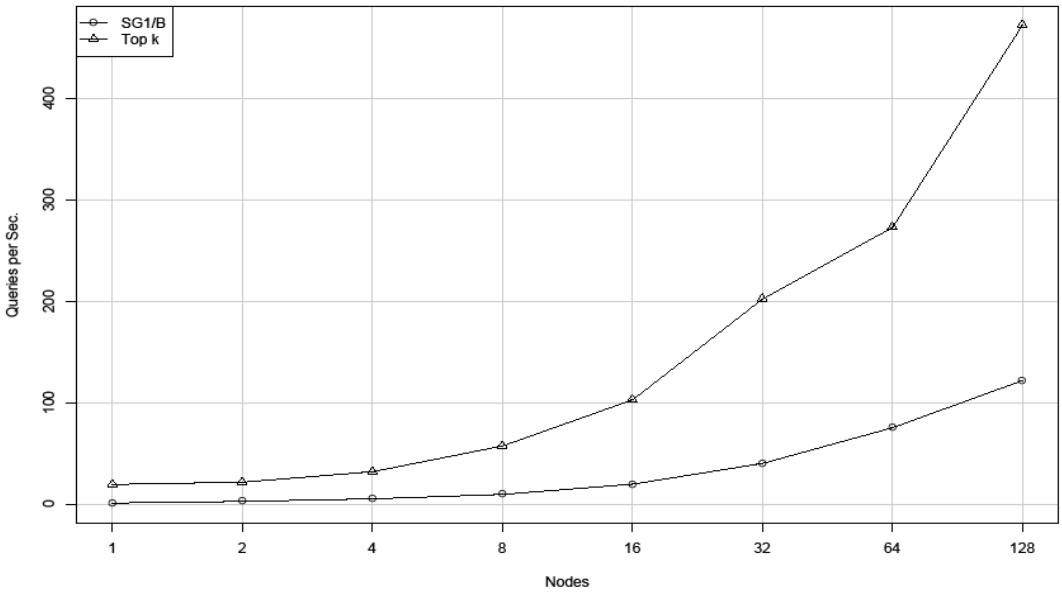


Figure 2: Evaluation of top k queries (on Arminius)

In our third branch of work, we have integrated our prototype with a prototype developed by Giovanni Cortese within the scope of the DELIS project. This system (called XGR) focuses on storing and querying XML documents in a P2P network. In contrast to our system, XGR does not distribute the documents over the network; they rather stay at their origin. Thus both systems have their strength and weaknesses. While BabelPeers is very strong in combining information and reasoning about this combined information, XGR is better in handling dynamic data that is updated often. Thus the idea was to combine both systems to get both benefits. Figure 3 shows the scenario. A p2p network is established where each node both runs the BabelPeers and the XGR system. A client has capabilities to query both systems. The BabelPeers system is used to store semantic meta-data, this is used to find the correct nodes that store the desired information in XML using XGR. From the results of the BabelPeers queries, new queries for the XGR system are designed. The resulting system combines the best of both worlds.

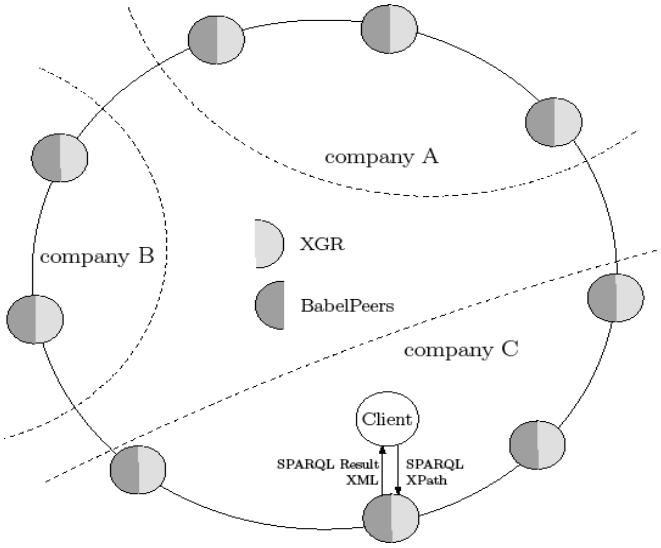


Figure 3: Running BabelPeers and XGR

We have published the results of our work in various papers; the first [10,11,12] focus on the overall system. [13] describes the top k strategy, while [14] describes the joint prototype with XGR. The issue of load balancing is discussed in [15] and [16]. The BabelPeers activities also resulted in the Ph.D. thesis of Felix Heine, see [18].

Resource Usage

P2P networks are difficult to evaluate. As they target large deployed systems with possibly thousands of nodes, it is normally not possible to test the system in the desired scale. Thus simulations of the network are a good way to gain insight into a system's properties. We ran various simulations using the Arminius cluster for the evaluation of our system. Additionally, we have deployed the system on the cluster to measure some properties on the real system. Although the scale is smaller than the target scale we could already see interesting effects like load-balancing influences. Thus we used Arminius both as a compute farm for the simulations (parameter studies) and as a truly parallel system (with the Ethernet interconnect) to test a P2P network with up to 128 nodes.

References

[1] Inmon, W.H.: *Building the Data Warehouse,* John Wiley & Sons, 2005

[2] Berners-Lee, T.; Hendler, J. and Lassila, O.: *The Semantic Web*, Scientific American, May 2001

[3] Steinmetz, R. and Wehrle, K.: *Peer-to-Peer Systems and Applications*, Springer, LNCS 3485, 2005

[4] Rowstron, A. and Druschel, P.: *Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems*, Proc. of the 18th IFIP/ACM International Conference on Distributed Systems Platforms, 2001

[5] Manola, F. and Miller, E.: *RDF Primer*, http://www.w3.org/TR/rdf-primer, 2004

[6] Brickley, D. and V. Guha, R.: *RDF Vocabulary Description Language 1.0: RDF Schema*, http://www.w3.org/TR/rdf-schema, 2004

[7] Ullmann, J.R.: *An Algorithm for Subgraph Isomorphism*, Journal of the ACM, 23:1, pages 31-42, 1976

[8] Kubiatowicz, J.: Extracting Guarantees from Chaos, Communications of the ACM, 46:2, pages 33-38, 2003

[9] Bloom, B.H.: Space/Time Trade-offs in Hash Coding with Allowable Errors., Communications of the ACM, 13:7, pages 422-426, 1970

[10] Heine, F.; Hovestadt, M. and Kao, O.: *Processing Complex RDF Queries over P2P Networks.* Workshop on Information Retrieval in Peer-to-Peer-Networks P2PIR 2005, November 4, 2005

[11] Heine, F.: *Scalable P2P based RDF Querying*. In First International Conference on Scalable Information Systems (INFOSCALE06), 2006.

[12] Heine, F.: *Emergent Schema Management for P2P-based Applications.* In: Roberto Baldoni, Giovanni Cortese, Fabrizio Davide, Angelo Melpignano (Ed.), Global Data Management, IOS Press, pp. 224-248, 2006

[13] Battré, D.; Heine, F. and Kao, O.: *Top k RDF Query Evaluation in Structured P2P Networks.* In: Wolfgang E. Nagel, Wolfgang V. Walter, Wolfgang Lehner (Ed.), Euro-Par 2006 Parallel Processing: 12th International Euro-Par Conference, Springer, pp. 995-1004, 2006

[14] Battré, D.; Cortese, G.; Heine, F. and Höing, A.: *Accessing XML Documents using Semantic Meta Data in a P2P Environment.* In: Siegfried Handschuh, Andreas Abecker, Tudor Groza (Ed.), The 20th International FLAIRS Conference, Special Track on Artificial Intelligence and Social Semantic Collaboration (SEMSOC 2007), 2007

[15] Battré, D. Heine, F., Höing, A. and Kao, O. *On Triple Dissemination, Forward-Chaining, and Load Balancing in DHT based RDF stores.* In: Sonia Bergamaschi, Sam Joseph, Jean-Henry Morin, Gianluca Moro (Ed.), Fourth International Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P) in scope of 32st International Conference on Very Large Data Bases, Springer, 2006

[16] Battré, D., Heine, F., Höing, A. and Kao, O. *Load-balancing in P2P based RDF stores.* In: Holger Wache, Heiner Stuckenschmidt, Bijan Parsia, Yuanbo Guo, Tim Finin, Dave Beckett (Ed.), Second International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2006) in scope of the International Semantic Web Conference, pp. 21-34, 2006

[17] Karger, D.R.; and Ruhl, M.: *Simple Efficient Load Balancing Algorithms for Peer-to-Peer Systems*. In: SPAA '04: Proceedings of the sixteenth annual ACM Symposium on Parallelismin Algorithms and Architectures. ACM Press. 2004

[18] Heine, F.: *P2P based RDF Querying and Reasoning for Grid Resource Description and Matching*. Dissertation, Paderborn, 2006

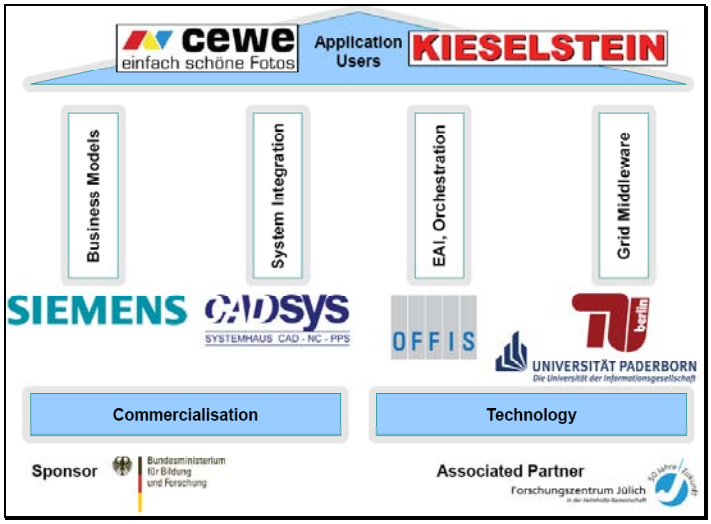
5.2.6 BIS-Grid: Betriebliche Informationssysteme: Grid-basierte Integration und Orchestrierung

Project coordinator	Prof. Dr. Holger Karl, PC², University of Paderborn
Project members	Jens Lischka, PC², University of Paderborn
	Holger Nitsche, PC², University of Paderborn
Supported by:	BMBF (Förderkennzeichen 01/G07005A)

General Problem Description

Business Information Systems (BIS) facilitate resource management within an organisation (ERP systems), efficient customer relationship management (CRM systems), and production and service support (PDM systems), to name a few. In order to create such information systems, often both existing and new subsystems have to be adapted and integrated. Thereby, workflow systems are employed to support the effective and efficient integration of organizational workflows by means of subsystem services orchestration. In doing so, Enterprise Application Integration (EAI) has gained a tremendous relevance to the realization of business processes that typically affect multiple information systems within one or multiple organisations. This integration of information systems is accomplished via orchestration in service-oriented architectures (SOA), which especially aim at effectively mapping the application domain to the technology. Since Grid Services are introduced with corresponding standards that are based on Web service technologies (such as WSRF), Grid technologies can also be used to build SOA. Thereby, Grid technologies and EAI have much in common since both technologies focus on integration problems within a heterogeneous environment – Grid technologies on resource level and EAI on application level.

In BIS-Grid [1], we intend to realise a horizontal Service Grid in the application domain of business information systems. The overall goal is to enable Grid technologies to be used for the integration of decentralised business information systems. This will be achieved by developing and providing organisational and technical extensions based on the current state of the art in Grid technologies, EAI, and SOA. On the organisational layer, we will develop new forms of inter-organisational collaboration and new business models. On the technical layer, we will extend WS-BPEL-based orchestration to allow the orchestration of state-full WSRF-based Grid Services. Thereby, we plan to integrate existing middleware systems that are already well established in the commercial sector and Grid-specific middleware systems such as UNICORE 6 [2].



Picture 1 BIS-Grid partners and there roles

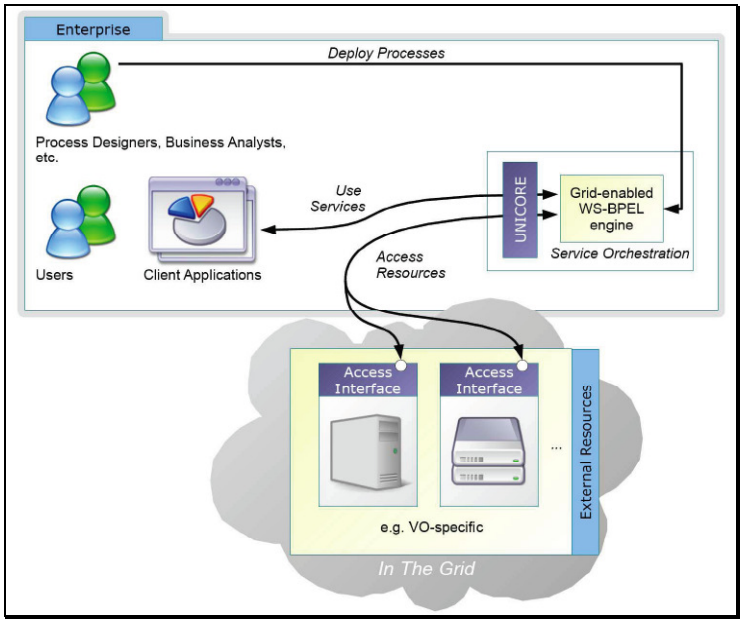
The achievement of these goals will be evaluated by means of two application scenarios at our industrial partners. We will evaluate the developed generic services exemplarily for the integration of ERP, PDM, and CRM systems, covering the major sectors of business information processing. Thus, the results of BIS-Grid provide a significant contribution for applying Grid technologies in the domain of real-world business information systems integration. To achieve an optimal exploitation of our results, we will also develop dedicated exploitation and marketing plans that especially aim at SMEs. Via Grid providing or Grid software consulting, SMEs will be capable of providing additional IT services to the market and thus will be able to improve their market position. As a consequence, Grid technologies gain access to a whole new application domain. To summarise, our intent is to extend the application potential of Grid technologies to explicitly support SMEs in business information processing.

The project is funded by the Bundesministerium für Bildung und Forschung (BMBF) as part of the German Grid Initiative Phase 2 “D-GRID 2” [3]. The project started on 01.05.2007, the kick-off meeting was held on 1st of June 2007 in Oldenburg, and the project is planned to terminate on 30.04.2010.
The project coordinator is OFFIS - Institut für Informatik FuE-Bereich "Betriebliches Informations-management" (BI) in Oldenburg. The PC² is one of the research partners and is responsible for the BIS-Grid integration into D-GRID.

Problem details and work done

As BIS-Grid is intended to be enterprise oriented, UNICORE was chosen as the supported Grid-middleware due to it's easy installation, the consistent design and usability for novices. At the time of writing the proposal, UNICORE6 with WSRF was announced for release in early 2007.
However the UNICORE6 development was delayed and at the start of the project UNICORE6 was only available as beta software. Right after project start we installed the UNICORE6 beta development version to support the project with a test platform. Since then multiple versions of the UNICORE middleware were deployed, at the moment version 6.0.1 (release).

As the performance of grid-services is relevant, for the design and specification of the BIS-Grid architecture, we have done an extensive test series. With a simple service running in UNICORE6 internal JETTY container [4] we measured execution time with different types of X.509 certificates for local and remote calls. The tests showed significant larger latencies compared to results from a typical web service. The results have to be studied in detail, however from the first impression the larger latency is introduced by the security mechanism used in the grid middleware.



Picture 2 - One possible BIS-Grid scenario

In August 2007 all D-GRID projects got a call for proposals for "D-GRID Sonderinvestitionen". The BMBF offered additional funding for new resources in D-GRID. We submitted a proposal for the BIS-Grid consortium and a funding of approx. €220.000 were granted. The new resource is an "enterprise class" resource to meet

the needs of enterprises. The system will be installed until end of the year 2007 at the PC^2 computing facility and is mainly for the BIS-Grid community. Spare CPU cycles are offered to other D-GRID users and communities.

Resource Usage

The ARMINIUS cluster was utilised through grid middleware to experiment with the newly introduced “grid-beans” in UNICORE6 compared to the traditional plug-in concept in UNICORE5. All jobs were sequential jobs and had a very short runtime.

References

[1] BIS-Grid project homepage, http://bisgrid.d-grid.de/
[2] UNICORE (Uniform Interface to Computing Resources) homepage, http://www.unicore.eu/
[3] The German Grid Initiative (D-Grid) description, http://www.d-grid.de/index.php?id=104&L=1
[4] JETTY homepage, http://docs.codehaus.org/display/JETTY/

5.2.7 GRIDCHESS IPCCC 2006 and WCCC 2007 Tournament Participations

Project coordinator	Prof. Dr.-Ing. Dietmar P.f. Möller, Department of Informatics, University of Hamburg Dr. habil. Ulf Lorenz, Department of Computer Science, University of Paderborn
Project members	Kai Himstedt, Department of Informaticsm University of Hamburg

NOTE: The work was done by externals in cooperation with members of the PC² core group and by using the available hardware at the PC², which was administrated and configured for our special requirements with great engagement by Axel Keller, Tobias Schumacher and Dr. Jens Simon.

General Problem Description

For our GRIDCHESS system an idea is used where a cluster, or a parallelized chess program running on a cluster respectively, forms a base component, and with a second parallel approach on top several such clusters can be used for a distributed game tree search. The top level parallel search method named Optimistic Pondering [1] is used to speculatively search ahead with the expected best moves of the opponent. If a modern chess program does its computation for the next move, it generates, as a side effect, quite a good prediction for several next moves of the game. This prediction allows us not only to make useful computations for the next move, but we can also predict the opponent's next move, and then already compute on the move thereafter. If our prediction is correct, we will have gained time for the tree search. If it is not correct, we will waste our resources. In fact, we build a kind of process-pipeline, following the principal variation of the underlying chess programs. The design is similar to a pipeline within a processor, where branch-prediction allows for better performance. Interestingly, this idea can be further expanded. Instead of using a pipeline of processing units, one can build a tree of pipelines as well, where several possible branches are examined simultaneously. The scheduling for Optimistic Pondering is based on the rarely changing information about the predicted next moves delivered by the parallel base programs and is well suited for Grid environments. Grid environments can normally not be used for a distributed game tree search due to the demands for high speed interconnection networks of the classic parallelization approaches.

The GRIDCHESS system combines in its prototypical realization the two levels of parallelization methods, using the Young Brothers Wait Concept (YBWC) [2, 3] as the state of the art approach for the parallelization of the base programs. As we had to coordinate resources which do not belong to one site only, as we used standardized communication interfaces, and as we coordinated computational power in non-trivial fashion, we have chosen the name GRIDCHESS for our system.

The public character of tournament participations is an ideal base to show how a new approach performs in real practice. More recently, “man vs. machine” matches like the one by Wladimir Kramnik against DEEP FRITZ [4], with a clear win for the chess program, confirm that the end of an era seems to be reached and computers are dominating now in this discipline. But of course “computer vs. computer” matches are still suitable to test new algorithms in the domain of game tree search, and besides there is a public and commercial interest in such events. The participation in the 16th International Paderborn Computer Chess Championship (IPCCC) in December 2006 [5] and the 15th World Computer-Chess Championship (WCCC) in June 2007 in Amsterdam [6] with GRIDCHESS under real computer match conditions and playing against several of the top computer chess programs of the world was chosen to analyze its behavior and performance in practice.

Problem Details and Work Done in the Reporting Period

THE GRIDCHESS system is based on a meta chess engine concept: Chess programs (so-called chess engines) are usually controlled by graphical user interfaces (GUIs) (e.g. XBoard and WinBoard [7], Arena [8] or the ChessBase GUI [9]) via standardized engine protocols, and the first major component of the system is a proxy chess engine based on Hyatt’s well known CRAFTY [10] source code to behave to the GUI like a normal chess engine. The CRAFTY tree search, e.g. any kind of evaluation functionality, is not used and is in fact irrelevant for the proxy chess engine. But internal data structures, the protocol handling for the GUI communication and the time control ideas of CRAFTY were reused and adopted for the proxy chess engine to perform the Optimistic Pondering. Workers – controlled by the proxy engine – form the second major component and are able to control associated base chess engines running on clusters. As a simplified explanation of Optimistic Pondering here, one can imagine the workers forming a tree of pondering pipelines with expected opponent moves extracting this information from the principal variations (PVs) provided by the chess engines. A worker may be seen as a middleware component based on a small C++ framework used to handle different engine protocols achieving an utmost independence from concrete base chess engines. Letouzey’s FRUIT [11] and Gaksch’s TOGA (based on FRUIT) [12] are the strongest open source chess engines. It was therefore an obvious decision to use their code as

the basis to implement Feldmann's parallel Young Brothers Wait Concept at the single cluster level. YBWC is a state of the art approach to parallelize game tree search at the single cluster level and was very successfully used for systems like HYDRA [13, 14]. The principle of the YBWC is to defer the parallelization in each node up to the time when its eldest brother has completely been searched – i.e. the "young brothers wait". This implicitly avoids non-meaningful parallelization and the related search overhead. The scheduling is implemented by a "work stealing" mechanism to balance the load dynamically. The YBWC parallelized versions were named FRUITP and TOGAP.

Resource Usage

During the 16th International Paderborn Computer Chess Championship held from December 27 to 30, 2006, the PC² provided us with the computing power of 16 Intel Xeon cores. GRIDCHESS was configured to use 2 workers, each consisting of FRUITP using 8 cores. Even in the mix of playing against seven different top chess engines during the tournament the ponder hit rates achieved are evenly high, reaching nearly 60% averaged over all games. The top engines of the tournament were RYBKA [15] (1st) and SHREDDER [16] (2nd). Against these engines the highest ponder hit rates were achieved, which indicates a similar kind of playing style of FRUITP and these top engines. GRIDCHESS reached a 5th rank of 10 in the tournament which was also its debut.

The most outstanding event was the participation in the 15th World Computer-Chess Championship in June 2007 in Amsterdam. For this tournament the 16 Intel Xeon cores of the cluster at the PC² and additional 64 AMD Opteron cores of the cluster provided by the Regional Computer Centre at Hamburg University [17] were available during the tournament. GRIDCHESS was configured to use 5 workers, each consisting of TOGAP using 16 cores, in a small dynamic tree of pondering pipelines to increase the ponder hit probability by expecting several opponent moves. Expecting several good moves can increase the ponder hit rate noticeably (up to 90% or more against 6 of 11 opponents). GRIDCHESS reached a score of 7.0 in 11 games and shares the 4th place of 12 with the 12-fold world champion SHREDDER. SHREDDER just won the speed chess tournament discipline in Amsterdam and became "World Computer Chess Speed Champion". Beating SHREDDER in the main tournament could impressively show the performance of GRIDCHESS under real tournament conditions.

So far we have used just 5 workers for GRIDCHESS. How to arrange a larger number of workers in the tree to get the best overall performance is a subject of our actual examinations.

References

[1] Himstedt; K.: An Optimistic Pondering Approach for Asynchronous Distributed Game-Tree Search. In ICGA Journal. Vol. 28 No. 2 (2005):77-90.

[2] Feldmann, R.: Game Tree Search on Massively Parallel Systems. PhD Thesis, University of Paderborn, Paderborn, 1993.

[3] Feldmann, R.; Mysliwietz, M. ans Monien, B.: Studying Overheads in Massively Parallel Min/Max-Tree Evaluation. In 6th ACM Annual symposium on parallel algorithms and architectures (SPAA'94), pages 94-104, New York, NY, 1994.

[4] ChessBase: Kramnik vs DEEP FRITZ: Computer wins match by 4:2. http://www.chessbase.com/newsdetail.asp?newsid=3524

[5] 16th International Paderborn Computer Chess Championship (held from December 27 to 30, 2006): http://wwwcs.uni-paderborn.de/~IPCCC/ipccc2006.html

[6] 15th World Computer-Chess Championship (held from June 11 to 18, 2007 in Amsterdam): http://www.grappa.univ-lille3.fr/icga/events.php

[7] XBoard and WinBoard graphical user interfaces for chess: http://www.tim-mann.org/xboard.html

[8] Arena, graphical user interface (GUI) for chess engines: http://www.playwitharena.com/

[9] ChessBase: http://www.chessbase.com

[10] CRAFTY (Hyatt's home page): http://www.cis.uab.edu/info/faculty/hyatt/hyatt.html

[11] FRUIT - pure playing strength: http://www.fruitchess.com/

[12] TOGA II: http://www.superchessengine.com/toga_ii.htm

[13] HYDRA: http://www.hydrachess.com

[14] Donninger, C. and Lorenz, U.: The Chess Monster Hydra. In Proc. of 14th International Conference on Field-Programmable Logic and Applications (FPL), pages 927-932, Antwerp - Belgium, 2004.

[15] RYBKA: http://www.rybkachess.com

[16] Meyer-Kahlen, S.: Shredder Computer Chess. http://www.shredderchess.com/

[17] Regional Computer Centre, University of Hamburg Home Page: http://www.rrz.uni-hamburg.de/RRZ/e.index.html

5.3 Distributed High Performance Computer Graphics

5.3.1 Computational Steering of Interactive and Distributed Reality Applications

Project coordinator	Prof. Dr. Odej Kao, PC², UPB / TU Berlin
Project members	Dipl.-Ing. Cord Bauch, L-LAB, University of Paderborn Dr.-Ing. Jan Berssenbrügge, HNI, University of Paderborn Dr. Mathias Fischer, HNI, University of Paderborn Dipl.-Inform. Stefan Lietsch, PC², University of Paderborn Dipl.-Inform. Henning Zabel, C-LAB, University of Paderborn
Work supported by	Ministry of Innovation, Science, Research and Technology of the State of North Rhine-Westphalia

General Problem Description

In this article we present a system that transfers the well-known computational steering paradigm to interactive and distributed Virtual Reality applications. Those are often used in areas like rapid prototyping and all kinds of vehicle simulation. The distribution has many different purposes and affects various subsystems of a VR application. Most of the currently existing systems are very specialized and have a proprietary design for data-exchange and coupling of the components. We propose a more flexible approach by designing a computational steering framework that is well-adapted to the needs of highly interactive and distributed VR systems. Thereby we achieve higher reusability and scalability for the steering component itself as well as the possibility to exchange and compare subsystems. As a proof-of-concept we adapted an existing driving simulator to the proposed computational steering framework and discuss the advantages and difficulties in the second part of the paper.

Introduction

With today's ever increasing computational power and the possibility to simulate various real-world processes, scientists and engineers have a powerful tool to do efficient research and development. One big step towards common availability of High Performance Computing and Simulation was the introduction of cluster systems

which today almost completely replace proprietary High Performance Computers. By the utilization of cheap commodity hardware and the coupling over fast networks, current cluster systems provide a huge amount of computational power at comparably low costs. To make use of the vast amount of data such systems produce, new ways of visualization had to be found. With a similar approach as for parallel computing, visualization clusters were developed that provide enough performance in terms of polygon throughput and scalability by clustering standard graphics hardware. With those powerful tools at hand the simulation of complex real-world processes became feasible. The next step is to not only simulate but also interact with these processes, which leads to the idea of computational steering.

Computational steering is a well-known and valuable method in science as well as in some industrial environments. It helps controlling large scale software projects by coordinating parameter settings of the running program and visualizing the computed results immediately. Computational steering offers even more potential in combination with Grid computing. The coupling and steering of multiple simulations running on different sites allows to process vast amounts of data and opens new perspectives for scientists all over the world. However a lot of work still needs to be done to develop appropriate tools and frameworks for various tasks. The most important, current architectures are presented and briefly described in chapter *Related Work*. They fit for many applications in High Performance Computing such as Computational Fluid Dynamics (CFD), Molecular Dynamics (MD) or crash simulation. All of them have in common that they are not interactive by nature. Interaction is basically limited to changing simulation parameters such as the position of an object in a flow channel.

Our focus lies on the research of highly interactive distributed applications namely distributed Virtual Reality systems. This is quite a new but very promising field in High Performance Computing. For a long time Virtual Reality applications were limited in terms of scalability, performance and flexibility. Proprietary systems on specialized hardware were used to simulate fixed use cases such as driving or flight simulators or material simulation for rapid prototyping. With a flexible framework for distributed Virtual Reality on a computing and visualization cluster we provide a basis for highly flexible and scalable VR systems that can be customized and compared by the use of various modules. One exemplary application is a distributed driving simulator called Virtual Night Drive (see [2] and [4]). It serves as a demonstrator for the proposed framework.

Related Work

Computer simulations help scientists to study the behavior of complex systems and to speed up design and development in advanced fields of engineering. With simulations running in environments where accessibility is exchanged for the

uttermost computational power; controlling and surveillance of such simulations requires more effort and thought than on conventional computers. Therefore a collection of methods and techniques has been developed to accomplish what is known as *computational steering*. During the last decade a number of frameworks have been published which address the task of computational steering. The most prominent solutions are: SCIRun [PC1995], CavernSoft G2 [PJ2000] and Cumulvs [7]. The computational steering methods of these frameworks however are mostly concerned with the steering of long running simulations and applications, and unfortunately, none of the discussed frameworks satisfies the requirements of a flexible steering framework for highly interactive applications.

To demonstrate the opportunities of computational steering in VR applications we utilized an existing VR driving simulator (see Figure 1). The Virtual Night Drive simulator was developed in cooperation with a company that produces automotive headlights. Thus, the main focus lies on the realistic simulation of these lights. Since the simulator is partly used to do psychological tests with human probands it also has to offer a car-like steering unit (steering wheel, pedals, clutch etc.). Additionally, since the luminance distribution of headlights is very complex a shader-based approach was developed to provide a very high level of realism. To prepare the application for computational steering it was split up into simulation, visualization and input components and appropriate interfaces were designed. The various theoretical aspects in this paper will be exemplified with this sample application.

Figure 1 The distributed Virtual Night Drive simulator

Extending Computational Steering for Distributed VR

Use cases of Computational Steering in distributed Virtual Reality

Computational steering initially refers to the steering and visualization of a single simulation. This basic definition is very strict and has only limited flexibility in the application of computational steering in areas different from HPC. Traditionally, there is only one simulation, one Visualization and one side of user interaction. By extending each of these domains, three additional use cases can be identified that are needed for the successful transfer of Computational Steering to distributed Virtual Reality applications.

Collaborative Computational Steering

In order to successfully collaborate on a scientific distributed simulation, it is necessary to extend the traditional paradigm of computational steering. Any scientist involved should be able to access a running high performance simulation independently of his location. In this scenario it is possible for one researcher to work on the level of detail of a simulation model and for another to study its particular performance figures. Figure 2 shows an example of a collaborative computational steering setup. Multiple sites connect to one high performance simulation, each with its own visualization and methods for user interaction. The simulation can provide several parameters for steering by the user. With multiple users steering the same application, care must be taken to guarantee consistency of steering parameters. In order to avoid conflicts during the access to a parameter, it should be possible to lock certain parameters through locking mechanisms.

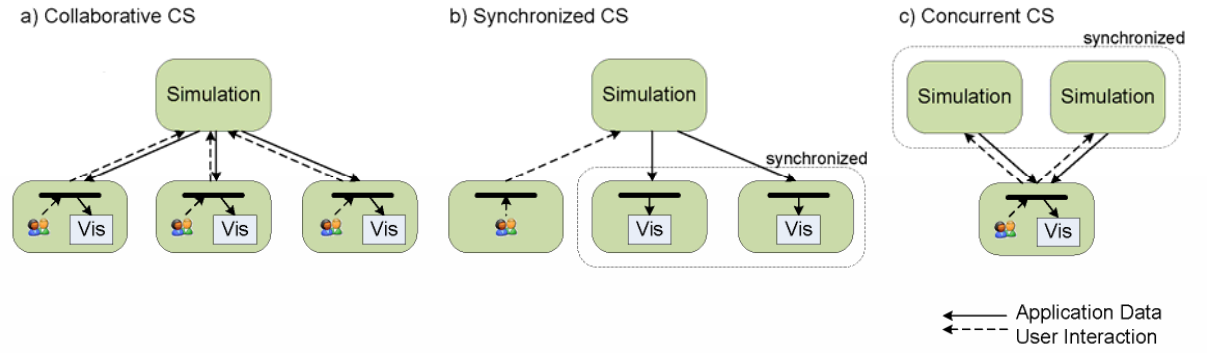


Figure 2 Three additional use cases of computational steering of distributed virtual reality systems

Collaborative Computational Steering especially makes sense in distributed Virtual Reality since many of the existing applications include multi-user scenarios by nature. In a driving simulator for example several users can control cars in the same virtual world.

Synchronized Computational Steering

This variation of traditional computational steering is a direct requirement of two recent and widespread visualization technologies: very high-resolution displays and 3D data visualization (see e.g. Figure 3). Both cases require decent frame synchronization across multiple visualization nodes. High-resolution visualization is required by users working with data that is too extensive and detailed to fit onto conventional screen, e.g. chip designers, city planners. Since the data throughput of high-resolution displays\ tiled displays vastly exceeds data rates offered by high end graphic cards, data is fed to a tiled display by multiple graphic nodes, each displaying a magnified partial view of the original scene. The side by side positioning of displays in a tiled wall makes a possible time offset between displays disturbing for the viewer. This is especially true for volatile visualizations with large objects crossing multiple tiles of a display.

Figure 3 Synchronized CS to render on a tiled display

The second emerging visualization technology where time offsets in visualization are more than a minor annoyance is 3D visualization. For a three-dimensional illusion of a data visualization it is necessary to create a pair of 2D images, of which each image represents a slightly different perspective of the same scene. This slight difference is set up in such a way that the separate presentation of each image to the eyes of the viewer creates an artificial depth perception. Figure 2b) shows an example of a computational steering environment where timely presentation of two visualizations is required. In this case, two nodes form a synchronization group in order to minimize time offsets. Different processor loads and large scheduling differences of the nodes' visualizations are avoided with the introduction of a third node, which is used to provide user input to the steered application.

Concurrent Computational Steering

Simulation usually is a sequential task; multiple simulations are run one after another with several parameter variations in order to study their effects. Though computational steering greatly helps to reduce the turnaround time of the parameter variation, simulation, and visualization loop, the conclusions of a simulation are still drawn based on the comparison of two or more successive simulation runs. Furthermore, it is quite difficult to perceive subtle effects of a changed parameter in a complex simulation. Researchers then have to resort to a quantitative analysis of the simulation results. In such cases, employing computational steering adds no value to a simulation study. In order to make computational steering more useful in comparative simulation scenarios, it is helpful for a framework to allow the concurrent execution of two or more simulations and with a subsequent combination of their outputs into one visualization. Figure 2c) shows an exemplary setup of a comparative simulation scenario and exposes two challenges which arise from concurrent computational steering. On the one hand, the user must be able to selectively change parameter values throughout all concurrent simulations or it must be defined which user steers which simulation. On the other hand, it would aid the usability of concurrent computational steering, if certain parameters could be treated as virtual global parameters. Setting the value of such a parameter would cause it to be set in all steered applications instantly, without further user interaction. Besides the requirement of methods for user input direction, concurrent computational steering requires synchronization of time among the concurrent simulations in order to make their respective results comparable.

Figure 4 shows a sample scenario of the Virtual Night Drive application where the extended use cases of computational steering are applied and combined. The scenario is the following: A user steers one car through the nightly scene which is displayed on a high-resolution tiled wall. To render the images for the 3x2 tiled wall three different graphics nodes (VND-VIS) are used. To ensure that every node renders the right view they are synchronized by using *Synchronized CS*. The input signal from the user's steering wheel is processed by an instance of a dynamics simulation (VND-SIM) which computes the position, direction and speed of the steered vehicle. Furthermore, three additional cars are introduced to simulate a populated scenery. They are controlled by simple AI simulations (AI-SIM). These four concurrent simulation threads (one user, three AI) are coordinated by *Concurrent CS*. Finally, the whole scenario is observed by a supervisor who has a general overview over the scene. He can control the environment of the test person and the AI simulated cars. Thus the supervisor and the user interact with the system through *Collaborative CS*.

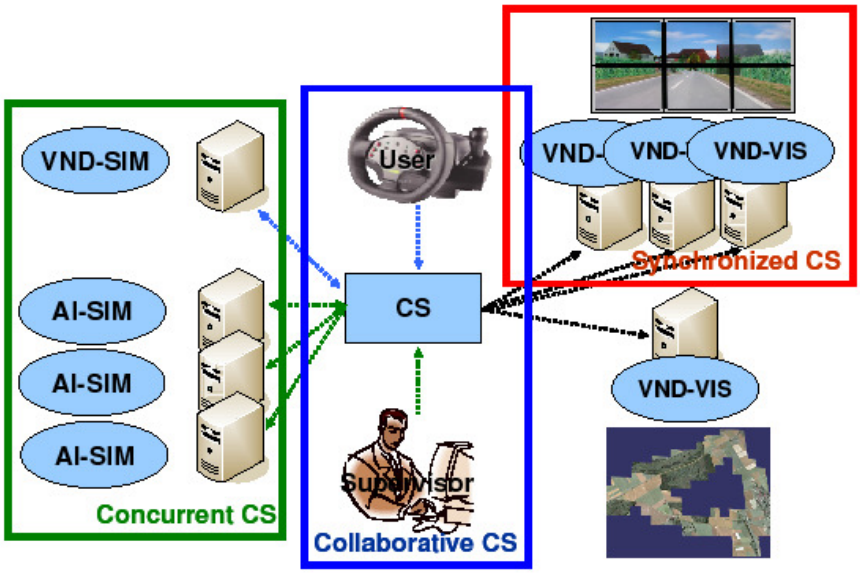


Figure 4: Extended use cases in the Virtual Night Drive application

Results, Conclusion and Outlook

The proposed framework has been implemented and tested and offers a flexible basis for various VR applications. The Virtual Night Drive simulator is the first application that makes use of most of the features of the platform and thereby achieves great flexibility. Through, utilizing the principles of computational steering the simulator is now able to support dynamic multi-user scenarios, various display setups and simultaneous simulation environments.

The framework also enabled new research possibilities in the field of distributed visualization. By using multiple nodes for the simulation of headlights and coupling them by synchronized CS we could breach given limitations of current graphics hardware. In more detail it is now possible to simulate a nearly unlimited amount of cars with headlights only depending on the number of visualization nodes available. All of them render as much lit cars as possible and send these images to one (or more) composer nodes which merges all results to one single view that is shown to the user. Figure 5 shows the idea of composing the synchronously rendered frames. Thereby the Renderer renders the unlit scene and all Shaders do the compute intensive, shader based light simulation. Additionally the Renderer and all Shaders have the same view on the scene. Since they are all synchronized through CS, all nodes have the same state of the simulations and a composition of all rendered frames is possible. This composition is done by the Composer node (cascaded Composer setups are planned to avoid the network-related bottleneck). More details and results can be found in [3].

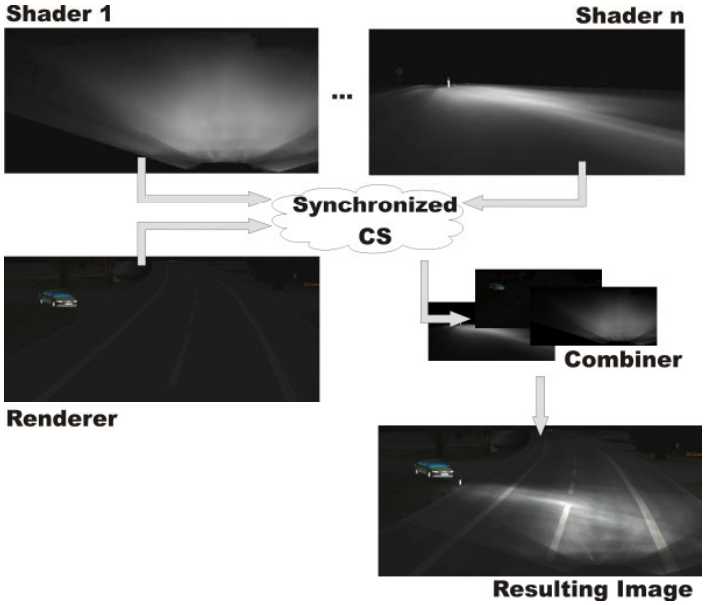


Figure 5 Composing synchronously rendered frames

By transferring computational steering techniques to the field of highly interactive simulations such as distributed VR applications we designed a platform that offers great flexibility and scalability to its users. The first steps were to analyze the differences between "traditional" computational steering and the steering of interactive simulations, to determine possible use cases and to identify the nature of the data to be transferred. Upon this analysis we designed a framework consisting of three components: a communication server (*Commuvit)* for the exchange of volatile state data, a publish\subscribe bus system *(D-Bus)* for the passing of local or global parameters, and a steering library that provides an API to integrate existing and new applications. The Virtual Night Drive simulator was implemented and optimized on the proposed architecture and serves as demonstrator for many of the described features. Through applying the extended use cases of CS to this application we created a universal system with support for multiple users, multiple simulations and distributed visualization. Other applications are currently adapted to the platform to benefit from its flexibility and scalability. To learn more about the technical and implementation details please read [1].

Resource Usage

This project uses resources of the Arminius cluster system at the PC². Especially the visualization nodes and the various displaying systems (Power Wall, Tiled Wall …) are used to run and display the visualization components. Moreover the computational steering server and some simulation components are running on the compute nodes.

References

[1] Lietsch, S., Berssenbruegge, J., and Zabel, H.., 2007: „Computational Steering Of Interactive And Distributed Virtual Reality Applications". In Proceedings of ASME 2007 IDETC and CIE, International Design Engineering Technical Conference and Computers and Information in Engineering Conference, ASME.

[2] Gausemeier, J.; Berssenbruegge, J., and Bauch, J., 2006: "A virtual reality-based night drive simulator for the evaluation of a predictive advanced front lighting system". In Proceedings of ASME 2006 IDETC and CIE, International Design Engineering Technical Conference and Computers and Information in Engineering Conference, ASME.

[3] Lietsch, S. and Berssenbrügge, J. 2007: "Parallel, Shader-Based Visualization of Automotive Headlights" In Proceedings of the 7th Eurographics Symposium on Parallel Graphics and Visualization (EGPGV07) - Short Paper

[4] Lietsch, S.; Zabel, H.; Berssenbruegge, J.; Wittenberg, V. and Eikermann, M., 2006. "Light simulation in a distributed driving simulator". In ISVC06: Advances in Visual Computing, Springer, pp. 343–353.

[5] Park, K. S.; Cho, Y. J.; Krishnaprasad, N. K.; Scharver, C.; Lewis, M. J.; Leigh, J. and Johnson, A. E., 2000: "CAVERNsoft G2: A toolkit for high performance teleimmersive collaboration". In Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST-00), K. Y. Wohn, ed., ACM Press, pp. 8–15.

[6] Parker, S. G. and Johnson, C. R. 1995: "SCIRun: A scientific programming environment for computational steering". In Proceedings of Supercomputing '95, ACM/IEEE.

[7] Kohl, J. A.; Papadopoulos, P. M. and II, G. A. G. 1997: "CUMULVS: Collaborative infrastructure for developing distributed simulations". In PPSC, SIAM. pp. 79–88.

5.3.2 A CUDA-Supported Approach to Remote Rendering

Project coordinator	PC²
Project members	Dipl.-Inform. Stefan Lietsch, PC², University of Paderborn Dr. Oliver Marquardt, PC², University of Paderborn,
Work supported by	Ministry of Innovation, Science, Research and Technology of the State of North Rhine-Westphalia

General Problem Description

In this article we present the utilization of advanced programming techniques on current graphics hardware to improve the performance of remote rendering for interactive applications. We give an overview of existing systems in remote rendering and focus on some general bottlenecks of remote visualization. Afterwards we describe current developments in graphics hardware and software and outline how they can be used to increase the performance of remote graphics systems. Finally we present some results and benchmarks to confirm the validity of our work.

Introduction

With today's ever increasing computational power and the possibilities to simulate and visualize more and more complex systems, it is obvious that efficient technologies are needed to make the results of such simulations and visualizations available for a broad audience. Remote rendering and remote visualization are techniques to fulfill this need for visual data. Since the early 90s researchers work on the topic of transmitting the rendered images or visual data to the clients of remote users to let them analyze and work with this data. There are many systems that focus on several different aspects of remote rendering/visualization. Some allow comfortable administration of remote servers and others focus on interactively showing remote users images that where rendered on powerful workstations. But all of those systems need to deal with vast amounts of data that needs to be transferred. This is because visual data is mostly pixel or voxel based. Therefore even single images which are displayed for only a fraction of a second can be consist of one or more Megabytes of data. To provide a smooth animation, 30 or more frames are needed every second. All have to get from the GPU of the server, over the network, to the GPU and finally to the display of the user. There are already many approaches to minimize the amount of data that needs to be transferred for example through prefetching certain data to the client or by introducing level of detail mechanisms and compression of all kind. But there is still no solution on how to enable users to interact with a remote system as if it was local. However, recent developments in

graphics hard- and software technology open up new possibilities to access, compress and transfer visual data on and to the graphics cards. The introduction of PCI Express for Graphics (PEG) for example cleared the way to efficiently download rendered images from the graphics card and send it over the network. This was a serious improvement over the asynchronous AGP-Bus used until then, which only granted very slow access to the relevant memory areas on the card. With the so called *Frame Buffer Extension* and the *Render Buffer Objects* of the OpenGL 2.0 specification [1] efficient methods to enable offscreen rendering and the read-back of rendered images were introduced. Besides, the only recently published CUDA framework by NVIDIA [2] opens up new ways on also using graphics hardware for compression or filtering techniques to reduce the size of the data that needs to be transferred already on the graphics card. An example of how all these techniques could be used to improve the abilities of remote rendering is given in this paper.

Remote Rendering/Visualization - An Overview

There are different classes of remote visualization for different tasks. In the following we will introduce three classes that group the most common existing systems. This should help to understand what the problems of the different classes are, and to which class the approach proposed in this paper belongs.

Client-Side Rendering

Systems where the application runs on a server but the graphics data (polygon meshes, textures, volumes) is sent to the client and rendered there belong to this class. The best known system in this class is a remote X-Server which allows to start X-based applications on a remote server as if they where local. The main disadvantage, however, is that the rendering power of the server system is not used at all while all rendering work is done by the client. This is fine for small desktop applications but insufficient for complex visualization applications that require certain rendering power. Another popular representative of this class is the Chromium [3] framework which is able to transfer the whole OpenGL stream from a server to a client. This is only one feature of Chromium but has the same problem as the remote X-Server, namely it does not use the servers rendering power. Chromium is a very flexible framework and is not only designed to do remote rendering. It also offers a lot of features to support all kinds of distributed and parallel rendering tasks. But for this work we only consider the remote rendering capabilities and classify them as Client-Side Rendering.

Server-Side Rendering for 2D and Administration

This class contains all systems that are mainly used for server administration and remote control. They render the images on the server side, compress them, and send them to the client where they can be viewed with simple viewers. They also work with low bandwidth connections and mostly show the whole desktop of the remote server. They perform well with simple 2D and little interactive applications (e.g. administration and settings dialogs) but they are insufficient for highly interactive 3D applications such as a driving simulator or an interactive 3D viewer. The most popular representatives of this class are the Virtual Network Computing Framework (VNC) [4] and Microsoft's Terminal Service architecture [5]. Both offer possibilities to adapt the remote visualization to the available bandwidth and client device. However they are not optimized to achieve high frame rates and low latency for interaction.

Server-Side Rendering for 3D application (with and without transparent integration)

The third class of systems is specialized on displaying interactive 3D applications on remote clients. The server (or servers) with a lot of graphics processing power renders a 3D application (e.g. OpenGL-based). Every frame is read back, compressed (lossy or lossless) and send to a client. Mostly, only the window of the 3D application is processed, to save bandwidth and processing power. Popular systems of this class are the SGI Viz-Server [6] and VirtualGL [7]. All of these systems are optimized to provide high frame rates with the available bandwidth. This is supported by the possibility to choose different resolutions and compression methods as well as certain level of detail mechanisms. Most systems allow a transparent usage of existing applications (mostly OpenGL-based). This guarantees a comfortable and universal utilization. On the opposite, there are systems that are tightly coupled to a certain application to further increase performance by adjusting the rendering process dynamically to fit the needs of remote visualization. The Invire prototype presented in this report belongs to this class and will support both transparent and non-transparent integration. Furthermore it utilizes advanced hard- and software mechanisms to achieve high performance remote visualization.

Advances in Computer Graphics Hard- and Software

As mentioned before graphics hardware significantly changed in the last decade. The GPUs evolved from highly specialized and slowly clocked graphic processors to highly parallel, multi-purpose, fast co-processors which in some tasks outperform the CPU by far. This fact inspired a lot of programmers to try to port their application on the graphics hardware to gain significant performance increases. However the main problem was that the general purpose applications needed to be transferred into a graphics domain since the graphics cards could only be programmed through

graphics APIs such as OpenGL or DirectX. There where efforts to develop more general APIs, mainly driven by the GPGPU[1] consortium. However one of the most important step towards general usage of graphics hardware was the introduction of the Compute Unified Device Architecture (CUDA) by NVIDIA in 2006 (see [2]).

CUDA

CUDA is a combination of software and hardware architecture (available for NVIDIA G80 GPUs and above) which enables data-parallel general purpose computing on the graphics hardware. It therefore offers a C-like programming API with some language extensions. The architecture offers support for massively multi threaded applications and provides support for inter-thread communication and memory access. The API distinguishes between *host* and *device* domains and offers access to fast caches on the device side. The implemented method of thread partitioning allows the execution of multiple CUDA applications (*kernels*) on one GPU. Each kernel on the host device has access to a *grid* of thread *blocks*. A block consists of a batch of threads that can be synchronized and is organized by one-, two- or three-dimensional IDs. This allows uniquely identifying each thread and assigning tasks to each thread. All threads inside one block have access to a fast shared memory space to exchange data.

Another feature of the CUDA architecture is the interoperability with graphic APIs (OpenGL and Direct3D) which allows using, for example, rendered images as input to CUDA kernels. Since this data already resides on the graphics device it only needs to be copied on the device to be processed by CUDA. This offers great possibilities for e.g. online image compression, which is one topic of this report (see section *Compression*).

Render Buffer Objects

Another important technique in this field is the introduction of Frame Buffer and Render Buffer Objects in the OpenGL 2.0 specification [1]. They replace the slow pbuffer mechanisms for off-screen rendering. This technique offers, among others, a fast way to render to specified memory regions other than the framebuffer. This can be used e.g. to implement rendering servers for a remote rendering framework as proposed in this paper.

[1] www.gpgpu.org

Invire - A Concept for an Interactive Remote Visualization System

In this section we introduce the concept for a new system called Invire (INteractive REmote VIsualization). It belongs to the third class of remote rendering systems and focuses on high interactivity, maximized performance, and best visualization results. We therefore utilized the new advances in graphics technology described above to improve compression and readback performance. The goal was to achieve adequate frame rates (above 20 FPS) and unrestricted interactivity for the remote visualization of high-resolution (1000x1000 pixels and above) 3D applications. We choose to use lossless compression techniques first to maximize graphics quality and because of their high potential for parallelization. We also work on integrating lossy compression methods which allow guaranteeing fixed bandwidth utilization.

The main limitation, however, still is the network connection between server and client. Uncompressed image data is to large to be transferred over current internet or even LAN connections. An example shows that to achieve 25 FPS for a remote visualization with a resolution of 1000x1000 pixels we would need a network connection that has a bandwidth of 100 MB/s:

*Bandwidth =size of one frame * frames per second*
*= (1000*1000*4 byte/frame) * 25 frame/second*
= 100 MByte/second.

Even Gigabit Ethernet is limited to about 80 MB/s in practical usage. That leads to the conclusion that a compression rate of more than 0.1 is needed to realize our goal on a system with a Fast Ethernet connection (100 MBit/s).

The Architecture

The overall architecture of Invire (as shown in Figure 1) is kept simple to maximize the pure remote visualization performance. The server side offers an *Invire Plugin* which can easily be integrated in existing OpenGL applications. This allows the passing of parameters from the host application to Invire. We also work on a transparent integration possibly on the basis of VirtualGL. The *Invire Plugin* reads the current OpenGL context into a Renderbuffer Object which is passed to the *Compression* facility. This part of the software is implemented as modular collection of compression algorithms that can be exchanged arbitrarily. This allows us to compare the different methods and eventually combine them to achieve higher compression rates. After the compression of a rendered frame, it is passed to the *TCPServer* where a header is generated. The header contains information about the image size, compression. Afterwards header and compressed data is send to the *Invire Client*. It receives the compressed frame and passes it to the *Decompression* facility together with the information from the header. After decompression the frame is displayed by the client.

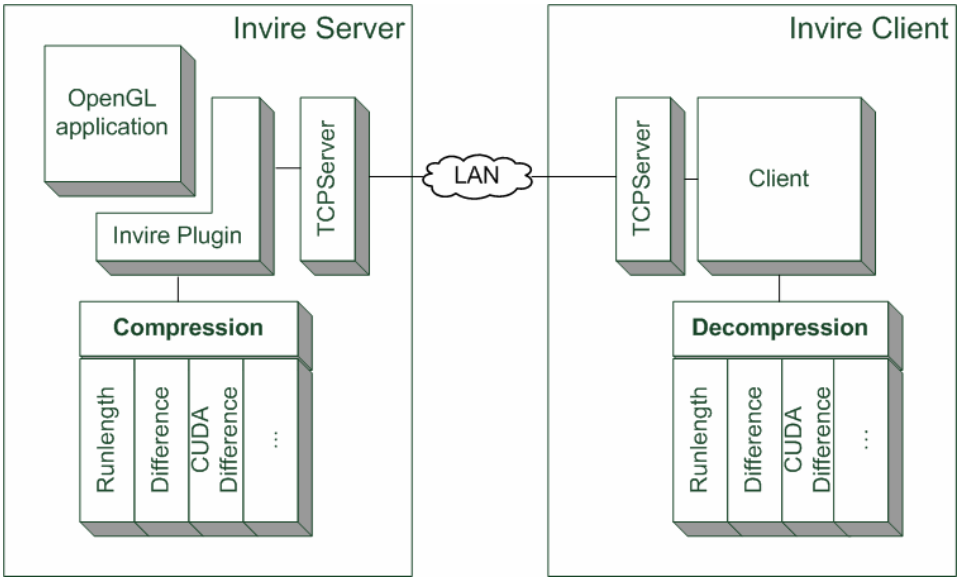


Figure 1: Invire components and architecture

Image Transfer

To transfer the images a TCP socket connection is established between server and client. This socket remains active until either client or server cancels the transfer. The advantage of TCP sockets is that the correct order and the integrity of the image data is guaranteed. After a socket connection is established the protocol overhead is minimal and allows a good utilization of the available bandwidth.

Image Readback and Compression

The most important parts of Invire are the readback and compression methods. We implemented a simple run-length (RLE) and difference encoding using standard programming techniques to have a basis for comparison (for more information on data and image compression see [SA2004]. We shortly present the two basic compression methods and afterwards describe our CUDA-based difference compression algorithm in detail. The use of relatively simple and lossless compression methods has several reasons. First of all, they are very fast since both RLE and difference compression only need to go over the input once. Additionally the difference compression has a very good potential for parallelization, as shown in the following. However, the main problem of lossless compression is that the achievable compression rate strongly depends on the input frame. If it is highly arbitrary in terms of consecutive colors or frame-to-frame difference then compression ratios can even be above 1.0. Nevertheless, typical visualization data (CAD, scientific visualization, VR) is often very regular in one or both terms. Therefore one can expect good compression rates for this application. There are also more sophisticated compression methods (e.g. statistical such as Huffman Coding or dictionary based

such as LZ) but they cannot meet the requirements of interactive applications because of their comparably long run times or high memory consumption.

Run-length encoding

After reading back the image data of the current context to a given location in (host) memory, the run-length algorithm goes over the array, counts consecutive pixels with same color values and writes the sum followed by the actual color information into a new array. The decompression can be done by writing the amount of pixels with the same color in a new array consecutively. Afterwards it can be displayed by the OpenGL function *glDrawPixels()*. The RLE algorithm can encode and decode *n* pixels in *O(n)* time.

Difference compression with index

Another basic technique for lossless image compression is the difference compression. The current and the last frames are compared pixel wise and only the pixels that are different are saved. Additionally the position of the pixels that changed is needed to decompress the current frame. This can be most efficiently done by an index which maps one bit to every single pixel of a frame. If the bit is 1 the pixel has been changed and the saved pixel value at the position *#of preceding 1s in index* is needed to update the pixel of the last frame. This requires *O(n)* time to compress and decompress *n* pixels.

Difference compression with index using CUDA

The difference compression with index method described before is very suitable for parallel execution (on *k* threads with *n>k* pixels). Especially creating the index, as well as the copying of the pixel data can be done in *O(n/k)* time. Therefore we decided to implement this method in parallel and choose to use the CUDA architecture. It is able to process the data directly on the graphics hardware and allows highly parallel execution. Since CUDA offers several SIMD multi processors we decided to split the frames into *m=n/k* blocks which can be processed independently. This helps to optimize the workload on the graphics hardware. Each block consists of *k* threads which are working in parallel on one multiprocessor and which have access to a fast, shared memory. Each thread processes one pixel which is assigned to it by its thread and block index (*thid* and *bid*). In a first step the index is generated by simply comparing corresponding pixels of the last and the current frame. When all threads have finished, a parallel compaction method (the second step) based on the stream compaction algorithm by [9] is invoked on the shared memory. This algorithm requires *O(log k)* time in parallel to compute the amount of empty spaces to its left for every item of the *s_result* array. It needs to run for all *m* blocks. With this information the pixels can now be stored in an array without empty spaces which is copied to a position in the global result array which is determined by the block index (*bid*). Additionally the number of stored pixels is written to a global

array (*g_changedPixels*). After all blocks finished the second step, the *g_changedPixels* array is used to compute the absolute position in memory for each blocks partial result (step 3). This is also done in parallel based on the scan algorithm presented by [10] which sums up all items to the left of the current value in *O(m/k * log m/k)* time in parallel. Finally, the partial results of the blocks are copied to the calculated memory locations and then the final compressed frame and the index are readback from the graphics hardware to be send to the client. The decompression at the client side is done sequentially as described above or can be parallelized equal to the compression using CUDA if available. Finally the whole algorithm can be run in *O(2*(m)+ m * log k + m/k * log m/k)* time in parallel with *n= number of pixels*, *k= number of threads*, *m= number of blocks* and *n=m*k*.

Benchmarks and Results

To proof the theoretical concept and to compare the described compression algorithms we prototypically implemented the Invire system and tested it in a sandbox environment. The server part runs on a computer equipped with a CUDA ready Geforce 8800 GTS graphics card by NVIDIA. It has 12 multiprocessors, a wrap size[2] of 32. That leads to *k=12*32=384* threads that can run concurrently. As testing scenario we chose two OpenGL based applications (see Figure 2) The first is a simple rotating teapot on black ground and the second is a driving simulator called Virtual Night Drive (VND) [11]. The VND is specialized on simulating automotive headlights at night and uses the shaders of the graphics card to calculate the luminance intensity pixel wise. The first teapot application is highly regular and has a uniform black background, whereas the VND is relatively arbitrary through its textured and lighted scene and its unpredictable movement.

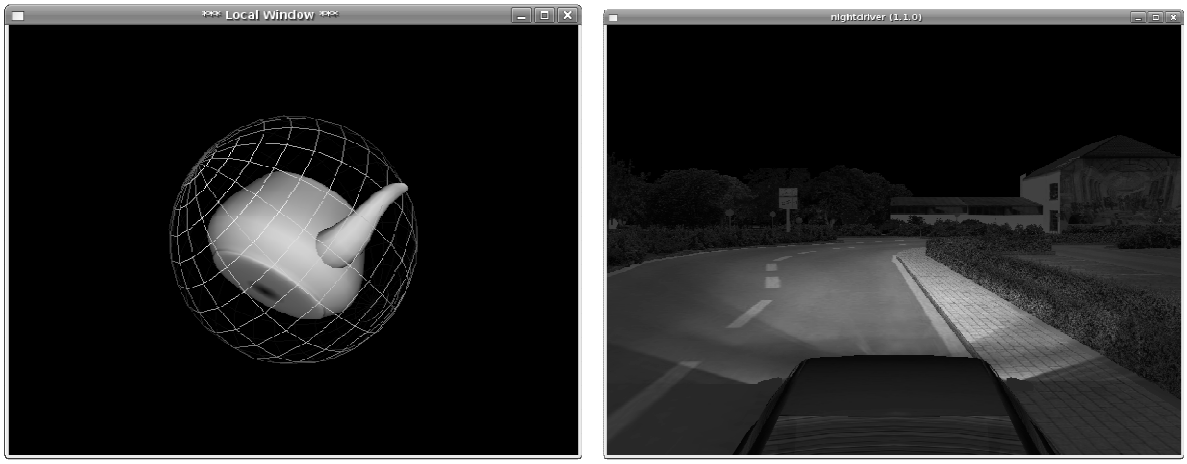


Figure 2: Test cases: simple teapot, Virtual Night Drive with headlight simulation

[2] Number of threads that are executed in parallel on one multiprocessor

The charts in Figure 3 show some interesting results. For both test application we measured the average frame rate for various resolutions and compression methods. Figure 3 A) shows the teapot application with resolutions from 16*16 pixels up to 1024*1024 pixels. For resolutions below 256*256 the overhead for compressing and decompressing the frames is obviously to big so that no compression method performs better than just sending uncompressed images. But from 256*256 on the size of the transferred images is the limiting factor. The RLE and the difference without CUDA method perform nearly similar and are a bit faster than the version with CUDA up to the resolution of 512*512. There the difference coding with CUDA takes the lead and outperforms the other compression methods by a factor of about two (RLE 14 FPS, difference 10 FPS and difference with CUDA 21 FPS). This can be explained by the computational overhead the CUDA-based algorithm introduces (mainly the compaction methods). But for high resolutions this overhead amortizes and we can nearly double the frame rate. The VND application shows similar results. The interesting thing here is, that all algorithms produce slower frame rates because of the poor compressibility of the input data. But still the CUDA-based approach provides the best results for the highest resolution. In comparison with another remote graphics system of the same class -VirtualGL- Invire performed slightly worse (about *20%* fewer FPS in the VND application) but that can be explained of the use of lossy compression versus lossless compression techniques.

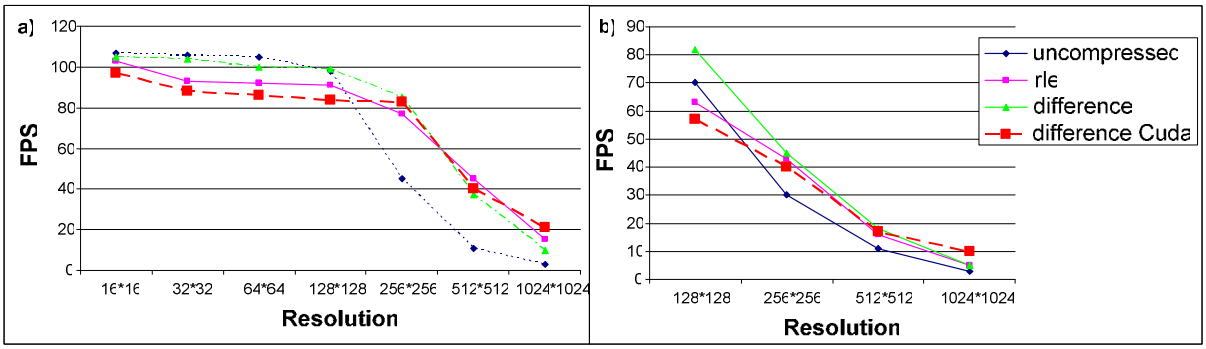


Figure 3: Benchmarking results with A) Teapot and B) Virtual Night Drive

Conclusion and Outlook

In this paper we presented a remote rendering system that takes advantage of recent advances in computer graphics hard and software. We could show that the speedup of a parallel compression method surpasses the resulting computational overhead for high resolutions (1024*1024 pixels and above). However, the implementation that was used for the performance benchmarks still is in a prototypical stage. There are some programming optimizations to do and especially the integration of high quality

lossy compression techniques promises good results. To learn more about the technical and implementation details please read [12].

Resource Usage

This project partially uses resources of the Arminius cluster system. However, most development and testing work is done on special machines equipped with Cuda enabled graphic cards.

References

[1] Segal, M., Akeley, K.: “The OpenGL Graphics System: A Specification Version 2.0”. Silicon Graphics (2004)

[2] NVIDIA: “NVIDIA CUDA - Compute Unified Device Architecture”. Website: http://developer.nvidia.com/object/cuda.html (2007)

[3] Humphreys, G., Houston, M., Ng, R., Frank, R., Ahern, S., Kirchner, P., Klosowski, J.: “Chromium: A stream-processing framework for interactive rendering on clusters”. ACM Transactions on Graphics 21 (2002) 693{702

[4] Richardson, T., Stafford-Fraser, Q., Wood, K., Hopper, A.: “Virtual network computing”. IEEE Internet Computing 2 (1998) 33-38

[5] Microsoft: “Microsoft Windows Server 2003 Terminal Services”. Website: http://www.microsoft.com/windowsserver2003/technologies/terminalservices/default.mspx (2007)

[6] SGI: “SGI OpenGL Vizserver - Visual Area Networking”. Website: http://www.sgi.com/products/software/vizserver/ (2007)

[7] VirtualGL: “The VirtualGL Project”. Website: http://www.virtualgl.org/ (2007)

[8] Salomon, D.: “Data Compression: The Complete Reference”, 3rd Edition. Springer (2004)

[9] Horn, D.: “Stream Reduction Operations for GPGPU Applications”. In: GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation. Addison-Wesley Professional (2005) 573-583

[10] Harris, M.: “Parallel Prefix Sum (Scan) with CUDA”. Website: http://developer.download.nvidia.com/compute/cuda/sdk/website/projects/scan/doc/scan.pdf (2007)

[11] Lietsch, S., Berssenbruegge, J., Zabel, H., Wittenberg, V., Eikermann, M.: “Light simulation in a distributed driving simulator”. In: Proceedings of the 2nd International Symposium on Visual Computing, 2006. Volume 4291 of Lecture Notes in Computer Science., Berlin Heidelberg New York, Springer-Verlag (2006) 343-353

[12] Lietsch, S., Marquardt, O., “A CUDA-Supported Approach to Remote Rendering”. To be published in Proceedings of the 3rd Inernational Symposium on Visual Computing (ISVC2007)

5.3.3 Parallel Image Reconstruction for Positron Emission Tomography on Multi-Core

Project coordinator	Prof. Dr Marco Platzner, PC², University of Paderborn
Project members	Dr. Jens Simon, PC², University of Paderborn Dipl.-Inform. Tobias Beisel, PC², University of Paderborn Dipl.-Inform. Stefan Lietsch, PC², University of Paderborn
Supported by:	Heart and Diabetes Center North Rhine-Westphalia

General Problem Description:

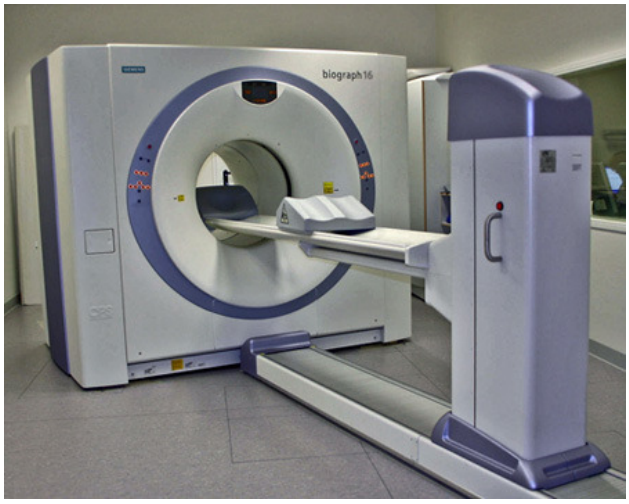


Figure 1 Siemens PET-Scanner (Source: Universitäts-Klinikum Regensburg)

The Positron Emission Tomography (PET) is an important part of internal medicine, allowing non-invasive examinations of metabolism, particularly diagnosing cancer and tumor. Figure 1 shows an image and Figure 2 the schematic setup of a PET scanner. As the interior of the human body can not be simply “photographed”, other methods of image generation are needed. In this case images are generated through reconstructing positions of radioactive fluids in the human body. As shown in the schematics, the collapse of a radioactive atom results in the emission of two electrons going in exactly opposite directions. These events are recorded by the scanner and stored as a so called “Lines Of Response” (LOR). These LORs are the basis for the image reconstruction using complex mathematical algorithms. .

Additionally, to generate three dimensional models of human organs many image slices are reconstructed and afterwards composed to a 3D object. Computing these algorithms for a big amount of images demands a considerable large amount of time to be completed, especially if quality requirements are set to a high level. In this project the desired acceleration of the reconstruction time is accomplished by parallelizing an existing implementation of reconstruction algorithms. The developed realization was done to be performed on easy to use multi-core systems by using a message passing approach. The results deliver a speedup of at least 2 for reconstructing the sample datasets.

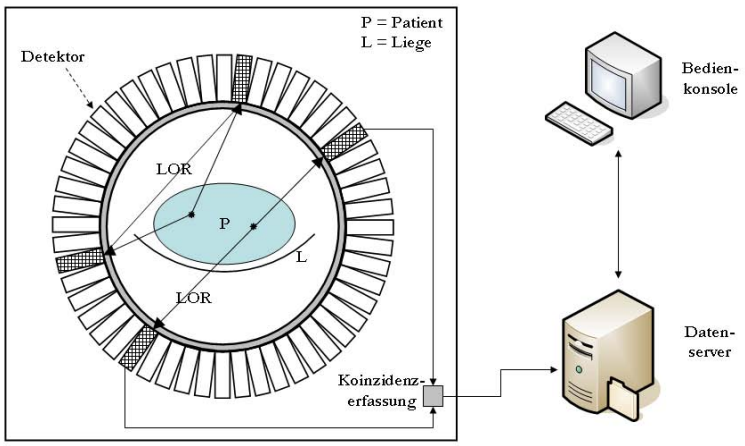


Figure 2 Schematic setup of a PET scanner (P = Patient, L = Stretcher, Koinzidenzerfassung = Coincidence Detection, Datenserver = Storage Server, Bedienkonsole = User Terminal)

Project Sumary

The primary objective of this project was to achieve a significant speedup of selected image reconstruction algorithms provided by the STIR framework (Software for Tomographic Image Reconstruction) [2]. The selection of the algorithms was influenced by the needs of the project partner. The hospital HDZ utilizes a PET scanner mainly for diagnosing deceases of the heart. Therefore, two algorithms – an analytic called “Filtered Back Projection” (FBP2D) and an algebraic one called “Ordered Subset Maximum A Posteriori One-Step Late” (OSMAPOSL) – are used to reconstruct the images of a Siemens PET scanner. To accelerate the well-approved algorithms a parallelization approach was taken. With utilization in clinical everyday life in mind, the decision was to design a system for affordable and easy to maintain multi-core machines instead of one for big cluster systems. The basic parallelization approach, however, is sufficient for both architectures.

To have a starting point both algorithms were analyzed in depth and several criteria for the parallelization where extracted. There are two different modes fore storing the raw data – list mode and viewgram- mode – which require different pre-processing

steps before the algorithms can work on them. After another pre-processing step which computes a sensitivity image once for every reconstruction, the actual reconstruction algorithm starts. To design a parallel reconstruction software a decision between thread based and message passing based parallelization had to be taken. Additionally it had to be taken into account that the parallel reconstruction algorithms should be integrated into the STIR framework again to utilize its extended functionalities and provide other users with the enhanced code. Finally, the fact that the sequential algorithms gain their performance to a certain amount through caching effects must be considered when distributing data.
These considerations lead to a conceptual design with the following design parameters:

- Only viewgram mode is natively supported (list mode data can be converted to viewgrams)
- The parallelization of the pre-processing steps (especially the computation of the sensitivity image) will not be implemented since the expected performance increase is not worth the overhead it would generate
- The FBP2D algorithm uses an OpenMP / thread based parallelization approach since its short run times would not justify the initial overhead of e.g. a MPI based approach
- The OSMAPOSL algorithm uses an MPI based approach to achieve maximum scalability and performance on different systems (multi core machines as well as clusters)
- The distribution of data and work packets is realized as a work pool system, where every worker client receives tasks from a work pool by a supervising master. This allows good load balancing and the utilization of cache hit effects on every client.

The conceptual design was implemented and tested on different machines. The main focus laid on an Intel Clovertown system (8 CPU cores, 8 GB RAM) and an AMD Opteron Sytsem (4 CPU cores, 4GB RAM). Additionally, the MPI based OSMAPOSL algorithm was tested on 20 nodes of the Arminius cluster system (Dual Xeon CPU, 4GB RAM). Several benchmarks where performed and the best overall speedup (from sequential to parallel version on the same machine) could be achieved on the 8 core Clovertown system. When utilizing 6 cores for the workers and one core for the dispatcher an average performance speedup of 2,6 is possible. On the Opteron system the average speedup was even higher (about 3,2), but the overall runtime of the sequential execution was 2-3 times longer than on the Clovertown (due to architectural differences and fewer RAM). On the Arminius cluster an average speedup of 8 could be achieved when using 20 worker nodes. But again the overall runtime of the sequential version was twice as long as on the Clovertown. The

optimization of the work distribution for caching did not show significant effects on the overall run time, since cache misses are quite rare for the small testing data set. But for larger data sets much higher cache miss rates are expected which can be reduced through the implemented optimization.
The parallelization of the FPB2D algorithms could achieve a speedup of 2.4 but is only of limited importance since the sequential runtime is already very short (about 5 seconds for one image).
In conclusion the developed parallelization approach works quite well for the given scenario and a relatively good speedup could be achieved for several systems. The Clovertown system gained the best results in terms of overall run time and cost/performance ratio. Even better results were prohibited by the limitations of the STIR framework and the need of small working packets of the iterative reconstruction algorithm. But for an everyday clinical usage, an overall speedup of about 20 from the system that is currently used to the parallel Clovertown system is significant. For more details on the actual implementation and theoretical basis of this project please refer to [1].

Resource Usage

This project uses resources of the Arminius cluster system and the Clovertown test systems for thread/OpenMP and MPI based parallelization of medical image reconstruction.

References

[1] Beisel, T.: “Entwurf und Evaluation eines parallelen Verfahrens zur Bildrekonstruktion in der Positronen-Emissions-Tomographie auf Multi-Core-Architekturen”. Master Thesis UPB (2007)

[2] Thielemans, K.:”STIR - General Overview, Version 1.3”. Website: http://stir.irsl.org/ (2007)

5.4 Models and Simulation

5.4.1 Numerical computation of three dimensional flow problems

Project coordinator	Prof. Dr. Burkhard Monien, PC², University of Paderborn
	Dr. Stephan Blazy, PC², University of Paderborn
Project members	Dr. Oliver Marquardt, PC², University of Paderborn
Work supported by	German Science Foundation (DFG) – SFB 376

General Problem Description

Research in the area of fluid flow is essential for many different practical natural science procedures and industrial applications. These are typically design problems which can occur in the area of automotive and aircraft engineering. Also the study of basic chemical and physical phenomena is of high interest to achieve a better understanding of fundamental processes. Many of these problems can often be mathematically modeled by the well known Navier--Stokes equations. The numerical solution of the three-dimensional Navier--Stokes equations on unstructured meshes is one of the most important problems in mathematical computation and simulation. To obtain meaningful solutions the approximation usually involves a large number of unknowns. Thus, it can only be solved on massively parallel computer systems in a reasonable amount of time. Parallel computer resources, like cluster systems of vector nodes or PC-based nodes, make it possible to investigate such computation-intensive problems.

Our massively parallel numerical simulation environment *padfem²* [1] utilized the method of characteristics combined with a pressure correction technique for solving the three dimensional incompressible Navier--Stokes equations. This algorithm decomposes the Navier--Stokes equations into three simpler sub problems: the transport, the diffusion and the pressure correction step, where the transport step is based on the method of characteristics. This approach reformulates the Navier—Stokes equations with the help of the Lagrangian derivative. Here, an ordinary differential system (ode) of equations needs to be solved in place of the nonlinear (convective) term. The solutions of the ode--system are the characteristic curves of the given velocity field. The diffusion (resolvent) step is given by a well conditioned elliptic boundary value problem for the velocity field. As a consequence of the splitting approach, the resulting velocity field violates the incompressibility condition. Therefore, the solution of a Poisson problem for the pressure correction is needed to

project the actual velocity field onto the space of divergence free functions. Both, elliptic partial differential equations from step two and three, can be solved with standard preconditioned conjugate gradient solvers or multigrid methods.

For the transport step, it is necessary to locate a particle position and the enclosing element (tetrahedron) from the unstructured mesh. This particle characterises the footpoint of the corresponding characteristic curve. For solving this problem, an efficient point location algorithm on unstructured meshes is needed particularly with regard to parallel computing. The spatial approximation is based on the finite element method using a mini element [1] approach, which is a stable finite element in the sense of the LBB-(Ladyshenzkaja-Babuska-Brezzi) condition. The flow problems are defined within the German Science Foundation high priority programs "Flow Simulation on High Performance Computers" and "Analysis, Modelling and Calculation of Mixing Processes with and without Chemical Reaction".

Problem details and work done

All considered flow problems are governed by the Navier--Stokes equations, where $v = v(t,x)$ describes the velocity, $p = p(t,x)$ the pressure, t the time and x the spatial variables. The cinematic viscosity ν is expressed by the dynamic viscosity μ mu and the density rho by $\nu = \mu / \rho$. Using the incompressibility condition and the assumption, that the density is constant, the Navier--Stokes equations reads:

$$\frac{\partial}{\partial t} v + (v \cdot \nabla) v = \nu \Delta v - \frac{1}{\rho} p, \quad \text{div } v = 0 \text{ in } \Omega \tag{0.1}$$

with given initial- and boundary conditions:

$$v = v_0(x) \text{ on } \Gamma_{\text{in}}, \ (-pI + \nu \nabla v) n = 0 \text{ on } \Gamma_{\text{out}}, \ v = 0 \text{ on } \Gamma / (\Gamma_{\text{out}} \cup \Gamma_{\text{out}}).$$

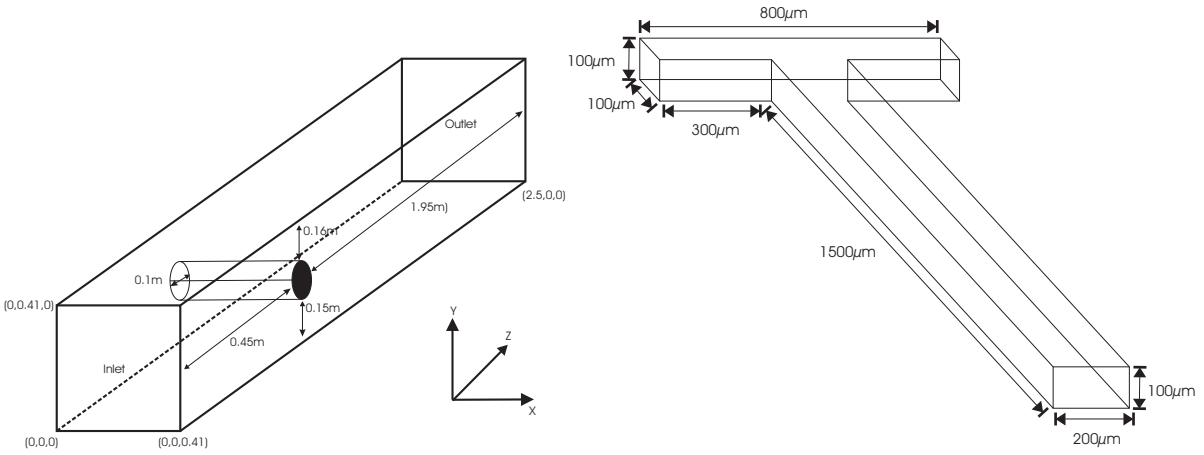


Figure 1: Computational domains

The first problem is a flow around a cylinder with circular cross-section defined within the German Science Foundation high priority program “Flow Simulation on High Performance Computers” [5]. The computational domain of this benchmark problem illustrated in Figure 1 is denoted by Ω with boundary $\Gamma = \partial\Omega$. The dimensions of the cylinder, the surrounding channel and the flow parameters are described in the following table:

Diameter D	Height H	Broadness B	Length L	Density ρ	Dyn.Viscosity μ
0.1m	0.41m	0.41m	2.5m	1kg/m³	0.001m/s²

The inflow velocity field u_0 ensues a Reynolds number of Re=20, where $u^0 = u_0^{\mathrm{Cyl}}(y,z) = 16Uyz(H-y)(H-z)/H^4$, where $U = 0.45\mathrm{m/s}$. This leads to a mean velocity $\overline{u} = 0.2$ m/s.

The second problem is a flow through a t-shaped micro-mixing channel defined within the German Science Foundation high priority programs "Analysis, Modelling and Calculation of Mixing Processes with and without Chemical Reaction" [1]. Here, the inflow velocity fields are given and ensue a Reynolds number of Re=186. This flow problem comprises an additional transport problem for the species. For this, one has to compute in advance a stationary solution of the Navier--Stokes system. Afterwards, with the known velocity field the instationary transport of the species is computed. The dimension parameter of the computational domain, see Figure 1, are shown in the next table:

Inflow	Outflow	L-top	L-outflow	Density ρ	Dyn.Viscosity μ
100µm x 100µm	100µm x 100µm	800µm	1500µm	998.2 kg/m³	0.001003m/s²

Note that the entire length of the channel is given by 1600µm. The inflow profile for both inlets are prescribed with U=1.4m/s by

$$u^0 = u_0^{\mathrm{Mix}}(x,z) = \pm \frac{U}{0.2138} \sum_{k,l=0}^{4} \frac{\sin((2k+1)\pi x/d)\sin((2k+1)\pi z/d)}{(2k+1)(2l+1)((2k+1)^2+(2l+1)^2)}.$$

The additional transport equation required for the chemical species c is described by the equation

$$c_t(t,x) + \big(\, \mathrm{v}(t,x)\cdot\nabla\big) c(t,x) = D\Delta c(t,\mathrm{x}) \text{ in } \Omega,$$

with boundary conditions $c_0(x) = 1 \text{ on } \Gamma_{\mathrm{in}}^{\mathrm{l}}, c_0(x) = 0 \text{ on } \Gamma / (\Gamma_{\mathrm{in}}^{\mathrm{l}} \cup \Gamma_{\mathrm{out}})$. For this test problem the pressure decay along the mixing channel, the mixing quality and the variation of the species are of interest. The mean values for the species distribution are computed on different cut planes of the t-shaped mixing channel. The numerical approximation scheme and the adaption details are described in [1].

Numerical Results

The numerical simulation has been carried out with our parallel adaptive finite element framework padfem². First the results of the cylinder benchmark problem are discussed. In this case the drag- and lift--coefficient, as well as the difference of the pressure on the cylinder--boundary are of interest. Reference values were already presented in several papers to this problem [4,5]. The computation of the drag coefficient c_d and the lift coefficient c_l depends on the drag and lift force given by the following boundary integrals,

$$F_d = \int_S \rho\nu \frac{\partial u_{\tau_1}}{\partial_n} n_y - p n_x \, dA, \quad F_l = \int_S \rho\nu \frac{\partial u_{\tau_1}}{\partial_n} n_x + p n_y \, dA.$$

The coefficients are then defined by

$$c_d = \frac{2F_d}{\rho \bar{U}^2 DH} \text{ and } c_l = \frac{2F_l}{\rho \bar{U}^2 DH}.$$

For the computation of the boundary integrals, a volume based approach developed by John [JOHN02] is used. Actually, the Reynolds number of Re=20 characterises a stationary flow around the cylinder. Therefore a stationary solver is used in [JOHN02, BRRI05]. Since padfem² currently does not support a solution technique of this type, a stationary condition is achieved with appropriate long simulations of the instationary process. The transient behavior of the mean values can be observed in Figure 4, 5 and 6. One can recognize that the stationary state is accomplished after 6 seconds of the simulation time. The real computation time for the simulation steps is presented in Figure 12. The pressure difference $\mathrm{P}_D = \mathrm{P}_F - \mathrm{P}_B$ is computed at the points $P_F = (0.45, 0.2, 0.205)$ and $P_B = (0.55, 0.2, 0.205)$. The reference intervals for the mean values are given by $P_D \in [0.165, 0.175]$, $c_d \in [6.05, 6.25]$, $c_l \in [0.008, 0.01]$ and the reference values from [4] by $c_d = 6.185327, c_l = 0.0094009, p_d = 0.1708754$.

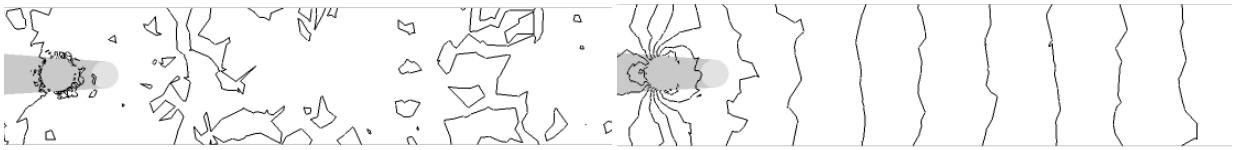


Figure 2: Iso-lines of the P1/P1 and P1-Bubble/P1 approach

For the numerical approximation the stable mini--element P1Bubble/P1 and a non-stable linear-linear P1/P1 finite element approach are used. The P1/P1 element is only used, in order to determine whether disturbances arise. Figure 1 show the iso-lines of the pressure. One can recognize that the P1/P1 element clearly points out several disturbances in the pressure in contrast to the mini element. The table below presents the relative error to the reference values and the mesh dimensions. The used meshes consist of a uniformly discretised (M1), static local refined (M4, M5) and adaptively refined meshes (M2, M3) using the error estimator from [1].

Type/Mesh	M1	M2	M3	M4	M5
Nodes	119265	39039	108368	10849	10849
Edges	774367	253491	723028	66143	66143
Faces	1278615	418182	1213510	104629	104629
Volumes	623513	203730	598850	49335	49335
Err-Cd	6.03E-03	1.61E-02	1.11E-02	2.71E-01	3.12E-01
Err-Cl	4.23E-00	7.12E-02	7.05E-02	6.00E-01	5.10E-00
Err-PD	1.11E-01	1.32E-01	8.11E-02	2.44E-02	3.91E-02

Table 1: Mesh dimensions and relative error to the reference values

The best result is obtained with the finest adaptive mesh M3. The adaptive computation results of the drag coefficient and the pressure difference are nearly the same as for the uniformly refined mesh. But one receives clearly better results for the lift coefficient using less than one third of the unknowns in case of mesh M2.

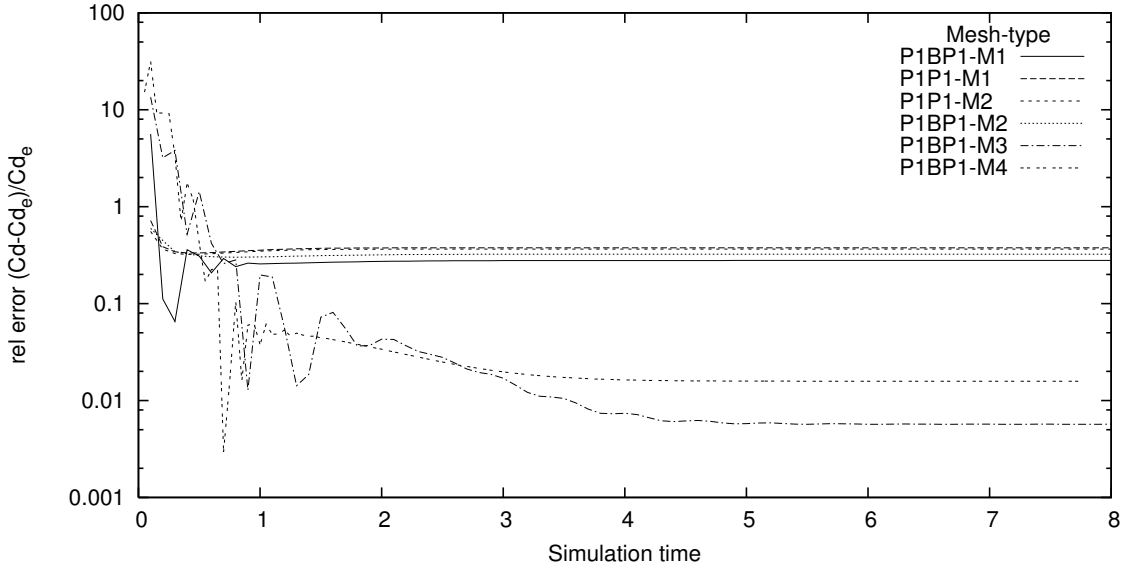


Figure 3: Relative error of the drag coefficient.

As expected, the error is reduced with an adaptive refinement of the mesh. If this behavior remains using the projection scheme further calculations must be investigated. Mesh M4 and M5 served essentially only the comparison of the pressure iso-lines using the stable and non stable finite element approximation.

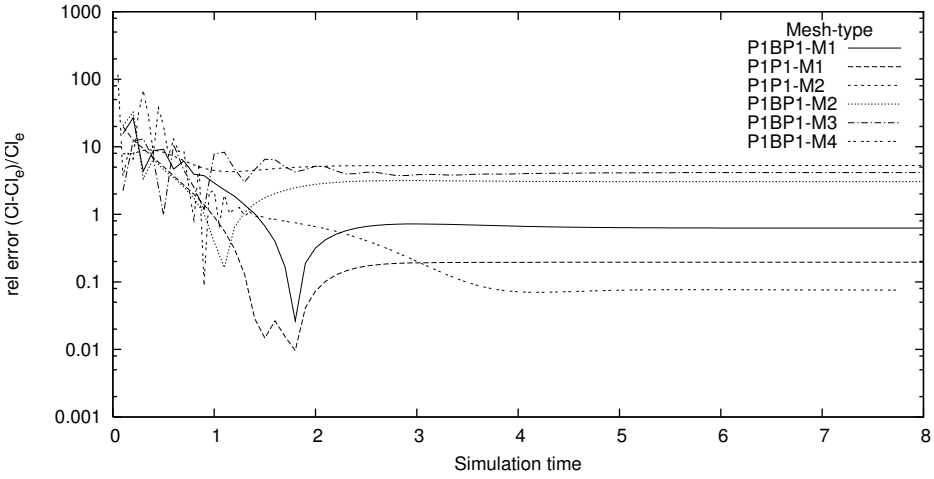


Figure 4: Relative error of the lift coefficient.

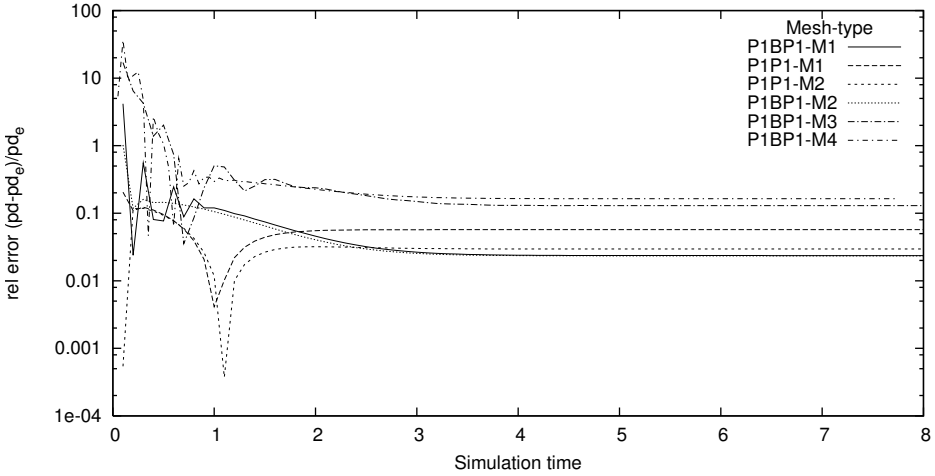


Figure 5: Relative error of the pressure difference.

However, acceptable values were also obtained in this case. Only the lift coefficient is badly approximated with mesh M1 and M5. Overall, the error exhibits the accuracy up to an order of 10^3. This result is comparable to the P^1/P^0 or Q^1/Q^0 computations carried out in [4]. It seems that both, an accordingly exact approximation of the cylinder, as well as a gradual refinement along the outflow channel, have a large influence on the computation of the mean values. A higher accuracy can be obtained with the help of another finite element approach, for example the Taylor-Hood element.

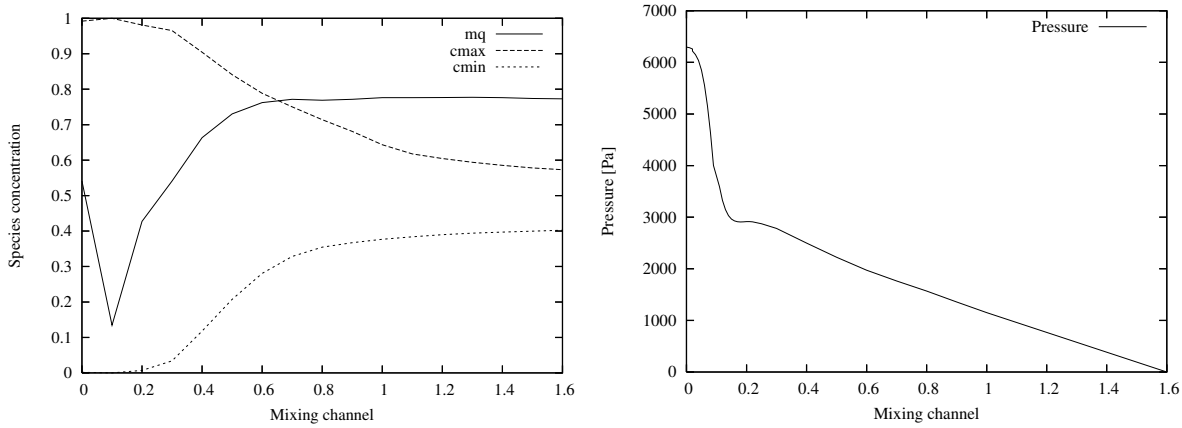


Figure 6: Pressure and Species distribution along the outflow channel.

Next the results of the micro-mixing problem are depicted, see also [2]. Figure 5 (left picture) presents different mean values. The dotted curves show the minimum and

maximum distribution of the species along the micro-mixing channel and the solid one the mixing quality. The mixing quality is computed by the formula

$$mc = 1 - (\langle (c - \langle c \rangle)^2 \rangle) / (\langle c \rangle \langle c_{max} - c \rangle).$$

The computed quality reached nearly 80%. In Figure 5 (right picture) one can see the pressure mean values along the micro-mixing channel. An impression of a stationary and instationary case is shown in Figure 6.

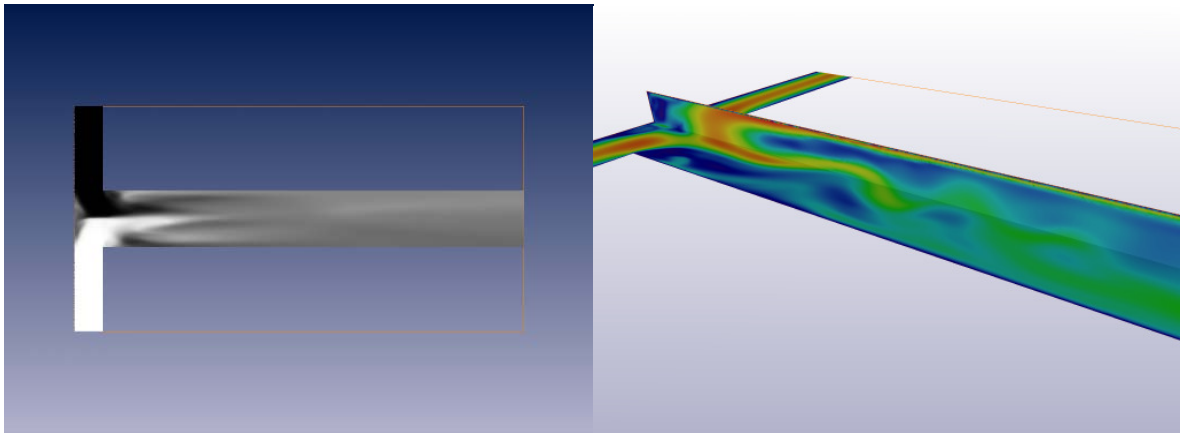


Figure 7: Species distribution for a stationary and instationary state.

Resource Usage

All computations were done on a 200 compute-node cluster-system with a high speed Infiniband-Network installed at the Paderborn Center for Parallel Computing. Each node consists of a INTEL Dual Xeon 3.2 GHz EM64T system board with 4 GB main memory. The peak performance of the cluster is 2.6 TFlops and allows a memory usage up to 896 GB. In the following some characteristic simulation data using 60 processors of this cluster--system are presented. In the parallel case, an overlapping technique for the partitioning (halo) is used by padfem², i.e. each partition knows the first neighbor elements from its neighboring partition. The data overhead for the halo construction depends on the original sequential mesh, but decreases if the mesh-size increases.

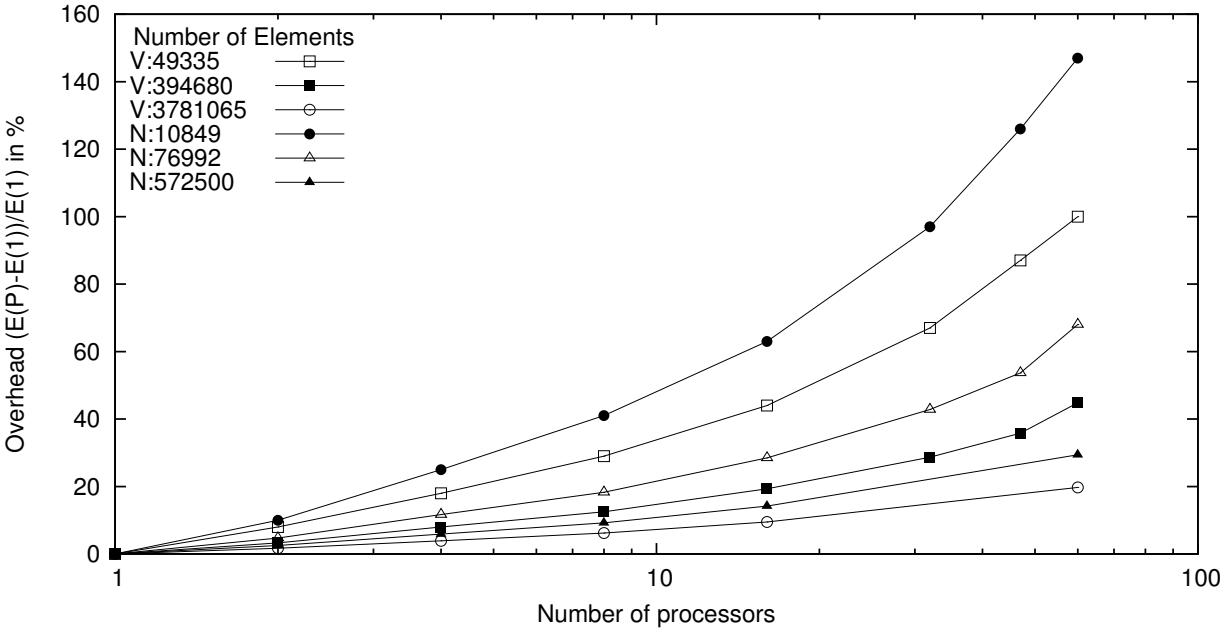


Figure 8: Halo-Overhead.

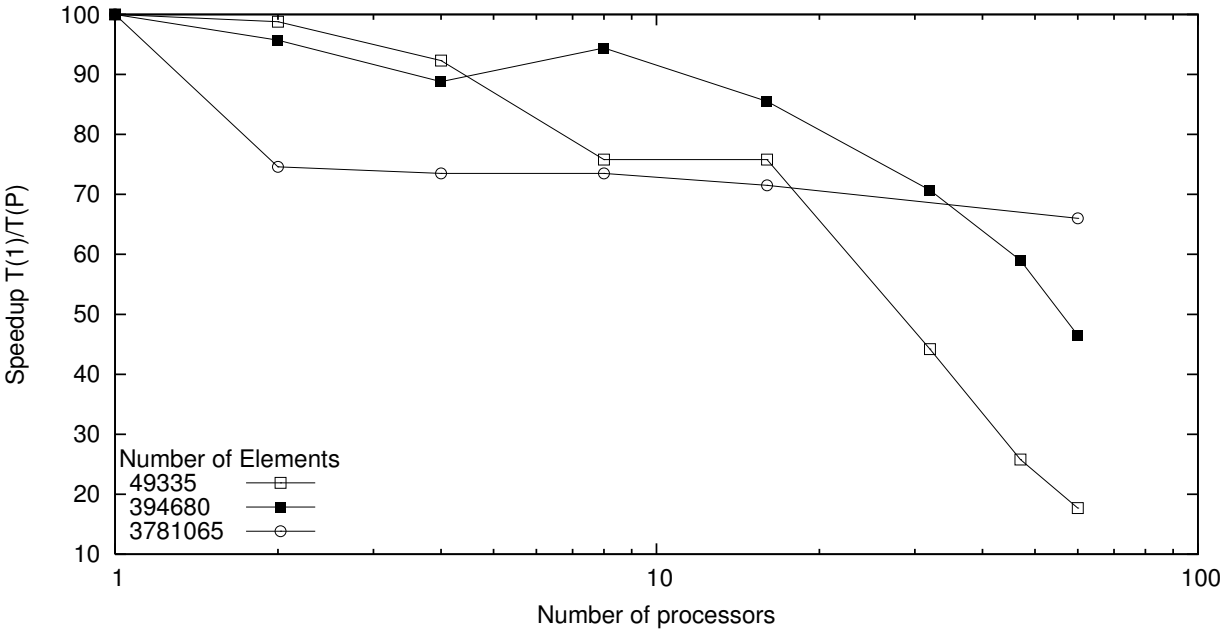


Figure 9: Speedup for the transport step.

A mesh, which containing up to 4 million volumes consists of 19% halo volumes approximately. But a mesh with only 50000 volumes reduplicates them using 60 processors, see Figure 8. Of course, it makes no sense to solve such small problems with a large number of processors. Figure 11 shows, which number of processors still produces acceptable results.

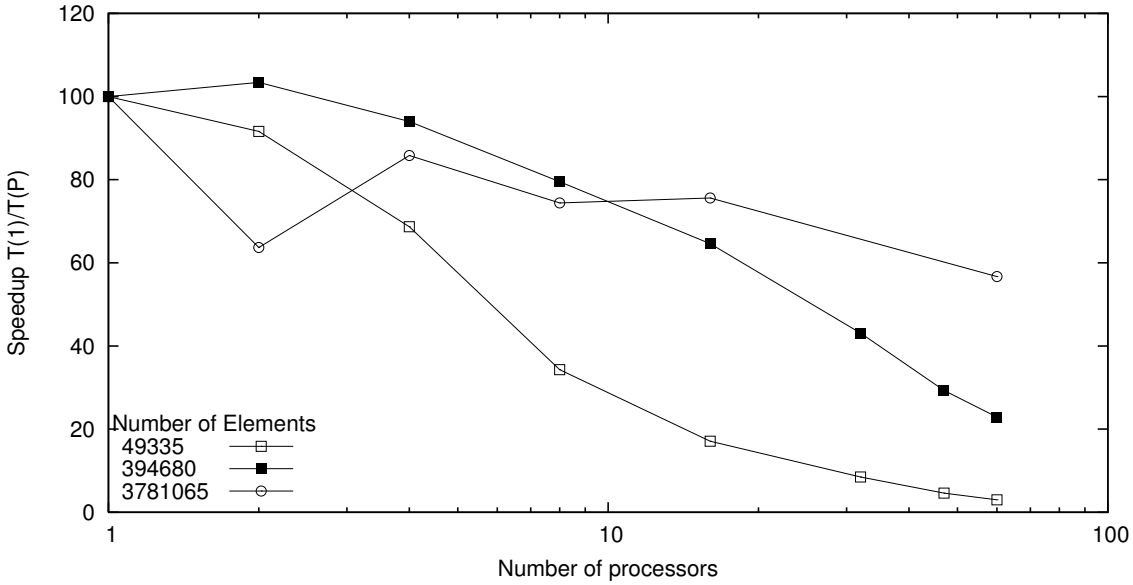


Figure 10: Speedup for the resolvent step.

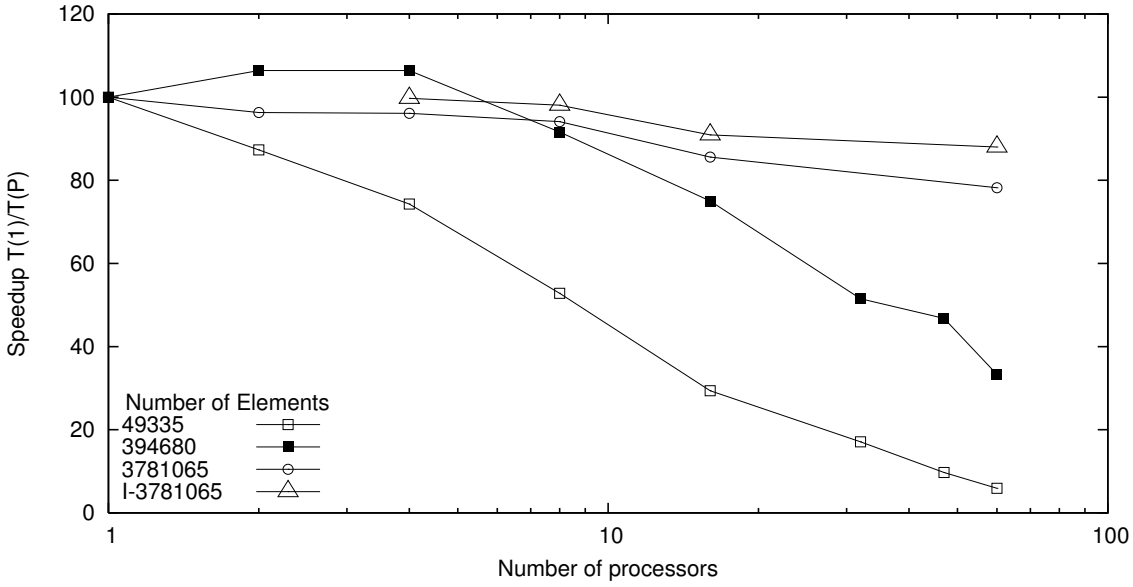


Figure 11: Speedup for the pressure correction.

The speedup for each sub--step of the projection scheme is presented in Figure 9-11. One can observe that a significant degrading of the speedup with an increasing number of processors arises for small problem sizes. In this case the communication effort of the conjugate gradient algorithm exceeds explicitly the computational cost per iteration.

The values for the largest mesh include an extrapolated value for the sequential case, computed on an equivalent system with large memory. Due to this, the incremental speedup is presented additionally. The predicted value reaches 78%, whereby the incremental speedup is approximately 88% using 60 processors. Figure

12 presents the time in seconds needed for solving the sub--steps of the projections scheme for a time step, where TP means transport step, RS resolvent step and PR pressure correction step for three different mesh sizes using 49335, 394680 and 3781065 elements. From the measurements it is evident that small problems can be solved efficiently with up to 16 processors, whereas for the largest problem 60 processors could be used with 88% efficiency.

A standard preconditioned conjugate gradient solver is used for solving the linear systems of equations. As expected, the most expensive part is the solution of the pressure correction term. It's clear, that a multigrid method is more suitable for this problem. Hence, padfem² an interface to the high performance solver package hypre [4], which provides an algebraic multigrid solver. The current version of padfem² supports only the sequential version. However, for sufficient large problems the strength of the multigrid procedure shows up. The solution of the pressure correction step can be accelerated thereby about factor 6. The implementation of the parallel version is still under construction.

Furthermore, the costs of the footpoint search from the transport step increases dramatically if the spatial step--size decreases. Figure 9 shows that the transport step runs still efficiently with 60 processors but needs already more than two times longer as the pressure correction step.
This behavior can be changed, if the global time step size is decreased, but then, the global simulation time increases. Therefore, a new variant of this step is under construction.

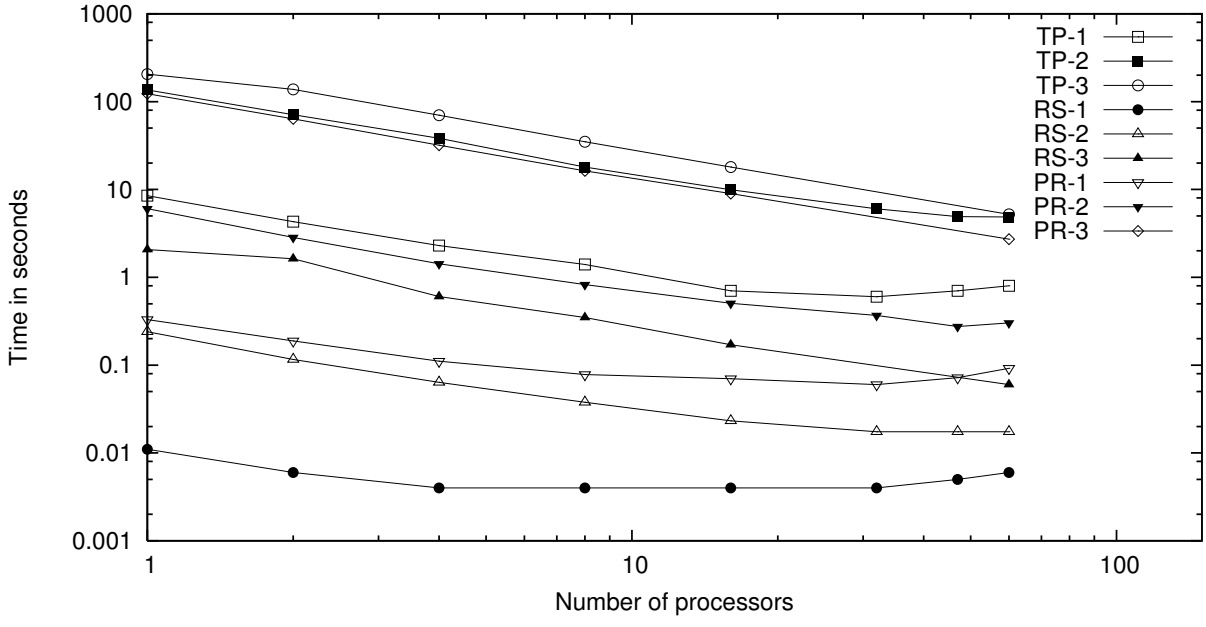


Figure 12: Simulation time for a complete time step.

References

[1] Blazy, S., and Marquardt, O.: Parallel finite element computations of three dimensional flow problems using padfem. *International Journal of Parallel, Emergent and Distributed Systems*, Volume 22, Issue 4, 2007, Pages 257 – 274.

[2] Bothe, D., Hoffmann, M., Schlueter, M., Rabiger, N., Blazy, S., Stemich, C. and Warnecke, H.-J. 2003: Experimental and numerical investigations of t-shaped micro-mixers. Proc. 11th European Conference on Mixing, VDI-GVC, pp. 269--276.

[3] Falgout, R.D., Jones J.E. and Yang, U.M., 2006, The Design and Implementation of hypre, a Library of Parallel High Performance Preconditioners. In Numerical Solution of Partial Differential Equations on Parallel Computers, A.M. Bruaset and A. Tveito, eds., Springer-Verlag, 51, pp. 267-294.

[4] John, V.: 2002, Higher order finite element methods and multigrid solvers in a benchmark problem for the 3d navier-stokes equations. Int. J. Numer. Meth. Fluids 2002}, 40, pp. 775--778.

[5] Schafer, M. and Turek S.: 1996, Benchmark computations of laminar flow around a cylinder. Flow Simulation with High-Performance Computation II, E.H. Hirschel ed., Notes on Numerical Fluid Mechanics}, Vieweg, Braunschweig, {\bfseries 52}, pp. 547--566.

5.4.2 Tools for the design of coating molecules for cellulose fibres

Project coordinator	Prof. Dr. Gregor Fels, University of Paderborn
Project members	Oliver Stüker, University of Paderborn

General Problem Description

Natural polymers like polysaccharides and proteins are of particular importance for our every day life. As a result, there is a yearly production on our planet of more than 10^{12} t of cellulose, starch and other biomaterials. It is, therefore, not surprising that there is a long tradition of cellulose usage by men, for instance in form of cotton and paper and as natural raw material for the production of textiles and plastics, e.g. rayon, viscose and celluloid. Cellulose is a green product and, consequently, is also of increasing interest for industrial use. Industry is actively engaged in improving the surface properties of cellulose, for instance for use in textiles, in order to give the surface a water resistant, dirt rejecting but yet air permeable property. This characteristic can be achieved by coating the cellulose surface with an appropriate material, i.e. with a chemical that is covalently or non covalently bound to the cellulose surface.

The underlying interaction between the cellulose and the coating molecules in this process is still far from being understood so that basic tools for the analysis of this interaction are strongly desired. If, in addition, these tools are computationally based, one would have the chance for an initial theoretical screening of various coating materials in terms of their behavior at the cellulose surface, before doing real experiments in the laboratory, certainly another “green aspect” cellulose use. We have developed a general method that allows a semi quantitative correlation between functional groups of coating molecules with sheets of cellulose layers as they occur for instance in textiles.

Problem details and work done

Cellulose is built from D-glucose units connected by β-1,4-glycosidic linkages as very long linear and unbranched chains (Figure 1). This is in contrast to the other prominent natural glucose polymer, starch, which shows an α-1,4-glycosidic linkages of the glucose monomers. While the α-1,4-glycosidic linkage of starch results in helical structures, the β-glycosidic linkage (red circles in Figure 1) as found in cellulose forces the glucose polymer to form a long stretched molecule that can aggregate into sheets of single cellulose stands. This results in a uniform distribution

of -OH groups on the outside of each chain which in turn are used to zip two or more cellulose chains together by hydrogen bonds. Finally many cellulose chains of this type can zip together to give a highly insoluble, rigid, and fibrous polymer.

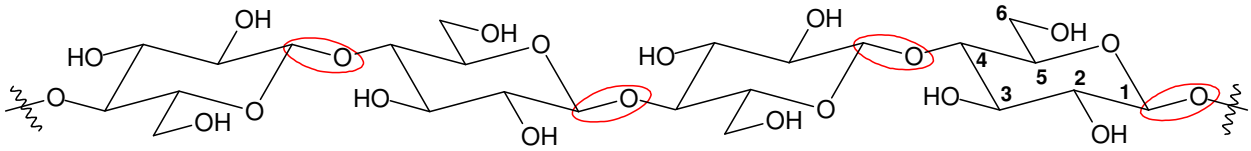


Figure 1 Chemical structure of a single cellulose strand made from four single glucose units covalently connected by β-1,4-glycosidic linkages (red circles)

From the known four types of cellulose (I, II, III_I, III_{II}, IV_I und IV_{II}) cellulose II is the most important modification for industrial use, in which layers of cellulose chains are oriented in an anti parallel fashion with respect to the layer above and below.
In order to investigate the cellulose-ligand interactions, we have constructed a theoretical cellulose II crystal from 40 single cellulose chains, arranged in 8 layers of 5 parallel cellulose chains, each constructed from 8 glucose rings (see Figure 2).

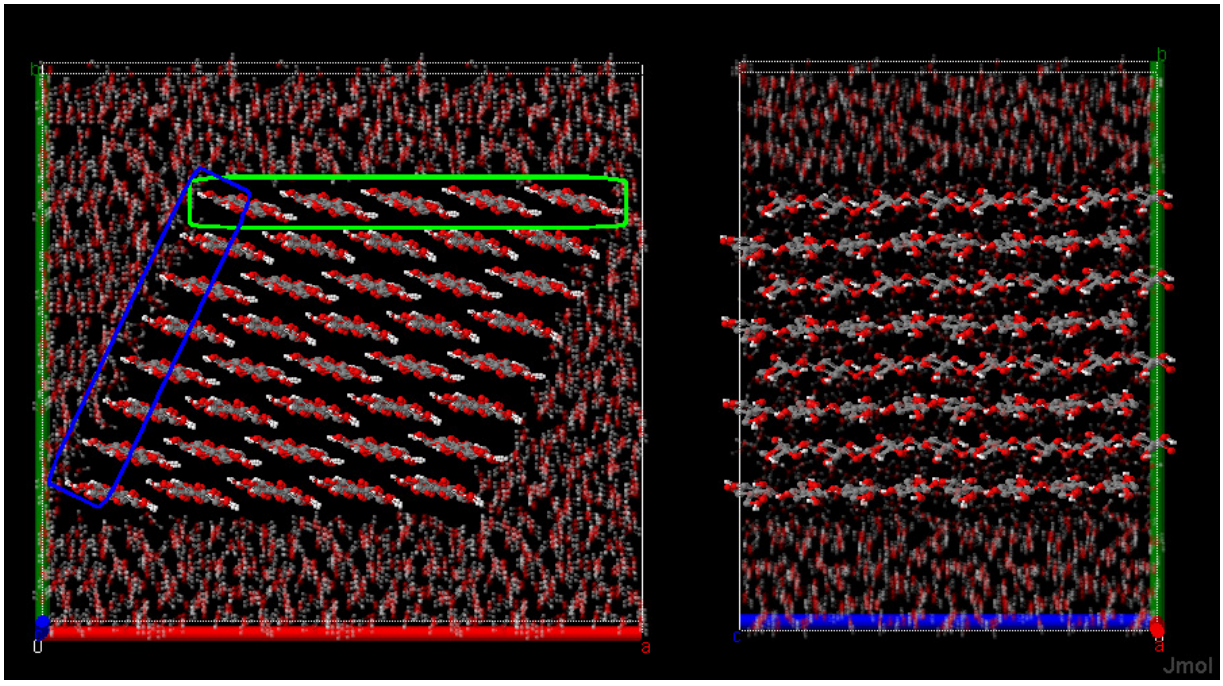
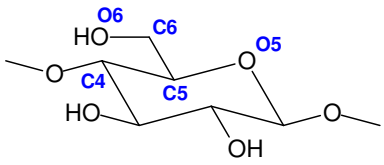


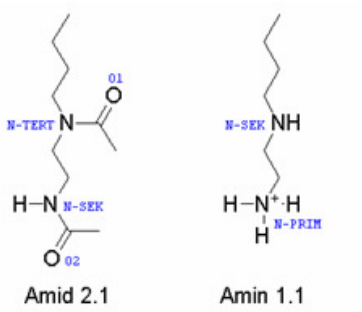
Figure 2: Simulation-box with 40 cellulose chains (8 layers with 5 chains each). Each chain consists of 8 glucose rings. The box is filled with a total of 3269 water molecules (transparent) and is shown as a projection of the X-Y-plain (left side), and the Y-Z-plain (right side). The chains are oriented parallel to the Z-axis (blue), while the rings are laying in the X-Z-plain (not shown). The resulting theoretical cellulose crystal represents a polar (blue border) and an non polar surface (green border).

This structure then was imbedded in an initial simulation box of X=7,00 nm, Y=5,50 nm and Z=4,10 nm size. The box was filled with a total of 3269 water molecules and was equilibrated several times in 10 ns Gromacs MD-runs. The resulting cellulose system was evaluated by comparing parameters of single chain cellulose crystal structures with the corresponding data of the model. In particular we have looked at the angle χ (O5-C5-C6-O6) and χ' (C4-C5-C6-O6) which depict the conformation of the free hydroxymethyl-group (see the inserted monomeric structure), as well as at the so called Puckering-parameter [1], which describes the conformation of the glucose rings. The analysis shows that the structural features of the 22 cellulose chains representing the interface to the bulk water do not differ substantially from those 18 chains buried inside the hypothetical crystal and that the resulting χ- and χ'-angles as well as the Puckering parameter are in good agreement with crystallographic data of cellulose II single chain structures [2, 3].



To investigate the behavior of small molecules (ligands, i.e. potential coating molecules) at the polar and the non polar cellulose surface, we have chosen the charged amine 1.1 and the uncharged amide 2.1 (see insert next page) as model compounds for the development of a quantitative method that describes the cellulose-ligand interaction. For both ligands two cellulose-ligand systems were set up in which the ligands were placed in an 1 nm distance either from the polar Y-Z-plain (blue border in Figure 2) or the non polar X-Z-plain (green border in Figure 2), respectively. These systems were then submitted to molecular dynamics simulation runs, twice for 250 ps and once for 1 ns, with reporting of the coordinates and the energies every 0.5 to 1.0 ps.

The structural differences between the diamine (amine 1.1) and the diamide (amide 2.1) result in a drastic change in their behavior at the cellulose surface i.e. in how these compounds interact with atoms at the cellulose surface. Although the ligands in the MD runs were initially placed 1 nm away from the surface, they quickly reach the cellulose surface by diffusion, where they interact predominantly with the OH-groups of the cellulose surface. In case of amide 2.1, the ligand perfectly sticks to the polar surface for the rest of the simulation time. The interaction between amide 2.1 and the non polar surface is less strong, as the ligand is able to leave the surface again after short residence. In contrast, amine 1.1. neither interacts with the polar nor with the non polar cellulose surface, as it walks at random on the surface as it does in the surrounding bulk water. Figure 3 shows the 250 ps trajectories of ligand movement on the polar cellulose surface. The position of the cellulose atoms are frozen in their t = 0 ps position, while the positions of the



amide 2.1 and amine 1.1 are shown in time resolved spectral color codes starting from t = 0 ps with a red color to the final position after t = 250 ps in blue. As can be seen, amide 2.1 (Figure 3 left) quickly reaches the surface where it is locked at a given location. Amine 1.1 (Figure 3 right), however, although also reaching the surface early on in the trajectory, never stays at a given location but rather moves along the surface and into the bulk water.

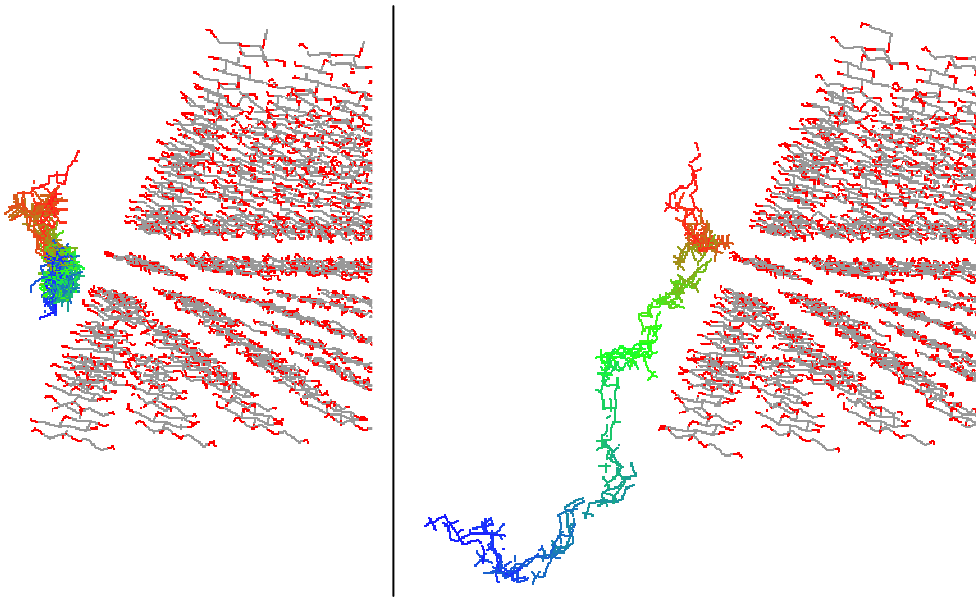


Figure 3: Trajectories of ligand movement of amide 2.1 (left) and amine 1.1 (right) at the polar cellulose surface. The Figure depicts the ligand positions in time-steps of 5 ps with a time resolved spectral color code from red (t=0 ps) towards blue (t=250 ps) with a constant picture of the cellulose at t=0 ps.

In order to quantify the interaction behavior, we have calculated the radial distribution function, RDF, between heteroatoms of the ligands and the oxygen-atoms of the OH-groups on the cellulose surface. The RDF-function in this case describes the relative frequency of the appearance of one or more oxygen atoms of the cellulose surface in a given distance “r to r+Δr” of a reference atom, i.e. a heteroatom of the ligand. The RDF function is described by

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi \cdot r^2}$$

where $<\rho_B(r)>$ is the particle density of the cellulose-oxygen atoms around a ligand atome, and $<\rho_B>_{local}$ is the particle density of ligand heteroatoms (averaged over all spheres around the cellulose oxygen atoms in question). The results are listed in

Table 1, in which the function g(r) is numerically displayed as histograms of the frequency versus the radius r, averaged over all time steps of the trajectories.

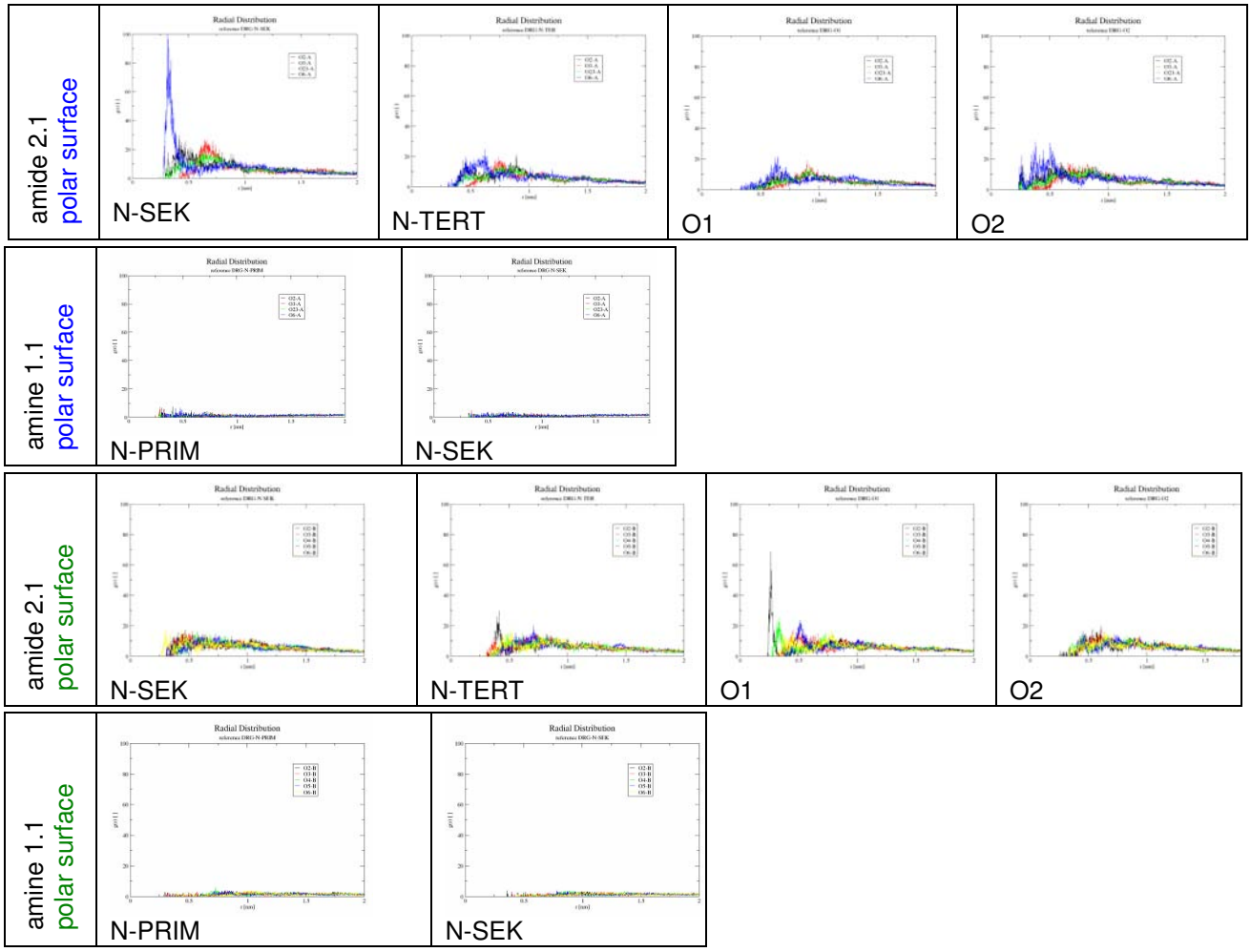


Table 1: RDF diagrams fort he polar an non polar cellulose surface: Ligand names are shown in the first column. The next 2-4 columns of a given row show the RDF-functions of ligand heteroatoms under investigation interaction with oxygen atoms at the cellulose surface. The color codes for the cellulose-oxygen atoms are
- for the polar surface: O2 (black), O3 (red), O2 und O3 (green) und O6 (blue)
- fort he non polar surface O2 (black), O3 (red), O4 (green), O5 (blue) und O6 (yellow).
The x-axis always is scaled from 0 to 2 nm.

The RDF diagrams clearly show that in case of amide 2.1 there is a strong interaction with the polar cellulose surface at a distance between 0.25 and 0.5 nm of ligand heteroatoms and oxygen atoms at the cellulose surface (row 1 in Table 1). In particular this interaction, which should be hydrogen bonding predominantly involves the secondary amide nitrogen of the ligand and O6 of the cellulose, i.e. a hydroxymethyl-oxygen at the cellulose surface (row 1, column 1, blue line). Amide 2.1 also shows a moderately strong interaction with the non polar surface, this time however governed by the interaction between the carbonyl-oxygen atom at the chain

terminal and the O2-oxygen at the surface (row 3, column 3). In contrast, the ionic ligand amine 1.1 almost shows no interaction, as the corresponding RDF-function are all close to the base line, both, at the polar and the non polar surface, respectively (row 2 and 4, respectively).

The method described allows for a quantification of the interaction between specifically defined ligand atoms with specifically defined atoms at the cellulose surface and, thereby, paves the way toward the screening of functional groups of ligands, i.e. of potential coating molecules, for their interaction with the cellulose surface. In addition, the methods yields a structural model of the special arrangement of these functional groups with respect to the surface atoms of the cellulose. The procedure is not limited to cellulose surfaces, and should, therefore, be a general tool for the the investigation of surface coating for all kind of natural, semisynthetic and synthetic polymer material, provided the molecular structure of the surface is known or – alternatively – can be modeled by computational techniques.

Resource Usage

All MD-simulation were calculated using GROMACS [4, 5], version 3.3 compiled with ScaMPI Message-Passing-Interface and using the Infiniband interconnect on the Arminius cluster at the PC^2 of the University of Paderborn. The forcefield calculation where done with the Gromos96-forcefield G45a3 including the G45a4-extension for carbohydrates [6], using a flexible SPC-water model. The topological parameters of the two ligands where generated using the Dundee-PRODRG2.5 Server (Beta) [7] and adjusted if necessary.

References

[1] Cremer, D. and J.A. Pople, *General definition of ring puckering coordinates.* J. Am. Chem. Soc., 1975, *97*, 1354-1358.

[2] Langan, P., Y. Nishiyama, and H. Chanzy, *X-ray Structure of Mercerized Cellulose II at 1 Å Resolution.* Biomacromolecules, 2001, *2*, 410-416.

[3] Langan, P., et al., *Synchrotron X-ray structures of cellulose Iβ and regenerated cellulose II at ambient temperature and 100 K.* Cellulose, 2005, *12*, 551-562.

[4] van der Spoel, D., et al., *GROMACS: Fast, flexible, and free.* Journal of Computational Chemistry, 2005, *26*, 1701-1718.

[5] Lindahl, E., B. Hess, and D. van der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Journal of Molecular Modeling, 2001, 7, 306-317.

[6] Lins, R.D. and P.H. Hunenberger, *A new GROMOS force field for hexopyranose-based carbohydrates.* J Comput Chem, 2005, *26*, 1400-1412.

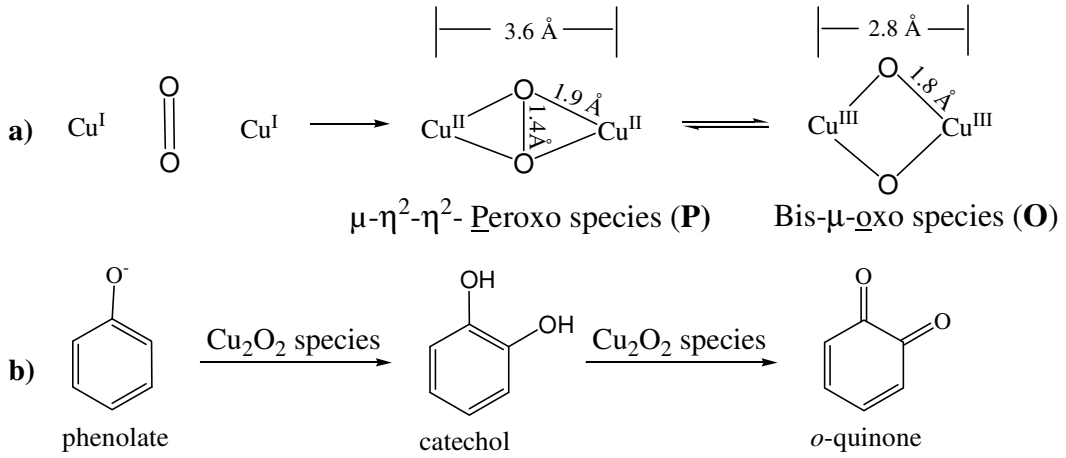
[7] Schuttelkopf, A.W. and D.M. van Aalten, *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes.* Acta Crystallogr D Biol Crystallogr, 2004, *60*, 1355-1363.

5.4.3 Density Functional Methods using Gaussian03 in Bioinorganic Chemistry

Project coordinator	Prof. Dr. Gerald Henkel, University of Paderborn
Project members	Dr. Sonja Herres-Pawlis, University of Paderborn
Supported By	DAAD, University of Paderborn

General Problem Description

Binding and activation of dioxygen by copper centres play an important role in the catalytic cycle of an array of enzyme systems. Tyrosinase, catechol oxidase and hemocyanin are the members of the type 3 dinuclear copper protein family in which each copper ion is ligated by three histidine imidazoles, holding molecular oxygen in a side-on binding mode to give a μ-η^2:η^2-peroxodicopper(II) complex (**P**) as the common active-oxygen species (Scheme 1a). Tyrosinase is a ubiquitous binuclear copper enzyme, found in fungi, plants, and animals, that catalyses the hydroxylation of phenols to catechols and the oxidation of catechols to quinones (Scheme 1b) [1].



Scheme 1: a) Oxygen activation by copper complexes,

b) Hydroxylation of phenolates and oxidation of catechols

The oxygenated form of tyrosinase contains a **P** species with an intact O–O bond, that hydroxylates phenols through a mechanism consistent with an electrophilic aromatic substitution. Because biological intermediates beyond the **P** species are unknown, the sequence of intimate steps of bond cleavage and formation remains unclear. Contrastingly, in synthetic chemistry, a great number of examples of an

isomer of the **P** species with fully reduced O-O bond have been described and it was shown that these so-called **O** species also are able of hydroxylating phenolates as well.[2] In the actual discussion, **O** components are suggested to act as hydroxylating agent but their high activity prevents them from being detected spectroscopically. Hence, density functional theory can contribute to explain the relative stability of these species, their ability of interconversion and their reactivity towards phenolic substrates. There has been extensive theoretical work focused on the **P**/**O** equilibrium. Very large variations in the predicted relative stabilities of **P** and **O** species have been reported with different levels of theory in various systems [3].

In our approach, a purely σ-donating amine functionality and a sterically demanding guanidine functionality with π-donating/accepting ability are combined in one ligand system (hybridguanidine systems). The calculation of **P**/**O** energies of hybridguanidine stabilised copper/oxygen species and their spectroscopical features are calibrated with experimental data.

The development of biomimetic model complexes for oxygen activating and transferring copper enzymes targets a better understanding of their structure-function-relationship. The study of these oxygen-activating processes is expected to supply new insights into the biologically relevant properties and the reaction processes of copper enzymes for the development of effective and environmental friendly oxidation catalysts.

Problem details and work done

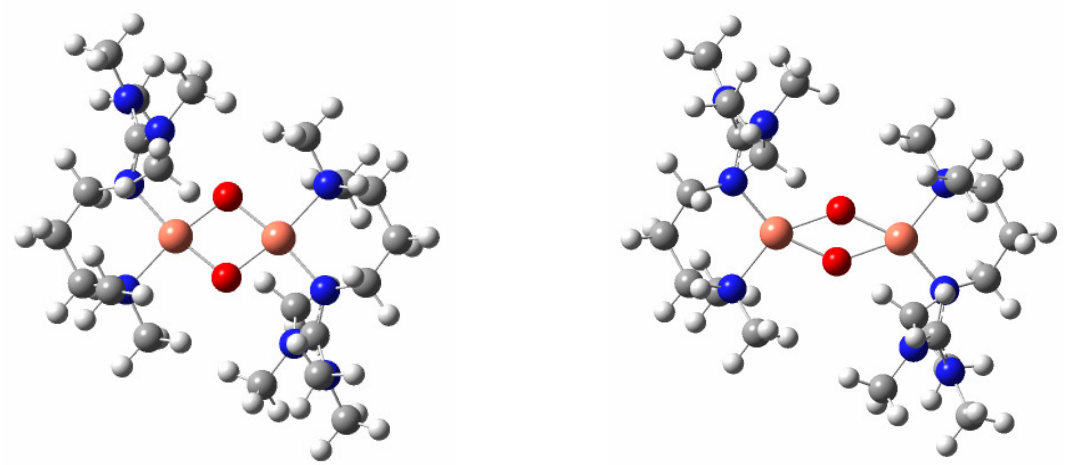
Calculation of the oxo and peroxo cores

Density functional theory (DFT) calculations were performed using the Gaussian 03 program. Optimized geometries were calculated using the B3LYP exchange-correlation functional. Tight SCF convergence criteria (10^{-8} a.u.) were used for all calculations. Frequency calculations were performed to ensure that the stationary points were minima.

The geometries of the **O** and **P** species were optimised using several basis sets starting with the crystal structures of the bis (μ-hydroxo) dicopper system. For the optimisation, the B3LYP functional was used which was shown to reproduce experimental data reliably. The **O** species were optimised with the restricted B3LYP functional using double-zeta and triple-zeta basis sets. In some calculations, a triple-zeta basis set was used on Cu, O and N atoms and a double-zeta basis set on C and H atoms, this approach (abbreviated „gen“) has been found to reproduce experimental data of comparable systems quite well [4]. The **P** species were optimized using the unrestricted B3LYP functional and the broken-symmetry formalism. Relevant metrical parameters are presented in Table 1 and the optimised structures are depicted in Figure 2.

Species	Basis Set	Cu…Cu	O…O	Cu-O_{avg}	Cu-N_{gua}	Cu-N_{amine}
O	6-31G(d)	2.762	2.274	1.789	1.912	2.004
	6-31G+(d)	2.797	2.298	1.810	1.937	2.039
	6-311G(d)	2.831	2.247	1.808	1.925	2.017
	gen	2.809	2.306	1.817	1.941	2.044
P	6-31G(d)	3.527	1.514	1.919	1.916	2.000
	6-31G+(d)	3.608	1.484	1.950	1.944	2.049
	6-311G(d)	3.574	1.500	1.938	1.933	2.021
	gen	3.637	1.482	1.964	1.948	2.054

Table 1: Metrical parameters of the **O** and the **P** species in Å



$[(TMGdmap)_2Cu(III)_2(\mu\text{-}O)_2]^{2+}$ (**O** species) $[(TMGdmap)_2Cu(II)_2(\mu\text{-}O_2)]^{2+}$ (**P** species)

Figure 2: Optimised structures of the **O** and the **P** species (B3LYP, gen)

In the next step, the energy difference between these two species was determined in dependence of the basis set. The results of these calculations are summarized in Table 2. Additionally, a solvation model for THF was included for the geometry optimised with gen. The energetic difference of the **O** and **P** species which was directly determined by the difference of the energies of the optimised geometries predicts the real situation not very well. Only with the double-zeta basis and by use of the solvation model, a „realistic“ value is produced. A more sophisticated method makes use of single point (sp) calculations with the BLYP functional and gen basis

sets. This methodology reflects the experimental values for the selected basis sets very well.

Basis set	ΔE(kcal/mol)	ΔE(kcal/mol) [sp gen BLYP]
6-31G(d)	-10.9	-8.8
6-31G+(d)	+1.2	-8.2
6-311G(d)	-0.60	-7.3
gen	-0.15	-5.7
PCM gen	-5.5	-9.2

Table 2: Energy differences of the **O** and **P** species

Time-dependent DFT of the oxo species

On the basis of these results, the geometries of the **O** and **P** isomers were then used for subsequent time-dependent DFT calculations. Calculated elctronic transitions were transformed via the SWizard Program into simulated UV/Vis spectra [5]. Molecular orbital (MO) compositions and the overlap populations between molecular fragments were calculated using the AOMix program [6].

Figure 3 shows the calculated UV/Vis spectra in comparison to experimental spectra, and Table 3 lists the principal electronic excitations which contribute to the electronic transitions.

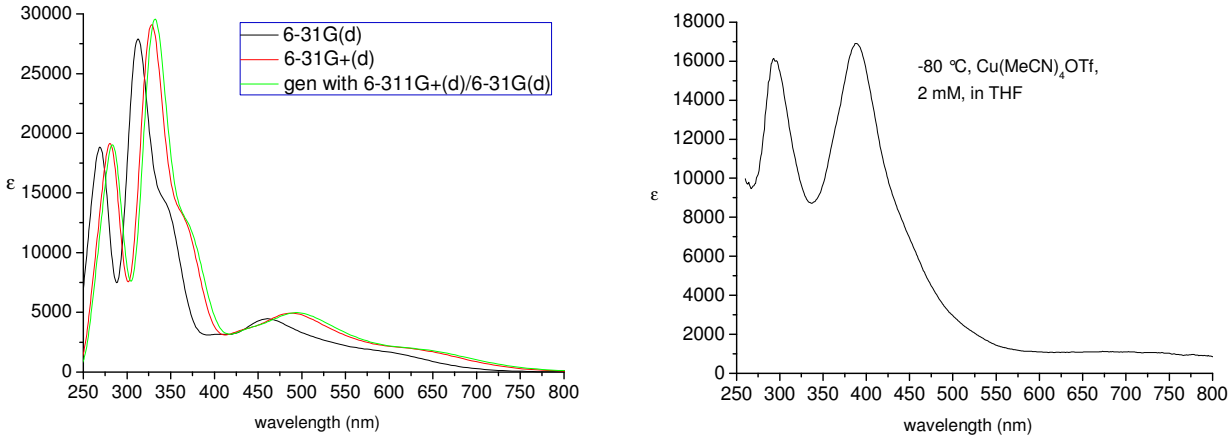


Figure 3: Comparison between calculated (left) and experimental (right) UV/Vis spectra

The TD-DFT calculations predict spectra which are in good agreement with the experimental spectra. The shift to the observed spectra depends on the basis set: with double-zeta basis set, the shift amounts up to 90 nm wheras with the triple-zeta the spectra are only shifted around 40 nm. In account of the complexity of the dinuclear antiferromagnetically coupled system, the TD-DFT calculations reproduce the real electronic transitions with high consistence.[7]

Wavelength	Oscillator strength	MO transitions
285 nm	0.146	Charge transfer of O, N_{gua} and N_{amine} in Cu_2O_2 core
331 nm	0.368	Charge transfer of N_{gua} in Cu_2O_2 core
372 nm	0.167	Charge transfer of O, N_{gua} and N_{amine} in Cu_2O_2 core
435 nm	0.034	Charge transfer of O in Cu_2O_2 core
490 nm	0.054	Charge transfer of N_{gua} (π system) in Cu_2O_2 core

Table 3: Principal electronic excitations

Especially the analysis of the contribution of the guanidine functionality to the bonding situation is an important result of these calculations. Figure 4 illustrates the orbitals with high guanidine contributions (HOMO=Highest Occupied Molecular Orbital), whereas Figure 5 shows the LUMO (Lowest Unoccupied Orbitals) and the LUMO+1 which are involved into the transitions.

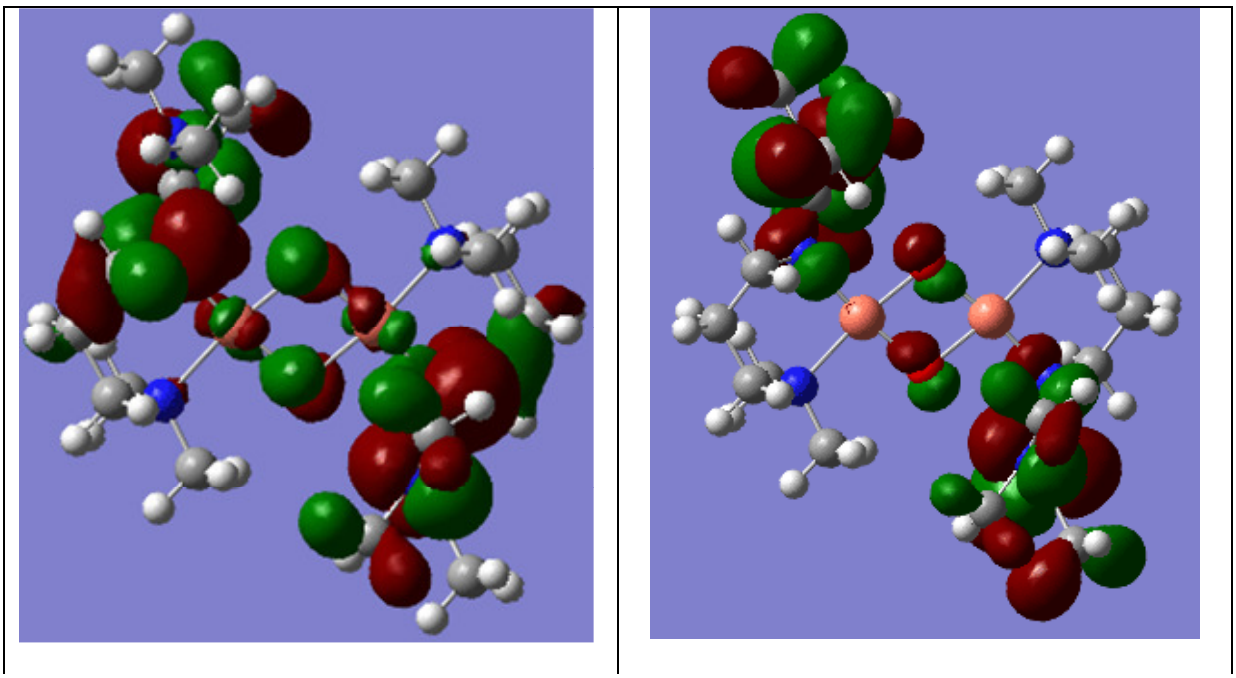


Figure 4: left: HOMO with 70% guanidine contribution, right: HOMO-2 with 70% guanidine contribution

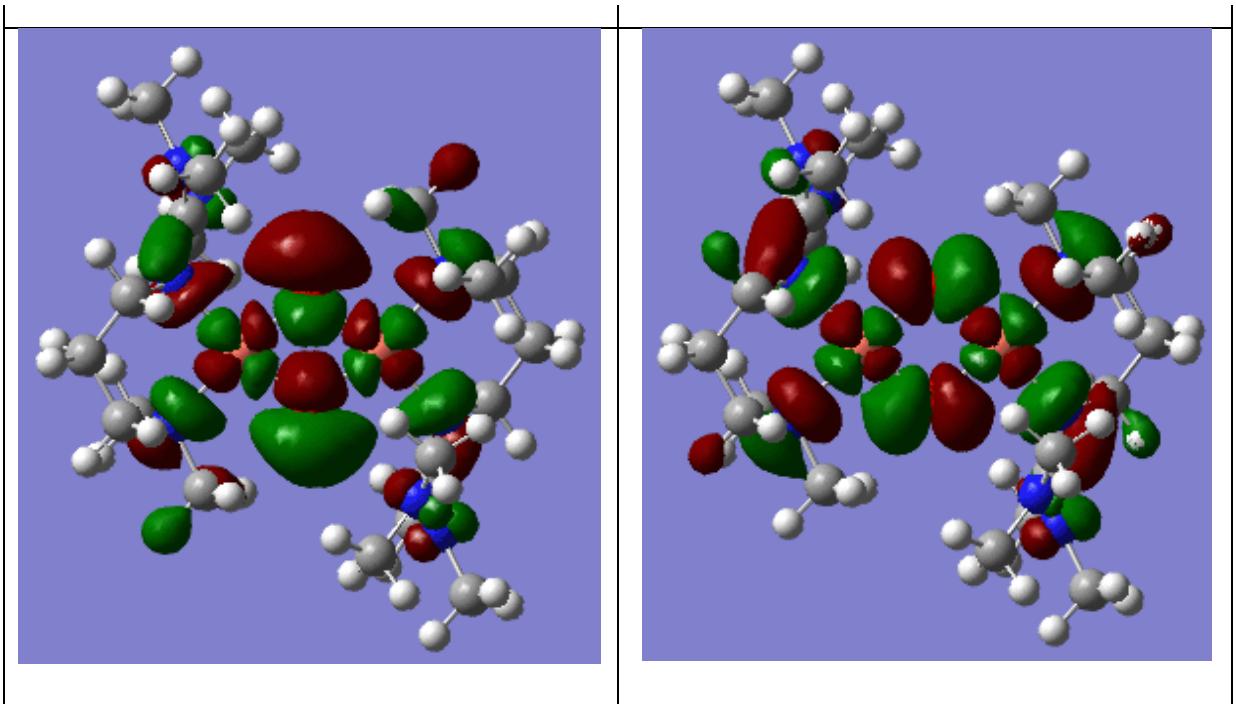


Figure 5: left: LUMO with 34 % Cu and 44 % O contribution, right: LUMO+1 with 37 % Cu and 34 % O contribution

The HOMO and the HOMO-2 are clearly dominated by the guanidine π system, whereas the LUMO.and the LUMO-1 have classical Cu_2O_2 character.[1] The asymmetry between amine and guanidine functionality causes the appearance of the new side band at 420 nm (exp. value) because the guandine σ donation has a different character than the amine σ donation.

Further hybrid ligands and their corresponding **O** and **P** species are also under investigation in order to understand the interplay of guanidine and amine functionality within these systems but these studies have not been concluded yet. In the next step, the coordination of phenolates to the relevant hybrid ligand stabilised **O** species will be simulated and analysed in comparison to pure amine systems.

Resource Usage

The calculations were executed at the ARMINIUS cluster at the PC^2. DFT calculations using Gaussian03 can only use up to 8 nodes parallely. Due to the high number of calculations which have to be accomplished, several calculations were done at the same time separately each with 4 double-core nodes using the Ethernet interconnect of the cluster. In this context, the ARMINIUS cluster provides outstanding computational power. Furthermore, the PC^2 facility gives us the possibility to use Gaussian03 as parallel version.

Geometry optimisations, energy calculations, frequency calculations and time-dependent density functional calculations were run as parallel processes on 4 nodes at the same time. The calculations were set up on a daily basis. Geometry optimisations including the energy calculation have an averaged duration of some days up to 3 weeks depending on the size of the molecule and the chosen basis sets. Normally, the calculations were set up for 6 days and then restarted if required. Frequency and time-dependent density functional calculations can not be restarted, such that the calculations have to be set up for 6 -12 days by experience. Usually, we use between 12 and 40 nodes at the same time.

References

[1] E. I. Solomon, U. M. Sundaram, T. E. Machonkin, *Chem. Rev.* **1996**, *96*, 2563.

[2] [a] H. Decker, R. Dillinger, F. Tuczek, *Angew. Chem.* **2000**, *112*, 1656; *Angew. Chem. Int. Ed. Engl.* **2000**, *39*, 1591.
[b] L.M. Mirica, M. Vance, D. Jackson-Rudd, B. Hedman, K.O. Hodgson, E.I. Solomon, T.D.P. Stack, *Science* **2005**, *308*, 1890.

[3] J.L. Lewin, D.E. Heppner, C.J.Cramer, *J. Biol. Inorg. Chem.* **2007**, *12*, 1221.

[4] P. Verma, T.D.P. Stack, in preparation.

[5] S.I. Gorelsky, SWizard program, http://www.sg-chem.net/.

[6] S.I. Gorelsky, A.B.P. Lever, *J. Organomet. Chem.* **2001**, *635*, 187.

[7] S. Herres-Pawlis, R. Haase, U. Flörke, G. Henkel, P. Kang, P. Verma, T.D.P. Stack, *Inorg. Chem.* **2007**, submitted.

5.4.4 Theoretical and numerical Investigation of nonreactive and reactive fluid mixing in an T-shaped micro-mixer

Project coordinator	Prof. Dr. Dieter Bothe, RWTH Aachen Prof. Dr.-Ing. Hans-Joachim Warnecke, PC², University of Paderborn
Project members	Dipl.-Chem. Carsten Stemich, University of Paderborn M.Sc. A. Lojewski, RWTH Aachen
Supported by	Deutsche Forschungsgemeinschaft (DFG)

General Problem Description

The large area-to-volume ratio of micro-reactors gives prospect of better yield and selectivity than for conventional designs, since diffusive fluxes of mass and heat scale with the area, while the rate of changes are proportional to the volume. Indeed, theoretical considerations of the scaling behavior support the fact that micro-reactors allow for faster chemical reactions and provide better thermal control [1,2]. In applications of Chemical Reaction Engineering, the previous mixing step is crucial, especially for fast reactions.

To avoid large pressure drops and to obtain defined flow conditions, laminar flow fields are chosen whit a secondary flow to generate large contact area. In such a laminar but complex flow, the mixing state is determined by the interplay of convective and diffusive transport phenomena [1,2]. Hence, thorough CFD-simulations of yield and selectivity of such reacting flows require the resolution of the finest length scales of both the velocity and the concentration fields. In order to reduce the numerical complexity we employ a mathematical model of the relevant convection-diffusion-reaction equations which is similar to the parabolized Navier-Stokes system. The model allows resolving the finest length scales to asses the effects of down scaling on conversion and selectivity.

In this project, a T-shaped micro-mixer with rectangular cross sections as illustrated in Figure 1 is investigated. With regard to chemical reaction, we perform our simulations with quasi-instantaneous neutralization reactions and fast irreversible reactions of second order. All calculations were carried out with the CFD program FLUENT 6.2, which solves the Navier-Stokes equations and additional transport equations in discretised form based on the Finite-Volume method.

Problem Details and Work Done in the Reporting Period

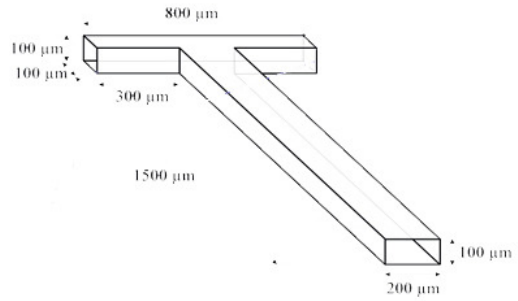


Figure1 Mixing zone of the T1-micro-reactor - investigated by 3D simulations.

We investigate the reactive fluid mixing of different geometries of a T-shaped micro-mixer with rectangular cross sections. In all cases the micro-reactor consists of two quadratic inlet- channels and a mixing-channel which has the same depth as the inlet-channels. The width of the micro-channel is two times of the depth of the inlet-channel. Due to the required resolutions of the velocity and the concentration field we performed 3D-CFD-simulations for a reduced domain of the micro-reactor with shortened inlet-channels and shorted channel – Table 1, [1,2].

The hydrodynamics of liquid flow inside a micro device with one of these geometries is adequately described by the incompressible Navier-Stokes equations. As boundary condition we chose a fully developed channel flow at the

micro-reactor	mixing-channel	inlet-channel
T1	200 × 100 × 1600	100 × 100 × 300
T2	400 × 200 × 2400	200 × 200 × 600
T3	600 × 300 × 3200	300 × 300 × 900

Table 1 Geometries of mixing zone of the investigated micro-reactors. All length values are given in µm.

inlets, the no slip condition at fixed walls and a pressure boundary condition at the outlet. Previous work has shown, that in case of a stationary flow the most efficient mixing is obtained in the so-called engulfment flow regime ($140 < Re < 240$ for the micro-reactor T1) [1, 2]. For further investigations of reactive mixing we focus on the engulfment flow regime. The simulations for the described mixing zone show, that convective mixing takes place in the first part of the investigated domain. Due to the complex flow field, which shows domains backflow is occurring in, 3D-simulations are required for this part of the micro-reactor.

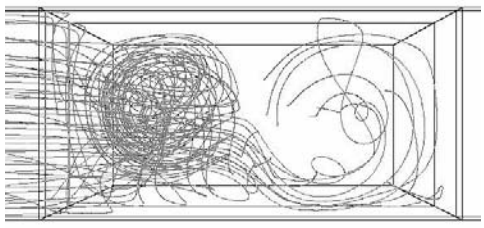


Figure 2 Flow trajectories crossing the symmetry plane in engulfment flow. View into the mixing-channel.

For species transport we use a Dirichlet boundary condition at the inlets and a Neumann boundary condition at fixed walls and at the outlet. For reacting flows, these equations have to be complemented with source/sink terms modeling the chemical kinetics. In case of fast chemical reactions the latter term leads to several numerical difficulties, but for quasi-instantaneous reactions a simplification can be found. For this purpose, only the difference of the two species concentrations is considered. The resulting equation has no source term and is of the form $\partial_t\phi + \mathbf{u}\cdot\nabla\phi = \nabla \mathrm{D}(\phi)\nabla\phi$ with $\phi = c_A - c_B$ in case of an irreversible second order reaction like the neutralization between HCl and NaOH. The obtained results from numerical simulations show good accordance with experimental data provided by the group of Prof. Räbiger (Institute of Environmental Process Engineering, University of Bremen). In the experiments, a pH-sensitive indicator was used to visualize the occurring transport phenomena. This indicator was calculated in CFD-simulations as a further scalar.

In case of low Schmidt numbers like those for HCl (Sc = 300) and NaOH (Sc = 470), we are able to resolve the finest length scales for both the velocity and the concentration fields by using a Cartesian grid with up to four levels of adaptation and about 30 million grid cells.

However, while the smallest length scales of the velocity field are about $\lambda_{vel} \approx 5\mu m$, the concentration fields can show considerably finer structures with typical smallest dimensions given by the Batchelor length scale $\lambda_{conc} = \lambda_{vel} / \sqrt{Sc}$ [1, 2]. For the micro-channel T1 and values of Sc ≈ 1000 for the Schmidt number, typical for aqueous media, the resulting finest length scales are $\lambda_{conc} \approx 0.1$ µm. These are significantly smaller then those of the velocity field. Under these conditions we are only able to resolve the length scale of the velocity field on a 3D-grid for a reduced mixing zone. Since chemical reactions take place in the whole micro-channel and because we want to understand the influence of the state of mixing on the chemical reaction, CFD-simulations need to be performed for the full micro-channel, resolving the length scales of the concentration field. To attain the required resolutions and to extend the simulations on the entire micro-channel we employ a simplified mathematical model by using some properties of the chosen laminar flow conditions.

Under stationary laminar flow conditions, the axial velocity component always points downwards along the axial direction already shortly behind the entry into the mixing-channel. The occurring transition of the flow field from an engulfment flow to a channel flow allows handling the axial flow direction as a pseudo-time variable, so that the evolution of the concentration profile can be computed on successive cross sections, thereby following the main axial flow direction. Considering the stationary

flow conditions and neglecting species diffusion in axial direction, the following axial evolution model is derived from the 3D species equation [3].

$$\partial_z c_i + \begin{pmatrix} \frac{u}{w} \\ \frac{v}{w} \end{pmatrix} \cdot \nabla_{x,y} c_i = \nabla_{x,y} \cdot \left(\frac{D_i}{w} \nabla_{x,y} c_i \right) + \frac{r_i(c_i,\ldots,c_m)}{w}, \quad (1)$$

where c_i is the molar species concentration, *x, y* denote the cross sectional and *z* the axial coordinate and *u*, *v*, *w* are the corresponding velocity components. Finally, D_i is the species' diffusivity and *r* is the rate function of the chemical reaction. The model is complemented by homogeneous Neumann boundary conditions. Let us note that while the full velocity satisfies the no slip condition at the lateral boundaries, we observe in the 3D-simulations that the ratios *u*/*w* and *v*/*w* attain finite limits at these walls. The implementation of this model into FLUENT 6.2 requires the reformulation of Eq. 1 as an equation with divergence structure. This leads to

$$\partial_t \rho c_i + \nabla_{x,y} \cdot \left(\begin{bmatrix} u/w \\ v/w \end{bmatrix} \rho c_i - \frac{\rho D_i}{w} \nabla_{x,y} c_i \right) \frac{r_i}{w} + \rho c_i \nabla_{x,y} \cdot \begin{bmatrix} u/w \\ v/w \end{bmatrix} - \nabla_{x,y} \left(\frac{\rho D_i}{w} \right) \cdot \nabla_{x,y} c_i . \quad (2)$$

This equation is solved on a 2D-grid employing a second order spatial discretization and a first order time discretization. To avoid too small time steps because of large stiffness of the differential equations due to fast chemical reactions, we split the time step into a transport part followed by a reaction part. In the latter, the changes in concentrations due to the chemical reaction are calculated analytically for each step [3]. This can be done in case of simple reaction systems but has to be replaced by an appropriate approximation – possibly by means of an inner numerical computation loop using time substeps – in the general case.

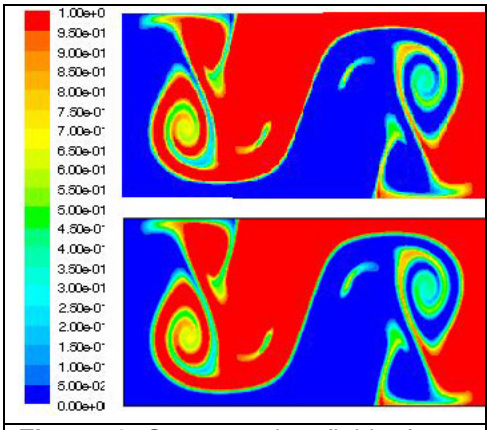


Figure 3 Concentration field of one educt on cross-section – 3D approach at 150 µm in flow direction above and initial concentration field for the 2D approach below

First test calculations with this approach were performed with an artificial flow and concentration field and were compared with simulations executed on a 3D-grid under the same conditions. The excellent agreement between the two approaches [3] allowed extending the 2D one to velocity and concentration fields obtained from 3D-simulations. Therefore we read out velocity data from every cross section from a simulation on a 3D-grid and used it as basic data for a linear interpolation of the velocity field for the 2D-approach. For cross sections, lying in a domain of the micro-channel, which were not investigated by a 3D- simulation the required velocity field for the 2D- approach was received by an approximation using the exponential decrease from secondary flow field to fully developed duct flow. To resolve the concentration field we determined the cross section on which no backflow occurs for the first time. Using highly refined grids (adaptive refinement, 33 Mio cells, 32 nodes) we were able to resolve the concentration field for the micro-channel up to the located cross section. The redistribution of the achieved concentration data on the uniform grid used for the 2D-approach gave the initial concentration field for every species. To simulate the entire micro-reactor a combination of computations on 3D- and 2D-grids is required. The 3D-approach is necessary for the domain, in which backflow is detected (yellow marked part in Figure 4). The 2D-approach is used to simulate the part of the micro-reactor in which no backflow is occurring (red marked part in Figure 4). The combination of both techniques allows the simulation of the entire micro-reactor by resolving all involved length scales. First simulations to reveal scaling effects on the rate of chemical reaction using both approaches have just been started. Results will be shown in the next report.

Resource Usage at PC²

The numerical investigations of the T-shaped micro mixer with rectangular cross sections have been performed with the commercial CFD-Tool, FLUENT 6.2. The necessary resolution which requires a computational domain with up to 33 million cubic grid cells in the mixing-channel can only be reached with high parallelization. Therefore, calculations with locally refined grids were done using the ARMINIUS FSC Xeon-DP cluster with up to 32 nodes. In case of the 2D-model the required resolutions are obtained by using 4 to 8 nodes in a parallel computation.

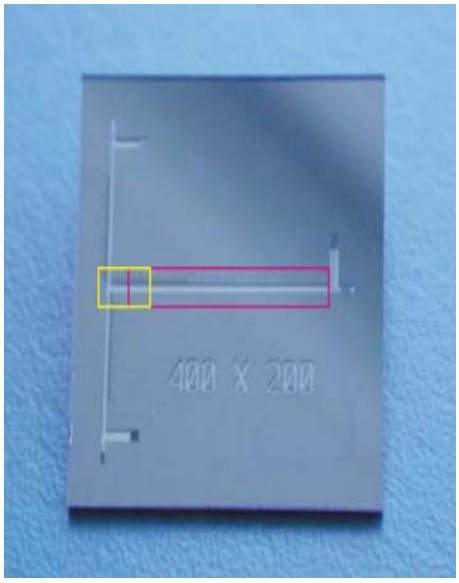


Figure 4 Simulation of the entire micro-reactor – yellow part 3D-model, red part 2D-model

References

[1] D. Bothe, C. Stemich, H.-J. Warnecke: *Fluid mixing in a T-shaped micro-mixer,* Chem. Eng. Sci 61, 2950-2958 (2006).

[2] D. Bothe, C. Stemich, H.-J. Warnecke: *Mixing in a T-shaped microreactor: scales and quality of mixing,* pp 351-357 in 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering (W. Marquardt, C. Pantelidis, eds.), Garmisch-Partenkirchen (Germany), Elsevier 2006.

[3] D. Bothe, A. Lojewski, H.-J. Warnecke: *Direct numerical simulation of reactive mixing in a T-shaped micro-reactor*, 5th Joint ASME/ JSME Fluids Engineering Conference, FEDSM2007-37507, San Diego (USA) July 30 – August 2, 2007

5.4.5 Simulation of spin coupled iron sulphur systems

Project coordinator	Prof. Dr. Marx, Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum
Project members	Dr. N. Nair, Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum
Work supported by	DFG (DFG MA 1547/7)

General Problem Description

Iron–sulfur clusters are among the most ubiquitous, functionally versatile, and ancient prosthetic groups in proteins [1,2]. One of the major classes of mobile electron carriers in contemporary biology is ferredoxin (Fd), which contains a fourfold cysteine coordinated [2Fe–2S] core [1,2]. These clusters are characterized by complex spin coupling between the iron atoms [3]. In the electronic ground state of the oxidized form of bioinorganic [2Fe–2S] clusters, the Fe3+ metal centers are typically in a local high–spin d5 configuration (i.e. *S*(Fe) = 5/2 for ferric Fe(III) centers) whereas they themselves are antiferromagnetically coupled to yield an overall low–spin state due to dominant superexchange interactions [3]. In particular, the electronic ground state of [2Fe–2S] clusters in Fd has genuine multi–reference character and makes computational studies challenging. Usually a so-called broken-symmetry description of the electronic structure is employed, which is however a highly spin contaminated state and does not represent the ground state [4].

A particularly interesting system is Fd from cyanobacterium Anabaena PCC7119 whose crystal structure was recently obtained at very high resolution in both the oxidized and the reduced forms [5]. Most interestingly, it was found that two alternative conformations of Cys46 exist, which is one of the four cysteinyl ligands of the [2Fe–2S] cluster. In the oxidized state Cys46 adopts the so called “CO–in” conformation, where the peptide oxygen points in the direction of the cluster, while upon reduction the backbone flips to the “CO–out” conformation [5].

In order to understand iron–sulfur proteins it is inevitable to consider temperature effects, protein conformational motions, solvent interactions and, most importantly, it is mandatory to properly describe the antiferromagnetic electronic ground state. To overcome the shortcomings of the broken-symmetry description, we introduce a general computational technique by which the antiferromagnetic ground state structure and the dynamics of (binuclear homovalent) transition metal clusters in proteins at realistic conditions can be obtained efficiently in the framework of mixed quantum/classical MD in combination with an extended spin–projection formalism. We compare our results with the broken symmetry deteriminent and study the magnetostructural dynamics for a model iron-sulfur cluster in vacuo. Using this

procedure, we are currently investigating the dynamics of the oxidized [2Fe–2S] cluster in fully solvated *Anabaena* PCC7119 Fd. We expect that this will provide insights into the relationships of magnetic properties and structural dynamics including the conformational influence of the protein environment.

Problem details and work done

In order to properly describe the forces acting on the nuclei in the antiferromagnetic electronic ground state of binuclear iron-sulfur clusters, and thus to overcome the shortcomings of the broken-symmetry description, we introduce a general spin projection technique.

At the heart of the proposed approach is the observation that the weak spin-coupling occurring in systems like the binuclear transition metal clusters can be described with the Heisenberg-Dirac-Van Vleck Hamiltonian [3]. This generates a spin-ladder of eigenstates with total (coupled) spin quantum number S ranging from $S_{max}=S_A+S_B$ (high-spin, HS) down to $S_{min}= S_A-S_B$ (low-spin, LS), with an interval $E(S)-E(S-1)=-2JS$, where J is a constant called exchange spin coupling constant. Employing this, we could derive a general formalism for the LS energy, from a broken-symmetry (BS) one determinant singlet state energy and HS energy. As the forces on the nuclei are the partial derivatives of the total energy with respect to the nuclear coordinates, we are also able to calculate the forces on the nuclei at the LS state.

Inspired by this projection formalism, we have also put forward a Lagrangian for performing efficient Car-Parrinello dynamics and implementation together with the quantum chemical-molecular mechanical methods (QM/MM) for studying proteins. We call our method as extended broken symmetry (EBS) approach.

After the implementation of the EBS technique in the CPMD code [6,7], we have performed thorough testing of our method and implementation. For a model cluster, $[Fe_2S_2(SH)_4]^{2-}$, we verified the stability of our Car-Parrinello Lagrangian. On the simulated timescale of 5 ps (microcanonical) simulation the fluctuations are within 4.0×10^{-6} a.u, and drift in the total energy was very small (See Figure 1), demonstrating that the dynamics is stable.

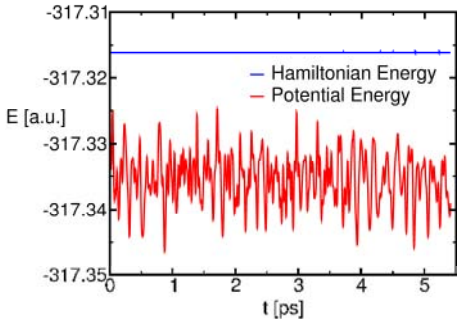


Figure 1: The conserved Hamiltonian energy (blue) on the scale of the potential energy (red). All energy values are reported in a. u.

The first question to address is, what an effect the EBS method has on the lowest potential energy structure of the iron sulfur cluster compared to the usual BS method and how these structures relate to the high spin conformation. Figure 2 shows shows a superposition of the corresponding cluster structures obtained by geometry optimization. Apparently, the BS and the EBS structures are very similar whereas the cluster is more open in the high spin state.

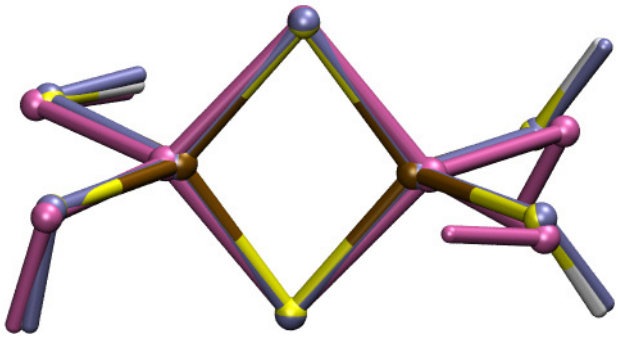


Figure 2: $[Fe_2S_2(SH)_4]^{2-}$ structures obtained by structure optimizations in vacuo in different spin states. The structure corresponding to the BS state is ice-blue, the HS structure is mauve and for the EBS structure the iron, sulfur and hydrogen atoms are shown in brown, yellow and white, respectively.

However the *J* values were -227, -390 and -435 cm^{-1} for the HS, BS and EBS states, respectively. Following to these simulations, we have analyzed the vibrational frequencies at the BS and the EBS states; vibrational motions at the ground state were found to be blue shifted compare to that at the BS state. These results show that at the ground state not only the structure and magnetic properties, but also the dynamics is different.

In order to relate the dynamics of *J* to the structural dynamics of the iron-sulfur cluster within the BS and the EBS schemes, Car-Parrinello AIMD simulations were performed using both approaches. The same trend as obtained by structure optimizations is seen in the averages of structural quantities and *J* calculated from molecular dynamics trajectories. The average of *J* is -356 and -403 cm^{-1} for the BS and EBS case, respectively, and a more compact average structure is obtained within the EBS scheme (see Figure 3). The corresponding averages of the Fe-Fe and the S-S distances within the core are different for the conventional and the extended broken symmetry schemes, indicating different minima on the free energy surface (FES). Additionally, since the variances are also different, the shapes of the underlying FES in the sampled regions differ as well. The rather broad probability distributions of *J* show also the necessity of considering dynamical behavior of the cluster when studying the magnetic properties of such systems: by fluctuations *J* can change by a factor of two.

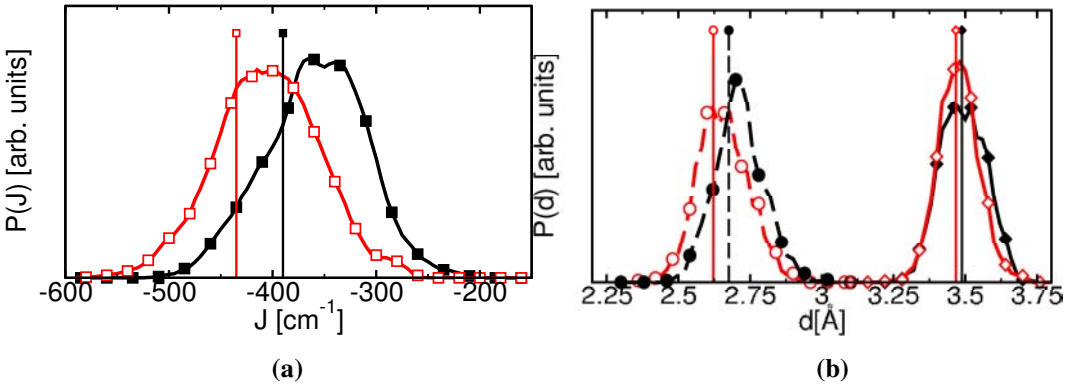


Figure 3: Probability distribution functions of (a) the Fe-Fe (broken line, circles) and S1-S2 distance (solid line, diamonds) and (b) of the coupling constant *J* (solid line, squares) obtained from AIMD of the $[Fe_2S_2(SH)_4]^{2-}$ cluster in vacuo using the BS (black, filled symbols) and EBS (red, open symbols) schemes. The corresponding values obtained for the minimum energy structures are shown as vertical bars with a symbol.

To understand the magnetostructural dynamics, we have taken the power spectrum of *J* from the MD simulation and estimated the cross correlation of the projected trajectory of the iron-sulfur core on all normal modes defined based on the ideal symmetry of the cluster (see Figure 4).

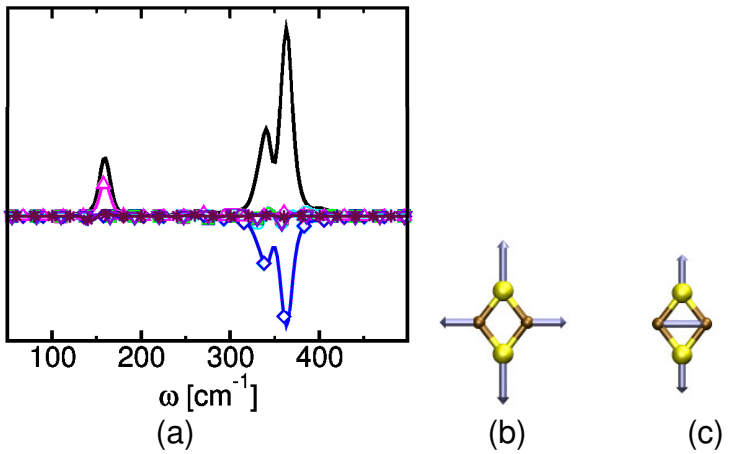


Figure 4: Power spectrum of *J* (black curve) together with the frequency dependent correlation between *J* and the individual vibrational components of the Fe2S2 core dynamics obtained from the EBS-AIMD of the cluster in vacuo; $A_{g,A}$ (magenta), $A_{g,D}$ (blue), other colors are the rest of the modes. Ideal symmetry adapted modes (b) $A_{g,D}$ and (c) $A_{g,A}$. Yellow spheres are sulfur atoms and grey spheres are iron atom**s.**

There, mainly the two A_g modes, the totally symmetric angle vibrations of the angles ($A_{g,A}$) and of the distances ($A_{g,D}$), account for all features exhibited by the dynamics of the exchange coupling constant. The influence of the individual modes on the coupling constant can be understood by recalling the nature of the superexchange interaction. Along the two A_g modes both superexchange paths are always affected

in a symmetric manner. For the $A_{g,D}$ mode for example, both paths simultaneously become shorter or longer, thus enhancing or weakening the superexchange interaction. Along other modes the two superexchange paths are either both disrupted, or not affected at all.

We are now working on studying similar dynamical magnetostructural relations for the iron-sulfur cluster in the Anabaena protein environment (see Figure 5). In the protein environment the symmetry of the cluster is distorted and there are many hydrogen bonds to the bridging sulfur atom of the cluster. Thus we expect a different magnetostructural behavior for this protein. QM/MM simulations using the CPMD/Gromos [6,7] interface code is underway.

Resource Usage

We have used the ARMINUS Xeon cluster for the computations. Jobs were submitted using the batch system, using a maximum of 41 processors at a time. The interconnect was infiniband. The program

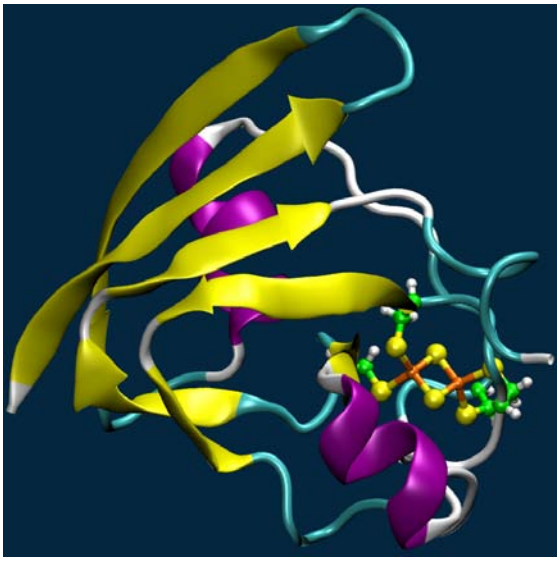


Figure 5: Anabaena ferredoxin; ball and sticks are used to show the quantum part of the system used in the QM/MM calculations. Color code: carbon (green), yellow (sulfur), white (hydrogen), orange (iron)

CPMD [6,7] used for these calculations is very well parallelized using MPI. During the production runs, the cluster was used on a daily basis.

References

[1] Hall DO, Cammack R, Rao KK (1971) Nature 233:136-138.
[2] Rees DC, Howard JB (2003) Science 300:929-931.
[3] Kahn O (1993) Molecular Magnetism (VCH, New York).
[4] Adamo C, Barone V, Bencini A, Broer R, Filatov M, Harrison NM, Illas F, Malrieu JP, de P R Moreira I (2006) J Chem Phys 124:107101.
[5] Morales R, ChronMH, Hudry-Clergeon G, P´etillot Y, Norager S, Medina M, Frey M (1999) Biochemistry 38:15764-15773.
[6] Hutter et al J CPMD, IBM Corp 1990-2007, MPI fuer Festkoerperforschung Stuttgart 1997-2001, see also http://www.cpmd.org.
[7] Marx D, Hutter J (2000) Modern Methods and Algorithms of Quantum Chemistry, vol. 3, ed. Grotendorst J, FZ Juelich (NIC), pp. 301-449

5.4.6 Mechanochemistry of Thiolates on Gold Surfaces

Project coordinator	Prof. Dr. Dominik Marx, Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum
Project members	Dr. Jordi Ribas, Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum
Work supported by	Alexander von Humboldt Fellowship

General Problem Description

Traditional ways to introduce excess energy in order to trigger chemical reactions is to use temperature, photons and electrical currents which leads to thermochemistry, photochemistry and electrochemistry, respectively. Recently, the capability to manipulate molecules on the atomic scale by Atomic Force Microscopy techniques made it possible to induce and control chemical reactions and to construct new nanoscale architectures by applying external mechanical forces[1]. In this project we want to study these phenomena by means of first principles calculations and *ab initio* molecular dynamics simulations[2,3]. The system we have chosen to carry out our theoretical study is thiolate molecules adsorbed on gold surfaces because this system is not only well accessible to experiments but also because previous theoretical work has shown that it is possible to draw gold nanowires when alkylic thiolates are pulled off a stepped gold surface or off small gold clusters.[4] One of the main goals of our project is to simulate which nanostructures and which chemical phenomena take place when other more complex thiolates rather than the alkylic ones are pulled off a gold surface. For instance we are interested in aromatic, bidentate and tridentate thiolates. In addition to that, we also want to compare their mechanochemical pathway(s) of detachment to the thermal reaction mechanisms. Overall, our project has implications for the stability and degradation of molecule/metal junctions as used, for instance, in coatings and molecular electronics.

Problem details and work done

COMPUTATIONAL DETAILS

In this project we have considered the Au (111) surface as representative of extended gold systems. To model the surface we have used periodically repeated slabs of four or six layers of gold atoms (Figure 1). The four-layer and the six-layer slabs consists of twelve and thirty gold atoms per layer, respectively. We have used either one surface or the other depending on the size of the thiolate adsorbed onto it. We have included a vacancy on the first layer of both surfaces in order to simulate a

surface with defects, since it is experimentally well known that the Au (111) surfaces are usually characterized by a large atomic roughness with both adatoms and vacancies [1]. The thiolates we have worked with are the four thiolates depicted in Figure 2.

To mimic a typical mechanical pulling experiment involving a thiolate attached to a gold surface the coordinates of the gold atoms of the bottom layers of the slab are frozen and the top methyl group of the thiolate is constrained to move in a plane parallel to the bottom layers of the slab. Stepwise static pulling in small increments (0.2 Angstroms) is first applied and then the system is optimized with the mentioned constraints. These static calculations have been done with the *Quantum ESPRESSO* package [2], which is a plane wave/pseudopotential implementation of Density Functional Theory[3]. The PBE functional [4] together with ultrasoft pseudopotentials [5] with a cutoff of 25 Ry have been used. On the other hand, in order to assess which is the effect of the temperature in this pulling experiments we have started to carry out metadynamics simulations [6] for the dithiolate molecule with the CPMD code [7], which is another plane wave/pseudopotential implementation of Density Functional Theory (DFT), particularly designed for *ab initio* molecular dynamics. [3]. The metadynamics simulations allow one to compute multidimensional free energy surfaces as a set of collective variables using coarse-grained non Markovian molecular dynamics. The collective variables used in our simulations are two: the distance of the dithiolate to the bottom layer of the gold surface, and the coordination number of the sulphur atoms of the dithiolate with respect to the gold atoms.

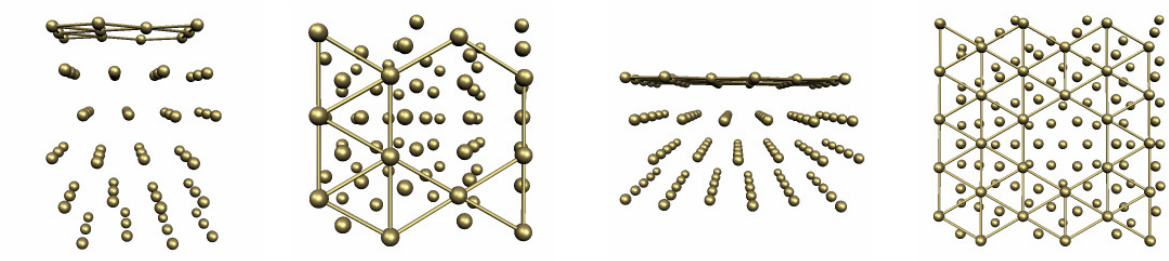


Figure 1: Two views of the six-layer gold slab (on the left) and two views of the four-layer slab (on the right)

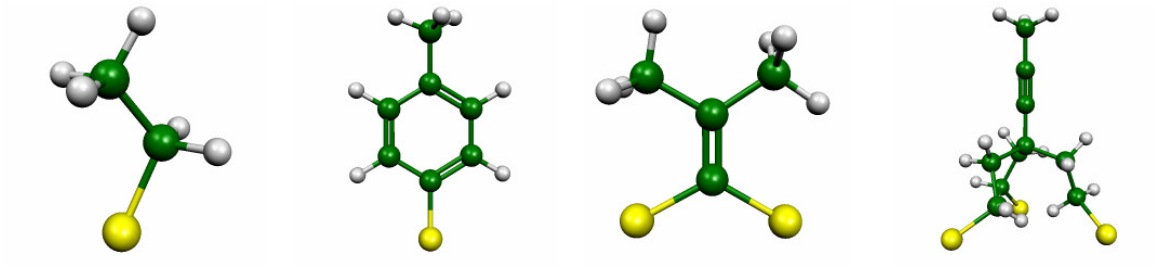


Figure 2: Thiolates used in our pulling simulations. From left to the right: an alkylic thiolate, an aromatic thiolate, a bidentate thiolate and a tridentate thiolate.

RESULTS

The simulations performed so far with the alkylic and aromatic thiolates indicate that the mechanochemical behaviour of the two species is different. As can be seen in Figure 3, when an alkylic thiolate is pulled off the gold surface a gold-gold bond is eventually broken, i.e., the alkylic thiolate is capable of pulling a gold atom off the surface. On the contrary, when an aromatic thiolate is pulled off the surface, the sulphur-gold bond is the bond which breaks, i.e., the aromatic thiolate is not capable of pulling any gold atom off the surface. Electronic structure calculations have allowed us to rationalize such different mechanochemical behaviour, the reason being a weakening of the sulphur-gold bond in going from the alkylic thiolate to the aromatic one.

It is also worth mentioning that neither the alkylic nor the aromatic thiolate are able to form complex nanostructures (or to generate a big reconstruction of the surface) when they are pulled off the Au (111) surface with one vacancy. The calculations carried out so far seem to indicate that in order to generate complex nanostructures in such a surface it is necessary to pull off bidentate and tridentate thiolates. The static pulling simulations of bidentate and tridentate thiolates are not yet finished and therefore in Figure 3 it can only be seen some of the intermediate nanostructures that have been generated during such pulling simulations.

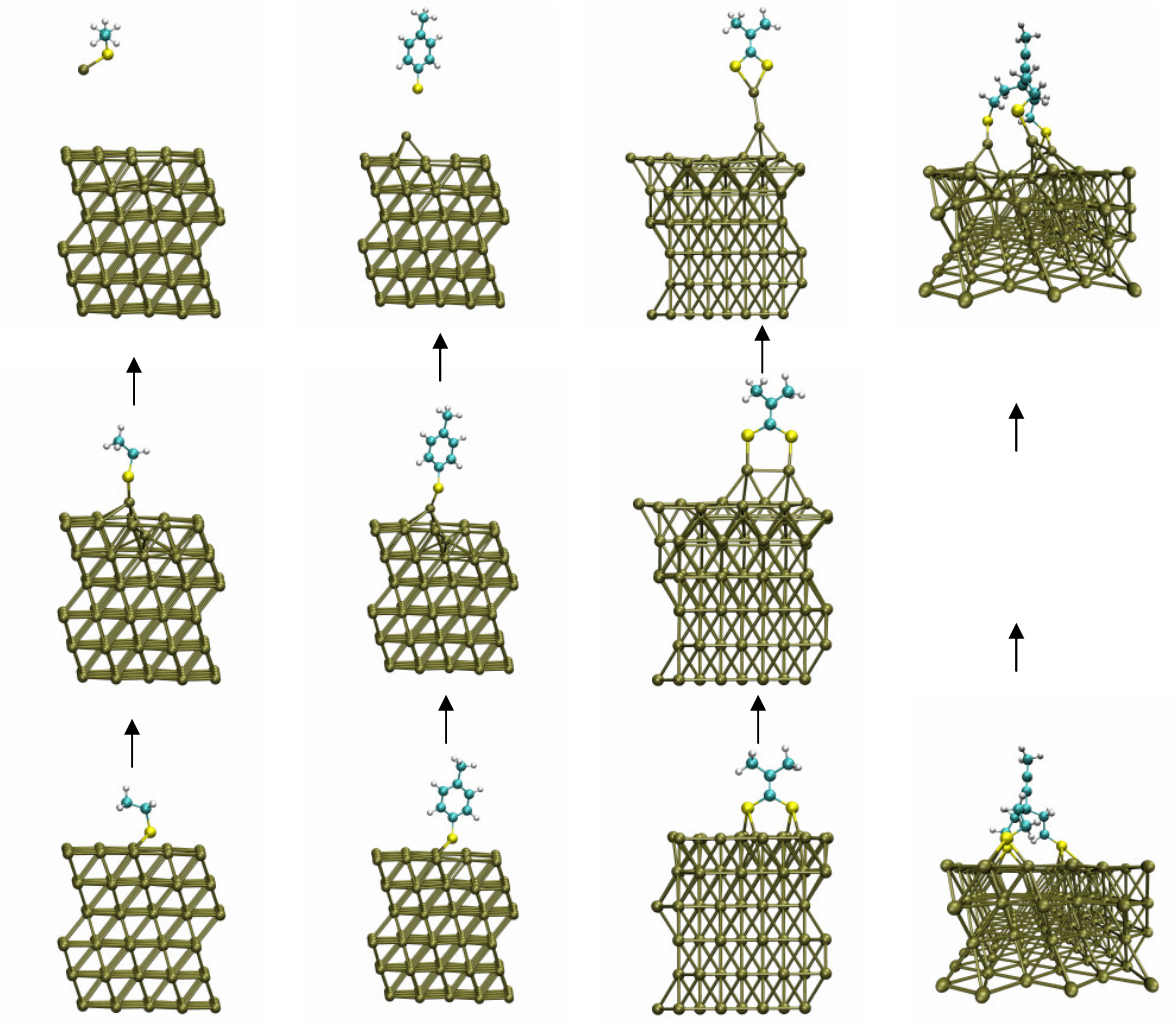


Figure 3: Nanostructures obtained in the static pulling simulations.

As we have said in the Computational Details section we are currently doing a metadynamics simulation for the bidentate thiolate in order to assess the effect of the temperature in the pulling experiments. We do not have yet any conclusive result in this respect.

FUTURE WORK

Regarding the different mechanochemical behaviour of the alkylic and aromatic thiolates, we plan to study which is the effect of adding substituents to the aromatic ring. In particular, we want to test if it is possible to weaken or strengthen the sulphur-gold bond by tuning the electronic structure of the aromatic ring.

On the other hand, we want to finish the static pulling and metadynamic simulations with bidentate and tridentate thiolates. We also plan to study the effect of the

electronic structure of the bidentate/tridentate thiolates on the mechanochemical properties of the system.

Overall, we want to explore to what extent it is possible to generate and manipulate nanostructures at will in thiolate-gold systems and which is the role played by the temperature and the subsequent dynamical effects.

Resource Usage

The simulations of this project are run on the processors of the ARMINIUS cluster (400 processors 64-bit INTEL Xeon with Infiniband communication). Both packages we are using (mainly CPMD and also Quantum-Espresso) are truly parallel and efficiently use the Infiniband interconnect. We usually run jobs with 16 processors running in parallel on a daily basis, but the codes easily scale up to much larger processor numbers if available for true throughput/capacity simulations. We are using the resources of the ARMINIUS cluster because of the compute power of the nodes in conjunction the fast Infiniband interconnect. This is crucial for being able of doing these calculations whereas clusters with only GigaBit communication could not be used for this purpose.

References

[1] Beyer, M. K. and Clausen-Schaumann, H. *Chem. Rev.* 2005, 105, 2921

[2] Car, R. and Parrinello, M. *Phys. Rev. Lett.* 1985, 55, 2471

[3] Marx, D. and Hutter, J. “Ab initio Molecular Dynamics: Theory and Implementation” in *Modern Methods and Algorithms of Quantum Chemistry*, Editor: J. Grotendorst (NIC, Forschungzentrum Jülich 2000). See www.theochem.rub.de/go/cprev.html

[4] (a) Krüger, D.; Fuchs, H.; Rousseau, R.; Marx, D.; Parrinello, M. *Phys. Rev. Lett*. 2002, 89, 186402; (b) Krüger, D.; Rousseau, R.; Fuchs, H.; Marx, D. *Angew. Chem. Int. Ed.* 2003, 42, 2251.

[5] Mazzarello, R.; Cossaro, A.; Verdini, A.; Rousseau, R.; Casalis, L.; Danisman, M.F.; Floreano, L.; Scandolo, S.; Morgante, A.; Scoles, G. *Phys. Rev. Lett.* 2007, 98, 016102.

[6] www.quantum-espresso.org

[7] Marx, D. and Hutter, J. “Ab initio Molecular Dynamics: Theory and Implementation” in *Modern Methods and Algorithms of Quantum Chemistry*, Editor: J. Grotendorst (NIC, Forschungzentrum Jülich 2000). See www.theochem.rub.de/go/cprev.html

[8] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* 1996, 77, 3865.

[9] Vanderbilt, D. *Phys. Rev. B* (Rapid Communications), 1990, 41, 7892.
[10] Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. USA.* 2002, 99, 12562.
[11] Hutter, J. et al. CPMD V3.9; Copyright IBM Corp 1990-2001, Copyright MPI für Festkörperforschung Stuttgart 1997-2001. see www.cpmd.org

5.4.7 Molecular self-organization on solid surfaces studied from massive parallel first-principles calculations

Project coordinator	Prof. Dr. Wolf Gero Schmidt, University of Paderborn
Project members	Prof. Dr. Ronei Miotto, University of Sao Paulo, Brazil Dr. Eva Rauls, University of Paderborn Dr. Uwe Gerstmann, University of Paderborn Msc. Stephan Blankenburg, University of Paderborn Msc. Stefan Wippermann, University of Paderborn Msc. Björn Lange, University of Paderborn Dipl.-Phys. Andreas Hermann, Massey University, New Zealand Dipl.-Phys. Christian Thierfelder, Massey University, New Zealand Dipl.-Phys. Susan Biering, Massey University, New Zealand
Work supported by	DFG SCHM 1361/8, DFG SCHM 1361/9, DFG SCHM 1361/10

General Problem Description

Organic molecules are very promising building blocks for electronic devices due to the possibility of tailoring molecules with particular properties, the tunability of their characteristics, and the efficiency and flexibility of deposition methods. Their functionality with respect to molecular electronics, nanodevices, and molecular recognition is intensively investigated. For years already, molecular materials have been used in solar cells, gas sensors, heterojunctions, and ultrafast optical switches. Because they have typically dimensions of a few nanometers, molecules are the ultimate limit of electronic devices.

Self-organization of organic molecules appears as one of the most promising approaches to the further miniaturization of electronic devices. This so-called bottom-up approach contrasts with the exponentially increasing fabrication costs of further down-scaling the lithographic processes in the top-down approach for device manufacturing. The latter approach already has led to atomic dimensions (the gate oxide thickness of the presently produced transistors of the 65 nm generation amounts to only 1.2 nm, i.e., about 4-5 atomic SiO_2 layers!) and is bound to lead soon to fundamental physical limits. The rich variety of living structures that are all based on different combinations of a few molecular building blocks, i.e., amino acids, proves the usefulness and robustness of the bottom-up approach for producing

complex structures. However, we are only beginning to understand how the mechanisms of molecular recognition and self-assembly could be exploited for actual device production. In order to investigate the molecular self-organization, suitable model systems need to be found that allow studying the molecular interactions reproducibly and with high accuracy. Surface adsorbed molecules are an obvious choice. They are accessible to sophisticated surface analysis tools such as scanning tunneling microscopy (STM) as well as electron diffraction techniques, infrared and other optical spectroscopies. Often, graphite or metal surfaces are used as substrates, because in these cases the substrate induced perturbations of the molecular properties are minimal and the molecule seems to swim freely on the substrate, as illustrated in Figure 1 for the case of adenine adsorbed on graphite(0001).

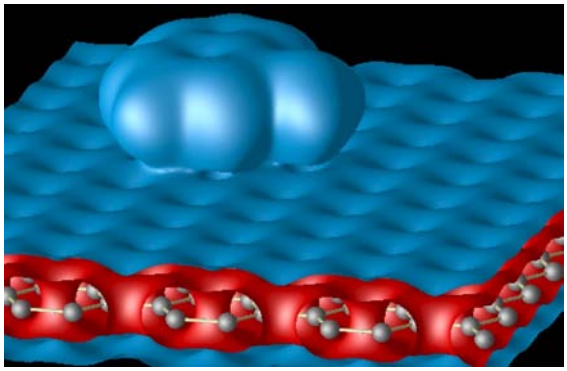


Figure 1: Isosurfaces of the calculated total valence charge density of adenine adsorbed on graphite (0001). The positions of the carbon atoms in the uppermost graphene sheet are indicated. From Ref. [1].

Our work aims at identifying the driving forces that lead to phenomena of molecular self-organzation and self-assembly of surface adsorbed molecules. To this end, we perform numerical simulations that describe the molecule-molecule and molecule-substrate interaction fully quantum mechanically. Due to the high number of degrees of freedom (typical calculations deal with hundreds of atoms, i.e, thousands of electrons) both numerically efficient algorithms and powerful parallel computers are needed.

Problem details and work done

In the last year, we focused in particular on the long-range chiral (enantiomeric) recognition that was discovered experimentally (but not at all understood!) in the behavior of adenine and phenylglycine adsorbed on Cu(110) [2]. In general, enantiomeric interactions are explained within the "three-point" contact model, shown schematically in Figure 2. In this model, developed by Easson and Stedman in the thirties of the last century already, stereochemical differences in reactivity are due to the differential bonding of enantiomers with three nonequivalent binding sites.

Discrimination occurs when one isomer can simultaneously interact with all of the sites, while its enantiomorph cannot.

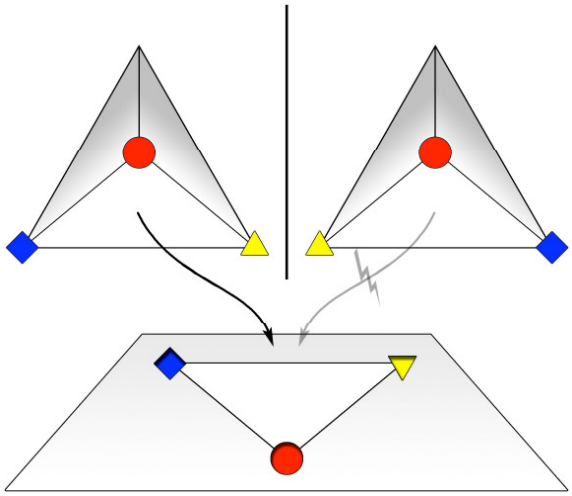


Figure 2: Schematic illustration of the "three-point" contact model for chiral discimination: The molecule on the left matches the three nonequivalent binding sites, in contrast to its mirror-imaged enantiomorph on the right.

The enantiomeric interactions found to govern the co-adsorption of adenine and phenylglycine on Cu(110) [2] are remarkable not only because the interplay of nucleic acid bases and amino acids is of fundamental importance for many biological processes. It also presents the first direct observation of diastereoisomeric interactions due to chiral recognition between dissimilar molecules. Moreover, the chiral discrimination acts at a distance of up to 2 nm, i.e., seems to evade the explanation within the Easson and Stedman model of chiral recognition. Based on first-principles calculations we succeeded to explain the experimental findings.

Let us start by briefly summarizing the experimental findings. Chen and Richardson [2] observed that adenine deposited on Cu(110) at room temperature forms ordered one-dimensional molecular dimer chains that grow along the lateral [+-1,2] directions (given with respect to the [-110] and [001] Cu crystal orientations, see Figure 3). Co-adsorbed phenylglycine shows a strong chiral preference in its interaction with these chains: S-phenylglycine attaches to [1,2] oriented chains, whereas R-phenylglycine decorates chains aligned along [-1,2]. The STM images show double rows of phenylglycine molecules that run parallel to the adenine dimer chains. The microscopic interpretation of the [1,2] oriented chain structure is shown in Figure 3.

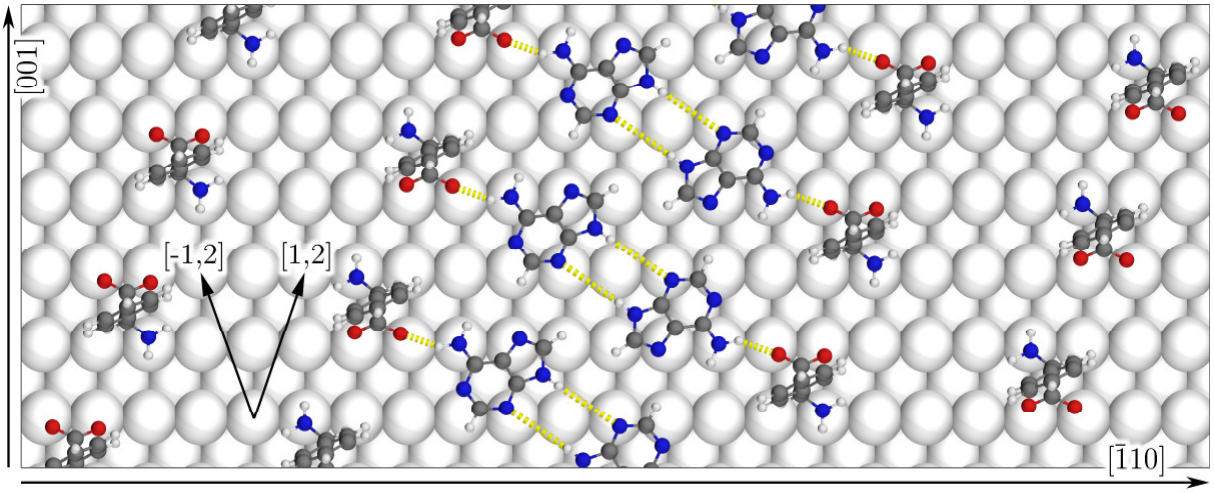


Figure 3: Molecular model derived in Ref. 2 for phenylglycine co-adsorbed with adenine forming dimer rows along the [1,2] direction on Cu(110). Hydrogen bonds are indicated with yellow lines.

We started with a detailed analysis of the substrate-molecule interaction for Cu(110) adsorbed phenylglycine [3,4] and adenine [5] molecules. The adenine-Cu(110) interaction is governed by mutual
polarization and Coulomb attraction. The adsorption of phenylglycine on Cu(110), on the other hand, leads to covalent bonding. Despite this difference, the adsorption characteristics of adenine and phenylglycine share one feature that is important in the present context: The energy barriers that hinder lateral movements of the molecules on the Cu(110) surface substrate are considerable, up to 0.5 and 1.0 eV for adenine and phenylglycine, respectively. In the present case, the lateral constraints are strengthened by hydrogen bonds between the carboxyl group of the first-row phenylglycine molecules and the adenine amino group, as well as within the adenine dimers, see Figure 3. Our calculations for the co-adsorbed species phenylglycine and adenine [6] revealed a long-range electrostatic interaction that is enabled by substrate-molecule charge transfer. By decomposing the amino acid in smaller building blocks we find the Coulomb repulsion between the phenylglycine amino-group and the DNA base to be responsible for the enantiospecific adsorption, see Figure 4. The substrate-mediated charge transfer acts in fact as a chiral selector, while the direct intermolecular interactions such as hydrogen bonds do not. The calculations show (i) that electrostatic forces acting over large distances can constitute at least one of the interactions in the "three-point" contact model for enantioselectivity and (ii) that the substrate may in fact catalyze molecular recognition and self-assembly.

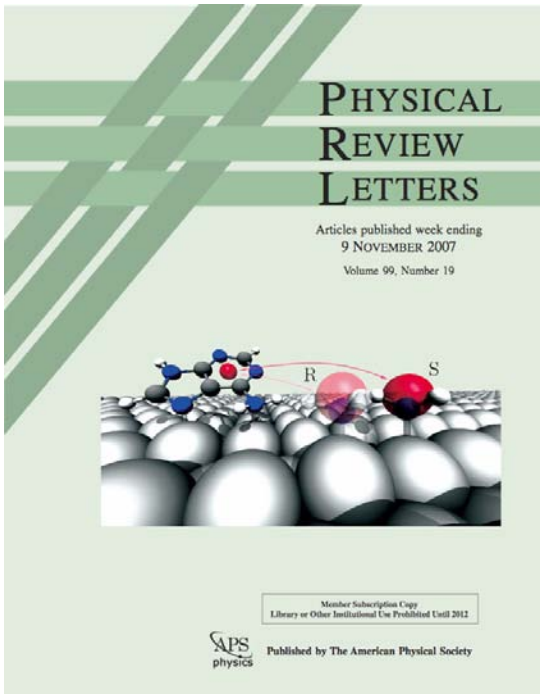


Figure 4: Schematic illustration of the Coulomb interaction due to electron transfer from the substrate to the admolecules. For clearer presentation only one pair of charges is shown for ammonia in S- (full color) and R-phenylglycine configuration (shaded color). Our article [6] was published in the prestigious Physical Review Letters and was

Resource Usage

The calculations were done using grants of computer time from the Paderborn Center for Parallel Computing PC^2 (ARMINIUS: FSC Xeon-DP Cluster and PLING: HP rx2600 Itanium Cluster) and the Höchstleistungs-Rechenzentrum Stuttgart (NEC SX-8).

References

[1] Schmidt, W.G.; Seino, K.; Preuss, M.; Hermann, A.; Ortmann, F. and Bechstedt, F.: “Organic molecule adsorption on solid surfaces: chemical bonding, mutual polarisation and dispersion interaction“, Appl. Phys. A 85, 387 (2006).

[2] Chen, Q. and Richardson, N.V.: “Enantiomeric interactions between nucleic acid bases and amino acids on solid surfaces“, Nature Mater. 2, 324 (2003).

[3] Blankenburg, S. and Schmidt, W.G.:“Steric effects and chirality in the adsorption of glycine and phenylglycine on Cu(110)“ Nanotechnology 18, 424030 (2007).

[4] Blankenburg, S. and Schmidt, W.G.:“Adsorption of phenylglycine on copper: Density functional calculations“, Phys. Rev. B 74, 155419 (2006).

[5] Preuss, M.; Schmidt, W.G. and Bechstedt, F.:“Coulombic Amino Group-Metal Bonding: Adsorption of Adenine on Cu(110)“, Phys. Rev. Lett 94, 236102 (2005); selected for Virtual Journal of Nanoscale Science & Technology 11(25) (2005).

[6] Blankenburg, S. and Schmidt, W.G.:“Long-Range Chiral Recognition due to Substrate Locking and Substrate-Adsorbate Charge Transfer“,Phys. Rev. Lett. 99, 196107 (2007); selected for Virtual Journal of Nanoscale Science & Technology 16(21) (2007)

5.4.8 Computational Surface Chemistry

Project coordinator	Prof. Guido Grundmeier, Department for technical and macromolecular chemistry, University of Paderborn
Project members	Dipl.-Ing. Markus Valtiner, Max-Planck-Institut für Eisenforschung, Düsseldorf

General Problem Description

Almost all functional metals, which are of high industrial and technological importance, are covered by a native oxide film. The stability of these oxide films determines the performance of materials as these thin films passivate and thereby protect the surface from further oxidation. Moreover, molecular adhesion forces at interfaces between organic coatings or adhesives and engineering metals are governed by physisorption or chemisorption of macromolecules or adhesion promoting additives on these passive film surfaces[1]. In this context the stability of passivating oxides is of great interest for an effective corrosion protection[2]; an understanding of the surface chemistry on a molecular level is important for polymer adhesion on oxide covered metal substrates[3]. In particular, as zinc is a widely used metallic coating for steel, ZnO films are of extremely high technical as well as commercial interest. Consequently, it is crucial to understand the complex relations of the adhesion chemistry on ZnO surfaces on a molecular level by combining experimental as well as theoretical investigations. A combined approach of theory and experiment will allow a much more aimed future development of adhesives and conversion treatments for corrosion protection.

Towards a fundamental understanding, adhesion has to be investigated experimentally on the different single crystal surfaces of ZnO, which allows an evaluation of the data in terms of the crystallographic/atomic structure of the respective surfaces. Moreover, the surfaces used in experiments must be well-ordered and even single crystalline under atmospheric conditions. This particularly important aspect, i.e using single crystalline substrates, allows experimental studies, which are in close agreement with the inevitable idealisations of ab-initio simulation approaches. A combined approach – or even a bridging – of theory and experiment is extremely important for a detailed understanding of interfacial processes on a molecular level. Within the experimental course of our work we have developed several methods to prepare single crystalline ZnO(0001) substrates for studying the adhesion and stability of zinc oxides. Currently adhesion studies are experimentally performed by pulling single molecule strands with different chemical functionalities

(e.g. $-PO(OH)_2$, -COOH, $-NH_2$, -OH) off the respective single crystalline surfaces by means of AFM force-distance spectroscopy. This approach enables experimental setups, which are in very close agreement with the idealisations of materials modelling. These experimental results will be related and compared to density functional theory (DFT) based theoretical results. Within the project "Computational Surface Chemistry" the respective theoretical adhesion studies are performed on the ARMINIUS cluster at the University of Paderborn.

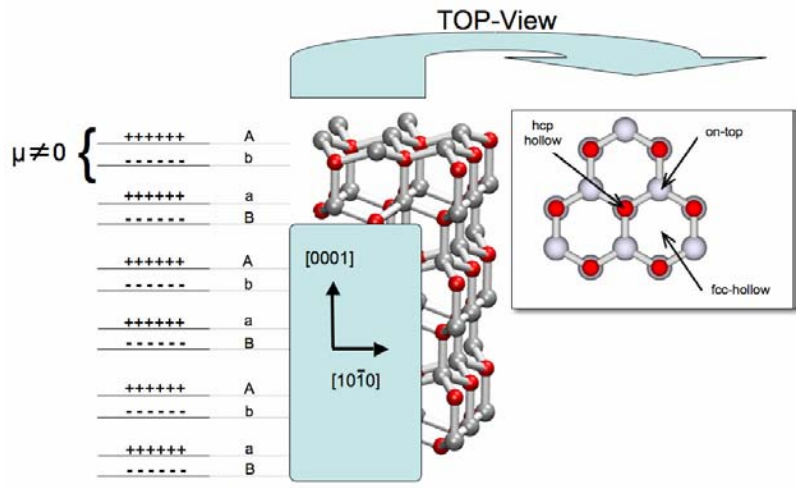
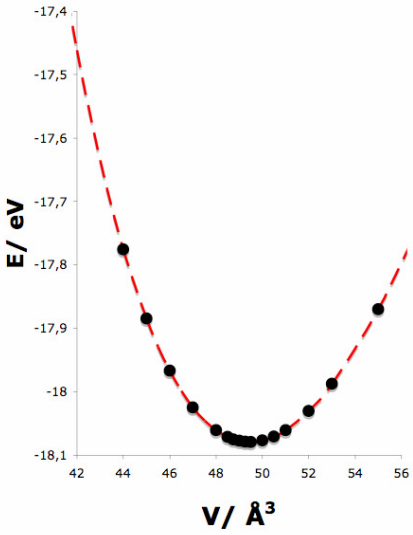


Figure 1: Within this figure the model of the ZnO is depicted. As can be seen clearly the (0001) surface consists of a hexagonal closed packed layer of Zn2+ atoms, which leads to a dipole moment normal to the surface. In contrast all the other low index surfaces like e.g. the 1010 surface have a stochiometric number of Zn:O = 1:1 at the surface.

Table 1: Computed and experimental values of the structural parameters for bulk ZnO, a and c are the lattice vectors of the hexagonal wurtzite structure. B_0 is the bulk modulus as obtained from fitting the Energy versus Volume plot to a 3rd order Birch-Murnaghan equation of state.

	PW91	**Expt.**
a(Å)	3.281	3.25
c(Å)	5.297	5.207
c/a	1.614	1.602
B_0(GPa)	131	143

Figure 2: Calculated Energy versus Volume Curve (Points) and a fit to the 3rd order Birch-Murnaghan equation of state (Curve).



Problem details and work done

The herein investigated single crystal ZnO(0001)-Zn surfaces have been extensively studied under UHV conditions within the last decades[4-6]. In contrast the literature on well-defined and crystalline ZnO(0001)-Zn surfaces under ambient conditions and in electrolyte solutions is not very comprehensive . In principle the ZnO(0001)-Zn surface were expected to be unstable due to the polar nature. Along the c-direction the ZnO crystal consists of an alternating sequence of hexagonal closed packed layers of Zn^{2+} and O^{2-} ions (see Figure 1). This leads to a dipole moment perpendicular to these planes. Within the bulk these dipole moments completely compensate each other. A perfect surface cut normal to the c-axis of the ZnO leads to ZnO(0001)-Zn surfaces on one side of the slab and ZnO(0001)-O surfaces on the other side of the slab (see Figure 1) and the dipole moment on the surface can not be compensated. The resulting polar surfaces are called tasker-type 3 surfaces and were expected to be unstable due to the dipole moment normal to the surface[7]; consequently reconstruction was expected. The interesting feature of Zn-terminated ZnO(0001)-Zn surfaces was that no reconstruction has been observed, which was discussed quite controversially in the literature. Arguing with an ionic model it is clear that the perfect Zn-terminated surface itself would be positively charged due to the nature of the polar ZnO(0001)-Zn surface. The surface Zn^{2+} ions have three oxide ions as nearest neighbours. Thus, formally ½ e^- per unit cell area is necessary to compensate the positive charge and quench the dipole moment normal to the surface. (1) Electron transfer from the O-terminated to the Zn-terminated side, (2) formation of ¼ ML Zn- and O-vacancies on the respective side as well as (3) adsorption of ½ ML of hydrogen and hydroxides on the respective sides were discussed as potential stabilisation mechanisms. Experimentally, stabilisation by the formation of Zn-vacancies[4, 5] via a high density of steps (oxygen terminated) of UHV prepared surfaces and hydroxide stabilisation[8] of wet chemically prepared surfaces (alkaline etching) were proven to be effective stabilisation mechanisms for ZnO(0001)-Zn surfaces.

Within the last year the group of Prof. Grundmeier started to investigate the polar ZnO(0001) surfaces also with density functional approaches as a complementary tool to the experimental approaches. Within the work the VASP code is used. To calculate the DFT exchange and correlation contributions to the total energy, we have used the Generalized Gradient Approximation (GGA) as parameterised by Perdew and and Wang — also known as PW91. The interaction between valence and core electrons is described by projector augmented plane waves as supplied with the VASP code The Kohn–Sham equations are solved using the block Davidson scheme and the optimization of the atomic geometry is performed by means of a conjugate-gradient scheme was applied. Within the super cell approach of VASP, the Brillouin zone sampling was performed using a Monkhorst-Pack grid with the

Table 2: Calculated total energies of the ZnO(0001) surface covered with half a monolayer of hydroxides. The fcc-hollow turned out to be favourable in energy.

Adsorption-site	**Coverage**	**Total-Energy [eV]**
on-top	1/2 ML	-318.37
hcp-hollow	1/2 ML	-318.40
fcc-hollow	1/2 ML	-319.45

gamma–point included. An electronic smearing as given by the Fermi-Dirac distribution was applied with 0.1 eV broadening.

As a first step in every surface caluclation the bulk properties of the respective material, i.e. ZnO in this project have to be evaluated using the DFT appoach of choice. The optimized bulk parameters for the ZnO can be seen in Table 1. In Figure 2 the Energy versus Volme curve is plotted. The parameters agree quite well with the experimental data.

In the current stage of the project, the ZnO(0001) surface are simulated. To describe the surface, the so-called supercell approach is employed. As already discussed, the ZnO(0001) surfaces need to be stabilized in some way. As we are working experimentally in electrolytes, the dominating stabilization mechanism for the ZnO(0001) surface is the adsorption of ½ monolayer of hydroxides from the solution. As depicted in Figure 1, there are several possible high symmetry adsorption sites for the OH on the surface – atop of a surface Zn-atom, at the hcp-site or at the fcc-site. We calculated all the possible adsorption sites. The fcc-hollow turned out to be favourable in Energy as can be seen in Table 2. Therefore it is assumed that this is the structure, which is closest to the real experimental surface structure. Consequently this surface structure is now used to evaluate adsorption energies.

Currently and within the next year, the project is focusing on studying the adsorption of various functional groups, like amines, carbonates, phosphonates …, on the ZnO(0001)-Zn-OH surface. A comparison with the experimental data will allow on the one hand judging the theoretical results. On the other hand the synergistic combination of theory and experiments will allow getting a deeper inside into the adsorption mechanisms on a molecular basis.

Resource Usage

All calculations were performed on the ARMINIUS cluster with the VASP abinito simulation code. For the VASP calculations, depending on system size (i.e. number of atoms, size of Fourier grid, k-Point sampling, basis set etc…) calculations were performed with 6-10 processors. As the project is in the very first steps, the system is currently used approximately on an "every two weeks" basis, but usage on a daily basis is planned. For the next reporting period a massive use of the system with up to 32 processors is planned for a few large systems.

References

[1] Grundmeier, G.; Schmidt, W.; Stratmann, M., Corrosion protection by organic coatings: electrochemical mechanism and novel methods of investigation. Electrochimica Acta 2000, 45, (15-16), 2515-2533.

[2] Grundmeier, G.; Reinartz, C.; Rohwerder, M.; Stratmann, M., Corrosion properties of chemically modified metal surfaces. Electrochimica Acta 1998, 43, (1-2), 165-174.

[3] Grundmeier, G.; Stratmann, M., Adhesion and De-adhesion mechanisms at polymer/metal interfaces: Mechanistic understanding based on in situ studies of buried interfaces. Annual Review of Materials Research 2005, 35, 571-615.

[4] Dulub, O.; Diebold, U.; Kresse, G., Novel stabilization mechanism on polar surfaces: ZnO(0001)-Zn. Physical Review Letters 2003, 90, (1).

[5] Dulub, O.; Boatner, L. A.; Diebold, U., STM study of the geometric and electronic structure of ZnO(0001)-Zn, (000(1)over-bar)-O, (10(1)over-bar0), and (11(2)over-bar0) surfaces. Surface Science 2002, 519, (3), 201-217.

[6] Woll, C., The chemistry and physics of zinc oxide surfaces. Progress in Surface Science 2007, 82, (2-3), 55-120.

[7] Tasker, P. W., Stability of Ionic-Crystal Surfaces. Journal of Physics C-Solid State Physics 1979, 12, (22), 4977-4984.

[8] Valtiner, M.; Borodin, S.; Grundmeier, G., Preparation and characterisation of hydroxide stabilised ZnO(0001)–Zn–OH surfaces. Physical Chemistry Chemical Physics 2007, 9, (19), 2297-2436.

5.4.9 Novel Simulation Methods for Electro-Hydrodynamics

Project coordinator	Prof. Dr. Friederike Schmid, University of Bielefeld
Project members	Jens Smiatek, University of Bielefeld
Supported by	Volkswagen Stiftung

General Problem Description

We investigate the dynamics of charged macromolecules in salt solutions, taking full account of the electrostatic and the hydrodynamic interactions. We attempt to quantify the conditions under which hydrodynamic interactions and image charge effects play a significant role. This shall then be applied to selected problems related to electrophoresis in microfluidic systems and polyelectrolytes in free solution. A special focus thereby lies on the hydrodynamic screening. Furthermore it is planned to investigate polyelectrolyte transport phenomena through structured microchannels in general.

In the last years much effort has been spent on the development of new microfluidic devices. Recent success in reaching the nanometer scale has stimulated rising interest in fluid mechanics on submicrometer scales. Coarse-grained computer simulations are powerful theoretical tools for investigating the dynamical properties of such confined flows.

A main interest in modern biological physics lies in developing special separation devices for polyelectrolyes like DNA or biological matter in general. Applying an external electric force on polyelectrolytes unfortunately does not lead to size separation due to electrohydrodynamic screening [1]. Therefore special separation devices like micro-structured channels had to be invented to overcome this situation. Steric confinement is one of the possibilities which is able to lead to size-dependent separation by different trapping mechanisms.

The physical mechanism of trapping can be understood with respect to steric hindrance and a loss of free energy due to unfolding of the polyelectrolyte [2]. The mean radius of gyration, which is defined as the average distance of one monomer to the center of mass regulates the size of these separation devices which lies in the range of micro- to nanometers. Computer simulations in the last years explored the physical effects behind these separation mechanisms. Due to electrohydrodynamic screening the underlying hydrodynamics had not to be considered in these simulations. The effects of the solvent are purely described in a phenomenological approach corresponding to Brownian- or stochastic dynamics.

We expect that the explicit consideration of hydrodynamics and electrostatics will give rise to a range of new phenomena.

Problem details and work done

We plan to investigate electrohydrodynamic effects like screening and its influence on the dynamical behaviour of charged polymers in solution in general. Despite the first pioneering work by Smoluchowski, Debye and Hückel some aspects are still yet not fully understood. To our knowledge a thorough treatment of hydrodynamic and electrostatic interactions in mesoscopic coarse-grained simulations is still missing. Our goal is to fill this gap and to investigate those new fascinating aspects.

For this we use the method of Dissipative Particle Dynamics [3] to investigate and to simulate the underlying hydrodynamics.

A main focus lies thereby in exploring new separation techniques which are only realisable by a much more detailled knowledge of the underlying physical mechanisms. Focussing on polymers in microchannels, which are modelled by a simple bead-spring model with excluded volume, several additional physical effects have to be regarded like the electroosmotic flow. The electroosmotic flow appears if a surface comes in contact with liquid. Most materials get charged by the ionisation or dissociation of surface groups or by the adsorption of ions from solution onto a previously uncharged surface. By applying an external electric force on the counterions in the solution, a solvent flow can be induced [4].

Combining these surface effects with the underlying hydrodynamic and electrostatic effects would be a challenging task for a better understanding of polymer separation dynamics in general.

Therefore we developed a novel way of implemeting partial-slip boundary conditions in mesoscopic simulations methods [5]. This algorithm allows to simulate fluid flow behaviour at surfaces where the fluid velocity does not totally vanish in contrast to the famous no-slip boundary condition. Within our approach it is possible to apply a tunable partial-slip condition which means that the fluid velocity does not totally vanish at the postion of the hydrodynamic boundaries. This flow behaviour is mainly characterised by an effective parameter which is called “slip length”. Large slip lengths do mean larger velocities at the hydrodynamic boundaries. We apply this method to the special case of counterion-induced electroosmotic flow which is shown in Illustration 1. The special case of counterion distribution in a microchannel can be calculated analytically by the well-known Poisson-Boltzmann-Theory. The results for different slip lengths are shown in Illustration 2.

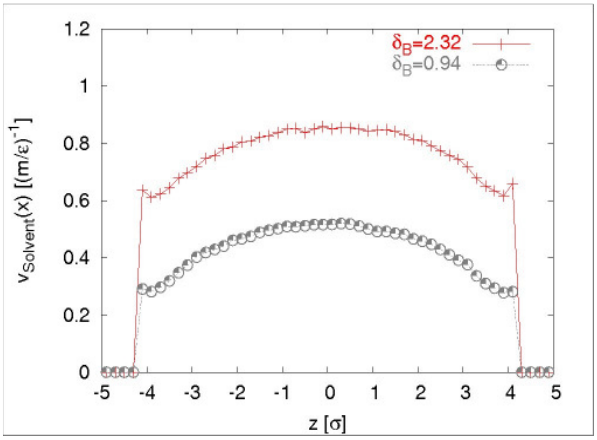


Illustration 1: Counterion-induced electroosmotic flow of the solvent for different slip lenghts.

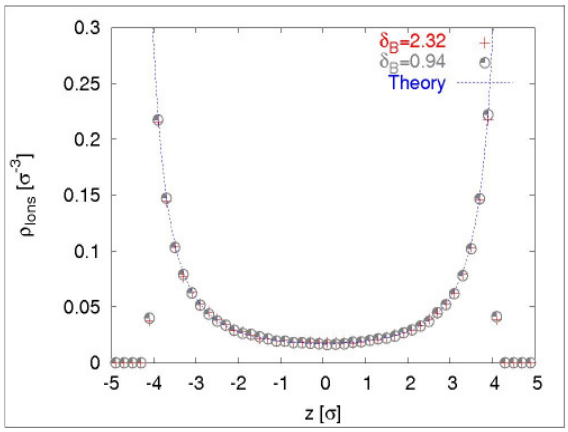


Illustration 2: Counterion distribution in a channel with charged walls for different slip lengths in comparison to Poisson-Boltzmann-theory.

As the results indicate the Poisson-Boltzmann-Theory describes the simulation results correctly and the counterion distribution is not disturbed by an electric force applied in transverse direction. Only the solvent flow is heavily influenced by different slip lengths.

After modeling the boundaries of the channel it is planned to simulate charged polymers in microchannels. The monomers are represented as coarse-grained beads which are connected by stiff springs. All particles are modeled by Dissipative Particle Dynamics [3] in a pure stochastic approach which means that any kind of conservative interactions are neglected.

In contrast to the solvent a conservative Lennard-Jones force acts on the monomers of the polyelectrolyte which mimics the excluded-volume effect. The monomers are connected to each other by the well-known FENE-Potential to avoid entanglement effects. All monomers carry a net charge of -1. The counterions and the salt ions are also simulated by Dissipative Particle Dynamics in combination with a purely repulsive Lennard-Jones-force which avoids a collapse of differently charged particles. The overall boundary condition is charge neutrality. Much of the recently published work explicitly avoids treatment of the salt. The main task of these studies is to investigate the electrophoretic mobility of polyelectrolytes in external electric fields in the pure counterion case.

In contrast to that we want to focus on the underlying mechanisms of hydrodynamic screening in mesoscopic simulation models in general. Different length scales are given if salt or only counterions are present. The Gouy-Chapman length explicitly describes the screening length in the case if only counterions are present whereas the Debye-Hückel-length for the salt case is mentioned in the literature as the crucial parameter for the screening length. Therefore it is planned to involve the

polyelectrolyte, the counterions and the salt in our mesoscopic simulations to achieve a complete understanding of the underlying physical effects.
Within the topic of electrophoresis in microchannels a collaboration project is planned with the "Frankfurt Institute of Advanced Studies" (FIAS) in which we want to reproduce some recently published experimental results [6]. Applying the partial slip boundary conditions and the results of the electroosmotic flow together with the polyelectrolyte in solution in the salt case, the verificaton of the experiment is a challenging task to test our simulation methods. By combining surface charge effects together with electrophoresis in small channels an unusual migration behaviour can be observed. Further effects to investigate are the role of hydrodynamics in specially structured microchannels. A lot of previous work was done and reported in [7,8].
To summarise the simulations in [7,8] had been carried out by a stochastic dynamics approach where hydrodynamic effects were neglected due to screening. Screening appears in salty polyelectrolyte solutions when an external electric force is applied. The electrophoretic mobility is then dependent on the number of beads instead of the hydrodynamic radius if no screening occurs. This Rouse regime can be simulated by a Langevin simulation and is well explained by a local force picture. Focussing on the mechanisms of separation in structured microchannels, the simulations reproduced the experimental results in good agreement.
Another point worthwile to mention is that the microdevices get more and more miniaturised which means that the above mentioned approach is no longer valid. This means that the Langevin simulation technique is not applicable to reproduce experimental results in the nanometer scale because of its phenomenological character. On the submicrometer scale the solvent has to be taken explicitly into account to reproduce the experiments which justifies our approach. As it was mentioned entropic trapping and steric hindrance leads to size-dependent separation of polymers. By explicit consideration of solvent effects the underlying separation mechanisms can be simulated in more detail.
Investigating these effects with ions and present hydrodynamic interactions as in our case, will lead to a more detailed prediction in the application of efficient separation devices in general.

Resource Usage

We run our simulations on the Fujitsu-Siemens Computers hpcLine-Cluster “ARMINIUS” with the open source software package “ESPReSo” (An **E**xtensible **S**imulation **P**ackage for **Re**search on **So**ft Matter) [9,10] which is fully parallelised. The parallelisation is MPI-based and we use the the computing cluster daily because of the immense accelaration of computing time by running the simulations in parallel.

References

[1] J.-L. Viovy, Rev. Mod. Phys. **72,** (2000), 813
[2] M. Streek, F. Schmid, T. T. Duong, A. Ros, Phys. Rev. E **71**, (2005), 011905
[3] P. J. Hoogerburgge, J. M. Koelman, Europhys. Lett. **19,** (1992), 155
[4] J. Smiatek *et al.,* Poster presented at the International Soft Matter Conference, Aachen, October 2007
[5] J. Smiatek, M. P. Allen, F. Schmid, "Tunable slip boundaries for coarse-grained simulations of fluid flow", submitted to European Physical Journal (2007)
[6] J. Mathe, J.-M. Di Meglio, B. Tinland, arXiv:cond-mat/0610328v2 [cond-mat.soft], submitted to JCIS
[7] M. Streek, F. Schmid, T. T. Duong, A. Ros, J. Biotechnology **112,** (2004), 79
[8] T. T. Duong, M. Streek, R. Ros, F. Schmid, J. Brugger, D. Anselmetti and A. Ros, Microelectronic Engineering, **67-68,** (2003), 906
[9] ESPResSo-Homepage, http://www.espresso.mpg.de (2002-2007)
[10] A. Arnold, B. A. Mann, H.-J. Limbach and C. Holm, Comp. Phys. Comm. **174,** (2005), 704

5.4.10 Lipid-mediated protein interactions in lipid bilayers

Project coordinator	Prof. Dr. Friederike Schmid, Department of Physics, University of Bielefeld
Project members	Dipl.-Phys. Beate West, Department of Physics, University of Bielefeld
Supported by	DFG (Sonderforschungsbereich 613)

General Problem Description

Biomembranes play a central role in all biological cells. The biomembrane consists of a bilayer of amphiphile lipids, into which membrane proteins and other macromolecules are inserted. These proteins, like for example receptors, enzymes and ion channels, are the biochemically active components. The lipid protein composition of a bilayer has a fundamental effect on its phase behavior and physical characteristics, e.g. the flexibility and curvature. The proteins and lipids are not homogeneous distributed onto the membrane, but divide the membrane in domains with different protein and lipid fractions. Domains with a high concentration of certain lipids so-called rafts, interact with protein rich domains [4,12]. A biological function is assigned to these rafts; they locate or sort proteins.

The indirect interaction effects obtained by the lipids contribute significantly to the entire interaction between membrane proteins [3,6,7,10,11]. The direct electrostatic interactions in biologically relevant constellations are predominantly shielded by the aqueous environment of the membrane [2,15].

Apart from direct protein-protein interactions there are different factors that can induce indirect interactions [6]:

1) Hydrophobic mismatch will occur if the length of the hydrophobic part of the transmembrane protein doesn't fit to the thickness of the membrane. The membrane will contract or uncompress in the contact area of the protein.
2) Interactions between proteins can be induced if the proteins affect the local curvature of the membrane. For example the proteins bend the membrane and/or they change the bending rigidity of the membrane.
3) The proteins influence the lateral structure of the lipid membranes. Lipids that are near the proteins lose their freedom of translation or conformation.

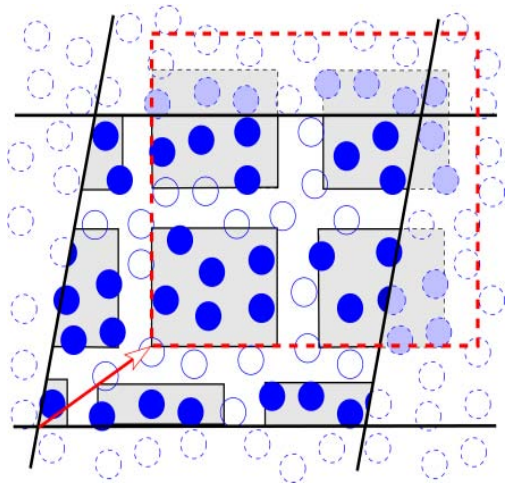
Problem details and work done

For the simulation of the lipid bilayer we use an idealized bead-spring model [5]. Each lipid consist of six tail beads and one slightly larger head bead. Adjacent beads in a lipid chain are bound together by a FENE-potential. Additionally, chains are given a bending stiffness by a bond-angle potential. Beads that are not connected with each other interact via a truncated and shifted Lennard-Jones potential. Head-head and head-tail interactions are purely repulsive and tail-tail interactions also have an attractive part.

The lipid bilayer is surrounded by a solvent environment [8,9]. This environment is represented by explicit solvent beads, which behave like unbounded head beads, except for the fact that they don't interact with each other.

The proteins are modelled as cylinders whose diameter correspond to that of an α-helix. They are free to move in the lipid bilayer. The proteins interact with each other and the lipids in the xy-plane via a Lennard-Jones kind of potential. The interaction between tail beads and proteins is repulsive and attractive. All other interactions are purely repulsive.

The system is simulated by using Monte-Carlo methods at constant pressure and temperature with periodic boundary conditions.



The program is parallelized using a geometrical decomposition scheme [13,14]. The idea is to define active regions (light blue areas). The distance between these regions corresponds to the maximum interaction range. Each processor gets one of this active regions. Only beads inside the active region will be moved during a Monte-Carlo move; moves that lead out of the active region will be rejected. The offset of the active regions (red arrow) is regularly moved randomly.

We simulate a lipid bilayer with two proteins in the liquid phase for different protein-tail interaction strengths ε_{pt}. From the pair distribution function g as a function of the protein-protein distance inferences can be made regarding the protein-protein interaction:

$$w = k_B T \ln g$$

w is the effective pair potential, k_B is the Boltzmann constant and T is the temperature. The pair distribution function is obtained by an umbrella sampling method as a function of the protein-protein distance. Illustration 1 shows the pair distribution function for different protein-tail interaction strengths. When looking at the graphs, we see that the proteins prefer distances where a layer of one and more lipids are between them. The curves get lower with increasing interaction strength. With the protein-tail interaction strengths we can induce a hydrophobic mismatch effect. The stronger this interaction the larger this effect gets. This hydrophobic mismatch effect causes the drawdown of the curves.

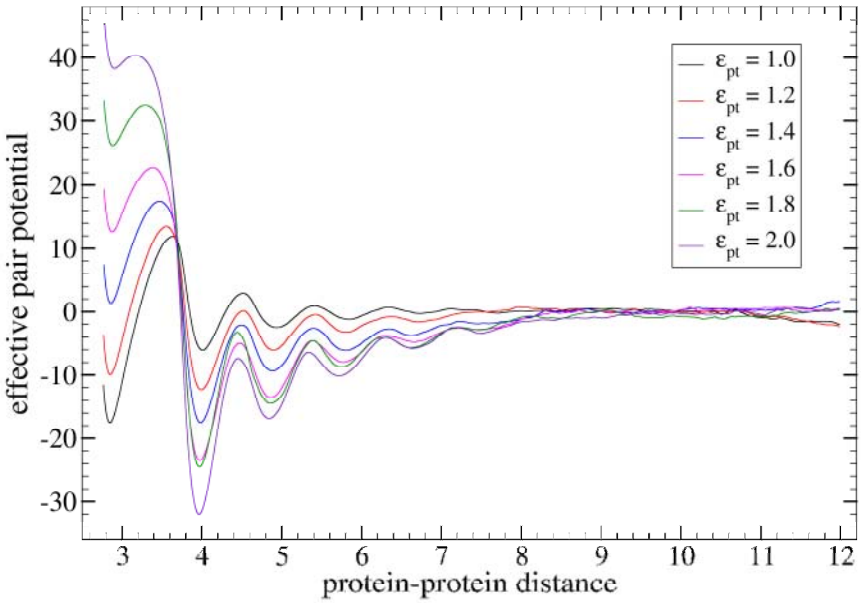


Illustration 3: Effective pair potential between two proteins in the fluid phase

At the moment we investigate the deformation profile of one protein in the fluid phase and compare it with the predictions of the elastic theory [1]. We are interested in the change of the free energy when a protein is inserted into the lipid bilayer. Therefore we use the method of thermodynamic integration: In different simulation runs the bead-protein interaction energy U_p is rescaled by a factor of $0 \le \alpha \le 1$. The free energy is the integral over α of the mean value of U_p.

Another point of interest is the effective pair potential between proteins in the gel phase. The problem here is that the lipids are tilted in this phase. Therefore we have to do an umbrella sampling depending on the protein-protein distance and the angle between the proteins and the average tilt orientation of the lipids.

Resource Usage

Our simulations run on the cluster **ARMINIUS**. Our program is parallelized with MPI using the scheme explained in the last section. We use the computing cluster daily because of the immense computing time that we need for our simulations. Depending on the system size we want to investigate we need between 2 and 4 nodes.

References

[1] H. Aranda-Espinoza, A. Berman, N. Dan, P. Pincus, S. Safran, *Interaction between Inclusions Embedded in Membranes*, Biophysical Journal **71**, 648, 1996

[2] N. Ben-Tal, B. Honig, *Helix-helix interactions in lipid bilayers*, Biophysical Journal **71**, 3046, 1996

[3] M. Bloom, E. Evans, O.G. Mouritsen, *Physical properties of the fluid lipid-bilayer component of cell membranes: a perspective*, Quart. Rev. Biophys. **24**, 293, 1991

[4] D.A. Brown, E. London, *Functions of lipid rafts in biological membranes*, A. Rev. Cell. Dev. Chem. **14**, 111, 1998

[5] D. Düchs, F. Schmid, Phase behavior of amphiphilic monolayers: Theory and simulations, J. Phys: Cond. Matt. **13**, 4853, 2001

[6] T. Gil, J.H. Ipsen, O.G. Mouritsen, M.C. Sabra, M.M. Speretto, M.J. Zuckermann, *Theoretical analysis of protein organization in lipid membranes*, Biochimica and Biophysica Acta **1376**, 245, 1998

[7] P. Lague, M.M. Zuckermann, B. Roux, *Protein inclusion in lipid membranes: A theory based on the hypernetted chain integral equation*, Faraday Disc. **111**, 165, 1998

[8] O. Lenz, F. Schmid, *Structure of Symmetric and Asymmetric Ripple Phases in Lipid Bilayers*, Phys. Rev. Lett. **98**, 058104, 2007

[9] O. Lenz, F. Schmid: *A simple computer model for liquid lipid bilayers,* J. Mol. Liquids **117(1-3)**, 147, 2005

[10] S. May, *Theories on structural perturbations of lipid bilayers*, Curr. Opn. Coll. Interf. Sci. **5**, 244, 2000

[11] O.G. Mouritsen, M. Bloom, *Models of lipid-protein interactions in membranes*, Quarterly Reviews of Biophysics: Biomol. Struct. **22**, 145, 1993

[12] Rietveld, K. Simons, *The differential miscibility of lipids as the basis for the formation of functional membrane rafts*, Biochimica and Biophysica Acta **1376**, 467, 1998

[13] F. Schmid, D. Düchs, O. Lenz, B. West, *A generic model for lipid monolayers, bilayers, and membranes*, Computer Physics Communications **177(1-2)**, 168, 2007

[14] Uhlherr, S.J. Leak, N.E. Adam, P.E. Nyberg, M. Doxastakis, V.G. Mavrantzas, D.N. Theodorou, *Large scale atomistic polymer simulations using Monte Carlo methods for parallel vector processors*, Computer Physics Communications **144(1)**, 1, 2002

[15] D. Walther, P. Kuzmin, E. Donath, *Brownian dynamics simulation of the lateral distribution of charged membrane components*, Eur. Biophys. J. **24**, 125, 1996

5.4.11 Dynamic Vesicle Formation of Amphiphilic Copolymers

Project coordinator	Prof. Dr. Friederike Schmid, Department of Physics, University of Bielefeld
Project members	Dr. Xuehao He, Department of Physics, University of Bielefeld
Supported by	Alexander von Humboldt Foundation

General Problem Description

Amphiphilic block copolymer in solution can form various morphologies of micelle, such as spherelike, rodlike or lamellar micelles. Among them, a specific type, named vesicle structures, are biologically significant due to their extensive applications in drug delivery systems and artificial cells.[1] In biology, vesicles store, transport, or digest cellular products and wastes. It is no doubt that deep understanding to formation of different vesicles will be helpful to find artificial method to control structures and realize the functionality of vesicle.

In the past, experimental methods and skills, such as time-resolved dynamic (DLS) light, small-angle neutron (SANS) and X-ray (SAXS) scattering were applied to study vesicle formation of the mixture of cationic and anionic surfactants. The mechanism of spontaneous formation of vesicle was analyzed by scattering data. Transmitting electron micrograph (TEM) can provide the direct image of morphology of aggregation but difficultly capture the dynamic process for fast nucleation and the evolution of intermediary aggregation. Basic problems, e.g., whether vesicle is thermodynamically stable structures and whether the formation of vesicle is controlled by thermodynamics, are still unsolved. Computer simulation provides an important way to peek insight of this process. Monte Carlo Simulations, Brownian Dynamic, Dissipative Particle Dynamics (DPD) and even Molecular Dynamics (MD) with atomistic details have been applied to investigate the vesicle formation of lipid or small surfactant molecules. The formation process was discovered that, first, small cluster and sphere micelle formed, then small micelle coalesced to small disk micelle (small bilayer structure), finally bilayer curved and closed to vesicle. The important intermediate state, disk micelle (small bilayer structure) was discovered. These experiments and simulations focused on vesicle formation of lipid molecules or small surfactant in water solution.

Compared to lipid or small surfactant molecules, amphiphlic block copolymer has longer chain. Generally, solvent-phobic part is composed of several hundred carbon units, and the length of solventphilic part is also up to several tens units (while hydrophobic part of lipid molecules consists is short chain up to 10-24 carbon units and hydrophilic part is only one large molecular group). Another difference is that the

molecular interaction of amphiphilic copolymer in selective solvent is available to tune at a wide range. The aggregation process is relatively slow due to low diffusion of longer chain and weak segregation. This system is very large and complex, which includes several ten thousand to million atoms. Particle models mentioned above (MD or DPD) are not suitable to study such system due to great computation time and simulation size for long chain structure and dilute system. An effective algorithm is vital to simulate the spontaneous vesicle formation of amphiphilic block copolymer and analyze thermodynamic stability of polymer vesicle structure.[2] The research of vesicle formation of amphiphilic copolymer can provide deep insight to the mechanism of self-assembly of macromolecular micelles.

Problem details and work done

A modified dynamic self-consistent field method (External Potential Dynamic, EPD) can be used to simulate the dynamic process of vesicle formation. Assume that one component amphiphilic diblock copolymer P, and solvent S are mixed in the system of volume V. Every copolymer chain has two parts, solvent-phobic block A and solventphilic block B. The volume fractions of segments A and B in system are f_{A} and f_{B}, and the volume fractions of copolymer and solvent in solution are f_{P} and f_{S}, respectively. In polymer self-consistent field theory,[3, 4] one considers the statistics of a copolymer chain in a set of effective chemical potential fields, ω_I, I represent block species A, or B. These chemical potential fields, which replace the actual interactions between different components, are conjugated to the segment density fields ϕ_I of block species *I*. Similarly, solvent molecules are considered to be in an effective chemical potential field ω_{S} that conjugates to the solvent density field ϕ_{S}. The free energy function (in units of $k_{\mathrm{B}}T$) of the system:

$$F = -f_S \ln(Q_S/V) - \frac{f_P}{N}\ln(Q_P/V) + \frac{1}{V}\int dr[\chi_{AB}\phi_A\phi_B + \chi_{AS}\phi_A\phi_S + \chi_{BS}\phi_B\phi_S - \omega_A\phi_A - \omega_B\phi_B - \omega_S\phi_S + \frac{\kappa_H}{2}(\phi_A + \phi_B + \phi_S - 1)^2] \quad (1)$$

where *N* is the length of copolymer chain in the unit of coarsed-grain (every unit is corresponding to several tens monomers), Q_{S} is the partition function of solvent in field ω_{S} and Q_{P} is the partition function of a single chain of diblock copolymer in field ω_{A} and ω_{B}. χ_{IJ} is the Flory-Huggins interaction parameter between species i and j to show molecular interaction. First two terms are the contributions of entropies from solvent and polymer, and last term comes from the contribution of compression energy, κ_{H} is the coefficient of compression energy. Self-consistent field theory included basic character of linear structure of long chain polymer, and its free energy contained contribution of entropy.

$$\frac{\partial \omega_I(r)}{\partial t} = -D_I \nabla^2 \left(\frac{\delta F}{\delta \phi_I} + \eta_I\right)$$
$$\langle \eta_I(r,t) \rangle = 0, \qquad (2)$$
$$\langle \eta_I(r,t)\eta_J(r',t') \rangle = \beta \delta_{IJ} \delta(t-t')\delta(r-r')$$

External potential dynamic (EPD) method developed from density dynamics of polymer dynamic self-consistent field theory (DSCFT) are applied to simulation the dynamic of vesicle formation.[5] Using EPD modal, a real thermodynamic process was revealed without increasing computation cost for complicated correlation coefficient in dynamic self-consistent field method. Meanwhile, nonlocal kinetic coupling, i.e., rouse dynamics of polymer, is inherently included in such framework (eq.2). η is random thermal potential.

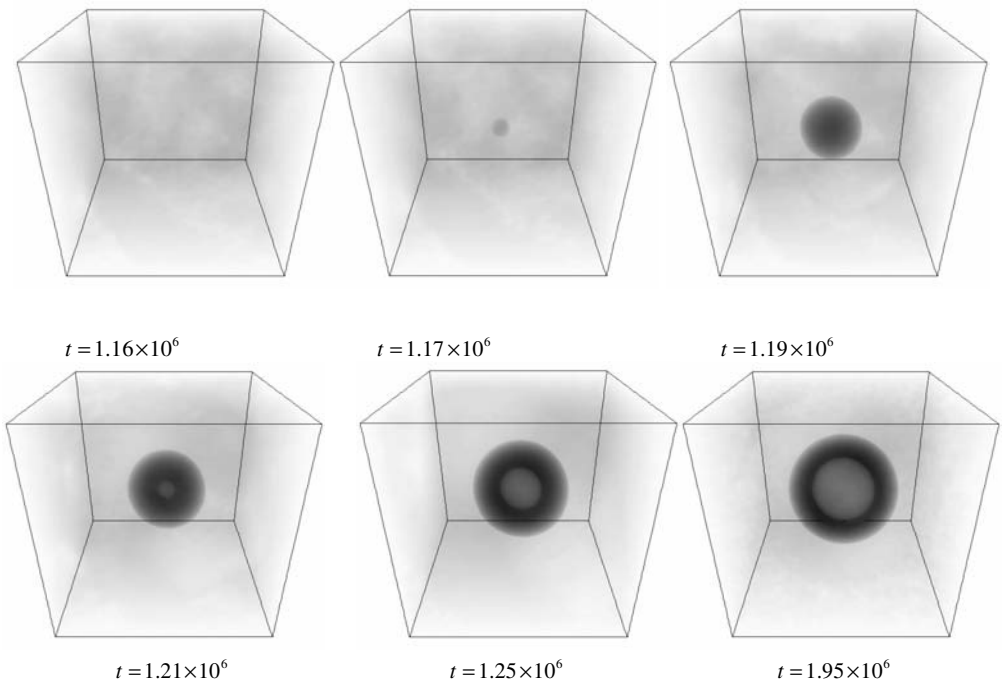


Figure 1: Aggregation morphology of amphiphilic diblock copolymers in dilute solution in 3D at various times. Only the density distribution of A is shown. The Flory-Huggins parameters are $\chi_{BS} = -0.45,\ \chi_{AS} = 1.2, \chi_{AB} = 1.05$,

It is discovered that when the system was quenched into the unstable two phase region inside the spinodal curve. A new pathway of spontaneous vesicle formation appears: First, spinodal decomposition sets in and the fluid acquires a weakly modulated structure. After an incubation time, the composition fluctuations of this background pattern trigger the nucleation of spherical micelles. In a third step, copolymers from the solution slowly aggregate to the micelles, they grow and

become semivesicles (bigger spherical micelles with a solventphilic core). Finally, solvent particles diffuse into the semivesicles, and they become full vesicles. It is discovered that the solventphilic parts of the copolymers play a crucial role for the transition from semivesicles to vesicles. This new mechanism of vesicle formation is very different with the conventional one proposed in the past.[6]

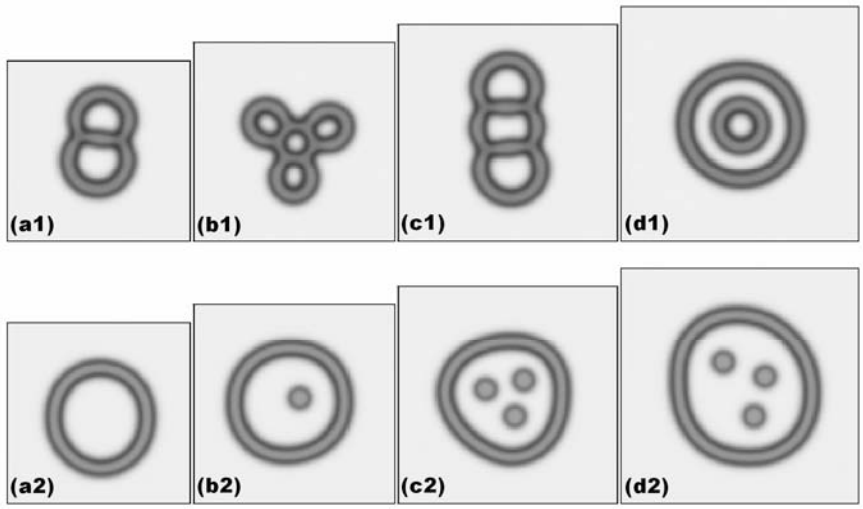


Figure 2: Manifold vesicles at the various initial number density of stable nucleus c_{ini} and Interaction parameters χ_{BS}, Black and grey stand for solventphilic and solvent-phobic regions. χ_{BS} = -0.3 (top group), χ_{BS} = -0.6 (bottom group): $c_{ini} = 9.0\times10^{-4} R_g^{-2}$ (a1, a2); $c_{ini} = 7.44\times10^{-4} R_g^{-2}$ (b1,b2); $c_{ini} = 6.25\times10^{-4} R_g^{-2}$ (c1 and c2); $c_{ini} = 5.33\times10^{-4} R_g^{-2}$ (d1,d2). Black and grey as in Figure 1

Based on this new mechanism of block polymer vesicle formation, a simple prenucleation method allows to control the size and structure of copolymeric vesicles.[7] Various vesicle morphologies such as twin or triplet vesicles, large compound vesicles, large vesicles containing entrapped smaller vesicle or spherical micelles can be obtained by varying the initial number density of stable nucleus c_{ini} and the solventphilicity of solvent-philic copolymer block, χ_{BS} (Figure 2). R_g is the mean gyration radius of polymer chains in the unperturbed state. The new technique provides a possible way to control the self-assembly of metastable vesicles of amphiphilic diblock copolymers in real applications.

In summary, a new mechanism of spontaneous vesicle formation was discovered and a prenucleation technique to control the vesicle self-assembly is proposed.[6, 7] Unilamellar vesicles were obtained at moderate nuclei densities c_{ini}, and their size can be related to c_{ini}. At low values of c_{ini}, a rich spectrum of large complex compound vesicles was discovered, including twin or triplet vesicles, entrapped vesicles and vesicle with embedded sphere micelles. The shapes and morphologies of the vesicles were shown to result from the dynamical competition of macro-aggregation and microphase separation, and they depend not only on the interaction parameters, but are also controlled by c_{ini}. The prenucleation method provides a new

way to control the size and structure of self-assembled vesicles, which should be useful for applications. The technique can easily be extended, e.g., by mixing colloidal particle to the solution, or adding non-sphere-like micelles as prenucleation germs.

Resource Usage

ARMINIUS Compute Cluster in PC^2 supercomputer of Paderborn University was used for this project. The program is parallel under the use of up to 12 nodes (24 processes) for every job at the most of cases. The program is compiled and runs in the environment of Scampi, Intel Compiler and FFTW2 package with MPI. The resources was used weekly based on the situation of work queue submitted and project research proceeding.

References

[1] D. E. Discher, A. Eisenberg, *Science* 297, 967, 2002;
[2] X. H. He, H. J. Liang, L. Huang, C. Y. Pan, *J. Phys. Chem. B* 108, 1731, 2004
[3] F. Schmid, *J. Phys. Cond. Matt.* 10, 8105, 1998.
[4] M. Müller, F. Schmid, *Adv. Polym. Sci*. 185, pp1-85, 2005.
[5] N. M. Mauris, J. G. E. M. Fraaije, *Macromolecules* 107, 5879, 1997;
[6] X. H. He, F. Schmid, *Macromolecules* 39, 2654, 2006.
[7] X. H. He, F. Schmid, *Macromolecules* 39, 8908, 2006

5.4.12 Numerical Simulation of Fluid Flow and Heat Transfer in Thermoplates

Project coordinator	Prof. Dr.-Ing. Jovan Mitrovic, University of Paderborn
Project members	Dipl.-Ing. Boban Maletic, University of Paderborn
Supported by	Thermal Process Engineering and Plant Technology, University of Paderborn

General Problem Description

As heat transfer devices, thermoplates are encountered in several branches of engineering practice, e.g. as condensers or evaporators in thermal process technology and cooling technique. In comparison to shell-and-tube heat exchanges, their installation is relatively simple, and the periphery is drastically reduced.

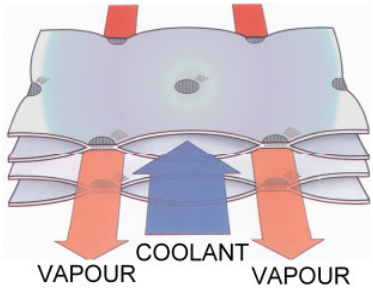


Figure 1: Fluid flow arrangement in thermoplates (DEG-Engineering).

A thermoplate consists of two metallic sheets, which are spot-welded over the whole surface according to an appropriate pattern, whereas the edges – except for connecting tubes – are continuously seam-welded. By applying a hydro-form technique, a channel having a complex shape is established between the sheets, as shown in Figure 1. One fluid is conducted through this channel, the other one through the channel created by two neighbouring thermoplates that are assembled in parallel at certain spacing thus making a thermoplate heat exchanger.

The objective of the numerical experiments is to numerically obtain the sets of the geometrical parameters that, in interaction with process parameters, should pave the way for optimal heat transfer of the inside fluid which is assumed to pass the thermoplate as a single phase (coolant in Figure 1). The geometry of the simulated three-dimensional domain is shown in Figure 2. It consists of a strip of a thermoplate channel, the latter being unbounded in spanwise (z) direction. The semi-circles represent the welding spots, which are arranged in a staggered manner

In the simulations, the geometrical parameters, such as the welding pattern, the

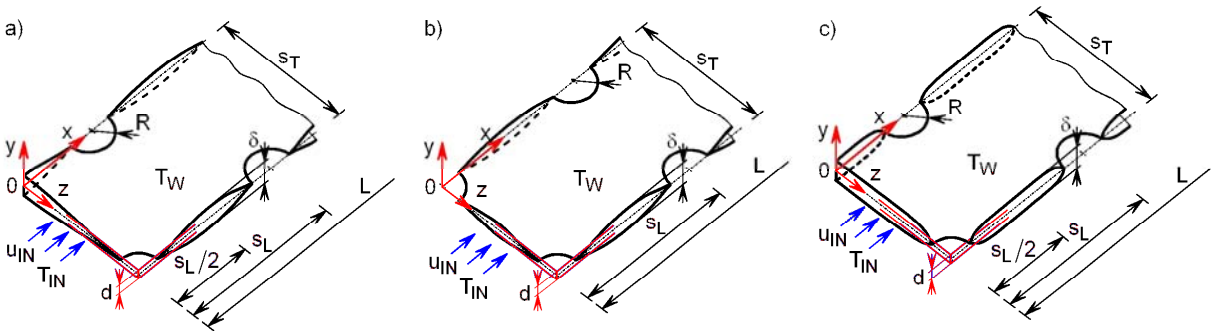


Figure 2: Geometry and dimensions of the three-dimensional simulated domains.

surface shape, the streamwise welding spots pitch, s_L, the transversal pitch, s_T, and the maximal distance between the metallic sheets, δ, have been varied, whereas the radius, R, of the welding spots and the channel length, L, were kept constant. To obtain some deeper insights into the flow and temperature fields in the developing region, a relatively short length, L, of the strip was chosen first. This length was then gradually increased up to $L = 500$ mm, and, at low fluid velocity, the fully developed region was included into the simulation domain. The Reynolds number, Re, that is, the fluid inlet velocity, u_{IN}, and the distance, δ, has also been varied. The fluid inlet temperature, T_{IN}, and the wall temperature, T_W, were taken to be constant.

The fluid flow was considered to be laminar, incompressible, at steady-state and three-dimensional. Water of constant physical properties is adopted for the numerical experiments. The velocity and the temperature fields are governed by the equations of continuity, momentum and energy.

In the numerical simulations, the commercial software StarCD was employed. The calculations have largely been performed on the PC^2 system for parallel computing.

Problem details and work done

The velocity field

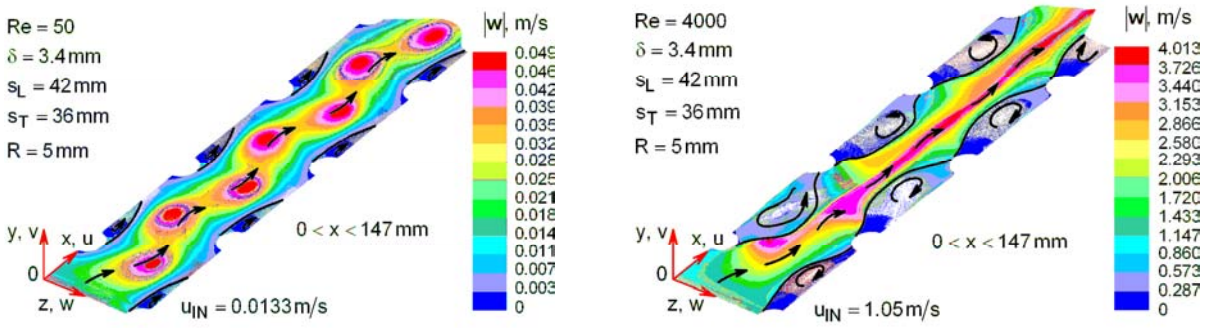


Figure 3: Velocity field in the plane $y = 0$ at $Re = 50$ and $Re = 4000$.

Figure 3 shows the velocity field in the x0z symmetry plane near to the channel inlet ($x < 147$ mm) at the smallest and at the largest Reynolds number Re adopted in the numerical experiments for a specific model whose parameter are also given in Figure 3. At the Reynolds number of $Re = 50$, the velocity field is largely smooth and the flow quiet with relatively narrow separation zones behind the welding spots. However, the velocity field is extremely heterogeneous, revealing a strong velocity variation both in transversal (z) and axial (x) direction, which ranges almost up to the

factor of 4 with respect to the fluid inlet velocity. With the strict laminar flow, the fluid is detained in the recirculation zones, and the surface area corresponding to these zones is expected to be less effective for heat transfer.

At the Reynolds number, $\mathrm{Re} = 4000$, fluid separation establishes between the welding spots with reattachment at the contour of the neighbouring, downstream spots, and a comparatively large portion of channel is occupied by recirculation zones. These zones are responsible for local fluid acceleration. In the middle of the channel, there is a meandering fluid core that is bounded by the recirculation zones and that only touches the welding spots.

Heat transfer

The distribution of the heat flux as the vector intensity $|\dot{\mathbf{q}}|$ is illustrated in Figures 4. The heat flux is seen to decrease in streamwise (x) direction, becoming smaller in the recirculation zones than in the central part of the strip. The heat flux fluctuates significantly not only in spanwise, but also in streamwise direction. The positions of the maxima of $\dot{q}(x)$ nearly coincide with the positions of the welding spots.

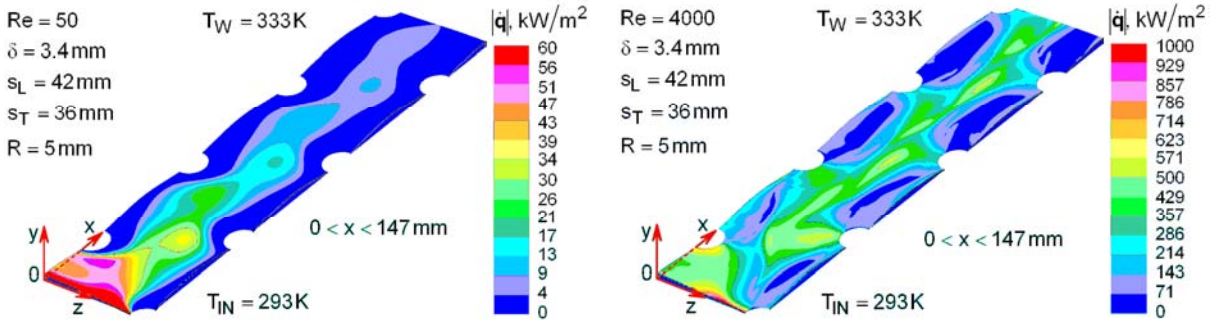


Figure 4: Wall heat flux field at $\mathrm{Re} = 50$ and $\mathrm{Re} = 4000$.

Based on the calculated values of the heat flux, $|\dot{\mathbf{q}}|$, and the fluid mean temperature, $T_b(x)$, the heat transfer coefficient at the wall, $h_W(x)$, is calculated. Its distribution at $Re = 50$ is shown in Figure 5. The quantity $h_W(x)$ behaves damped oscillatory along the channel, decaying at the same time both in amplitude and magnitude. The positions of the maxima of $h_W(x)$ nearly coincide (they are shifted in the upstream direction) with the axial positions of the welding spots, which was already observed in Figure 6.

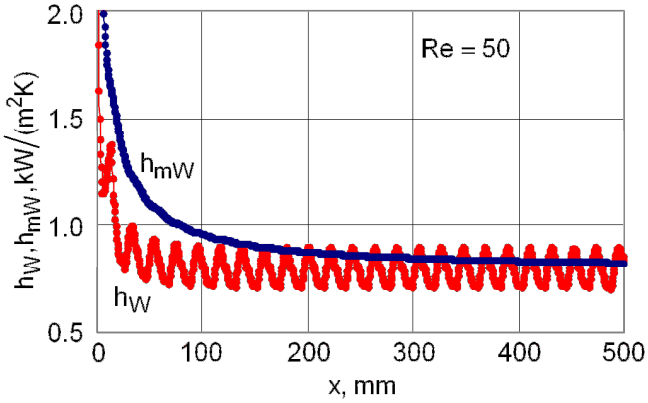


Figure 5: Distribution of the heat transfer coefficient

The Nusselt number of the thermoplate is shown by the numerical experiments to depend on the Reynolds number and the dimensionless geometrical parameters δ/L, s_T/L and s_L/L, and a function of the form:

$$Nu_m = C_1 \left(\frac{\delta}{L}\right)^{n_1} \left(\frac{s_T}{L}\right)^{n_2} \left(\frac{s_L}{L}\right)^{n_3} Re^{n_4} Pr^{n_5}, \quad (1)$$

would be appropriate to describe the numerical results in certain ranges of the parameters varied. The coefficients C_1 and n_1 to n_5 are calculated based on the results of the numerical simulations.

Conclusions

Thermoplates are efficient heat transfer devices, which are used in several branches of engineering practice. The complex geometry of the inside channel of such a plate provides the flowing fluid a pronounced three-dimensional character. Our comprehensive and at present not finished investigations aim at obtaining a better understanding of the inside fluid flow and heat transfer, in order to develop some

relationships, which could serve as indicators when trying to optimise the channel geometry with respect to the thermo-fluid characteristic of the thermoplate.
Depending on the geometrical and process parameters, the heat transferimprovement in the thermoplate ranges nearly up to the factor of 4.

Acknowledgments
We would like to acknowledge friendly support by the PC^2 team

References

[1] J. Mitrovic and B. Maletic, Numerical simulation of fluid flow and heat transfer in thermoplates, Proceedings of 13th International Heat Transfer Conference IHTC-13, Sydney, 2006.
[2] J. Mitrovic and B. Maletic, Numerical simulation of fluid flow, heat transfer and pressure drop in thermoplates, Proceedings of 5th International Conference on Heat Transfer , Fluid Mechanics and Thermodynamics HEFAT2007, Sun City, Südafrika, 2007.
[3] J. Mitrovic and B. Maletic, Numerical simulation of fluid flow and heat transfer in thermoplates, International Journal of Heat Exchangers, 2007, in Press.
[4] R. Fan, Numerische Untersuchungen zum Wärmeübergang und Druckabfall in Thermoblechen, Bachelorarbeit, University of Paderborn, 2007.
[5] Y. Li, Numerische Untersuchungen zur Hydrodynamik und zum Wärmeübergang in Thermoblechen, Bachelorarbeit, University of Paderborn, 2007.
[6] W. Heckel, Numerische Untersuchungen zur Wärmeübertragung in Thermoblechen, Studienarbeit, University of Paderborn, 2007.
[7] P. Grimm, Numerische Untersuchung des Wärmeübergangs an Thermoblechen, Studienarbeit, University of Paderborn, 2007.

5.4.13 Falling Films in Seawater Desalination

Project coordinator	Prof. Dr.-Ing. Jovan Mitrovic, University of Paderborn
Project members	Dr. rer. nat. Henning Raach, University of Paderborn
Supported by	European Commission – Project “Easymed”

General Problem Description

A falling film is a liquid layer that runs down a vertical wall driven by the gravitational force, Figure 1 top. Falling films are often employed in process engineering due to their small liquid volume, the large surface and the good heat transfer. In our case, we have falling films in seawater desalination, the process of transforming seawater in potable water by distillation. In order to increase the evaporation rate, horizontal wires are attached to the heating wall, Figure 1 bottom. The purpose of these wires is twofold. On the one hand, they homogenize the film preventing the formation of dry patches. On the other hand, they promote turbulence even at lower flow rates (i.e. lower Reynolds number Re). Therefore, they are called turbulence wires.

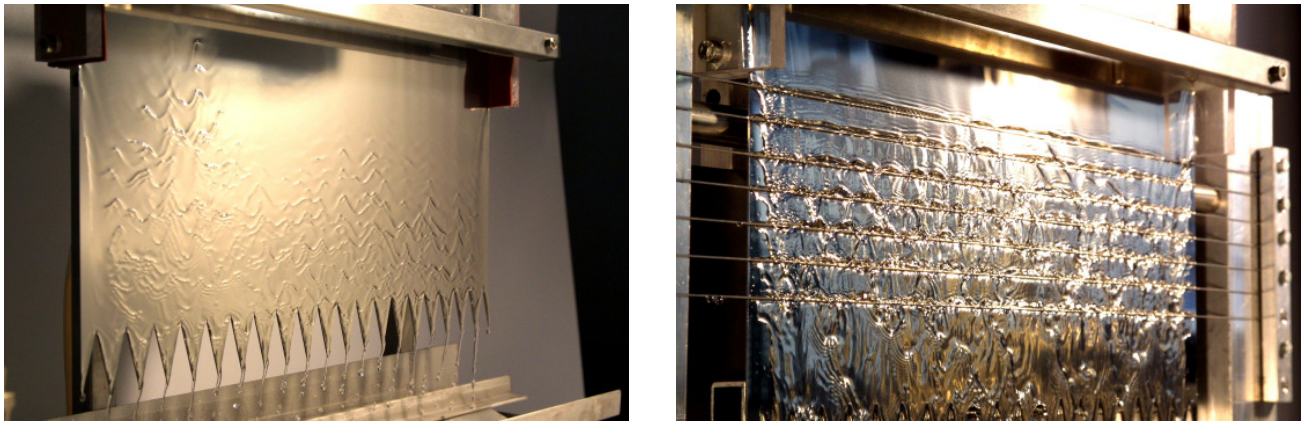
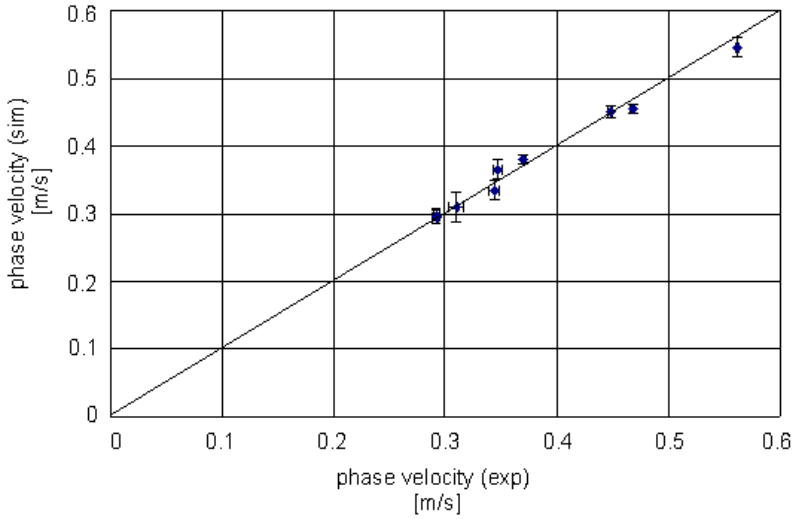


Figure 1: Wavy falling water film without (top) and with (bottom) turbulence wires

In the European project EasyMED, a desalination plant has been developed that is easy to operate and that makes use of energy efficiently [1]. Involved in this project were partners from France, Italy and Germany. Our concern was the simulation of the falling film in the evaporator and the optimization of the wire spacing. Two articles were published in the journal *Desalination* [2, 3] on this matter. The simulation was done by means of Computational Fluid Dynamics (CFD). Unfortunately, the state of the art software a few years ago did not permit to consider the action of the surface tension. This aspect shall now be taken into account. The CFD software of choice in this study is OpenFOAM [4] version 1.3.

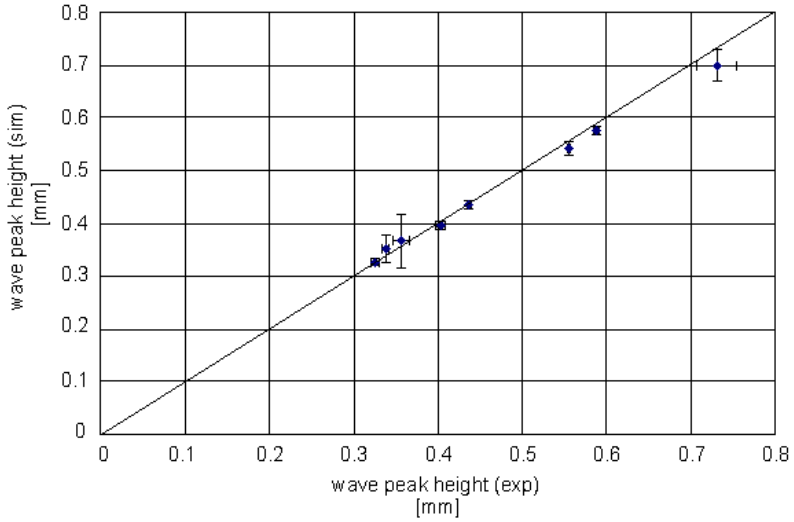
Problem details and work done

In order to use OpenFOAM in two-phase thermal flow, the Volume of Fluid application interfoam had to supplemented by a temperature solver.

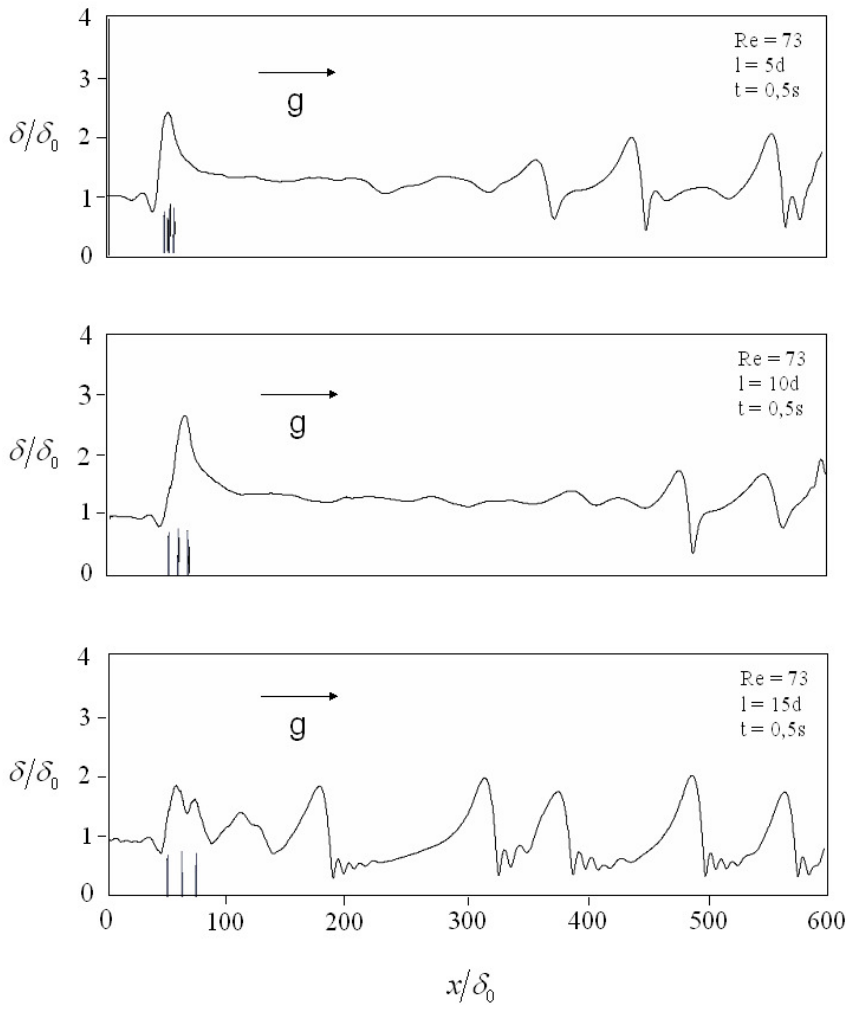


In a first step, the hydrodynamics of a wavy, laminar, two-dimensional film was checked. For that purpose, the simulated phase velocities and peak heights were compared to correlations found experimentally by Nososko [5]. The agreement was very good (cf. diagrams [6, 7]).

In a second step, turbulence wires were introduced and thermal simulations performed. It was the aim to find the optimal distance between the wires with respect to the highest evaporation rate. Numerical experiments with two and three wires were performed and the average interface temperatures were determined after certain time intervals. Hot as well as insulating wires were simulated. The stronger the perturbation by the wires, the higher was the interface



temperature, from which we concluded a bigger evaporation rate. The numerical experiments with two wires were not conclusive. From those with three wires, it was found, that a distance between the wires of 15 or 20 wire radii gave the best results.



The position of the gas-liquid interface is displayed for three different wire distances at Re=73. The perturbation of the falling film by the three wires is largest for a distance between the wires of 15 wire radii (not 10 or 5) after the long time period of 0.5 s. δ_0 is the thickness of a smooth film according to Nusselt (diagrams taken from [8]).

References

[1] http://www.easymed-eu.com

[2] Henning Raach, Jovan Mitrovic, Seawater falling film evaporation on vertical plates with turbulence wires, Desalination 183 (2005) 307-316 http://www.desline.com/articoli/6722.pdf

[3] Henning Raach, Jovan Mitrovic, Simulation of heat and mass transfer in a multi-effect distillation plant for seawater desalination, Desalination 204 (2007) 416-422 http://www.desline.com/articoli/8068.pdf

[4] http://www.openfoam.org

[5] Nosoko,T., Yoshimura, P.N., Nagata, T., Oyakawa, K..: Characteristics of two dimensional waves on a falling liquid film; Chemical Engineering Science, 51 (1996), 725-732

[6] Henning Raach, Sushan Somasundaram, Numerische Untersuchungen zu welligen Fallfilmen, Vortrag auf der GVC-Arbeitssitzung CFD in Würzburg, 6. März 2007

[7] Sushan Somasundaram, Validierung von OpenFOAM für wellige, laminare Fallfilme, große Studienarbeit, Universität Paderborn

[8] Sushan Somasundaram, Numerische Untersuchung des Wärmeübergangs in Fallfilmen mit Stolperdrähten, Diplomarbeit, Universität Paderborn

5.4.14 Diffusive Graph Partioning with Very High Quality

Project coordinator	Prof. Dr. Burkhard Monien, University of Paderborn
Project members	Dipl.-Inform. Henning Meyerhenke, University of Paderborn Dipl.-Math. Thomas Sauerwald, University of Paderborn Dr. Stefan Schamberger, University of Paderborn
Supported by	EU (Integrated Project IST-015964 AEOLUS), DFG (Sonderforschungsbereich 376, Graduiertenkolleg 693

General Problem Description

Graph partitioning is a widespread technique in computer science, engineering, and related fields. The most common formulation of the graph partitioning problem for an undirected graph G = (V,E) asks for a division of V into k pairwise disjoint subsets (partitions) of size at most ceil(|V|/k) such that the edge-cut, i.e., the total number of edges having their incident nodes in different subsets, is minimized. Amongst others, its applications include VLSI circuit layout, dynamical systems, and image segmentation. We mainly consider its use for balancing the load in a numerical simulation (e. g., fluid dynamics), which has become a classical application for parallel computers. There, our task is to compute a partitioning of the (dual) mesh derived from the domain discretization [14].

Implementations of the popular graph partitioning heuristic Kernighan-Lin (KL), combined with the multilevel paradigm, yield good solutions in very short time. Yet, the computed partitionings do not necessarily meet the requirements of all users. For example, the number of boundary vertices (vertices with a neighbor in a different partition) models the communication volume between processors in numerical simulations more accurately than the edge-cut. Moreover, the edge-cut is a summation norm, while often (e. g., for parallel numerical solvers) the maximum norm is of much higher importance. Finally, for some applications, the shape of the partitions, in particular small aspect ratios [4], but also connectedness and smooth boundaries, plays a significant role. Nevertheless, most partitioning-based load balancers do not take these facts fully into account. While the total number of boundary vertices can be minimized by hypergraph partitioning [3], an optimization of partition shapes requires additional techniques (e. g., [4,11]), which are far from being mature. Furthermore, due to their sequential nature, the heuristic KL is difficult

to parallelize. Although significant progress has been made [1,13,18], an inherently parallel graph partitioning algorithm for load balancing can be expected to yield better solutions, possibly also in shorter time.

These issues have led us to the development of the partitioning heuristic BUBBLE-FOS/C in previous work [8]. It is based on a disturbed diffusion scheme that determines how well connected two nodes are in a graph. Using this notion, it aims at very good partition shapes. This results in partitionings with nearly always connected partitions that also have short boundaries and good edge-cut values. Moreover, it contains a high degree of natural parallelism and can be used for parallel load balancing, resulting in low migration costs [9]. Yet, its partly global approach makes it too slow for practical relevance. That is why a significant acceleration without a degradation of the solution quality is part of this project.

Both theoretical and practical aspects of shape-optimizing graph partitioning have been advanced in this work. To understand BUBBLE-FOS/C better theoretically, we have proved its convergence (details omitted here). On the practical side we have developed a much faster new diffusive method for the local improvement of partitionings. The combination of BUBBLE-FOS/C and this new diffusive method within a multilevel framework with two different hierarchy construction algorithms is called DIBAP. This combined algorithm is much faster than BUBBLE-FOS/C and computes multi-way graph partitionings of very high quality on large graphs in a very reasonable amount of time. Our algorithm delivers better solutions than the state-of-the-art partitioning libraries METIS [5,6] and JOSTLE [17] in terms of the edge-cut and the number of boundary vertices, both in the summation and in the maximum norm. Certainly notable is the fact that DIBAP also improves for six benchmark graphs a large number (more than 80 out of 144) of their best known partitionings regarding the edge-cut. These six graphs are among the eight largest in a popular benchmark set [16].

Problem details and work done

Our previous work on shape-optimizing graph partitioning [8,9] has already indicated that this approach is able to compute high-quality partitionings. The main reason for its very high running time is the repeated solution of linear systems on the whole graph (or at least on an approximation of the whole graph, as in [10]). Yet, once a reasonably good solution has been found, alterations during an improvement step take place mostly at the partition boundaries. That is why we introduce in the following a local approach considering only these boundary regions.

Generally speaking, the Consolidation operation is used to determine a new partitioning from a given one. As illustrated in Figure 1 with an example of a path graph and k=3, one performs the following independently for each partition π_c: First, the nodes of π_c receive an equal amount of initial load $n/|\pi_c|$, while the other nodes' initial load is set to 0 (top). Then, a diffusive method is used to distribute this load within the graph (middle). To restrict the computational effort to areas close to the partition boundaries, we use a small number ψ of FOS [2] iterations for this. This is done by multiplying the diffusion matrix **M** with the load vector ψ times.

After the load is distributed this way for all k partitions, we assign each node v to the partition it has obtained the highest load from (bottom). This completes one Consolidation operation, which can be repeated several times to facilitate sufficiently large movements of the partitions. We denote the number of repetitions by Λ and call the whole method with this particular diffusive process TRUNCCONS (*truncated diffusion consolidations*), its algorithmic presentation is shown below. Although having some differences, it can be viewed as a k-way extension of Pellegrini's work [11].

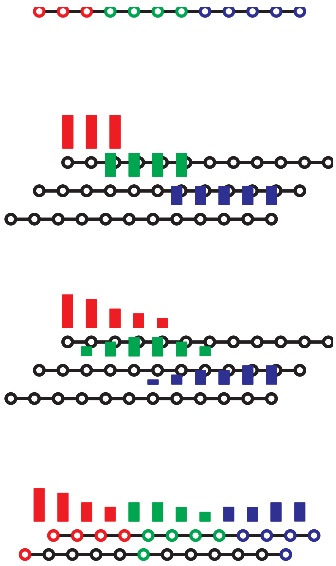


Figure 1: Consolidation scheme

Since an actual load exchange happens only at the partition boundary, not all nodes have to take part in this process. Instead, one keeps track of active nodes. During the course of the iteration, these nodes are the ones either directly at the partition boundary or whose current load differs from their initial load. By neglecting inactive nodes, the diffusive improvement process is restricted to local areas close to the partition boundaries and therefore greatly reduced in complexity. The choice of the initial load takes partition sizes into account and therefore improves the balance. For cases where this is not sufficient, additional balancing heuristics have been integrated.

```
Algorithm TruncCons(M, k, Π, Λ, ψ) → Π
01  for τ = 1 to Λ
02    parallel for each partition π_c
03      S = π_c; w_c = (0, ..., 0)^T /* initial load */
04      for each v ∈ S /* initial load */
05        [w_c]_v = n/|S|
06      for t = 1 to ψ /* FOS iterations */
07        w_c = M · w_c
08      for each v ∈ π_c /* assign nodes to partitions */
09        Π(v) = argmax_{1≤c'≤k} [w_c']_v
10  return Π
```

Now that we have a slow, but high-quality partitioner and a faster local improvement algorithm, we combine them to obtain an efficient multilevel graph partitioning algorithm that we call DIBAP (*Diffusion-based Partitioning*). The fine levels of its multilevel hierarchy are constructed by approximate

maximum weight matchings. Once the graphs are sufficiently small, we switch the construction mechanism to the more expensive algebraic multigrid (AMG) [15] coarsening. This is advantageous, because we use BUBBLE-FOS/C as the improvement strategy on the coarse levels and employ AMG to solve the linear systems that occur in this algorithm. That is why such a hierarchy needs to be built anyway. On the finer parts of the hierarchy, the faster TRUNCCONS is used as the local improvement algorithm. Since this does not involve the solution of linear systems, AMG is not required, so that it is much cheaper to use a matching hierarchy instead.

Extensive experiments have been performed to evaluate both BUBBLE-FOS/C and our new method. The former has been implemented in C for distributed memory parallelism using MPI. Such a setting is also targeted by DIBAP. Yet, its current C/C++ implementation supports only shared memory parallelism by POSIX threads so far. Due to space constraints, we focus in our presentation on DIBAP, whose main parameters are set to Λ=10 and ψ=14. Within DIBAP, we compute three different solutions with BUBBLE-FOS/C. The best one is chosen and improved by multilevel TRUNCCONS as described above. Its results are compared to the state-of-the-art partitioning libraries METIS and JOSTLE.

The test set comprises eight benchmark graphs, which are the largest (regarding |V|) in Walshaw's popular graph partitioning archive [16]. These graphs have to be partitioned into $k \in \{8,12,16,20,32\}$ subdomains. In summary, our experiments reveal the following results. Compared to METIS and JOSTLE, DIBAP is able to compute partitionings that have on average a lower edge-cut and fewer boundary nodes in total (around 7% improvement). This is insofar notable as the other two libraries explicitly minimize the egde-cut, unlike DIBAP. In the maximum norm, i .e. taking the maximum over all partitions, the improvements of DIBAP to METIS and JOSTLE are even higher. The average improvement to METIS w. r. t. the number of boundary nodes in the maximum norm – which can be considered a more accurate measure for communication in parallel numerical solvers than the edge-cut – is around 8%. The gain on JOSTLE is even approximately 12-13%. Furthermore, the number of disconnected partitions is much smaller for DIBAP than for METIS and JOSTLE.

To provide the reader with a visual impression on how DIBAP's results differ from those of METIS and JOSTLE, we include an 8-partitioning of the 2D graph biplane9, see Figure 2. Compared to the other two libraries, its boundaries are smoother, stair-like artifacts as in the two leftmost pictures hardly occur. As larger examples reveal more clearly, DIBAP's subdomains are in general somewhat rounder and more compact, i. e. with a smaller diameter. Also note that both other libraries generate a partition with two large disconnected node sets.

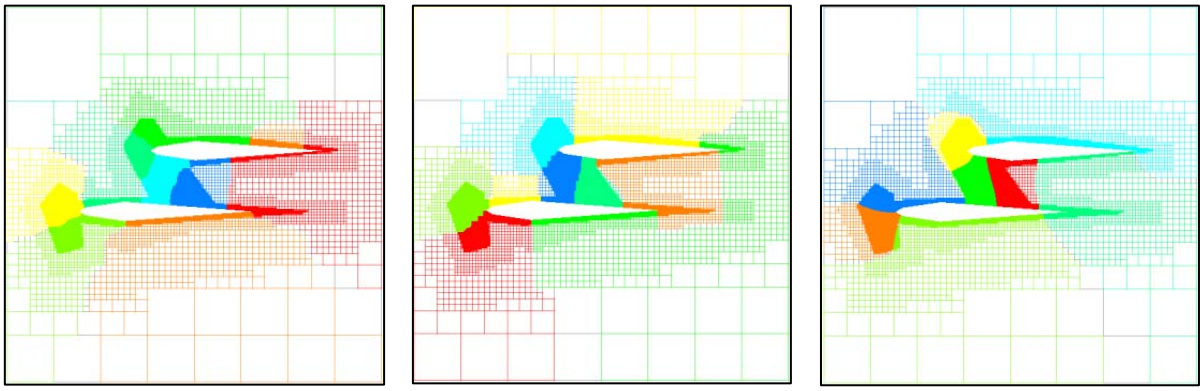


Figure 2: 8-partitioning of the FEM graph biplane9 of METIS, JOSTLE, and DIBAP (from left to right)

The average running times required by the programs to partition the “average graph” of the benchmark set is clearly the best for METIS (about one third of a second on a Core 2 Duo 6600 computer). JOSTLE is a factor of roughly 2.5 slower. Compared to this, the running times of DIBAP are significantly higher, one to two orders of magnitude. This speed gap is particularly high for larger k, which is mainly due to the fact that – in contrast to METIS and JOSTLE – DIBAP scales nearly linearly with k, i. e., doubling k results in a nearly doubled running time. A remedy of this problem is part of future work. After all, DIBAP constitutes a vast improvement over previous implementations using only BUBBLE-FOS/C for partitioning ([8], [12, p. 112]) with an acceleration factor of one to two orders of magnitude (increasing with the graph size).

Among other improvements, we plan for a distributed-memory parallelization of DIBAP. If one assumes a parallel load balancing scenario with k processors for k partitions, one may divide the sequential running times of DIBAP by k·e (where 0<e<=1 denotes the efficiency of the parallel program). Then, its parallel running time on k processors can be expected to be at most a few dozens of seconds, which is certainly acceptable.

Walshaw's benchmark archive also collects the best known partitionings for each of the 34 graphs contained therein, i. e., partitionings with the lowest edge-cut. Currently, results of more than 20 algorithms are considered. With each graph 24 partitionings are recorded, one for six different numbers of partitions ($k \in \{2,4,8,16,32,64\}$) in four different imbalance settings (0%,1%,3%,5%). Using DIBAP, we have been able to improve more than 80 of these currently best known edge-cut values for six of the eight largest graphs in the archive. The complete list of improvements with the actual edge-cut values and the corresponding partitioning files are available at the website of Walshaw's archive [16].

Resource Usage

Different computers have been used for our experiments. Our MPI parallel implementation of BUBBLE-FOS/C has been run on the ARMINIUS cluster, mostly using the fast Infiniband interconnect. DIBAP, the new algorithm, has not been implemented for distributed memory parallelism yet. That is why we have run it on different computers with multiple CPU cores and shared memory. Besides dual-core office machines and nodes of the ARMINIUS cluster, we have used two other systems with multiple processing units, both are employed at PC². One has four AMD Opteron CPUs, the other one two quad-core Intel Clovertown CPUs. These different architectures have provided an excellent environment for regular experiments.

References

[1] U. Catalyurek, E. Boman, K. Devine, D. Bozdag, R. Heaphy, and L. Riesen. Hypergraph-based dynamic load balancing for adaptive scientific computations. In Proc. 21st Intl. Parallel and Distributed Processing Symposium (IPDPS'07). IEEE, 2007.

[2] G. Cybenko. Dynamic load balancing for distributed memory multiprocessors. Parallel and Distributed Computing, 7:279-301, 1989.

[3] K. Devine, E. Boman, R. Heaphy, R. Bisseling, and U. Catalyurek. Parallel hypergraph partitioning for scientific computing. In Proc. 20th Intl. Parallel and Distributed Processing Symposium (IPDPS'06). IEEE, 2006.

[4] R. Diekmann, R. Preis, F. Schlimbach, and C. Walshaw. Shape-optimized mesh partitioning and load balancing for parallel adaptive FEM. J. Parallel Computing, 26:1555-1581, 2000.

[5] G. Karypis and V. Kumar. MeTiS: A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, [...], Version 4.0, 1998.

[6] G. Karypis and V. Kumar. Multilevel k-way partitioning scheme for irregular graphs. J. Parallel Distrib. Comput., 48(1):96-129, 1998.

[7] B. W. Kernighan and S. Lin. An efficient heuristic for partitioning graphs. Bell Systems Technical Journal, 49:291-308, 1970.

[8] H. Meyerhenke, B. Monien, and S. Schamberger. Accelerating shape optimizing load balancing for parallel FEM simulations by algebraic multigrid. In Proc. 20th IEEE Intl. Parallel and Distributed Processing Symposium (IPDPS'06). IEEE, 2006.

[9] H. Meyerhenke and S. Schamberger. Balancing parallel adaptive FEM computations by solving systems of linear equations. In Proc. Euro-Par 2005, volume 3648 of LNCS, pages 209-219, 2005.

[10] H. Meyerhenke and S. Schamberger. A parallel shape optimizing load balancer. In Proc. Euro-Par 2006, volume 4128 of LNCS, pages 232-242, 2006.

[11] F. Pellegrini. A parallelisable multi-level banded diffusion scheme for computing balanced partitions with smooth boundaries. In Proc. Euro-Par 2007, volume 4641 of LNCS, pages 195-204. Springer, 2007.

[12] S. Schamberger. Shape Optimized Graph Partitioning. PhD thesis, Universität Paderborn, 2006.

[13] K. Schloegel, G. Karypis, and V. Kumar. Parallel static and dynamic multi-constraint graph partitioning. Concurrency and Computation: Practice and Experience, 14(3):219-240, 2002.

[14] K. Schloegel, G. Karypis, and V. Kumar. Graph partitioning for high performance scientific simulations. In The Sourcebook of Parallel Computing, pages 491-541. Morgan Kaufmann, 2003.

[15] K. Stüben. An introduction to algebraic multigrid. In U. Trottenberg, C. W. Oosterlee, and A. Schüller, editors, Multigrid, pages 413-532. Academic Press, London, 2000. Appendix A.

[16] C. Walshaw. The graph partitioning archive. http://staffweb.cms.gre.ac.uk/~c.walshaw/partition/, 2007.

[17] C. Walshaw and M. Cross. Mesh partitioning: a multilevel balancing and refinement algorithm. SIAM J. Sci. Comput., 22(1):63-80, 2000.

[18] C.Walshaw and M. Cross. Parallel optimisation algorithms for multilevel mesh partitioning. J. Parallel Computing, 26(12):1635-1660, 2000.

6 Summary of References (alphabetical order)

[1] Abschlussbericht der Projektgruppe Gomputer, PC² November 2007

[2] Adamo, C.; Barone, V.; Bencini, A.; Broer, R.; Filatov, M.; Harrison, NM.; Illas, F.; Malrieu, JP. and de P R Moreira I (2006) J Chem Phys 124:107101.

[3] Aranda-Espinoza, H.; Berman, A.; Dan, N.; Pincus, P. and Safran, S.: *Interaction between Inclusions Embedded in Membranes*, Biophysical Journal 71, 648, 1996

[4] ARMINIUS-Cluster. http://www.upb.de/services/systems

[5] Arnold, A.; Mann, B.A.; Limbach, H.J. and Holm, C.: Comp. Phys. Comm. 174, (2005), 704

[6] Battré, Dominic; Hovestadt, Matthias; Kao, Odej; Keller, Axel and Voß, Kerstin: Transparent Cross-Border Migration of Parallel Multi Node Applications

[7] Battré, Dominic; Hovestadt, Matthias; Kao, Odej; Keller, Axel and Voß, Kerstin: Planning Based Scheduling for SLA-awareness and Grid Interaction

[8] Battré, Dominic; Heine, Felix and Kao, Odej: *Top k RDF Query Evaluation in Structured P2P Networks.* In: Wolfgang E. Nagel, Wolfgang V. Walter, Wolfgang Lehner (Ed.), Euro-Par 2006 Parallel Processing: 12th International Euro-Par Conference, Springer, pp. 995-1004, 2006

[9] Battré, Dominic; Cortese, Giovanni; Heine, Felix and Höing, Andre: *Accessing XML Documents using Semantic Meta Data in a P2P Environment.* In: Siegfried Handschuh, Andreas Abecker, Tudor Groza (Ed.), The 20th International FLAIRS Conference, Special Track on Artificial Intelligence and Social Semantic Collaboration (SEMSOC 2007), 2007

[10] Battré, Dominic; Heine, Felix; Höing, André and Kao, Odej: O*n Triple Dissemination, Forward-Chaining, and Load Balancing in DHT based RDF stores.* In: Sonia Bergamaschi, Sam Joseph, Jean-Henry Morin, Gianluca Moro (Ed.), Fourth International Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P) in scope of 32st International Conference on Very Large Data Bases, Springer, 2006

[11] Battré, Dominic; Kao, Odej and Voss, Kerstin: Implementing WS-Agreement in a Globus Toolkit 4.0 Environment, In Usage of Service Level Agreements in Grids Workshop in conjunction with The 8th IEEE International Conference on Grid Computing (Grid 2007), September 2007

[12] Battré, Dominic; Heine, Felix; Höing, André and Kao, Odej. *Load-balancing in P2P based RDF stores.* In: Holger Wache, Heiner Stuckenschmidt, Bijan Parsia, Yuanbo Guo, Tim Finin, Dave Beckett (Ed.), Second International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2006) in scope of the International Semantic Web Conference, pp. 21-34, 2006

[13] Beisel, Tobias.: “Entwurf und Evaluation eines parallelen Verfahrens zur Bildrekonstruktion in der Positronen-Emissions-Tomographie auf Multi-Core-Architekturen”. Master Thesis UPB (2007)

[14] Ben-Tal, N. and Honig, B.: *Helix-helix interactions in lipid bilayers*, Biophysical Journal 71, 3046, 1996

[15] Berners-Lee, Tim; Hendler, James and Lassila, Ora: *The Semantic Web*, Scientific American, May 2001

[16] Betriebskonzept für die D-GridInfrastruktur, http://www.d-grid.de/fileadmin/user_upload/documents/Kern-D-Grid/Betriebskonzept/D-Grid-Betriebskonzept.pdf

[17] Beyer, M. K. and Clausen-Schaumann, H. *Chem. Rev.* 2005, 105, 2921

[18] BIS-Grid, https://bi.offis.de/bisgrid/tiki-index.php

[19] BIS-Grid project homepage, http://bisgrid.d-grid.de/

[20] Blankenburg, Stefan and Schmidt, Wolf-Gero: “Steric effects and chirality in the adsorption of glycine and phenylglycine on Cu (110)“ Nanotechnology 18, 424030 (2007).

[21] Blankenburg, Stefan and Schmidt, Wolf-Gero:“Adsorption of phenylglycine on copper: Density functional calculations“,Phys. Rev. B 74, 155419 (2006).Bloom, Burton H.: Space/Time Trade-offs in Hash Coding with Allowable Errors., Communications of the ACM, 13:7, pages 422-426, 1970

[22] Blankenburg, Stefan und Schmidt, Wolf-Gero: “Long-Range Chiral Recognition due to Substrate Locking and Substrate-Adsorbate Charge Transfer“, Phys. Rev. Lett. 99, 196107 (2007); selected for Virtual Journal of Nanoscale Science & Technology 16(21) (2007)

[23] Bloom, M.; Evans, E. and Mouritsen, O.G.: *Physical properties of the fluid lipid-bilayer component of cell membranes: a perspective*, Quart. Rev. Biophys. 24, 293, 1991

[24] Blazy, Stephan and Marquardt, Oliver: Parallel finite element computations of three dimensional flow problems using padfem. *International Journal of Parallel, Emergent and Distributed Systems*, Volume 22, Issue 4, 2007, Pages 257 – 274.

[25] Bothe, D.; Hoffmann, M.; Schlueter, M.; Rabiger, N.; Blazy, S.; Stemich, C. and Warnecke, H.-J. 2003: Experimental and numerical investigations of t-shaped micro-mixers. Proc. 11th European Conference on Mixing, VDI-GVC, pp. 269--276.

[26] Bothe, D.; Stemich, C. and Warnecke, H.-J.: *Fluid mixing in a T-shaped micro-mixer,* Chem. Eng. Sci 61, 2950-2958 (2006).

[27] Bothe, D.; Stemich, C. and Warnecke, H.-J.: *Mixing in a T-shaped microreactor: scales and quality of mixing,* pp 351-357 in 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering (W. Marquardt, C. Pantelidis, eds.), Garmisch-Partenkirchen (Germany), Elsevier 2006.

[28] Bothe, D.; Lojewski, A.; Warnecke, H.—J.: *Direct numerical simulation of reactive mixing in a T-shaped micro-reactor*, 5th Joint ASME/ JSME Fluids Engineering Conference, FEDSM2007-37507, San Diego (USA) July 30 – August 2, 2007

[29] Brickley, Dan and Guha, Ramanathan *V.: RDF Vocabulary Description Language 1.0: RDF Schema*, http://www.w3.org/TR/rdf-schema, 2004

[30] Brown, D.A. and London, E.: *Functions of lipid rafts in biological membranes*, A. Rev. Cell. Dev. Chem. 14, 111, 1998

[31] Burchard, Lars-O. et al. The Virtual Resource Manager: An Architecture for SLA-aware Resource Management. In *4th Intl. IEEE/ACM Intl. Symposium on Cluster Computing and the Grid (CCGrid) Chicago, USA*, 2004.

[32] Burchard, Lars-Olof; Heine, Felix; Heiss, Hans-Ulrich; Hovestadt, Matthias; Kao, Odej; Keller, Axel; Linnert, Barry and Schneider, Jörg: The Virtual Resource Manager: Local Autonomy versus QoS Guarantees for Grid Applications

[33] Car, R. and Parrinello, M. *Phys. Rev. Lett.* 1985, 55, 2471

[34] Catalyurek, U.; Boman, E.; Devine, K.; Bozdag, D.; Heaphy, R. and Riesen, L.: Hypergraph-based dynamic load balancing for adaptive scientific computations. In Proc. 21st Intl. Parallel and Distributed Processing Symposium (IPDPS'07). IEEE, 2007.

[35] Chen, Q. and Richardson, N.V.: "Enantiomeric interactions between nucleic acid bases and amino acids on solid surfaces", Nature Mater. 2, 324 (2003).

[36] ChessBase: Kramnik vs DEEP FRITZ: Computer wins match by 4:2. http://www.chessbase.com/newsdetail.asp?newsid=3524

[37] ChessBase: http://www.chessbase.com

[38] Condor. http://www.cs.wisc.edu/condor

[39] CRAFTY (Hyatt's home page): http://www.cis.uab.edu/info/faculty/hyatt/hyatt.html

[40] Cremer, D. and Pople, J.A.: *General definition of ring puckering coordinates.* J. Am. Chem. Soc., 1975, *97*, 1354-1358.

[41] Cybenko, G.: Dynamic load balancing for distributed memory multiprocessors. Parallel and Distributed Computing, 7:279-301, 1989.

[42] Decker, H.; Dillinger, R. and Tuczek, F.:*Angew. Chem.* 2000, *112*, 1656; *Angew. Chem. Int. Ed. Engl.* 2000, *39*, 1591.

[43] Devine, K.; Boman, E.; Heaphy, R.; Bisseling, R. and Catalyurek, U.: Parallel hypergraph partitioning for scientific computing. In Proc. 20th Intl. Parallel and Distributed Processing Symposium (IPDPS'06). IEEE, 2006.

[44] Diekmann, Ralf; Preis, Robert; Schlimbach, Frank and Walshaw, Chris: Shape-optimized mesh partitioning and load balancing for parallel adaptive FEM. J. Parallel Computing, 26:1555-1581, 2000.

[45] Discher, D.E. and Eisenberg, A.: *Science* 297, 967, 2002;

[46] Donninger, C. and Lorenz, U.: The Chess Monster Hydra. In Proc. of 14th International Conference on Field-Programmable Logic and Applications (FPL), pages 927-932, Antwerp - Belgium, 2004.Empfehlungen zur Nomenklatur von

Ressourcen und Diensten im D-Grid,
http://www.d-grid.de/fileadmin/dgi_document/FG2/Kern-D-Grid/Dokumente/D-Grid-URI-NomenklaturV1.1.pdf

[47] Dulub, O.; Diebold, U. and Kresse, G.: Novel stabilization mechanism on polar surfaces: ZnO(0001)-Zn. Physical Review Letters 2003, 90, (1).

[48] Dulub, O.; Boatner, L. A. and Diebold, U.: STM study of the geometric and electronic structure of ZnO(0001)-Zn, (000(1)over-bar)-O, (10(1)over-bar0), and (11(2)over-bar0) surfaces. Surface Science 2002, 519, (3), 201-217.

[49] Düchs, D. and Schmid, F.: Phase behavior of amphiphilic monolayers: Theory and simulations, J. Phys: Cond. Matt. 13, 4853, 2001

[50] ESPResSo-Homepage, http://www.espresso.mpg.de (2002-2007)

[51] Falgout, R.D.; Jones, J.E. and Yang, U.M., 2006, The Design and Implementation of hypre, a Library of Parallel High Performance Preconditioners. In Numerical Solution of Partial Differential Equations on Parallel Computers, A.M. Bruaset and A. Tveito, eds., Springer-Verlag, 51, pp. 267-294.

[52] Fan, R.: Numerische Untersuchungen zum Wärmeübergang und Druckabfall in Thermoblechen, Bachelorarbeit, University of Paderborn, 2007.

[53] Feldmann, Rainer: Game Tree Search on Massively Parallel Systems. PhD Thesis, University of Paderborn, 1993.

[54] Feldmann, Rainer; Mysliwietz, Peter and Monien, Burkhard: Studying Overheads in Massively Parallel Min/Max-Tree Evaluation. In 6th ACM Annual symposium on parallel algorithms and architectures (SPAA'94), pages 94-104, New York, NY, 1994.

[55] Foster and Kesselman, C: (Eds.). The Grid 2: Blueprint for a New Computing

[56] FRUIT - pure playing strength: http://www.fruitchess.com/

[57] Future Generation Grids, Springer, pp 83-98, 2006

[58] Ganglia Monitoring System, http://ganglia.info

[59] Gausemeier, Jürgen; Berssenbruegge, Jan and Bauch, Jochen 2006: "A virtual reality-based night drive simulator for the evaluation of a predictive advanced front lighting system". In Proceedings of ASME 2006 IDETC and CIE, International Design Engineering Technical Conference and Computers and Information in Engineering Conference, ASME.

[60] Gil, T.; Ipsen, J.H.; Mouritsen, O.G.; Sabra, M.C.; Speretto, M.M. and Zuckermann, M.J.: *Theoretical analysis of protein organization in lipid membranes*, Biochimica and Biophysica Acta 1376, 245, 1998

[61] Gorelsky, S.I.: SWizard program, http://www.sg-chem.net/.

[62] Gorelsky, S.I.: A.B.P. Lever, *J. Organomet. Chem.* 2001, *635*, 187.

[63] Grundmeier, G.; Schmidt, W. and Stratmann, M.: Corrosion protection by organic coatings: electrochemical mechanism and novel methods of investigation. Electrochimica Acta 2000, 45, (15-16), 2515-2533.

[64] Grundmeier, G.; Reinartz, C.; Rohwerder, M. and Stratmann, M.: Corrosion properties of chemically modified metal surfaces. Electrochimica Acta 1998, 43, (1-2), 165-174.

[65] Grundmeier, G. and Stratmann, M.: Adhesion and De-adhesion mechanisms at polymer/metal interfaces: Mechanistic understanding based on in situ studies of buried interfaces. Annual Review of Materials Research 2005, 35, 571-615.H.

[66] Bal et al. Next Generation Grids 2: Requirements and Options for European Grids Research 2005-2010 and Beyond. ftp://ftp.cordis.lu/pub/ist/docs/ngg2_eg_final.pdf, 2004.

[67] Hall D.O; Cammack . and Rao K.K.: (1971) Nature 233:136-138.

[68] Harris, M.: "Parallel Prefix Sum (Scan) with CUDA". Website: http://developer.download.nvidia.com/compute/cuda/sdk/website/projects/scan/doc/scan.pdf (2007)

[69] Heine, Felix; Hovestadt, Matthias and Kao, Odej: *Processing Complex RDF Queries over P2P Networks.* Workshop on Information Retrieval in Peer-to-Peer-Networks P2PIR 2005, November 4, 2005

[70] He, X.H.; Liang, H.J.; Huang, L. and Pan, C.Y.: *J. Phys. Chem. B* 108, 1731, 004

[71] He, X.H. and Schmid, F.: *Macromolecules* 39, 2654, 2006.

[72] He, X.H. and Schmid, F.: *Macromolecules* 39, 8908, 2006

[73] Heckel, W.: Numerische Untersuchungen zur Wärmeübertragung in Thermoblechen, Studienarbeit, University of Paderborn, 2007.

[74] Heine, Felix: *Scalable P2P based RDF Querying*. In First International Conference on Scalable Information Systems (INFOSCALE06), 2006.

[75] Heine, Felix: *Emergent Schema Management for P2P-based Applications.* In: Roberto Baldoni, Giovanni Cortese, Fabrizio Davide, Angelo Melpignano (Ed.), Global Data Management, IOS Press, pp. 224-248, 2006

[76] Heine, Felix: *P2P based RDF Querying and Reasoning for Grid Resource Description and Matching*. Dissertation, Paderborn, 2006

[77] Herres-Pawlis, S.; Haase, R.; Flörke, U.; Henkel, G.; Kang, P. and Verma, P.: T.D.P. Stack, *Inorg. Chem.* 2007, submitted.

[78] Highly Predictable Cluster for Internet-Grids (HPC4U), EU-funded project IST-511531. http://www.hpc4u.org.

[79] Himstedt, Kai: An Optimistic Pondering Approach for Asynchronous Distributed Game-Tree Search. In ICGA Journal. Vol. 28 No. 2 (2005):77-90.

[80] Hoogerburgge, P.J. and Koelman, J.M.: Europhys. Lett. 19, (1992), 155

[81] Horn, D.: "Stream Reduction Operations for GPGPU Applications". In: GPU Gems 2: Programming Techniques for High-Performance Graphics and General-Purpose Computation. Addison-Wesley Professional (2005) 573-583

[82] Hovestadt, M.; Lerch, N.; Nitsche, H.; and Voss, K.: Cracow Grid Workshop, 2006 First Steps of a Monitoring Framework to Empower Risk Assessment on Grids,

[83] http://www.easymed-eu.com
[84] http://www.openfoam.org
[85] Humphreys, G.; Houston, M.; Ng, R.; Frank, R.; Ahern, S.; Kirchner, P.; Klosowski, J.: “Chromium: A stream-processing framework for interactive rendering on clusters”. ACM Transactions on Graphics 21 (2002) 693{702
[86] Hutter, J. et al: J CPMD, IBM Corp 1990-2007, MPI fuer Festkörperforschung Stuttgart 1997-2001, see also http://www.cpmd.org.
[87] Hutter, J. et al: CPMD V3.9; Copyright IBM Corp 1990-2001, Copyright MPI für Feskörperforschung Stuttgart 1997-2001. see www.cpmd.org
[88] HYDRA: http://www.hydrachess.com
[89] Infrastructure. Morgan Kaufmann Publishers Inc. San Francisco, 2004
[90] Inmon, William H.: *Building the Data Warehouse,* John Wiley & Sons, 2005
[91] JETTY homepage, http://docs.codehaus.org/display/JETTY/
[92] John, V.: 2002, Higher order finite element methods and multigrid solvers in a benchmark problem for the 3d navier-stokes equations. Int. J. Numer. Meth. Fluids 2002, 40, pp. 775--778.
[93] Kahn O (1993) Molecular Magnetism (VCH, New York).
[94] Karger, David R. and Ruhl, Matthias: *Simple Efficient Load Balancing Algorithms for Peer-to-Peer Systems*. In: SPAA '04: Proceedings of the sixteenth annual ACM Symposium on Parallelismin Algorithms and Architectures. ACM Press. 2004
[95] Karypis, G. and Kumar, V.: MeTiS: A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, [...], Version 4.0, 1998.
[96] Karypis, G. and Kumar, V.: Multilevel k-way partitioning scheme for irregular graphs. J. Parallel Distrib. Comput., 48(1):96-129, 1998.
[97] Kaufmann; Paul and Platzner, Marco: MOVES: A Modular Framework for Hardware Evolution. In *Proc. of the 2nd NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, pp.447-454, Aug. 2007. Received the Best Paper Award in the Evolvable Hardware Category.
[98] Kaufmann, Paul and Platzner, Marco: Toward Self-adaptive Embedded Systems: Multiobjective Hardware Evolution. *In Proc. Int'l Conf. on Architecture of Computing Systems (ARCS)*, March 2007. Springer, LNCS 4415.
[99] Kaufmann, Paul and Platzner, Marco: Multi-objective Intrinsic Hardware Evolution. In *Proc. of the MALD Int'l Conference*, Sept. 2006.
[100] Kernighan, B.W. and Lin, S.: An efficient heuristic for partitioning graphs. Bell Systems Technical Journal, 49:291-308, 1970.
[101] Kohl, J. A., Papadopoulos, P. M., and II, G. A. G., 1997. “CUMULVS: Collaborative infrastructure for developing distributed simulations”. In PPSC, SIAM. pp. 79–88.
[102] Krüger, D.; Fuchs, H.; Rousseau, R.; Marx, D.; Parrinello, M. *Phys. Rev. Lett*. 2002, 89, 186402; (b) Krüger, D.; Rousseau, R.; Fuchs, H.; Marx, D. *Angew. Chem. Int. Ed.* 2003, 42, 2251.

[103] Kubiatowicz, John: Extracting Guarantees from Chaos, Communications of the ACM, 46:2, pages 33-38, 2003

[104] Laio, A. and Parrinello, M.: *Proc. Natl. Acad. Sci. USA.* 2002, 99, 12562.

[105] Langan, P.; Nishiyama, Y. and Chanzy, H.: *X-ray Structure of Mercerized Cellulose II at 1 Å Resolution.* Biomacromolecules, 2001, *2*, 410-416.

[106] Langan, P. et al.: *Synchrotron X-ray structures of cellulose Iβ and regenerated cellulose II at ambient temperature and 100 K.* Cellulose, 2005, *12*, 551-562.

[107] Lague, P.; Zuckermann, M.M. and Roux, B.: *Protein inclusion in lipid membranes: A theory based on the hypernetted chain integral equation*, Faraday Disc. 111, 165, 1998

[108] Lenz, O. and Schmid, F.: *Structure of Symmetric and Asymmetric Ripple Phases in Lipid Bilayers*, Phys. Rev. Lett. 98, 058104, 2007

[109] Lenz, O. and Schmid, F.: *A simple computer model for liquid lipid bilayers,* J. Mol. Liquids 117(1-3), 147, 2005

[110] Lewin, J.L.; Heppner, D.E. and Cramer, C.J.: *J. Biol. Inorg. Chem.* 2007, *12*, 1221.

[111] Li, Y.: Numerische Untersuchungen zur Hydrodynamik und zum Wärmeübergang in Thermoblechen, Bachelorarbeit, University of Paderborn, 2007.

[112] Lietsch, Stefan; Berssenbrügge, Jan and Zabel, Henning: 2007: „Computational Steering Of Interactive And Distributed Virtual Reality Applications". In Proceedings of ASME 2007 IDETC and CIE, International Design Engineering Technical Conference and Computers and Information in Engineering Conference, ASME.

[113] Lietsch, Stefan and Berssenbrügge, Jan 2007: "Parallel, Shader-Based Visualization of Automotive Headlights" In Proceedings of the 7th Eurographics Symposium on Parallel Graphics and Visualization (EGPGV07) - Short Paper

[114] Lietsch, Stefan; Zabel, Henning; Berssenbrügge, Jan; Wittenberg, Veit and Eikermann, Martin: 2006: "Light simulation in a distributed driving simulator". In ISVC06: Advances in Visual Computing, Springer, pp. 343–353.

[115] Lübbers, Enno and Platzner, Marco:„ReconOS: A RTOS Supporting Hard- and Software Threads", In *Proceedings of 2007 IEEE International Conference on Field Programmable Logic and Applications (FPL'07)*, Amsterdam, August 27-29, 2007

[116] Lietsch, Stefan; Berssenbrügge, Jan; Zabel, Henning; Wittenberg, Veit and Eikermann, Martin: "Light simulation in a distributed driving simulator". In: Proceedings of the 2nd International Symposium on Visual Computing, 2006. Volume 4291 of Lecture Notes in Computer Science., Berlin Heidelberg New York, Springer-Verlag (2006) 343-353

[117] Lietsch, Stefan and Marquardt, Oliver: "A CUDA-Supported Approach to Remote Rendering". To be published in Proceedings of the 3rd Inernational Symposium on Visual Computing (ISVC2007)

[118] Lindahl, E.; Hess, B. and van der Spoel, D.: *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Journal of Molecular Modeling, 2001, 7, 306-317.

[119] Lins, R.D. and Hunenberger, P.H.: *A new GROMOS force field for hexopyranose-based carbohydrates.* J Comput Chem, 2005, *26*, 1400-1412.

[120] LSF - Load Sharing Facility. http://www.platform.com/products/wm/LSF/index.asp

[121] Manola, Frank and Miller, Eric: *RDF Primer*, http://www.w3.org/TR/rdf-primer, 2004

[122] Marx, D. and Hutter, J.: (2000) Modern Methods and Algorithms of Quantum Chemistry, vol. 3, ed. Grotendorst J, FZ Juelich (NIC), pp. 301-449

[123] Marx, D. and Hutter, J.: "Ab initio Molecular Dynamics: Theory and Implementation" in *Modern Methods and Algorithms of Quantum Chemistry*, Editor: J. Grotendorst (NIC, Forschungzentrum Jülich 2000). See www.theochem.rub.de/go/cprev.html

[124] Mathe, J.; Di Meglio, J.-M. and Tinland, B.: arXiv:cond-mat/0610328v2 [cond-mat.soft], submitted to JCIS

[125] Mauris, N.M. and Fraaije, J. G. E. M.: *Macromolecules* 107, 5879, 1997;

[126] May, S.: *Theories on structural perturbations of lipid bilayers*, Curr. Opn. Coll. Interf. Sci. 5, 244, 2000

[127] Mazzarello, R.; Cossaro, A.; Verdini, A.; Rousseau, R.; Casalis, L.; Danisman, M.F.; Floreano, L.; Scandolo, S.; Morgante, A.; Scoles, G. *Phys. Rev. Lett.* 2007, 98, 016102. www.quantum-espresso.org

[128] Meyerhenke, Henning; Monien, Burkhard and Schamberger, Stefan: Accelerating shape optimizing load balancing for parallel FEM simulations by algebraic multigrid. In Proc. 20th IEEE Intl. Parallel and Distributed Processing Symposium (IPDPS'06). IEEE, 2006.

[129] Meyerhenke, Henning and Schamberger, Stefan: Balancing parallel adaptive FEM computations by solving systems of linear equations. In Proc. Euro-Par 2005, volume 3648 of LNCS, pages 209-219, 2005.

[130] Meyerhenke, Henning and Schamberger, Stefan: A parallel shape optimizing load balancer. In Proc. Euro-Par 2006, volume 4128 of LNCS, pages 232-242, 2006.

[131] Meyer-Kahlen, Stefan: Shredder Computer Chess. http://www.shredderchess.com/

[132] Microsoft: "Microsoft Windows Server 2003 Terminal Services". Website: http://www.microsoft.com/windowsserver2003/technologies/terminalservices/default.mspx (2007)

[133] Mirica, L.M.; Vance, M.; Jackson-Rudd, D.; Hedman, B.; Hodgson, K.O. and Solomon, E.I.:T.D.P.: Stack, *Science* 2005, *308*, 1890.

[134] Mitrovic, Jovan and Maletic, Boban: Numerical simulation of fluid flow and heat transfer in thermoplates, Proceedings of 13th International Heat Transfer Conference IHTC-13, Sydney, 2006.

[135] Mitrovic, Jovan and Maletic, Boban: Numerical simulation of fluid flow, heat transfer and pressure drop in thermoplates, Proceedings of 5th International Conference on Heat Transfer , Fluid Mechanics and Thermodynamics HEFAT2007, Sun City, Südafrika, 2007.

[136] Mitrovic, Jovan and Maletic, Boban: Numerical simulation of fluid flow and heat transfer in thermoplates, International Journal of Heat Exchangers, 2007, in Press.

[137] Morales, R.; Chron, M.H.; Hudry-Clergeon, G.; P´etillot, Y.; Norager, S.; Medina, M. and Frey, M.: (1999) Biochemistry 38:15764-15773.

[138] Mouritsen, O.G. and Bloom, M.: *Models of lipid-protein interactions in membranes*, Quarterly Reviews of Biophysics: Biomol. Struct. 22, 145, 1993

[139] Müller, M. and Schmid, F.: *Adv. Polym. Sci*. 185, pp1-85, 2005.

[140] Nosoko,T.; Yoshimura, P.N.; Nagata, T. and Oyakawa, K.: Characteristics of two dimensional waves on a falling liquid film; Chemical Engineering Science, 51 (1996), 725-732

[141] NVIDIA: “NVIDIA CUDA - Compute Unified Device Architecture”. Website: http://developer.nvidia.com/object/cuda.html (2007)

[142] Paderborn Center for Parallel Computing, www.upb.de/pc2

[143] Park, K. S.; Cho, Y. J.; Krishnaprasad, N. K.; Scharver, C.; Lewis, M. J.; Leigh, J. and Johnson, A. E.: 2000. “CAVERNsoft G2: A toolkit for high performance teleimmersive collaboration”. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST-00), K. Y. Wohn, ed., ACM Press, pp. 8–15.PBS - Portable Batch System. http://www.openpbs.org

[144] Parker, S. G., and Johnson, C. R.: 1995. “SCIRun: A scientific programming environment for computational steering”. In Proceedings of Supercomputing ’95, ACM/IEEE.

[145] Pellegrini, F.: A parallelisable multi-level banded diffusion scheme for computing balanced partitions with smooth boundaries. In Proc. Euro-Par 2007, volume 4641 of LNCS, pages 195-204. Springer, 2007.

[146] Perdew, J. P.; Burke, K. and Ernzerhof, M.: *Phys. Rev. Lett.* 1996, 77, 3865.

[147] Preuss, Martin; Schmidt, Wolf-Gero und Bechstedt, Friedhelm: “Coulombic Amino Group-Metal Bonding: Adsorption of Adenine on Cu(110)“, Phys. Rev. Lett 94, 236102 (2005); selected for Virtual Journal of Nanoscale Science & Technology 11(25) (2005).

[148] Proceedings of the 26th Workshop of the UK Planning and Scheduling Special Interest Group (PlanSIG2007)

[149] Proceedings of the 7th Cracow Grid Workshop, Academic Computer Center CYFRONET AGH 2007

[150] Raach, Henning and Mitrovic, Jovan: Seawater falling film evaporation on vertical plates with turbulence wires, Desalination 183 (2005) 307-316 http://www.desline.com/articoli/6722.pdf

[151] Raach, Henning and Mitrovic, Jovan: Simulation of heat and mass transfer in a multi-effect distillation plant for seawater desalination, Desalination 204 (2007) 416-422 http://www.desline.com/articoli/8068.pdfRegional Computer Centre, University of Hamburg Home Page: http://www.rrz.uni-hamburg.de/RRZ/e.index.html

[152] Raach, Henning and Somasundaram, Sushan: Numerische Untersuchungen zu welligen Fallfilmen, Vortrag auf der GVC-Arbeitssitzung CFD in Würzburg, 6. März 2007

[153] Rees DC, Howard JB (2003) Science 300:929-931.

[154] Ressourcenverbund Nordrhein-Westfalen (in German), http://www.rv-nrw.de

[155] Richardson, T.; Stafford-Fraser, Q.; Wood, K. and Hopper, A.: "Virtual network computing". IEEE Internet Computing 2 (1998) 33-38

[156] Rietveld, K. Simons: *The differential miscibility of lipids as the basis for the formation of functional membrane rafts*, Biochimica and Biophysica Acta 1376, 467, 1998

[157] Rowstron, Antony and Druschel, Peter: *Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems*, Proc. of the 18th IFIP/ACM International Conference on Distributed Systems Platforms, 2001

[158] RYBKA: http://www. rybkachess.com

[159] Sahai et al: Specifying and Monitoring Guarantees in Commercial Grids through SLA. Technical Report HPL-2002-324, Internet Systems and Storage Laboratory, HP Laboratories Palo Alto, November 2002.

[160] Salomon, D.: "Data Compression: The Complete Reference", 3rd Edition. Springer (2004)

[161] Schafer, M. and Turek S.: 1996, Benchmark computations of laminar flow around a cylinder. Flow Simulation with High-Performance Computation II, E.H. Hirschel ed., Notes on Numerical Fluid Mechanics}, Vieweg, Braunschweig, pp. 547--566.

[162] Schamberger, Stefan: Shape Optimized Graph Partitioning. PhD thesis, Universität Paderborn, 2006.

[163] Schloegel, K.; Karypis, G. and Kumar, V.: Parallel static and dynamic multi-constraint graph partitioning. Concurrency and Computation: Practice and Experience, 14(3):219-240, 2002.

[164] Schloegel, K.; Karypis, G. and Kumar, V.: Graph partitioning for high performance scientific simulations. In The Sourcebook of Parallel Computing, pages 491-541. Morgan Kaufmann, 2003.

[165] Schmid, F.: *J. Phys. Cond. Matt.* 10, 8105, 1998.

[166] Schmidt, K.; Seino, M.; Preuss, A.; Hermann, F.; Ortmann, and Bechstedt, F.: "Organic molecule adsorption on solid surfaces: chemical bonding, mutual polarisation and dispersion interaction", Appl. Phys. A 85, 387 (2006).

[167] Schumacher, Tobias; Lübbers, Enno; Kaufmann, Paul and Platzner, Marco: Accelerating the Cube Cut Problem with an FPGA-augmented Compute Cluster, In Proceedings of the ParaFPGA Symposium, International Conference on Parallel Computing (ParCo), Aachen/Jülich, Germany, September 2007.

[168] Schuttelkopf, A.W. and van Aalten, D.M.: *PRODRG: a tool for high-throughput crystallography of protein-ligand complexes.* Acta Crystallogr D Biol Crystallogr, 2004, *60*, 1355-1363.

[169] Segal, M. and Akeley, K.: "The OpenGL Graphics System: A Specification Version 2.0". Silicon Graphics (2004)

[170] SGE - Sun Grid Engine. http://www.sun.com/software/gridware

[171] SGI: "SGI OpenGL Vizserver - Visual Area Networking". Website: http://www.sgi.com/products/software/vizserver/ (2007)

[172] Simon, Jens: PC² Benchmarking Center, http://wwwcs.uni-paderborn.de/pc2/about-us/staff/jens-simons-pages/benchmarkingcenter.html

[173] SLA Management in a Service Oriented Architecture, K. Djemame, M. Haji, and J. Padgett, In Proceedings of ICCSA'2005, Singapore, May 2005, Lecture Notes in Computer Science 3483, pp.1282-1291

[174] Smiatek, J.: *et al.,* Poster presented at the International Soft Matter Conference, Aachen, October 2007

[175] Smiatek, J.; Allen, M.P. and Schmid, F.: "Tunable slip boundaries for coarse-grained simulations of fluid flow", submitted to European Physical Journal (2007) Solomon, E.I., Sundaram, U.M., Machonkin, T. E *Chem. Rev.* 1996, *96*, 2563.

[176] Somasundaram, Sushan: Validierung von OpenFOAM für wellige, laminare Fallfilme, große Studienarbeit, Universität Paderborn

[177] Somasundaram, Sushan: Numerische Untersuchung des Wärmeübergangs in Fallfilmen mit Stolperdrähten, Diplomarbeit, Universität PaderbornSpringer, LNCS 3485, 2005

[178] Streek, M.; Schmid, F.; Duong, T.T. and Ros, A.: Phys. Rev. E 71, (2005), 011905

[179] Streek, M.; Schmid, F.; Duong, T.T.; Brugger J. and Ros, A.: J. Biotechnology 112, (2004), 79

[180] Streek, M.; Schmid, F.; Duong, T.T.; Brugger J.; Ros, A and Anselmetti, D.: Microelectronic Engineering, 67-68, (2003), 906

[181] Stüben, K.: An introduction to algebraic multigrid. In U. Trottenberg, C. W. Oosterlee, and A. Schüller, editors, Multigrid, pages 413-532. Academic Press, London, 2000. Appendix A.

[182] Tasker, P. W., Stability of Ionic-Crystal Surfaces. Journal of Physics C-Solid State Physics 1979, 12, (22), 4977-4984.

[183] The Alice Experiment, http://aliceinfo.cern.ch/
[184] The ARMINIUS Cluster at PC2: http://www.upb.de/pc2/services/systems/arminius The Globus Alliance, http://www.globus.org
[185] The D-Grid web portal, http://www.d-grid.de
[186] The German Grid Initiative (D-Grid) description, http://www.d-grid.de/index.php?id=104&L=1
[187] The Globus Toolkit, http://www.globus.org
[188] The Open Grid Forum, http://www.ogf.org
[189] The SFB Cluster at PC2. http://www.upb.de/pc2/services/systems/ic
[190] The SysMES project, http://www.kip.uni-heidelberg.de/wiki/SysMes/index.php
[191] Thielemans, K.:" STIR - General Overview, Version 1.3". Website: http://stir.irsl.org/ (2007)
[192] TOGA II: http://www.superchessengine.com/toga_ii.htm
[193] Ullmann, Julian R.: *An Algorithm for Subgraph Isomorphism*, Journal of the ACM, 23:1, pages 31-42, 1976
[194] UNICORE Forum e.V., http://www.unicore.org
[195] UNICORE at SourceForge, http://unicore.sourceforge.net
[196] UNICORE (Uniform Interface to Computing Resources) homepage, http://www.unicore.eu/
[197] Valtiner, M.; Borodin, S. and Grundmeier, G., Preparation and characterisation of hydroxide stabilised ZnO(0001)–Zn–OH surfaces. Physical Chemistry Chemical Physics 2007, 9, (19), 2297-2436.
[198] Van der Spoel, D., et al., *GROMACS: Fast, flexible, and free.* Journal of Computational Chemistry, 2005, *26*, 1701-1718.VirtualGL: "The VirtualGL Project". Website: http://www.virtualgl.org/ (2007)
[199] Vanderbilt, D.: *Phys. Rev. B* (Rapid Communications), 1990, 41, 7892.
[200] Verma, P.: T.D.P. Stack, in preparation.
[201] Viovy, J.-L.: Rev. Mod. Phys. 72, (2000), 813
[202] Walshaw, Chris: The graph partitioning archive. http://staffweb.cms.gre.ac.uk/~c.walshaw/partition/, 2007.
[203] Walshaw, Chris and Cross, M.: Mesh partitioning: a multilevel balancing and refinement algorithm. SIAM J. Sci. Comput., 22(1):63-80, 2000.
[204] Walshaw, Chris: and Cross, M.: Parallel optimisation algorithms for multilevel mesh partitioning. J. Parallel Computing, 26(12):1635-1660, 2000.
[205] Web Services Agreement Specification (WS-Agreement) A. Andrieux, et. Al, OGF, (GRAAP) WG, http://www.ogf.org/Public_Comment_Docs/Documents/Oct-2006/WS-AgreementSpecificationDraftFinal_sp_tn_jpver_v2.pdf
[206] Windows Compute Cluster Server 2003, http://www.microsoft.com/windowsserver2003/ccs/

[207] Woll, C.: The chemistry and physics of zinc oxide surfaces. Progress in Surface Science 2007, 82, (2-3), 55-120.

[208] XBoard and WinBoard graphical user interfaces for chess:http://www.timmann.org/xboard.html Arena, graphical user interface (GUI) for chess engines: http://www.playwitharena.com/

[209] 16th International Paderborn Computer Chess Championship (held from December 27 to 30, 2006): http://wwwcs.uni-paderborn.de/~IPCCC/ipccc2006.html

[210] 15th World Computer-Chess Championship (held from June 11 to 18, 2007 in Amsterdam): http://www.grappa.univ-lille3.fr/icga/events.php