PDC

Progress Report

2000 – 2003

PDC

Progress Report

2000 – 2003

ROYAL INSTITUTE
OF TECHNOLOGY

Cover Image:
The figure shows the velocity distribution for gas tungsten arc welding at 4 different welding speeds. Simulation by Minh Do-Quang, Mechanics Department, KTH on the IBM SP system at PDC.

Published by PDC at KTH.
PDC operates leading-edge, high-performance computing systems as easily accessible national resources and carries out research in soft-ware and tools for use and administration of such systems. The hardware resources and the costs of their operations are largely covered by funding from the Swedish Research Council and KTH. The research is funded from a variety of sources with the majority of research funds coming from the European Commission.
Publisher: Lennart Johnsson
Editors: Mats Höjeberg, Gert Svensson
Graphic Design and Layout: Maria Engström
Formula Typesetting: Anders Selander
ISBN 91-631-6106-0

Contact
Visiting address: Teknikringen 14, plan 4, KTH, Stockholm
Mail: PDC, KTH, SE-100 44, Stockholm, Sweden
E-mail: pdc-staff@pdc.kth.se
WWW: http://www.pdc.kth.se/
Phone: +46 8 790 78 00

Fax: +46 8 24 77 84

PDC

Progress Report

2000 – 2003

PHYSICS

CHEMISTRY

ENGINEERING SCIENCE

PDC PROJECTS

Foreword

Computing has become an integral part of science and engineering disciplines, and data warehousing and analysis are rapidly emerging as significant needs. Simultaneously, the rapid improvement of the capabilities of wide-area and global networks are enabling the creation of distributed resource and collaboration environments that are changing the way science, engineering and many other disciplines are pursued. These trends have been the source of major advances and changes at PDC.

The users of facilities operated and supported by PDC have continually produced groundbreaking research leading to national and international recognition of individuals and groups. The result of these successes has led not only to well deserved recognition but also growing dedicated computational and storage resources for these groups. An example of one such successful research collaboration is the Stockholm Bioinformatics Center (SBC), who chose to use PDC to assist in procurement of their resources, as well as electing to have the PDC staff maintain and operate the center computational and storage resources. In fact, the services PDC is offering successful consortia and research groups add significantly to the resources managed by PDC staff.

Collaboration among Swedish HPC centers increased during the period 2000 – 2003, in particular between NSC, HPC2N and PDC, the three centers operating systems for the Swedish academic community on behalf of the Swedish Research Council (VR). The centers collaborate directly in implementing allocations granted to Swedish researchers by the Swedish National Allocations Committee (SNAC) that began its efforts to evaluate user requests for HPC resources during 1999. The broadest collaboration took place within the SweGrid project, a still active endeavor including major users of high-performance computing and large-scale storage resources and HPC centers at all six major Swedish university sites, and SNIC, the Swedish National Infrastructure for Computing that commenced operations during 2003. The SweGrid project is establishing a Grid prototype in Sweden with each of the six University sites having a 100-node cluster as a resource in this Grid.

During 2000 – 2003, PDC also made significant progress in expanding the research activities at the center in methodologies and tools for next generation computing and data exploration environments and for the improved use of current environments. Internationally recognized experts joined PDC during this period. Increasingly, PhD students are now active in areas relevant to PDC as well as users of its facilities. The research is mostly realized through participation in community efforts towards the construction, operation and support of HPC environments, or towards the creation of standards. PDC's specific focus is on security including authentication, authorization and accounting, quality of service, performance tools and engineering, and data management.

A research partnership with IBM led to the establishment of the KTH LinuxLab. Hewlett Packard Corporation (HP) sponsored PDC's participation in the Open Source Gelato federation. Together with CERN and the UK e-Science program PDC founded the European Grid Support Center (EGSC) that has been joined by HP and the GridLab project. The perhaps strongest recognition of PDC's competence in Grid technology was the invitation from the Globus project to co-found the Globus Alliance together with Edinburgh University, Argonne National Laboratory and the University of Southern California. PDC actively contributed to the formation of the European Grid Forum and has been active in the Global Grid Forum after the US and European Grid Forums merged. During the period of this report PDC participated in five EU funded and several nationally funded projects

In December 2003, PDC moved into a newly renovated building on the KTH campus. The building also houses KTH Network Operations Center enabling an even closer cooperation with the Swedish and Nordic university networks than ever before. The new location offers PDC more than double the capacity for computer equipment and room for future expansion. The fact that the KTH Administration funded the renovation and relocation demonstrates clearly KTH's commitment to PDC. We are now well prepared for a continued expansion of both hardware and research activities.

We look forward to 2004 with great optimism having new facilities, funding for new equipment, interest from several groups in PDC services, and exciting competitively funded research projects under way.

Lennart Johnsson

Director

Overview

PDC was formed in January 1990 to act as a focal point and national forum for research on and use of parallel computers. During the ten years of operation preceding this report the center progressed in terms of size of its research, its impact on the Swedish research community and the capacity of the computational facilities. At that time, the focus was on scalable architectures. But, during this period PDC was in fact also the only Swedish provider of parallel vector resources. Ushering a new era in supercomputing, the Cray J932se was decommissioned in 2001. Finally, in November 2003, the last vector computer available to Swedish academic researchers was decommissioned.

PDC's role as the main provider of significant computing resources to the Swedish academic research community that it was given by HPDR in 1995 was honored by the Swedish Council for Planning and Coordination of Research (FRN), when HPDR was phased out in 1998. When the Swedish Research Council (VR) was formed in 2001 by merging several funding agencies it continued to fund PDC as the main center for providing the Swedish academic research community with High Performance Computing (HPC) resources, which was continued again by the Swedish National Infrastructure for Computing (SNIC) during 2003, its first year of operation. The formation of SNIC put a welcome end to four years of uncertainty of the handling of the government funding of HPC resources. SNIC extended the support to include HPC centers at all six major Swedish university sites. The PDC Director currently serves as the first Chair of SNIC's Strategic Technical Advisory Committee.

From 1999 through 2002 financing of HPC resources were through one year extension contracts, effectively preventing any long-term planning and major investments in HPC resources for VR at PDC, causing a relative decline in those resources compared to resources of successful research groups. But, at the end of 2002, PDC was in a position to make a modest investment in a new computational resource on behalf of VR thereby changing the main computational resource from a tightly coupled IBM SP cluster to a similarly tightly integrated HP Itanium2 cluster with high-speed interconnect and large node memories continuing the long standing tradition of cluster based computations and Linux and of introducing new technology to the Swedish academic community. At the very end of 2003, PDC received its first funding in several years for a resource upgrade. During 2004 PDC will again be in a position to provide a resource that should have a capacity that measurably exceeds that of any dedicated group resource.

PDC's success in providing quality dependable service has lead to independent groups seeking PDC's considerably expertise in systems maintenance and operations, and assistance in procurements. Since early 2002, the services provided by PDC to independent research groups have grown significantly, with the service provided to the Stockholm Bioinformatics Center (SBC), being the most significant. The SBC cluster has been of higher capacity than any other system at PDC since 2002. The user environment and services provided by PDC were chosen by SBC and the customization of both has served SBC well. Additionally, PDC continued to serve the KTH computational consortium Kallsup. The consortium upgraded its computational resources in 2001, at which time the capacity of the Kallsup resources were comparable to those operated by PDC for VR.

During the period covered by this report PDC also initiated production storage services for the Swedish Space Corporation's Odin satellite, and for the SSF funded Large-Scale Genotyping Laboratory at Uppsala University. Discussions regarding large scale storage services have been initiated with three other planned or early stage projects, the PAMELA satellite project, the AMANDA South Pole project, and the Human Proteome Resource project. Together these initiatives have storage needs estimated at a few hundred terabytes over the next few years.

However, planning, operating and maintaining systems are not the only aspects in which the center plays a leadership role. PDC's early involvement in the development and application of Grid technologies has resulted in the center now being an established participant in many internationally renowned projects and a principal player in some of the major European Grid projects. Additionally, through PDC's effort to explore Grids, PDC has also developed a unique relationship among Swedish HPC centers with Sunet, Nordunet, NetherLight, Starlight and the international high-performance networking community.

As a further step in PDC's evolution, the center moved into a newly renovated building on the KTH campus in November 2003. The facility includes a new computer room equipped for high availability operations and security. PDC is now co-located with the KTH Network Operations Center (KTHNOC), allowing not only for a favorable network provision and capability, but also for increased possibilities for advanced network experiments such as the use of dedicated s for end-to-end optical connectivity. The first Nordic demonstration of this new networking technology was carried out in August of 2003 in collaboration with KTHNOC, Nordunet, Surfnet, StarLight, SARA, Karolinska Institutet and the National Center for Microscopy and Imaging Research at the University of California, San Diego.

As witnessed by its unique history and extensive international collaborations, the center has clearly come along way from its inception. This is most dramatically illustrated when looking at the growth of the center's computational prowess from the end of 1996, when the capacity of systems at PDC was a mere 30 GFlops, whereas at the end of 2004, eight years later we expect the total capacity of systems at PDC to be close to 10 TFlops, a more than 300-fold increase. As a matter of reference, the well known Moore's law gives about a 50-fold increase over the same period of time.

This seventh Progress Report of PDC at Kungl. Tekniska Högskolan (KTH, Royal Institute of Technology) highlights some of the research and development activities of the center, services carried out for the Swedish academic community and research teams, and research carried out using PDC facilities during the period January 2000 until December 2003. Further, this report provides a greater context to the contributions to the science PDC's services have enabled in fields such as chemistry, material science, and fluid dynamics. A complete listing of projects can be found at the end of the report.

Research and development at PDC

The evolution in systems architectures and changing balance between computing and storage needs, and change in user base and its way of interaction with the computing and storage environment requires a development of the basic software environment used to manage and access the resources and development of tools and techniques for the efficient use of resources operated and maintained by PDC. A noticeable change during this period was the change from clusters based on proprietary technology in all aspects (processors, networks, operating systems) to clusters based on commodity technologies. The IBM SP uses proprietary processors (Power3), proprietary interconnect (the Colony switch), and proprietary OS (AIX). The SBC IA-32 cluster and the SNIC Itanium2 IA-64 cluster both use commodity processors and Linux (Open Software) for operating systems. The SBC cluster use Fast or Gigabit Ethernet for interconnect, while the Itanium2 use both Gigabit Ethernet and Myrinet. For the Itanium2 cluster Debian Linux is used. The experience gained from our LinuxLab was of great value in building the Linux distribution for the Itanium cluster. Other development activities driven by the needs of our service offering were the integration of accounting, scheduling and allocation systems. The research and development activities are focused on both current environments and environments expected to be of great significance three to five years into the future. The focus includes enhancements of user interfaces to the various resources as well as improved user information about their accounts, submitted jobs, and resources available, tools and techniques for performance engineering of applications, visualization, and computational steering, and various aspects of Grids, in particular security and accounting.

Development of the SNAC application system

Twice a year, the Swedish National Allocations Committee (SNAC) invites researchers to apply for time at HPC2N, NSC, PDC, and SweGrid. Since SNAC's inception, the process of gathering applications for computing and storage resources, distributing the applications for peer review, collecting the reviews, and distributing the final allocation decisions to researchers was achieved mainly using paper. In the spring of 2001, this changed. The SNAC working group (Lennart Johnsson, Peter Münger, Mats Nylén, and Per Öster) asked PDC staff member Mike Hammill to implement an electronic application system to replace the paper system.

A review of current open-source and commercial packages was first undertaken, using as a base a study done by the Association for Computing Machinery (ACM) on conference management software for potential use in the 80-odd conferences the ACM special interest groups put on yearly. In planning a Web tool for SNAC PDC also took into account its experience using such software tools. For example, to collect student projects during the yearly PDC/NGSSC Summer School in high-performance computing, PDC has used both Wimpe http://www.crhc.uiuc.edu/~nicol/wimpe/wimpe.html and CyberChair http://www.cyberchair.org/ with the latter offering the best experience. It was chosen as the basis of the new SNAC application system.

Extensive modifications were however made to CyberChair to make it fit SNAC's requirements. The main changes consisted in developing custom user interfaces for the SNAC applicants, the SNAC reviewers, and the SNAC working group. For example, the SNAC application is now a Web form that collects information from researchers in a format familiar to them from the paper-application days. However, this data is parsed by the system into approximately 300 queriable fields, making it much more useful than the old paper form. All user interfaces conform to the XHTML World Wide Web Consortium (W3C) standards so researchers with virtually any Web browser can submit an application. A backend to handle distributing, storing, and archiving the applications data was also created. The backend stores and serves the data in XML format, complete with a validating DTD.

Ease of use was a primary design goal. Since the beginning, Web-based help has been available, including e-mail or phone contact with the developer. In the fall of 2003, a full-fledged monitoring system was added so others with proper access privileges may also help assure that the system is running smoothly. During each SNAC application period, there are on average only three questions related to using the system itself.

During 2001-2003, there were six SNAC application periods and the electronic application system was used during them all. An average of 68 applications was submitted each period for a total of 406 applications. During this entire time, the system was down only 2 hours.

Performance engineering

PDC has a high level of competence in the field of performance analysis, tools and libraries. PDC staff has contributed substantially to the Performance API (PAPI) project [1], contributed to porting the performance counter patches for the Linux kernel to the AMD64 platform, and developed new tools for performance analysis [3,4]. All these tools are available to users at PDC, documented on the PDC web pages and used in exercises during the course "Introduction to High-Performance Computing" that is held yearly at PDC.

To stimulate both performance engineering of application codes and tools development PDC organized a well received workshop on performance tools in the spring 2003. Dissemination of PDC developed tools to the broader HPC community has been made through presentations in forums such as the Supercomputing conference [2], the Parallel Computing conference and Gelato Federation technical meetings. Much of the tools development is performed through collaborative efforts, such as through PDC's participation in the SEPtools EU funded project on performance tools for parallel program development, or through direct collaboration between PDC staff and developers in other projects, such as the PAPI project at the University of Tennessee's Innovative Computing Laboratory with Jack Dongarra as its Director.

Benchmarking of emerging computational platforms [8] is an integral part of the performance engineering effort. It is made both as a preparation for acquisitions by or for the various parties PDC serves, but also for the development of programming paradigms and tools for new environments. Public benchmarking codes such as stream2, NASPB, lmbench, SKAMPI and others, are used as well as SNIC user codes, and PDC developed benchmark codes [9,10]. PDC personnel have been active in the OpenMP 1.1 and 2.0 standard peer review process and continues to contribute to both the shared memory and distributed memory programming paradigms [5,6,7].

PDC through its good relations with computer vendors was in the position of having early access to AMD x86_64 systems running SUSE Linux. The system made available through AMD Italy allowed us to involve in the Linux perfctr kernel for this platform in close cooperation with the maintainer Mikael Pettersson, Uppsala University. With this kernel infrastructure in place, PDC staff made the port of PAPI to the AMD x86-64 platform.

References

1. http://icl.cs.utk.edu/papi
2. Moore, S. and Smeds, N. "Performance Tuning Using Hardware Counter Data". Half-day tutorial, IEEE Conference on SuperComputing, Denver, Colorado, Nov 10-16, 2001.
3. Ekman, P., i2prof performance metric extractor for IA64 processors. Software available at http://www.gelato.org
4. Ekman, P. and Mucci, P. iotrack application I/O tracer Software available at http://www.pdc.kth.se/~pek/iotrack/
5. Smeds, N. "OpenMP application tuning using hardware performance counters". OpenMP Shared Memory Parallel Programming: International Workshop on OpenMP Applications and Tools, WOMPAT 2003, Toronto, Canada, June 26-27. Springer LNCS, Vol. 2716, Springer Verlag, 2003.
6. Russell K. Standish, Clinton Chee, Nils Smeds, "OpenMP in the Field: Anecdotes from Practice". International Conference on Computational Science 2003. In Springer LNCS, Vol. 2660, pp 637-647, 2003.

7. G. Almasi, C. Archer, J. G. Castaos, C. C. Erway, P. Heidelberger, X. Martorell, J. E. Moreira, K. Pinnow, J. Rattermann, N. Smeds, B. Steimacher-burow, W. Gropp, B. Toonen. “Implementing MPI on the BlueGene/L Supercomputer”. In Euro-Par 2004 Parallel Processing 10th International Euro-Par Conference, Pisa, Italy, August 31-September 3, 2004, Lecture Notes in Computer Science, Vol. 3149. Danelutto, Marco; Vanneschi, Marco; Laforenza, Domenico (Eds.), 2004

8. Andersson, U. and Ekman, P and Öster, P. Performance and performance counters on the Itanium2 - A benchmarking case study. In ParCo2003 proceedings, 2003.

9. Andersson, U. Yee_bench-A PDC benchmark code, TRITA-PDC 2002:1, KTH, November 2002.

10. Andersson, U. The GemsTD benchmark suite, TRITA-PDC 2003:2, KTH.

System development work

Hierarchical storage management

During this period a major change was undertaken in the hierarchical storage management (HSM) system at PDC. As plans for the decommissioning of the KALLSUP Cray J932 machine hosting the old Data Migration Facility (DMF) HSM system went ahead IBM’s TSM-HSM software was picked as a replacement HSM system. However, after a six month effort to migrate to the new software the TSM-HSM software was abandoned due to persistent technical problems. A reinvestment in DMF was made on an Irix platform from SGI. The migration to DMF was successfully performed in two weeks.

At the end of 2003 DMF was managing 6TB of data in 1.5 million files on tape, with a 146GB disk cache.

Operating systems - Linux

During the period a major effort was also undertaken to prepare for a transition to Open Source operating system software (Linux). The experience gained from the LinuxLab was fundamental to the decision to use Debian for the Itanium2 cluster instead of for instance Red Hat or Suse distributions that by the time of the Itanium2 cluster installation would incur significant cost.

File systems

In the area of systems development the search for flexible, reliable, scalable and high-performance file systems for clustered environments is of high priority at PDC, like many other HPC centers. During 2002 and 2003 PDC has evaluated both the Parallel Virtual File System (PVFS) and the Lustre file system and in both cases, the LinuxLab cluster was used as testbed. Both file systems were found to be promising but not yet functional enough to deploy in our production environment. PDC will, however, keep following the development of both file systems closely since they are both making steady progress. Also, the demand from our users for these kinds of file systems is very big and will only increase as more and more of PDC's clusters will be running Linux in the future.

Scheduling

Easy, originating at Argonne National Labs, was introduced at PDC at the time the IBM SP system was acquired. The reason was simple - it worked. It did offer reliability, predictability, high utilization, and was simple to maintain. Over the years several other systems, with similar capabilities, has seen the daylight; i.e. Maui+PBS. But, we have found that maintaining and enhancing our modified EASY scheduler is the most cost effective approach to systems management.

PDC enhancements of Easy includes making Easy Kerberos and AFS-aware and a more general and configurable fifo/back-fill algorithm and a feature enabling users to make reservations of resources in future time. Furthermore, jobs can be coupled to each other so that they i.e. can execute in predefined order. PDC enhancements also include improved flexibility in configuring the system. For instance PDC staff can, while still maintaining high utilization, define future service windows etc.

Projects in the VR-cube facility

In the area of virtual reality and visualization the major PDC resource, the 6-surface VR-cube, was relocated because the space where it was housed, an old boiler room, would become part of a new campus library at KTH. This world's first six wall cave has received great international attention and a number of projects and activities involving a vast range of disciplines have taken place in this immersive environment. The EU-funded project VirtualFires has used the VR cube as its main visualization resource. A new EU-project called Uni-Verse, awarded in 2003 with a start in 2004, will develop a distributed VR system for graphics, audio and acoustic simulation. Free time in the VR cube has been used by artists who have developed a number of applications. During KTH's 175 year celebration some of these art pieces were shown to the public with considerable interest and large coverage in the media (Table 2). Table 1 below summarizes the number of demonstrations performed in the VR cube. Other research and developments projects in the VR cube can found in the following list:

- Kai-Mikael Jää-Aro, Mika Iltanen, Tamperi University of Technology, Ralph Schroeder, CTH, Anthony Steed, University College London, Collaborative Virtual Environments, 2000

- Kai-Mikael Jää-Aro, Holland Cheng, KI, Visualization of Biological Data, 2000-
- Kai-Mikael Jää-Aro, Sonification of Molecular Dynamics 2000-
- Niklas Jacobsson, CAD to CAVE, M.S. Thesis, 2000
- Erik Engquist, Steering and Visualization of Electro-Magnetic Simulations Using the Globus Implementation of a Computational Grid, 2000
- Per Ekman: Visualizing Glacier Ice Flow in an Immersive Environment, M.S. Thesis, 2000
- Kicki Fricsh, Comparative Study of the Working Efficiency Using Proprioception in Highly Immersive VR Environments, 2000
- Teresa Oróñez, Use of 3 D Visualization Platforms in Radiotherapy Treatment, M.S. Thesis, 2000
- Harald Barth, ANTI, A Network Transparent Infrastructure for VR devices, 2000-
- Kai-Mikael Jää-Aro, Bertil Dorch, SU, Visualization of Astro Physical Data 2001-2002
- David Hansson, Om öppenhet och slutenhet i arkitektur, M.S. Thesis, 2001
- Niklas Jacobsson, Mattias Claesson, An extensible simulation and navigation system "Navier", 2001
- Thomas Weber, Gert Svensson, A simulation environment for spatial sound, 2001
- Kai-Mikael Jää-Aro, Ambjörn Naeve, Gustaf Taxen, The Cybermath Collaborative Virtual Envionment, 2002-
- Anders Sandberg, Legba - Finding clusters and visualizing of graph data, 2002-
- Sarah Amandusson, Auditory display and the VTK sonification toolkit, M.S. Thesis, 2003
- Horacio Palomino, Visualization of 3D Numerical Simulations in Astrophysics using the VR-cube, M.S. Thesis, 2003
- Elias Sundström, Stefan Görtz Virtual-Reality Environment for Visualization of Unsteady Three-Dimensional CFD Data, M.S. Thesis, 2003

Table 1
Number of demonstrations in the VR cube

	KTH departments and administration	Academic courses	Other universities	Companies and organizations	Schools	International guests
Year	# of groups	# of courses	# of groups	# of groups	# of groups	# of visits
2000	11	13	8	47	6	7
2001	18	13	8	23	5	15
2002	16	15	3	21	4	9
2003	8	10	4	10	2	6
Total	53	51	23	101	17	37

Table 2
Artists in the VR cube

Bino and Cool

- YGGDRASIL Distributed VR art performance 2002
- VOUNI - Reconstruction of an ancient Cypriotic palace 2003

Samara Braga

- Checkmate 2002

Teresa Wennberg

- BRAINSONGS (Welcome to my brain) 2001

Other major R&D projects

Below is a listing of other major R&D projects at PDC during 2000 – 2003:

PDCTTN 1997 – 2000, EU-project.

A project to disseminate information about HPC in Small and Medium Sized Enterprises (SME) in Europe. A number of offices, Technology Transfer Nodes were set up including one at PDC. The PDCTTN also initiated and managed 6 development projects, similar to ordinary European RTD projects, together with Nordic SMEs.

JACO3 1998 – 2000, EU-project.

Java and CORBA based Collaborative Environment for Coupled Simulations, was a project creating an environment for distributed simulation in a Grid environment.

SEPTools, 1998 – 2000, EU-project
The SEPTools program was a tool integration project creating an integrated set of porting tools appropriate to the job of porting industrial Fortran codes to parallel hardware.

SiMPLe, 1999 – 2000
A collaboration between PSCI, Polhem and PDC where an existing research finite element code for welding and cutting problems was parallelized for distributed memory computing platforms.

BINS, 1999 – 2002, SSF funded
A joint project between PDC and KI investigating the use of high-performance computing and data storage to store, retrieve and analyze brain image data

ENSCUBE, 2000 – 2001, EU-project
An EU funded project investigating cooperative environments with emphasis on the need for the air craft manufacturing industry.

NeuroGenerator, 2000 – 2003, EU-project
EU funded follow-up project to BINS, building a database for brain images including automated analyzes and query tools.

ODIN 2000 – , Swedish Space Corporation
Storage services for Swedish Space Corp. for the ODIN research satellite. The ODIN project is a cooperation between four countries: Sweden, Finland, Canada and France.

VirtualFires, 2001 – 2004, EU-project
A EU-funded project investigating the use of high-performance computing and high-end visualization for studying the behavior of fires in environments such as long vehicle-tunnels.

Service Activities

The largest service activity at PDC is the integrated high-performance computing and storage services environment it provides the Swedish academic community, first on behalf of HPDR, then FRN, then VR followed by SNIC on behalf of VR. Computer systems maintenance and operations services have also been provided to the KTH Kallsup consortium, including the housing of the Kallsup equipment in the PDC machine room. The Kallsup consortium includes the departments of Chemical Engineering, Civil and Environmental Engineering, Energy Technology, Machine Design, Mechanics, Numerical Analysis and Computer Science, Solid Mechanics and Structural Engineering. The consortium needs in terms of system architecture and configuration have been sufficiently well aligned with the needs PDC serves for VR that the Kallsup resources have been integrated with the VR resources and hence both sets of users have had access to a larger shared resource than either party otherwise would have had.

A new significant group that entrusted PDC with the housing, operations and maintenance of its resources during the period of this report is the Stockholm Bioinformatics Center (SBC) a joint initiative between KTH, Stockholm University and Karolinska Institutet and the Foundation for Strategic Research (SSF). The center has expanded its computing resources about every 12 months and since the beginning of 2002 the theoretical peak performance of the SBC cluster has exceeded that of any of the other computing resources at PDC. This is however expected to change when the cluster acquired for the SNIC funds awarded at the end of 2003 is installed.

PDC is also a partner in the SweGrid project that started in earnest in 2003. The SweGrid project is a collaboration between several application groups and SNIC affiliated university HPC centers. As part of the SweGrid prototype grid PDC received a 100-node Intel P4 cluster in November of 2003 after PDC's relocation to its current location.

PDC has also offered storage services in various forms to various groups regardless of their geographic location. One particularly interesting service is the one offered to the SSF Large-Scale Genotyping Laboratory at Uppsala University. In this case PDC provides a storage service in a form in which the file systems and associated hierarchical storage system appears as a local resource to the users at the Uppsala laboratory. In today's Grid terminology this can be viewed as the laboratory and PDC forming a virtual organization in which users at the laboratory have transparent access to files at PDC.

PDC also provides storage services to the Swedish Space Corporation's Odin satellite Astronomical and Atmosphere research project. In this case data from the Odin satellite is sent from the satellite to a receiving station in Kiruna, and from there via Sunet, the Swedish University network, to PDC where it is accessible to members of this joint Swedish, Finish, French and Canadian project. In addition, PDC also has provided storage and backup services for the Department of Electrical Engineering at KTH and KTHNOC. We also have a joint back-up agreement with HPC2N through which the two centers provides backup services for each other. The service and service related development aspects are described in some detail below.

Hardware

A number of changes to the computational resources managed by PDC have taken place in the period 2000-2003. The governmental funding for national computational resources was flat through 2001, and after HPDR was abolished in 1998 there was no strongly focused attention to HPC needs from the government until 2003 when SNIC, the Swedish National Infrastructure for Computing commenced operations. Both aspects had a dampening effect on the vitality of high-performance computing for academic research. With SNIC becoming operational during 2003, PDC received its first three year staff contract for services to the Swedish academic community, and towards the very end of the year funds for procuring a tightly integrated compute cluster. Figure 1 shows the changes in platforms operated and maintained by PDC through the end of 2003, with the SBC cluster being the one of highest capacity at PDC from the beginning of 2002. This cluster is based on 32-bit processor architectures, as is the LinuxLab cluster. Platforms funded by

Figure 1
Development of theoretical peak performance of the computational resources available and maintained by PDC during the period 2000-2004. Dashed lines indicate machines that are taken out of production mode and run without maintenance contracts. In the case of IBM SP Strindberg and SGI Boye this is due to the cost savings enforced by the SNIC administration.

As the capabilities of the computing equipment grows the infrastructure in which the various platforms are integrated needs to be improved as well. In particular, data storage capabilities and the ability to move data between various computing platforms and elements of the storage hierarchy must be upgraded. In fact, the increased emphasis on data in many projects implies that critical elements of the infrastructure need to be improved more rapidly than the compute capabilities.

During 2000 – 2003 both internal and external routing capability has been upgraded. These upgrades were made to meet the demand of internal data exchange as well as to accommodate users to take full advantage of the SUNET backbone upgrade during the period. Internal networking infrastructure has moved from being based on FDDI and Fast Ethernet to be based on Gigabit Ethernet.

To meet the needs for a number of data intensive projects as well as comply with the storage requirements of SNIC-funded researchers the data storage facilities have been expanded, though the growth rate in the demands towards the end of 2003 caused a severe challenge in our abilities to service that need with the funds made available to us.

The major changes during 2000 – 2003 are summarized in Table 1.

Table 1
Major system configuration changes

Time	**System**	**Activity**
Q1 2000	VR cube	Relocation to Brinellvägen, KTH.
Q2 2000	Nighthawk	Installation of an 8 nodes IBM Power3, 8-way SMP, 222 MHz
Q4 2000	SBC	Installation of the first cluster dedicated for bioinformatics; 8 nodes Athlon 900 MHz 12 nodes dual PIII 866 MHz
Q2 2001	Kallsup2(40CPUs) Nighthawk (8 CPUs)	Installation of a 3 nodes IBM Power3 16-way SMP, 375 MHz for the KTH KALLSUP consortium
Q3 2001	HSM	SGI Onyx2 replaces the Cray J932 as HSM server
Q3 2001	Cray J932se	The KALLSUP Cray J932se is decommissioned
Q1 2002	SBC	Second phase of the bioinformatics cluster; 80 nodes Athlon XP 1400 MHz (1600+)
Q3-2002	HSM	The HSM server is upgraded to an SGI Origin300
Q4 2002	Networking	2 Extreme Black Diamond routers installed
Q1 2003	SBC	Third phase of the bioinformatics cluster 80 nodes Athlon XP 2166 MHz (3000+)
Q2 2003	SBC	32 nodes Athlon XP 2166 MHz (3000+)
Q3 2003	Lucidor	New major high-performance compute platform; 90 nodes dual Itanium2 900 MHz (McKinley)
Q4 2003	Relocation to new building	
	Fujitsu VX	Decommissioned
	IBM SP (Power2 and PowerPC based parts)	Decommissioned
Q4 2003	SweGrid	New compute platform for SweGrid with 100 2.8 GHz Intel P4 processor

Resource utilization and scheduling

During 2000 – 2003 the main compute resource for the Swedish academic community was the computer known as Strindberg, an IBM SP multi-processor platform co-financed by the Kallsup consortium and VR and shared between these two groups. The distribution of the total number of node hours delivered by PDC to SBC, Kallsup and the Swedish academic community through SNAC are shown in Figure 2.

Of the users receiving access through allocations from SNAC, users at KTH and the Stockholm University accounts for about 35% each with users at Chalmers accounting for about half as much and Uppsala University users accounting for slightly less. It is interesting to note that from 2000 – 2003 the relative fraction of local users has grown. Whether this is a consequence of an aging resource or due to other reasons is an interesting question.

Figure 2
Users of the IBM SP Strindberg at PDC during 2000 – 2003 in terms of node hours.

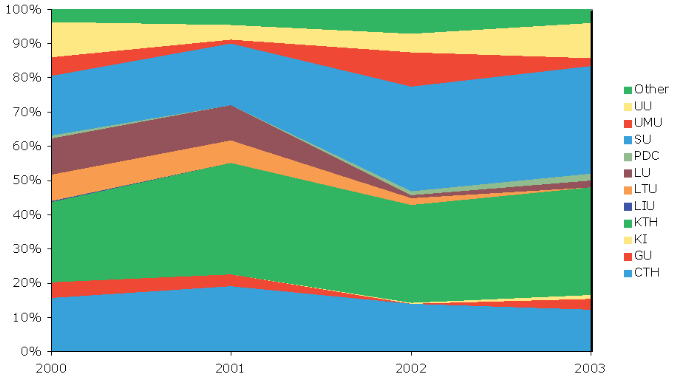


Figure 3
Relative time consumed by user categories for the IBM SP Strindberg at PDC.

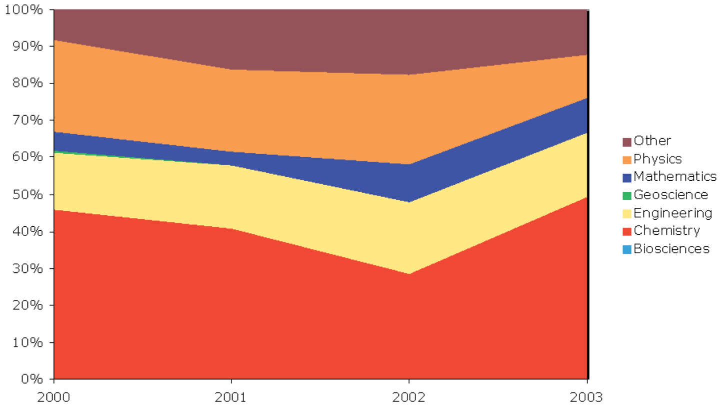


Figure 3 shows the distribution of compute time for the IBM SP Strindberg over scientific areas. Physics, chemistry and engineering sciences are the dominating areas. The most noticeable change is the decrease in resources used for phyiscs during 2003. However, physics represent about 35% of the node hours consumed on the Itanium cluster that was made available about mid 2003, as shown in Figure 4. Within engineering computational fluid dynamics stands for the major fraction of time. Mathematics actually represents the combined use by the applied mathematics and computer science disciplines.

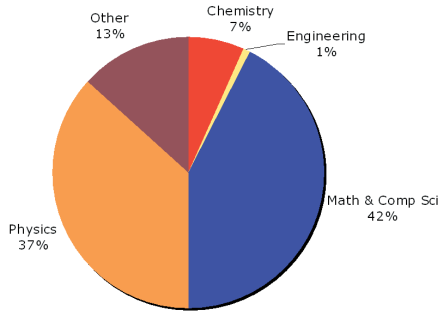
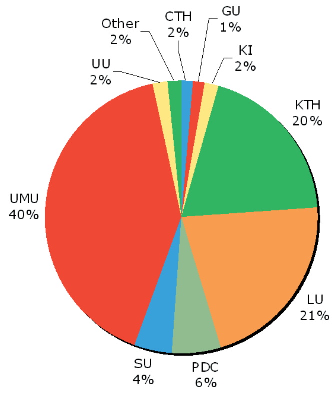


Figure 4
Use of the Itanium cluster installed during the spring of 2003.



For the Itanium cluster use it is interesting that for the new Itanium cluster, users from Uppsala University is the largest group consuming close to half the resource with users from Lund and KTH each responsible for about 20% of the resource consumption, unlike the case for the aging Strindberg system. Thus, for the new resource local users only account for about 25% of the consumption. We believe this is due to the Itanium cluster being a new and highly desirable resource.

Accounts and resource allocation

It is easy to become a user at PDC. PDC has allocations to accommodate a variety of users' needs. In particular, there are four types of allocations: test allocations, local research allocations, SNAC research allocations, and class allocations. Regardless of which type of allocation a user gets, he or she receives the same interface to the computing environment and the same high level of user support.

Typically, a new user first applies for a test account. With a test account, users receive 50 hours of wall-clock computing time per month for six months. This allows the user to perform preliminary tests on available hardware to find out what resources are needed to continue his or her research. If a user needs more than 50 but less than 1000 hours of computing timer per month, a user may apply directly to PDC for a local research allocation. These are more convenient to obtain than SNAC research allocations because they are granted continuously throughout the year; however, no more than 1000 hours may be obtained this way.

If a user needs more than 1000 hours of computing timer per month, he or she applies for a research allocation. Applications for research allocations are handled by the Swedish National Infrastructure for Computing (SNIC). This national application procedure for high-performance computing resources allows for coordinated review of proposals, optimized use of available resources, and a unified application procedure at one or more of its members centers (HPC2N, NSC, PDC, and SweGrid). Starting from the spring 2001, applications to SNIC have been handled by the online Web submission system developed for SNAC by PDC. Applications are taken twice yearly and the corresponding allocations begin January 1 and July 1 for six- or twelve-month durations.

Teachers and students represent a special class of PDC users. Their needs are often temporary but intense. Typically, groups ranging in size from a small handful to 60-70 people need accounts all at the same time, with priority access during lab hours, and sometimes with access to special software packages. Class accounts have been a long-standing tradition at PDC. During 2000-2003, there was an increased interest in these accounts. Accordingly, in March 2003, PDC created a new class-account application form to better handle the requests. With the new online Web application form, teachers can easily get accounts for everyone in their class by answering a few questions and submitting a list of their students.

When a person is granted an allocation at PDC, an account is created that grants access to permanent disk space, access to backups of this disk space, access to one or more of PDC's machines, and other resources that depend on the person's particular request or grant. The permanent disk space is in a distributed file system, AFS, which means that a user sees all his or her files regardless of which machine they log into. Access to a PDC machine allows access to all the non-licensed software installed on that machine. Typically, other resources include access to batch job-submission systems and long-term, high-volume storage (HSM). More particular resources include such things as access licensed software, parallel file systems, and the VR cube, just to name a few. Certain groups of users, such as Kallsup Consortium members and SBC researchers typically want access to a common set of PDC's resources.

Regardless of the type of allocation granted, or the group a researcher comes from, PDC follows a unified procedure for creating their accounts. For example, students' accounts are just like any other PDC account. Since today's students are tomorrow's researchers, when a student becomes ready to further pursue their computing interests, his or her PDC account is already ready. He or she only needs to secure a time allocation through an appropriate scientific review, just like any other researcher.

For user convenience PDC has strived to allow a user to have the same username on all systems to which they have access. During the first part of 2000-2003, PDC shared a username database with NADA, the academic department using PDC resources for many of its courses and the home department of many scientists using PDC resources for their research. This shared database made it possible for all NADA users to have the same username at PDC and vice versus. During the latter part of 2000-2003, KTH implemented a university-wide database making it possible for anyone getting an account on any machine at KTH to have the same username across all KTH machines. PDC registers all its users in this KTH-wide database. This database automatically generates a unique and easy-to-remember username for a user. If a user has another username they would prefer – for example, if they are coming from outside of KTH - the preferred name can be used as long as it is not already taken in the KTH-wide database. Once a username is obtained from the KTH-wide database, a PDC administrator executes a script that handles all the details of account creation. It creates the disk space for the user, registers them in the login password file, etc. The password is sent to the user via regular postal mail. For research accounts, inclusion in the correct accounting group is done automatically from the XML data provided by SNAC. For other users, this is handled manually after verification of information provided on the user's account application form. Table 2 show the courses using PDC resources and the number of students

Table 2
Enrollment in courses using PDC resources

2000

Enrolled	Title	Course Leader
69	PDC HPC Summer School 2000	Mike Hammill, PDC
17	Parallella Datorsystem	Mats Brorsson, IMIT, KTH
19	Computational Physics	Tore Brinck, Physical Chemistry, KTH
7	Chalmers HPC-course	Thomas Ericsson, Math, Chalmers
~30	NWChem/ECCE Workshop	Olav Vahtras, PDC

Total: 142

2001

Enrolled	Title	Course Leader
46	PDC HPC Summer School 2001	Mike Hammill, PDC
13	Parallella Datorsystem	Mats Brorsson, IMIT, KTH
~50	Computational Physics	Tore Brinck, Physical Chemistry, KTH
11	Numeriska metoder för parallelldatorer	Michael Hanke, Nada, KTH
9	Chalmers HPC-course	Thomas Ericsson, Mathematics, Chalmers
~10	HPC-course in computing	Lennart Johnsson, University of Houston
~60	10th ABC meeting at PDC	Tore Brink, KTH; Olav Vahtras, PDC
36	CAVE Programming Workshop	Kai-Michael Jää-Aro, PDC

Total: 225

2002

Enrolled	Title	Course Leader
56	PDC HPC Summer School 2002	Mike Hammill, PDC
~40	Parallella Datorsystem	Mats Brorsson, IMIT, KTH
~50	Computational physics	Tore Brinck, Physical Chemistry, KTH

Total: 146

2003

Enrolled	Title	Course Leader
69	PDC HPC Summer School 2003	Mike Hammill, PDC
16	NGSSC Grid computing course, summer 2003	HPC2N, NSC, PDC. Local coordinator: Olle Mulmo, PDC
43	Parallella Datorsystem	Mats Brorsson, IMIT, KTH
54	Computational Physics	Tore Brinck, Physical Chemistry, KTH
18	Scientific Computing Course	Björn Sjögreen, Nada, KTH
14	Numeriska metoder för parallelldatorer	Michael Hanke, Nada, KTH
29	Abaqus Workshop	Jan Hernelind, FEM-Tech AB; Ulf Andersson, PDC
20	Performance Measuring Workshop	Nils Smeds, PDC

Total: 263

Scheduling

Most PDC-systems are configured to maintain turnaround-time goals. Users can query the system for configuration information, historic and current job status as well as queue information. The same user interface is used for access to interactive nodes as for submission of batch jobs.

There is no hard quota enforcement in the PDC scheduler. When a quota assigned to an account is consumed, subsequent jobs are assigned to a "free" queue, i.e., a queue of jobs that when executed are not charged node-hours. The scheduler also supports setting dependencies between jobs, i.e., job X2 will not start before job X1 has started/is completed. It is also possible to schedule a job not to start before a certain time, which makes it possible to book nodes in advance for interactive work on dedicated nodes. Nodes in interactive mode are a shared resource that is immediately available. Quick turn-around time on short jobs is crucial when developing code.

The PDC scheduler supports different job classes and limits for job submission to these classes. These classes and limits are set by the system administrator with a typical setup being a weekend class for jobs that are between 15 and 60 hours, a night class (]4h,15h]), a day class (]1h,4h]), and a short class (]0m01s,1h]). During office hours a number of nodes are set aside for jobs shorter than 4 hours, and most nodes only run jobs longer than 15 hours during the weekend.

Job accounting is based on groups not individual users. Whenever the aggregate requested time of all queued jobs belonging to a certain job class and (group) account exceeds the max wall clock time limit the account becomes "saturated" for this job class. The next submitted job for this job class and account will then be held until saturation is revoked. A job that receives a hold status when submitted will not receive any queue number. Saturation can be revoked either by jobs starting or jobs being released.

PDC helpdesk

The PDC helpdesk is the primary point of contact for users of PDC resources. The helpdesk originated when PDC became the primary national computational resource for academic computing in 1996. The helpdesk can be contacted by users by either e-mail or telephone for two-way communications. The helpdesk also is responsible for the resource usage information on the PDC web site. In its original version, the helpdesk was staffed by KTH students during office hours who took care of answering phone calls and e-mail questions from users and directed the questions they were not able to answer themselves to suitable person at the regular PDC staff that could assist the user in addressing his or her problem.

The helpdesk student staff provided an excellent way for interested students to get involved in high-performance computing and allowed regular staff to concentrate on the more involved user problems as they quickly became aware of and recognized the more common and easily solved problems users might have in using the PDC resources. Several people at PDC and organizations that utilize PDC resources have been once recruited from the helpdesk. As students became more proficient they were encouraged to take part in more challenging tasks. Thus many early PDC projects have been completed in part with helpdesk personnel. Such projects include e.g. Windows versions of Kerberos travelkits and projects in the VR cube. The help desk students were also essential in staffing the VR cube.

Due to budget cuts the last years, students are no longer part of the helpdesk. Instead all helpdesk activity is taken care of by regular staff. In order to try to minimize the impact on regular operations, the helpdesk service is rotated among staff members on a weekly basis. The person in charge of the helpdesk responds to user questions directly, or when detailed expertise is needed, forwards the question to the right person at PDC. Questions arriving to the helpdesk via e-mail are automatically assigned an issue ticket to assure traceability and continued effort for problems that cannot be resolved immediately.

Most questions received at the helpdesk stem from new users at the systems. This is of course no surprise as each computational center in Sweden is slightly different from each other. Frequently asked questions are used to improve on the user guides available at the PDC web site that guide new users in the more common tasks. Guides are e.g. available for compiling and running programs on the different systems, accessing the storage services and using the AFS networked file system for cooperative work.

Outreach

The PDC Summer School

Moore's law predicts the speedup of computer hardware, but education is essential for adoption and efficient use of systems. One of the more visible examples of computer education at PDC is the PDC Summer School, "Introduction to High-Performance Computing." It is offered during the last two weeks of August every year – and the period from 2000-2003 was no exception. The goal of the PDC Summer School is to give students a strong conceptual foundation in high-performance computing as well as experience with practical aspects through lab work and a follow-up project.

To help the students make the most of the two-week course, a seasoned cadre of professors and professionals from around Sweden are joined by world-famous practitioners to teach the most important topics in HPC. The topics include parallel programming (OpenMP, MPI), modern computer architectures, parallel algorithms, efficient programming, Grid technology, tools (debuggers, profilers, tracers), and case studies.

These topics are covered by such well-known leaders in the field as Lennart Johnsson (University of Houston/KTH), Björn Engquist (Princeton/KTH), Mats Brorsson (KTH), Thomas Ericsson (Chalmers), and Erik Hagersten (Uppsala University/Sun). Typically eleven PDC staff members also teach each year. During 2000-2003, special presentations were given by such luminaries as Steve Tuecke, lead architect of the Globus Project, and Linus Torvalds, inventor of the Linux kernel.

Course time is roughly divided half and half between lectures and lab. For lab work, students receive personal PDC accounts with access to all production and experimental machines. During 2000-2003, students had access to, among other resources at PDC, the HP Itanium2 cluster (Lucidor), the IBM SP (Nighthawk/Strindberg), the Fujitsu VX/3 (Selma), the Cray J932se (Kallsup), the Silicon Graphics Onyx2 (Boye), and the VR-Cube. In 2003, PDC Summer School students were the first large group of test users who had access to the HP Itanium2 cluster while it was in its pre-production phase. Being at the leading edge of high-performance computing is part of what the class is all about. After the course is over, each student is expected to complete a project, which is often related to his or her research area. Typically, one to four students work together on a project. Projects are as varied as the students, with titles ranging from "FDTD code for solving the electromagnetic field in the inner ear of a mobile phone user's head" to "Molecular dynamics method for docking ligands to receptors." During 2000-2003, the most popular project topics – in descending order – were physics, computational science, and computational fluid dynamics.

PDC provides the students with a tutor to help oversee the completion of their projects. In addition to helping answer questions on the parallel implementation of the students' code, the tutor helps grade the work in conjunction with the course examiner, Jesper Oppelstrup, professor in Numerical Analysis and Computer Science at KTH. Questions about compilers, job submission, and other practical details are handled by the PDC help desk. This maximizes the help available to the students while minimizing the load on any one helper.

Typically, all students complete, and receive points for, their lab work. The after-course project is intended to require approximately two to three weeks of full-time work, and it represents a greater challenge. 127 of 217 students, or 60%, finished and received grades on their projects during 2000-2003. Of the remaining students, the majority chose to delay working on their projects and may turn them in at a later date. One of the many benefits of the course for the students is learning how to manage their projects and their time.

The students represent the next generation of researchers and are the life of the course. They come from a variety of programs, including the National Graduate School in Scientific Computing (NGSSC), the International Masters Program in Scientific Computing, and the Parallel and Scientific Computing Institute (PSCI). The course is part of the KTH curriculum and is open to KTH students as well as Ph.D. students world-wide. During 2000-2003, a total of 240 students took the PDC Summer School, averaging 60 students per year. The students' programs and affiliation are shown in Figures 1 and 2.

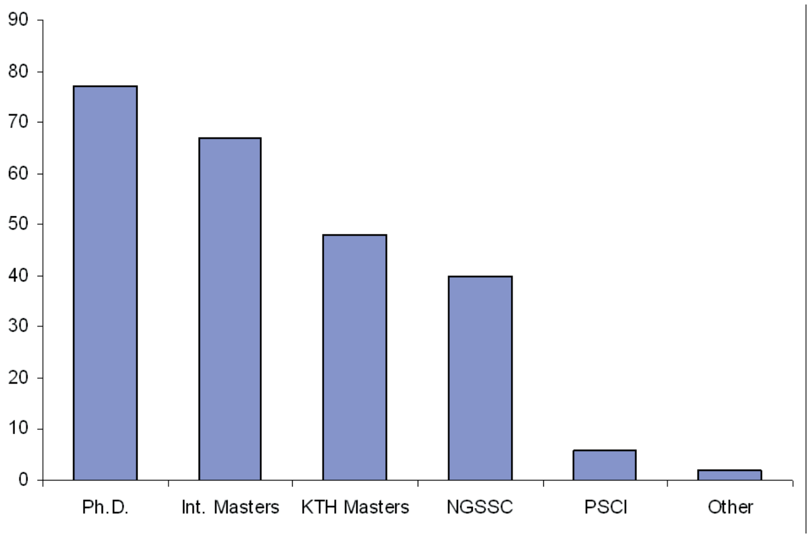


Figure 1
PDC Summer School students 2000-2003 by program

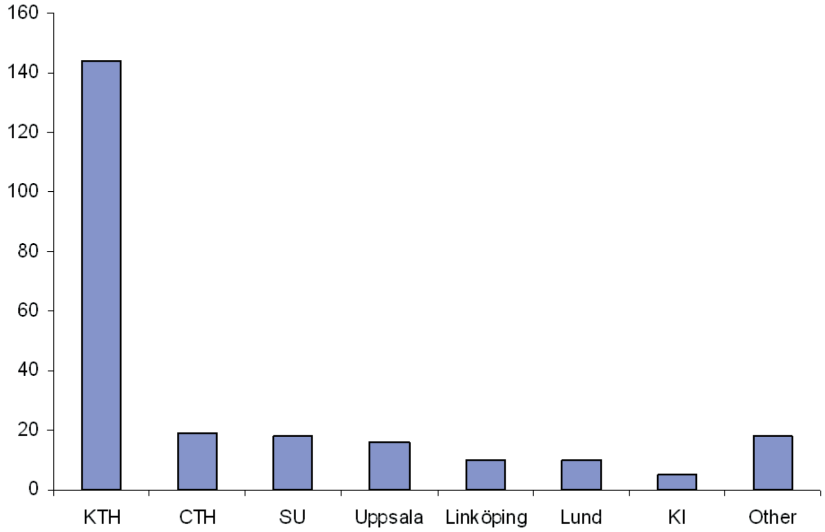


Figure 2
PDC Summer school students 2000-2003 by affiliation

In the course evaluations during 2000-2003 the course received an overall rating of 3.6 out of 4.0. The highest rating was given for computer resources, which received 3.7 out of 4.0. The following two quotes from the 2003 course evaluation captures the students perception of the course:

- "The course is very important for my work. I am happy that this course exists."
- "Well organized and structured. Good with a basic introduction and then more special fields!"

The student feedback has been very valuable in developing and evolving the course to its current form. Course organization has always been an important aspect of the course. Orchestrating the review of the after-course projects is a good example. After the course, students and their tutors may relocate off campus and indeed anywhere in the world. In order to facilitate the submission and review of projects, PDC took open-source conference-organization software and retooled it to allow students and tutors to submit their projects and reviews via the Web at any time from anywhere. During 2000-2003, an e-mail reminder system was added to help students and reviewers keep on track. Two other organizational improvements took place during this period: stricter deadlines were imposed for project completion, and the number of lab assistants was increased from four to six. The extra lab assistants were Ph.D. students in the department of numerical analysis at KTH, all of whom had themselves attended the PDC Summer School.

The teachers have been updating the content of the course to keep the students up-to-date with the evolution of HPC. During 2000-2003 more emphasis was put on the following topics:

- OpenMP
- Debugging and profiling tools
- Case studies, including more areas of science as well as deeper coverage of specific cases
- Grid technologies
- For student projects, creating a performance model to predict parallel behavior, comparing that to actual measured performance, and explaining the differences

The PDC Summer School is a win-win situation for PDC and the students. Not only do the students gain valuable knowledge that will help them with their research, but PDC benefits by encouraging more efficient use of its resources by the next generation of its users. Documentation and examples updated for the class are part of PDC's general user documentation available on the Web. This means that all of PDC's users, and indeed anyone interested in HPC, also benefit from the PDC Summer School.

During 2000-2003, funding for the PDC Summer School was provided by NGSSC, which is funded by the Swedish Foundation for Strategic Research, as well as KTH.

For more information, see the course Web pages at http://www.pdc.kth.se/training/SummerSchool/Overview/.

NGSSC Grid School

In 2003, the first NGSSC course on Grid computing was held at PDC. This course was created in collaboration between HPC2N, NSC, and PDC. The intent is to give the course on a rotating basis at the various HPC centers. The course was designed to give a broad overview of the concept of grid computing and state-of-the-art grid software, understanding and practical experience of how computational grids can be used to further research in science and technology, and to review some current and future directions of grid computing research and development.

The course was held at August 18-26, 2003. Attendance was 7 students from KTH, 3 from NGSSC, 2 from Umeå, 2 from Karolinska Institutet and 2 from Uppsala. The course consisted of 18 "hours" of lectures (45-minute sessions), of which 2 were co-allocated with the PDC Summer School, when Steve Tuecke, lead architect for the Globus project, gave a talk about the future of Grid computing.

In addition to lectures, 30 hours of planned laboratory time where assistants supervised. There were 5 assignments, ranging from learning how to submit a simple "hello world" job using Nordugrid command-line tools, to write a broker that would spawn around thousand 20-minute jobs in order to perform a brute-force attack on a common crypto algorithm and extract the encrypted message ("good job").

Of the 16 signed up for the course, 8 students (3 from KTH, 3 from NGSSC, 2 from Umeå) completed all assignments and received credits for their work. The other students did not complete the course due to either inappropriate prerequisite skills (not enough UNIX and programming experience) or lack of interest in completing the assignments (they attended to get a good overview of what Grid computing is all about).

Overall, the course received good feedback with a 3.2 average on a scale from 1 = poor to 4 = excellent. A formal evaluation questionnaire was not distributed, instead two debriefing sessions were held: one mid-way through the course (the course schedule was slightly modified based on that input), and one at the end of the course. The inputs were collected and summarized and have served as constructive criticism in the planning for the next Grid computing course, which will take place in Linköping in January 2005.

Events

Year 2000

Table 1
Events in 2000

April 13	Workshop on architecture at the VR Cube
August 20	PDC/NGSSC “Preparatory Programming Day”
August 21 – September 1	PDC/NGSSC Summer School
October 4	Prime Minister Göran Persson visits the VR cube
October 18 – 20	Ninth Grader’s Week
October 19 – 20	NWChem/ECCE Workshop
December 14 – 15	PDC Annual Conference. Simulation and Visualization in the Life Sciences

Other significant events with PDC contribution:

- Several activities to promote cooperation in the creation and operation of network activities and Grids, such as the organization of the First Egrid workshop in Poznan, Poland and the Fifth Grid Forum in Boston, Massachusetts and in planning meetings for the European DataGrid project.
- Inet2000 demonstration of computational steering from the exhibit floor in Yokohama and simulations driving the displays carried out at PDC and at the Texas Institute for Computer and Computational Sciences. The demonstration was carried out in collaboration with SUNET, Nordunet, Startap, APAN, and the iGrid2000 organizers.

Table 2
Events in 2001

Year 2001

February 13	KTH/IBM Open Source Linux Lab Inauguration
March 16	European Joint Research Council visits PDC
June 5	Startap Annual Meeting at PDC, ISOC 2001
August 4 – 5	Participation in the First US-UK Grid Workshop
August 20	Linus Torvalds visits PDC
August 20 – 31	PDC/NGSSC Summer School
October 1 – 4	CAVE Programming Workshop
October 18	Seminar at PDC, "Parallel processing including MPI, OpenMP and parallel tools," October 18, 2001
October 23	Seminar at PDC, "New Generalized Data Structures for Matrices Lead to a Variety of High Performance Algorithms" by Fred G Gustavson, IBM
October 23 – 25	Ninth Grader's Week – Basic HPC
December 4	Norwegian Supercomputing Board visits PDC
December 13	ABC Meeting. Annual Computational Chemistry Meeting.

Other significant activities with PDC contribution

- PDC staff served on the steering committee for the first EuroGlobus workshop
- The EU DataGrid project initiated.
- Participation in establishing the SweGrid interest group in cooperation with particle physics groups.
- Participation as an unfunded partner in the newly established NorduGrid project.
- Invited participation in the European Commission Grid Workshop in Brussels, first iVDgL planning meeting, and the first US-UK Grid Workshop.

Table 3
Events in 2002

Year 2002

January 9	Participation in the First iVdGL planning meeting
February 8	Visit by the Swedish Research Council's SweGrid study group
February 8	Baltic Art Workshop at PDC
April 15 – 17	20th NORDUnet Networking conference
April 22	Founding of Nordic Grid consortium with CSC and Parallab
August 19 – 30	PDC/NGSSC Summer School
October 21 – 25	Ninth Grader's Week

Other significant activities with PDC participation

- PDC co-founds the Nordic Grid Consortium with CSC, the Finnish IT Center for Science, Espoo, Finland, and Parallab at University of Bergen, Norway. http://www.nordicgrid.net/.

- PDC staff participated in several Grid initiation activities such as 2nd iVDgL planning meeting, the GridLab EU-project, and the Terabyte Challenge test bed led by Robert Grossman at the University of Illinois, Chicago.
- The SweGrid project funded by the KA Wallenberg foundation and initiation of the SweGrid project.
- Execution of the First LHC Computing Data Challenge.

Year 2003

Table 4
Events in 2003

March 18	European Grid Support Center was formed by PDC, CCLRC e-Science Centre and CERN
May 27	Performance measuring workshop at PDC
August 18 – 26	NGSSC Grid Computing course, "Grid Computing" at PDC
August 18 – 29	PDC/NGSSC Summer School, "Introduction to High Performance Computing" at PDC
August 24 – 27	Nordunet Conference. The first Nordic demonstration in collaboration with NorduNet, Surfnet, Starlight, Karolinska Institutet and the National Center for Microscopy and Imaging Research at the University of California, San Diego, of the utlity of λ- networking in tele-microscopy applications
September 2	PDC, EPCC and the Globus Project form the Globus Alliance
October 3	PDC joins HP Collaboration and Competency Network (HP-CCN) for Global File Systems, Computational & Data Grids and Scientific Visualization
October 15	Inauguration of Lucidor, PDC's major computing resource
October 13 – 15	Ninth-Grader's Week and Upper-Secondary-School-Student's Week
October 13 – 15	PDC hosts Gelato Federation fall Meeting
November 11 – 12	KTH Abaqus Workshop
November	PDC relocation to new building at KTH campus commences

Other significant activities with PDC participation:

- PDC co-founds the European Grid Support Center with the UK e-Science Program and CERN. http://www.grid-support.org.
- PDC staff serving on the Program Committee for the First International Summer School on Grid Computing 2003, Vico Equense, Italy, July 13--25, 2003. http://www.gridforum.org/L_News/grid_computing_school_b.htm.
- PDC becomes a member of the Gelato Federation http://www.gelato.org.
- PDC, EPCC and the Globus Project form the Globus Alliance. http://www.globus.org/.
- PDC joins HP Collaboration and Competency Network (HP-CCN) for Global File Systems, Computational & Data Grids and Scientific Visualization http://www.hp.com/techservers/hpccn/.

Spotlight on Grid Technology

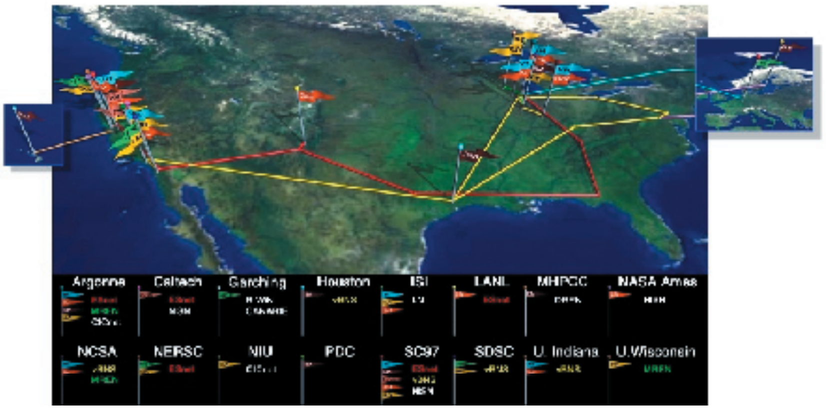
Background

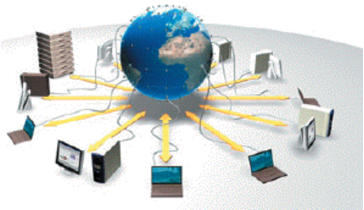
The concept of computer power being available in a ubiquitous form similar to electric power was put forward by Leonard Kleinrock in 1969: *"We will perhaps see the spread of 'computer utilities', which, like present electric and telephone utilities, will service individual homes and offices across the country"*. This was the year when he together with colleagues brought up the first ARPANET connection between UCLA and Stanford Research Institute with UC Santa Barbara and University of Utah added later in the year and BBN in Massachusetts added in March 1970.

The conception, research and implementation of Grid technology brought Klienrock's vision closer to becoming a reality. "Grids" as they are known today provide abstractions for resource sharing and collaboration across multiple administrative domains. The term *resource* includes physical resources (computation, communication, and storage), informational resources (databases, archives, and instruments), individuals (people and the expertise they represent), capabilities (software packages, brokering and scheduling services) and frameworks for access and control of these resources.

The formal concept of computational Grids as we currently know them was conceived by Tom DeFanti, Rick Stevens, and Larry Smarr when Tom DeFanti was responsible for the planning of the Exhibit at Supercomputing '95. The realization of the concept was the I-Way (Information Wide Area Year) demonstration that encompassed a dozen ATM networks, seven supercomputer centers, five virtual reality research sites, and over 60 applications groups.

GUSTO Computational Grid Testbed





The great success of this effort lead to the formation of the Globus project, which sought to build a software base for Grids based on the experience from the I-Wayeffort. The first incarnation of the Globus software was demonstrated on the GUSTO testbed at Supercomputing '97, with PDC as a participant.

In the Grid computing paradigm, collaborations take place across organizational boundaries in *virtual organizations*. Within a virtual organization, people with common interests and goals share their resources and data (storage resources, computational resources) with the goal of obtaining a coordinated and maximum total usage of the resources. A virtual organization can also be seen as a vehicle for reaching a common community goal (such as a research result) quicker and more efficient.

Examples of large-scale virtual organizations are research projects sharing output data from unique instruments, such as high energy physics accelerators, sensor networks, and satellites in high orbit. Another example of a smaller-scale virtual organization is that of a professor, his PhD students, and a couple of collaborating research fellows located at another institute or university. These examples highlight the diversity in both the scale and dynamics of virtual organizations and use cases for Grid technologies.

The evolution and maturity of the Grid computing concept and related technologies has evolved rapidly in the past decade. The technology front-end has been mainly pushed by the US-based Globus project, with which PDC has had a long and close collaboration. The open-source Globus Toolkit offers a set of middleware modules, the basic building bricks on top of which software distributions have been developed, such as the Nordugrid toolkit, the Virtual Data Toolkit used by the European Data Grid, and several commercial softwares.

Grids have had a significant impact on industry in the past years: companies such as IBM, Oracle, Sun, HP and others all offer and promote Grid solutions as a core part of their "on demand" and "adaptive enterprise" offerings. This has also accelerated the need for standardization and interoperability, and the creation of organizations such as the Global Grid Forum.

PDC research, development and support activities

PDC has taken part in several R&D related Grid efforts through the years. These projects cover the spectrum from applications to middleware and support of Grid middleware. PDC often plays an integral role in helping to shape and form new grid technology initiatives. For more information on current PDC research activities, see www.pdc.kth.se/grid. The project briefs below are presented with early projects presented first.

GEMSViz was a project in which the Nexus communication library from the Globus Grid middleware was used to develop a remote visualization and steering capability for a General Electro-Magnetic Solver (GEMS) developed at the Parallel and Scientific Computing Institute (PSCI), a Center of Excellence at KTH. The Nexus library was used to connect the flow of control information from the visualization process to the computation process and the flow of computed data in return. GEMSViz was demonstrated at iGrid2000 held in conjunction with iNET2000 in Yokohama, Japan. For the demonstration computing platforms at PDC and in Houston were used for the simulation and virtual reality equipment on the exhibit floor was used for visualization and steering. The Asia-Pacific Advanced Network (APAN), Nordunet and Internet2 were used for high-speed network interconnect. The GEMSViz software organization and the iGrid2000 testbed are shown in Figure 1. http:// www.pdc.kth.se/projects/GEMSviz.

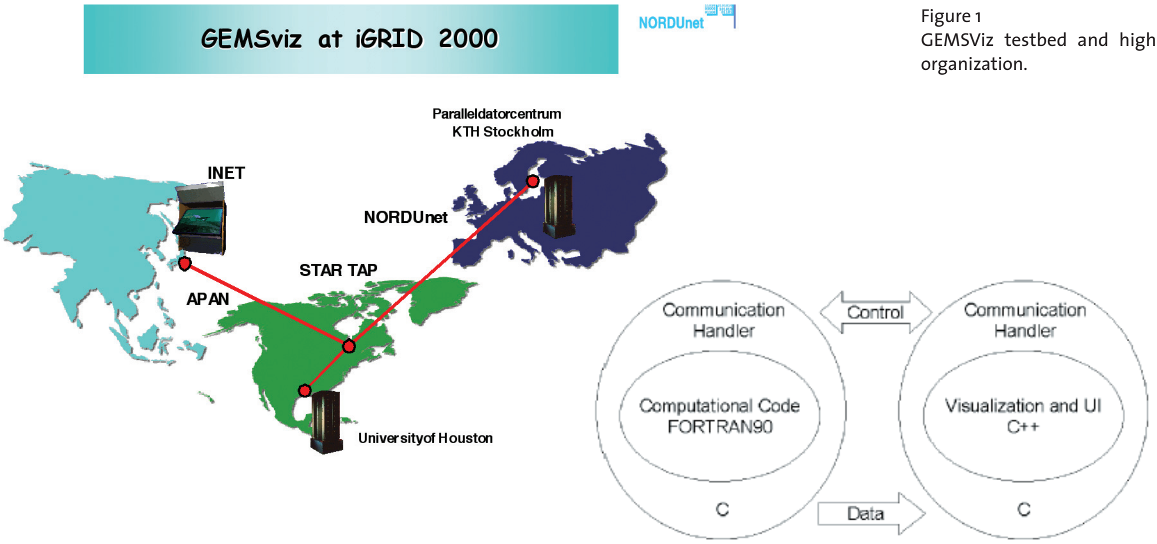


Figure 1
GEMSViz testbed and high-level software organization.

JaCo3, *Java and CORBA based Collaborative Environment for Coupled Simulations*, was an EU funded project creating an environment for distributed simulation in a Grid environment. The premise for the project was a heterogeneous networked computing environment including a wide range of communication capabilities and multiple interacting codes with different ownership. A code would only be executed at sites approved by its owner with relevant data and control information being exchanged between the codes at different sites for a correctly executed integrated simulation. The environment consists of a set of CORBA services that manage remote computing resources through the Internet. Figure 2 shows the JaCo3 software architecture and Figure 3 shows an application scenario. For more information, see, for instance, www.ercim.org/publication/Ercim–News/enw45/priol.html

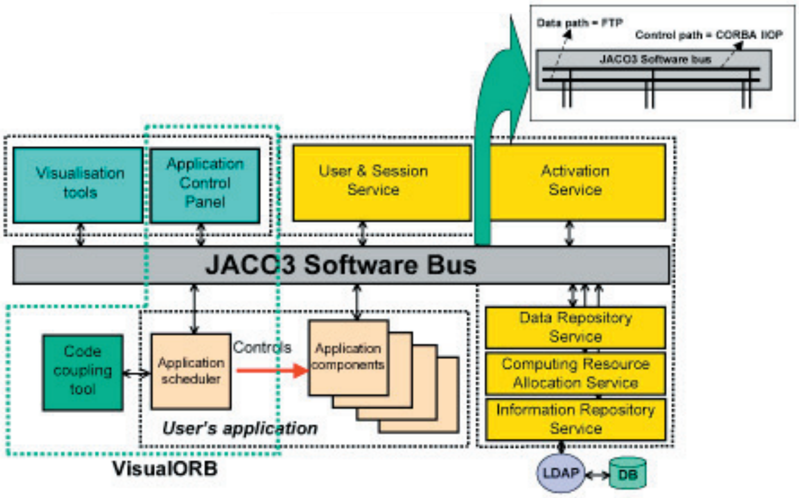


Figure 2
JaCo3 software architecture

ENSCUBE (European Numerical Simulation Services for SMEs) was an EU project for creating and establishing a platform for collaborative work and numerical simulation used to support the collaborative engineering workflow. The platform is a combination of off-the-shelves grid middleware collaborative tools and simulation environments. Alcatel, Vircinity, HRLS, CERFACS and PDC developed the environment for collaborative work and simulation jointly. The premise for the project was a heterogeneous networked computing environment including a wide range of communication capabilities. http://www.telecom.gouv.fr/programmes/eten/enscube.htm, http://www.vircinity.com/d+e/data0801/englisch/e_products/xe_fs_prod_Enscube.html

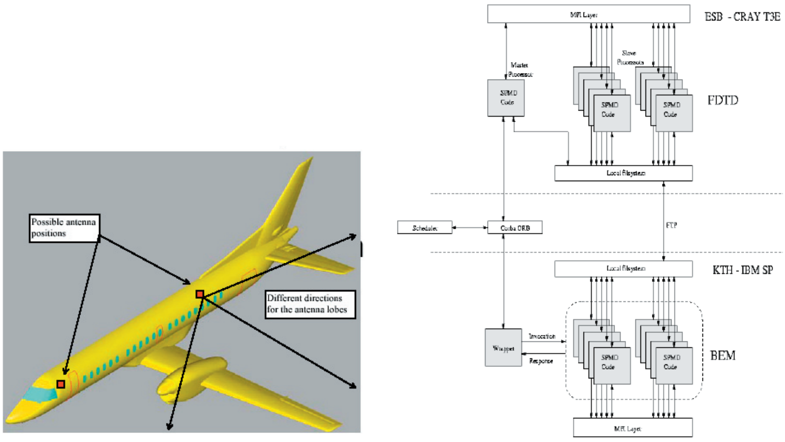


Figure 3
JaCo3 application scenario.

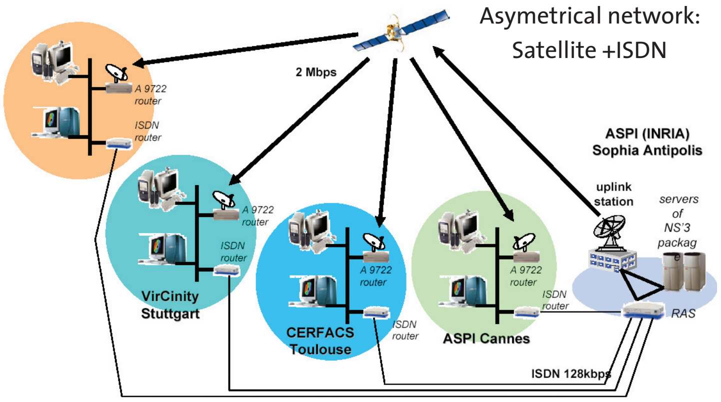


Figure 4
ENScube distributed secure simulation environment.

Remote Storage Service. One of the first Grid production services in the form of remote storing of data in which PDC engaged was storing data for the Odin satellite for the Swedish Space Corporation. Data received by an earth station in northern Sweden is transmitted over the Swedish University Network (SUNet) to PDC as illustrated in Figure 5 for this joint European-Canadian project.

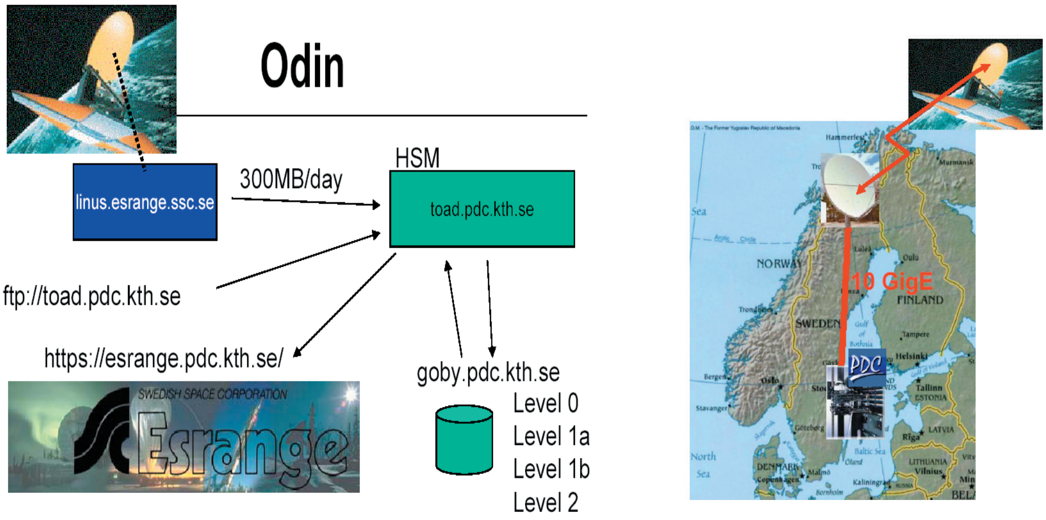
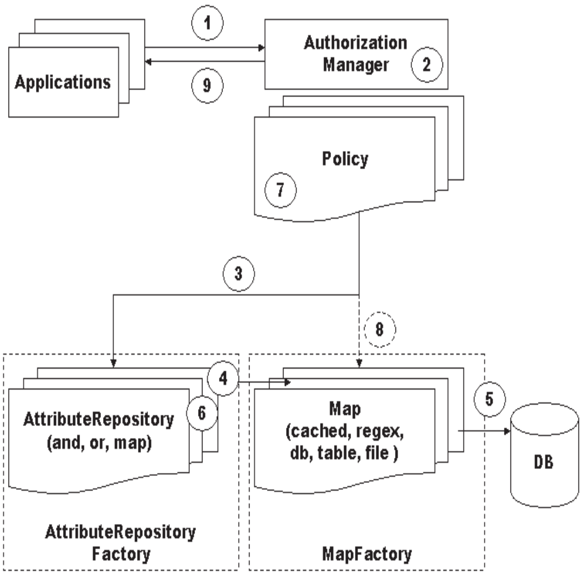


Figure 5
The Odin satellite remote storage service.

EDG. In the EU DataGrid project, www.edg.org, PDC participated by contributing to the efforts in developing the Security Architecture, and transparent access. The CERN Advanced STORage manager (CASTOR) was used as the point of departure for a Grid enabled distributed and hierarchical storage management service. PDC contributed to the overall security architecture and design, see Figure 6, and was responsible for developing and integrating data management software components for authentication and authorization.

Figure 6
The EDG authorization framework for Web and Grid services



Secure, complex data base queries. In late 2002, and in collaboration with Uppsala and Umeå Universities, PDC was awarded a research grant from VR for Grid research in relation to the Swegrid resources. During 2003, PDC initiated work with Uppsala University on the development of an architecture for processing database queries having computationally very expensive external functions that utilize Grid resources for user-defined computations over distributed datasets, while maintaining a sound security for process creation and inter-process communication.

Grid accounting. The Swegrid Accounting System, SGAS, is another project started within the scope of the VR Grid Research award. It is carried out in collaboration with Umeå University. The main problem considered is the accounting associated with a single resource allocation issued by the Swedish National Allocation Committee (SNAC) for SweGrid being consumed in portions at any of the SweGrid clusters having their own accounting systems. The SNAC allocated project *node hours* are awarded as node hours per month for a six or twelve month period. Each month one sixth (or one twelfth depending on the length of allocation period) of each project's node hours are withdrawn, unless the allocation has been consumed. Accounting is defined here as the phase in which the system audits the usage of system resources. The whole allocation may be used up at one cluster or partially at any number of clusters. The Grid users should not have to be concerned about which cluster runs his/her jobs (although some job descriptions may exclude all clusters but one). Furthermore, the accounting system must also address the issues related to the fact that accounts are created for individual users, but allocations are made to projects (virtual organizations). The SGAS software architecture is illustrated in Figure 7.

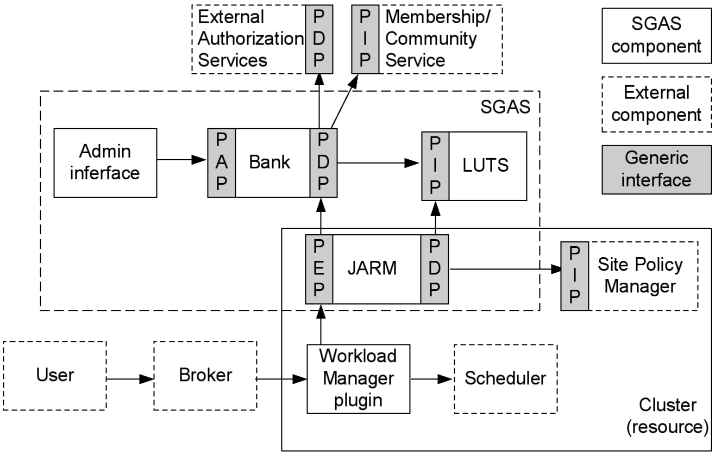


Figure 7
The SGAS software architecture for Grid accounting

PDC standardization and community efforts

EUROPEAN GRID SUPPORT CENTER

The European Grid Support Centre (EGSC) was established in response to a need seen by European projects for increased information of and influence on Grid middleware releases, in particular by American Grid middleware developers (specifically the Globus project), and a need by those developers who saw a rapid increase in support workload and bug reports originating from the European region. (At the time, there were roughly 20 EU sponsored Grid projects, plus a large number of national initiatives.) The bug resolving process often dragged out in time due to the time zone differences, effectively limiting the communications to a single round-trip of messages exchanged per day. Furthermore, the bug reports were often of low quality due to poor technical knowledge from the person reporting the bug. Thus, the European Grid Support Centre, EGSC, was created in February of 2003 by the UK e-Science Grid Support Centre, the LCG project at CERN, and PDC. The organizations had local in-house in-depth technical expertise, existing collaborations and – most importantly – the established trust and relations with several of the Grid middleware efforts in the US. In the fall of 2003, EGSC was joined by Poznan Supercomputing Center and HP.

EGSC is a forum in which current Grid projects in the European arena exchange deployment and support experiences of widely adopted Grid middlewares, and collaborate on problem resolution. EGSC strives to establish and foster long-term relationships and close collaborations with the developers from the relevant middleware projects. With those relationships in place, we can act as a funnel / front-end / mediator between the developers and projects on the European arena.

Globus Alliance. In September of 2003, PDC took a leadership role in helping to reshape and reinvigorate Grid research for the global community. The Globus project was reformed into the Globus Alliance, which PDC co-founded together with EPCC (Edinburgh) and the two original Globus teams at ISI and ANL, thus reaffirming PDC's position as a supporter of and leader in the field of core Grid technologies in Europe and the world.

The creations of both the EGSC and the Globus Alliance are an indirect result of the long-standing and close relationship between PDC staff and the Globus staff. PDC have continuously contributed to the stability and improvement of the toolkit, for instance by adding support for the AFS file system, and the integration of the toolkit's security features with a local Kerberos security infrastructure, such as the one used at PDC. This latter work was presented at the GlobusWorld conference in San Diego, January 2003.

The **Global Grid Forum** is a community-driven set of working groups that are developing standards and best practices for Grid and Peer-to-Peer computing. GGF caters to both "bottom-up" approaches, in terms of experiences sharing and community building among Grid practitioners and to a "top-down" Open Grid Services Architecture (OGSA) identifying a core set of services that are viewed as essential for many Grid systems and applications in both e-science and e-business, and the high-level functionality of and interrelationships among those core services.

PDC staff actively contributes to the GGF discussions and creations of GGF documents on requirements, best practices and standards, and participates in working group meetings.

NorduGrid. PDC participated as an unfunded partner in the Nordugrid project, a project funded by Nordunet within a program to stimulate applications of high-performance networking. PDC contributed to the project by installing versions of the NorduGrid software as needed and making suitable platforms available for development and test purposes. PDC has also helped to resolve issues, provide technological expertise, and to facilitate direct communication with, e.g.

key Globus personnel to clarify needs and requests for future improvements from the underlying Globus middleware on which the Nordugrid toolkit (recently renamed to ARC) depends.

Nordic Grid Consortium. In September of 2002, PDC founded the Nordic Grid Consortium together with Parallab (Bergen) and CSC (Helsinki), with a primary objective of establishing a network of experience sharing from a set of small projects to foster and improve the understanding and operative issues around the deployment of Grid technologies.

Atomistic Molecular Dynamics Simulations and First Principles Calculations in Materials Science

Anatoly B. Belonoshko[1, 2], **David Andersson**[1], **Pavel Korzhavyi**[1], **Levente Vitos**[1], **and Börje Johansson**[1, 3]

[1]Applied Materials Physics, Department of Materials Science and Engineering, The Royal Institute of Technology (KTH), Stockholm
[2]Condensed Matter Theory Group, Department of Physics, AlbaNova University Center, The Royal Institute of Technology (KTH), Stockholm
[3]Condensed Matter Theory Group, Department of Physics, Uppsala University

The studies we perform in our group heavily rely on computers and the scientific progress is highly dependent on the availability of computer power. In our calculations and simulations of materials we calculate energies of interaction using different methods of varying complexity, from ab initio calculations to semi-empirical comparably simple models. To reach a proper understanding it is often necessary to model a system consisting of a large number of atoms. In what follows, we give examples of our research performed recently using resources of PDC.

Introduction

The computer time required to solve a problem for N atoms scales non-linearly with N (in worst case as $O(N^3)$), it is understandable that some problems might require enormous computational efforts. The history of our approach is almost as old as computers. The first report on a molecular dynamics simulation was written by Alder and Wainwright in 1957. The purpose was to investigate the phase diagram of a hard sphere system, and in particular the solid and liquid regions. In a hard sphere system, particles interact via instantaneous collisions, and travel as free particles between collisions. The calculations were performed on a UNIVAC and on an IBM 704. The article *Dynamics of radiation damage* by J. B. Gibson, A. N. Goland, M. Milgram and G. H. Vineyard from Brookhaven National Laboratory, appeared in 1960, is probably the first example of a molecular dynamics calculation with a continuous potential based on a finite difference time integration method. The calculation for a 500-atoms system was performed on an IBM 704, and took about a minute per time step. Nowadays, such simulations can be routinely performed in a few minutes using a personal computer. In our strive for increasing complexity of the models, time and space scales of our simulations, we often reach the limit when we feel 'comfortable' with the time it takes to run a particular simulation. It is quite rare when a scientist is willing to run his/her code for months, not mentioning years, to get even a very valuable result. This is where parallel computations come into picture, drastically decreasing the time necessary for solving a problem. Needless to mention that we performed them being 'comfortable' thanks to the availability of fast parallel processing.

Carbides, nitrides and oxides

Compounds of plutonium. The physics and chemistry of the actinide elements form the scientific basis for rational handling of nuclear materials. Safe disposal of nuclear waste requires clear understanding of the behavior of uranium and plutonium in various forms of nuclear waste as well as of the interaction of these waste forms with the environment. We have developed a theoretical approach which allows one to obtain thermodynamic properties of plutonium-based alloys and compounds from first-principles electronic structure calculations based on density functional theory [1].

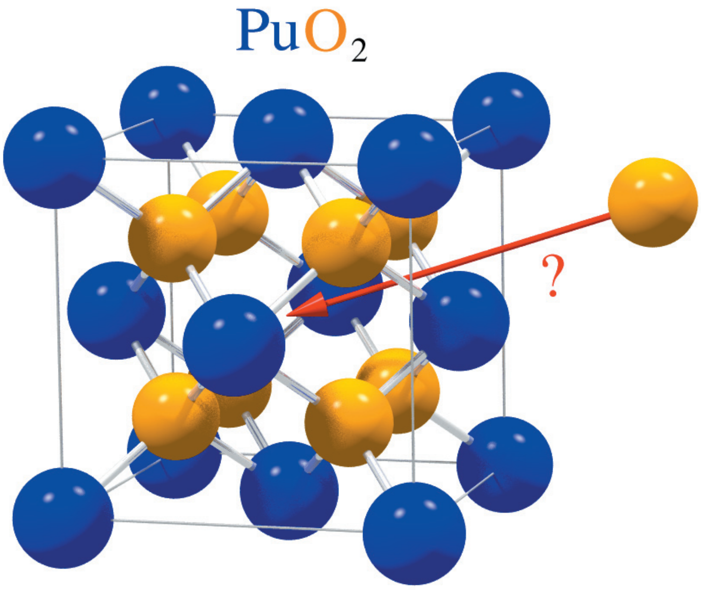


Figure 1
Can plutonium dioxide be oxidized further?

The corrosion behavior of nuclear materials is often unusual and difficult to interpret, both due to the complex electronic structure of the actinide elements and to the difficulties caused by their radioactivity. In recent experiments, most unexpectedly, plutonium dioxide was found to react with water to form higher oxides up to $PuO_{2.27}$, whereas PuO_2 was always thought to be the highest stable oxide of plutonium. Past experimental attempts to oxidize PuO_2 using atomic oxygen or ozone have been negative, so that no direct calorimetric measurements are possible. In this study, we obtain the lacking thermodynamic data from first-principles total energy calculations for compounds having the fluorite (CaF_2) related crystal structures. On the basis of the obtained theoretical results, combined with well-established thermochemical data, we have performed a thorough analysis of this complicated problem. The reactions of PuO_2 with either O_2 or H_2O to form PuO_{2+} are calculated to be endothermic, that is, in order to occur they require a supply of energy. However, our calculations show that, under certain conditions, this reaction may be driven by radiolysis of water through the formation of strong oxidants (e.g., hydrogen peroxide, H_2O_2) and their spontaneous reaction with PuO_2. This finding is important for better understanding the long-term behavior of nuclear waste [2].

Compounds of titanium. Titanium carbides and nitrides exhibit an exceptional combination of properties, such as chemical stability, great hardness, high melting point, excellent electrical and thermal conductivity. They are technologically important and they are for example used as coating material in cutting tools. In spite of their wide utilization there are still several unknown aspects of their properties. Vacancies, i.e. empty lattice sites, are known to account for the non-stoichiometry of titanium carbides, nitrides and oxides. The vacancies as well as dissolved oxygen strongly influence the properties of titanium carbonitrides. In our work, we have been using state-of-the-art first principles calculations in conjunction with thermodynamic modeling to study structural vacancies in titanium compounds.

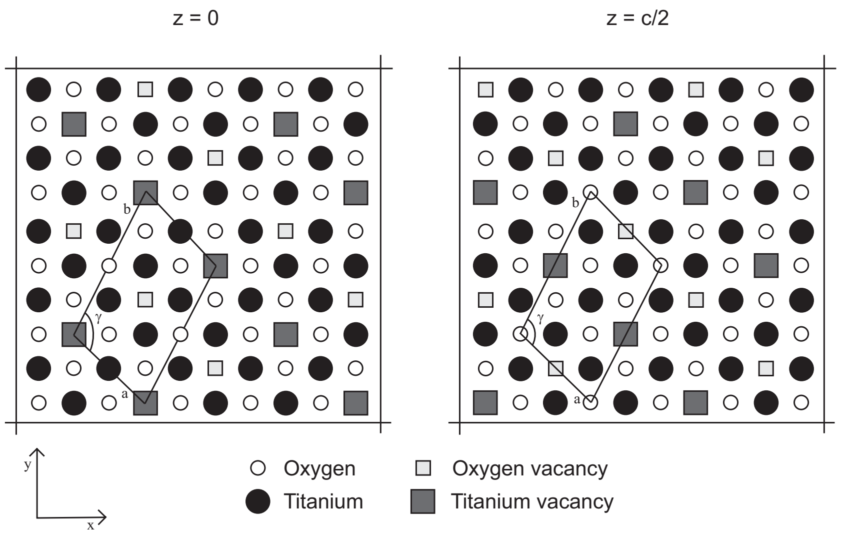
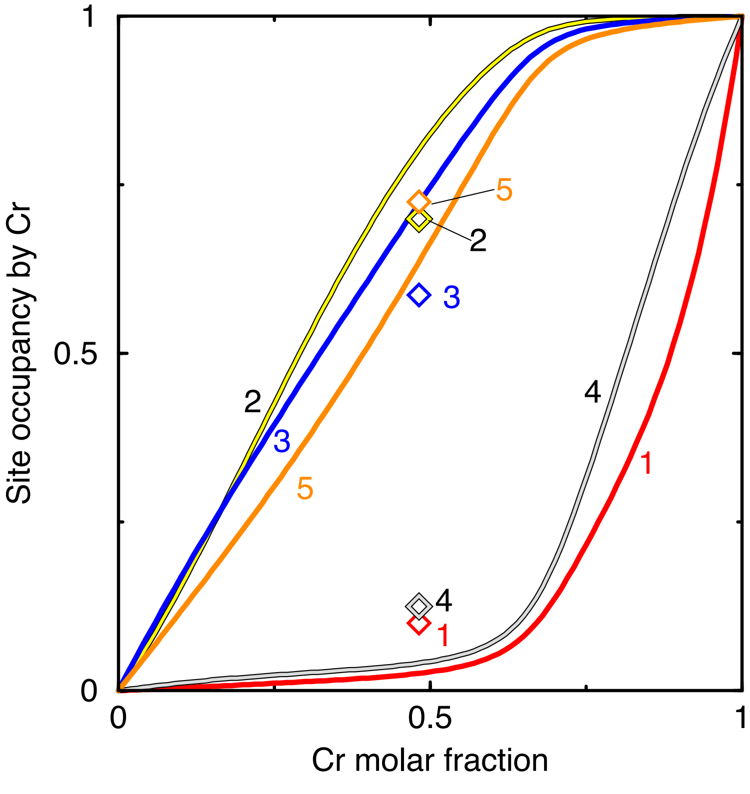


Figure 2
The structure and the monoclinic unit cell of TiO. The lattice parameters a, b, c, and γ of the monoclinic structure have been refined in this work using ab initio calculations.

The structure, stability and electronic properties of the low oxygen oxides of titanium, TiO_x with $1/3 \leq x \leq 3/2$, as well as the properties of non-stoichiometric titanium carbide TiC_x and nitride TiN_x have been studied by means of accurate first-principles calculations. Titanium monoxide represents the most difficult case since it contains large fractions of vacancies on both the titanium and the oxygen lattice sites. These structural vacancies are essential for understanding the properties and phase stability of titanium oxides. Structures with ordered arrangements of vacancies were treated with a plane wave pseudo-potential method, while calculations for structures with disordered vacancies were performed within the framework of the Korringa-Kohn-Rostoker Green's function technique. The theoretically refined structural parameters compare well with experimental data. The equation of state as well as the elastic properties are also derived for TiC_x, TiN_x, and TiC_x as a function of composition. A monoclinic, vacancy-containing, structure of stoichiometric TiO is confirmed to be stable at low temperature and pressure [3], see Figure 3.

Figure 3
The calculated occupancy of the five inequivalent lattice sites in the Fe-Cr sigma phase by chromium (lines). Experimental X-ray data are shown as symbols.



Fe-based sigma phases

Sigma-phases are complex tetragonal structures that form in transition metal alloys having the average number of s+d electrons in the interval 6.5 - 7.4. These phases, which may form during solidification or heat treatment, usually cause embrittlement of the alloys. A well-known example is the FeCr sigma-phase, whose formation has a deleterious effect on the mechanical properties of stainless steels. Like other sigma-phases, the FeCr phase possesses a high degree of chemical disorder among its five crystallographically inequivalent sublattices. Using a combination of *ab initio* total energy calculations with the compound energy formalism developed in CALPHAD, we were able to reproduce the experimentally observed site occupancy in the FeCr sigma-phase, see Figure 3. The calculated results on the electronic structure and thermodynamic properties of the FeCr sigma-phase are scheduled for publication [4].

Modeling of the sigma-phases is an important part of a more general activity of our group directed towards computer-aided design and optimization of alloy steels [5,6]. As these steels typically contain many alloying elements in addition to chromium, it is of interest to study other iron-based sigma phases. During the past year, we have completed the calculations for the Fe-Mo and Fe-V binary sigma phases. For each binary system, total energies of all the 32 end-member compounds (EMC) were calculated as a function of the volume-per-atom ratio and of the tetragonal c/a ratio. The thermodynamic properties of the Fe-Mo and Fe-V sigma phases, as a function of composition, have been studied using thermodynamic modeling (manuscripts in preparation).

Materials under extreme conditions

Our methods proved to be particularly useful in studies of matter under extreme conditions. One of such examples is our recent discovery of the body-centered-cubic phase of iron in the Earth's inner core [7]. A more recent example is simulation of shock compression of liquid D_2. The interest to this substance is easy to appreciate. First, hydrogen is the most abundant element in the Universe. Such giant planets as Jupiter and Saturn consist in large part of hydrogen and most of them at very high pressures and temperatures, the conditions currently available under the shock compression only. A more practical interest is due to the possibility of getting energy from fusion of deuterium atoms. However, experimental shock wave studies encountered certain problems. The results obtained in two US National Laboratories in Livermore (California) and Sandia (Albuquerque) contradict to each other. Our simulations have demonstrated that at high impact velocities the shock wave becomes smeared out [8]. Since the experiments in Livermore and Sandia use different techniques of measurements, they simply measure different parts of a shock wave. This leads to the difference in experimental results, which at present are extensively debated in literature [9] and scientific meetings.

Concluding, we wish to thank the personal of PDC who made these studies possible. As consumers of high-performance computational resources we would like to have alternatives, when choosing a particular computational platform to run our codes. Several of us learned it the hard way what it looks like when there is no alternative. We believe that the present policy in distributing resources among different centers is wise. We just wish that the resources themselves would be more abundant.

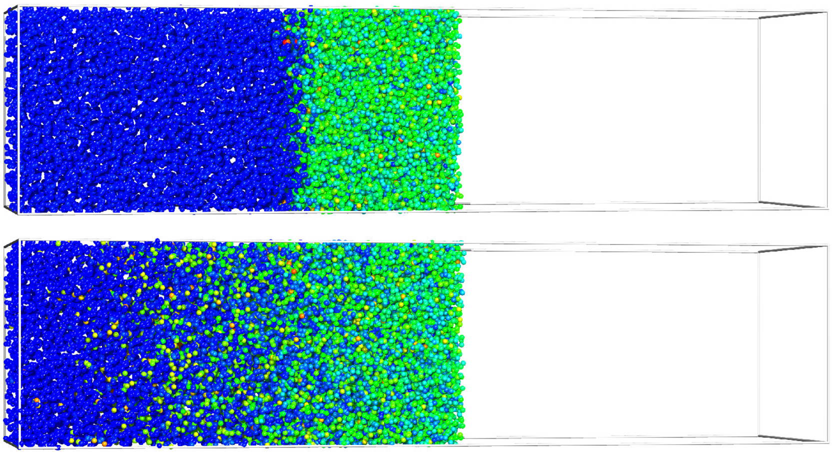


Figure 4
Snapshots of configurations of deuterium atoms under shock compression. The piston moves from right to left. The upper and bottom parts correspond to piston velocity of 10 and 40 km/s, correspondingly. The atoms are colored according to $v_x^2+v_y^2+v_z^2$, where v_i – i^{th} velocity component. The coldest atoms are blue, the green are hotter and red atoms are the hottest ones. While at 10 km/s there is a sharp border between compressed and uncompressed regions of deuterium, at 40 km/s the shock wave front basically disappears.

References

1. P. A. Korzhavyi, L. Vitos, and B. Johansson, *Thermodynamic properties of Pu-O-H compounds and alloys from density functional theory*, in Scientific Basis for Nuclear Waste Management XXVII, Materials Research Society Symposium Proceedings, Vol. 807 edited by V. M. Oversby and L. O. Werme (MRS, Warrendale, 2004), pp. 107-112.

2. P. A. Korzhavyi, L. Vitos, D. A. Andersson, and B. Johansson, *Oxidation of plutonium dioxide*, Nature Materials 3, 225 (2004).

3. D. A. Andersson, P. A. Korzhavyi, and B. Johansson, *Thermodynamics of structural vacancies in titanium monoxide from first principles calculations*, to be submitted.

4. P. A. Korzhavyi, B. Sundman, S. G. Fries, and B. Johansson, *Atomic and magnetic structure of the FeCr sigma phase*, manuscript in preparation.

5. B. Johansson, L. Vitos, and P. A. Korzhavyi, *Chemical composition-elastic property maps of austenitic stainless steels*, Solid State Sci. **5**, 931 (2003).
6. L. Vitos, P. A. Korzhavyi, and B. Johansson, *Austenitic stainless steels from quantum mechanical calculations*, Advanced Engineering Materials **6**, 228 (2004).
7. A. B. Belonoshko, R. Ahuja, and B. Johansson (2003), *Stability of the body-centered-cubic phase of iron in the Earth´s inner core*, Nature **424**, 1032 (2003).
8. A. B. Belonoshko, N. V. Skorodumova, A. Rosengren, S. Bastea, and B. Johansson, *Shock wave instability in deuterium at high impact velocities*, Phys. Rev. Lett. (submitted).
9. W. B. Hubbard, *The core problem*, Nature **431**, 32 (2004).

Ab Initio Studies of Defects in Indium Phosphide

CWM Castleton[I] **and S Mirbt**[II]
I. Materials and Semiconductor Physics Laboratory, The Royal Institute of Technology (KTH), Stockholm and Department of Physical Electronics/Photonics, Mitthögskolan, Sundsvall
II. Department of Physics, Uppsala University

In this report we will outline the need for and the role played by computation within the study of defects in semiconductors. We will then describe some of the current technical difficulties in this area, together with our approach to solving them: namely the use of larger supercells and finite size scaling. We will also describe examples of studies of individual defects in industrially relevant materials, in particular vacancy based defect complexes in p-type InP. A key underlying theme is the particular value that computing resources provided by PDC have to our work.

The role of computational semiconductor physics

A vast amount of our modern lifestyle depends upon the use of semiconductors. Silicon (Si) forms the basis of the technology used in everyday electronic devices ranging from computers to remote controlled toys to thermostats for central heating systems to mobile phones and televisions. Meanwhile, compound semiconductors such as gallium arsenide (GaAs) and indium phosphide (InP) are used to make so called "electro-optical" components: lasers and light emitting diodes on the one hand, light sensors and infra-red detectors (in TV remote controls, for example) on the other. Besides the lasers used to read off the information stored on CDs and CD-ROMs, opto-electronic components also lie at the heart of modern telecommunications. Lasers are needed to send the signals into fibre optic networks for both telephones (mobile or fixed) and the internet, whilst semiconductor based detectors receive the signals at the other end. The price, reliability and handling capabilities of these components are thus very important and a great deal of time and effort goes into developing them and understanding not only the properties of the components themselves but also of the materials from which they are made. The different semiconductors all have their own characteristic electrical and optical properties, and these need to be very precisely controlled and tuned for different applications or to form the different parts of a chip, laser or other device. This tuning is done by a process known as "doping", in which very small amounts (say, one atom in 10000 or more) of other elements are added to the material, for example, phosphorus (P) or boron (B) into Si, or zinc (Zn) into InP. The added atoms form "substitutional" defects in the host crystal (see diagrams in Figure 1) and are in this case known as dopants. They add charge carriers to the semiconductor – negatively charged electrons or positively charged "holes". (A hole is a place where an electron is missing, but it moves through the material rather like a positive electron.) Substitutional dopants are not the only types of point defects, however. Other elements may enter by mistake, either substitutionally or as "interstitials". "Vacancies" may form, or additional host atoms may turn up in the wrong places as interstitials or as "antisite" defects. The presence, energetics and structure, as well as stability, mobility and diffusion of all of these types of defects affect the electrical and optical properties and the effective lifetime of the resulting material and/or device. As a result they need to be understood in ever greater detail as the technology develops.

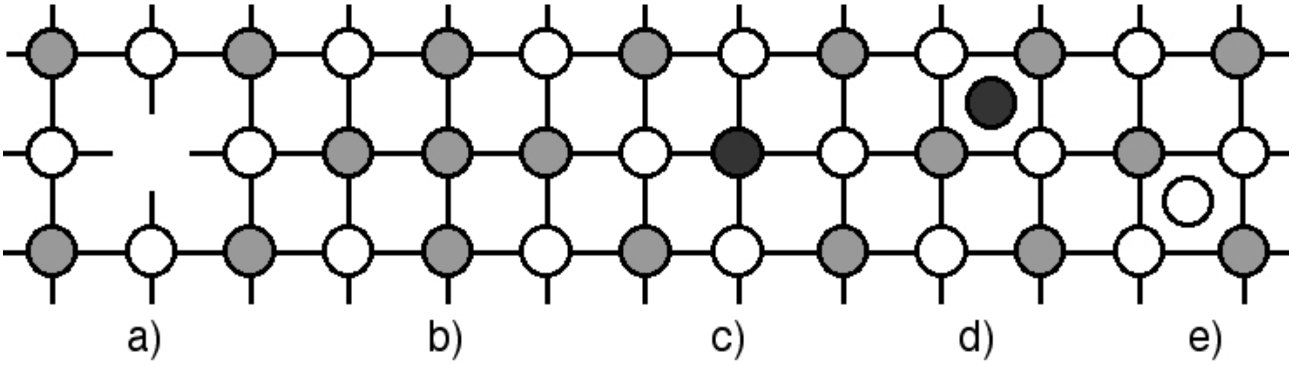


Figure 1
Types of point defect. The lattice here consists of two atom types: A (white circles) and B (grey circles) arranged alternately to give two sublattices, A and B.

a) a vacancy on the B sublattice: V_B
b) an antisite - a B atom on the B sublattice: B_A
c) impurity or dopant atom C on the A sublattice: C_B
d) & e) interstitials C_i and A_i in the gaps between lattice sites.

To this end, a great many studies have been made of defect formation, diffusion and structure over the years, but these have largely been experimental, or in terms of very simplified theoretical models and pictures. Accurate "first principles" theoretical calculations to predict the properties of the pure host materials have been possible for many years, since only one or two atoms need be involved in the calculations. ("First principles" or *"ab initio"* means that we predict the properties of a material, assuming nothing but the basic quantum mechanical equations describing all solid matter.) However, as soon as a defect is introduced, it becomes necessary to use many more atoms in the calculations - not just 10s but 100s or even 1000s, in order to accurately describe the interaction between the defect and the surrounding host atoms.

It is only in the last few years that the computer power available, particularly at facilities such as PDC, has really made this possible. There is a great deal of work now to be done, as we see the possibilities of not only explaining the results of experiments after the event, but even reliably predicting the properties of individual defects and defect complexes before expensive and time consuming experiments need to be done. Using first principles calculations as a virtual microscope, with atomic scale precision, we can obtain information about defect properties that simply is not accessible experimentally. We can tackle questions to do with the identification of particular defects, their shape, structure, binding energies, electrical and optical behavior, diffusion barriers etcetera far more directly and specifically than could be done previously.

The work we have been doing on the PDC machines over the last couple of years has focused upon two specific areas of this: a) developing scaling methods for extrapolating results of calculations involving small group of atoms (10s or 100s) to describe physical situations involving many more (1000s or 10000s or more) and b) examining the local structure and properties of defects in p-type InP. This latter project is related to the problem of "carrier compensation" in heavily zinc doped InP. Examples from both areas of our work are described below.

Outline of the calculational method used

The basic calculational step is to evaluate the total quantum mechanical energy, E_T^C, of a small group or cluster "c" of atoms of the bulk host semiconductor, and to then subtract this from the total energy of the same group with a defect added. This gives us the energy cost or "formation energy" for creating that defect. The charge state and local structure around the defect can be altered or "relaxed" to minimize the formation energy, thus providing us with a prediction of the structure and charge state that the defect would adopt in a real material. This in turn gives us information about its likelihood of forming, its electrical properties and its stability. Examining how these change as the defect is moved through the lattice gives us information about its diffusion and mobility, which can strongly affect the potential lifetime of a device made from the material in question.

To be more specific, we use first principles plane-wave pseudo-potential density functional theory (DFT) calculations with the local density approximation (LDA), using the VASP (Vienna Ab initio Simulation Package) code. We sometimes supplement this with the use of tight-binding models when interpreting the results. The formation energy is here defined formally as

$$E_d^C = E_T^C(defect^Q) - E_T^C(bulk) - \sum_X n_X \mu_X + Q\varepsilon_F \tag{1}$$

where $E_T^C(defect^Q)$ and $E_T^C(bulk)$ are the total energy of atom group "c" with and without the charge Q defect. To form the defect from the bulk material, n_X atoms of chemical potential μ_X were added or removed. The chemical potential for the electrons, ε_F, (roughly speaking, the Fermi level) is measured from the valence band edge. The values of E_d^C are then compared for different defects, defect complexes and defect structures. Comparing formation energies can indicate which native defects are likely to dominate due to thermal creation or formation under particular growth conditions. It can also indicate which sites are preferred by a particular dopant or impurity, either within a single material or between different parts of a heterostructure. Defect binding energies can also be rather important, and can be calculated by subtracting the formation energy of a complex from those of its substituents.

Scaling

The group of atoms labeled above simply as "c" can take two different forms. Calculations can be done using either a) clusters of atoms with open boundary conditions or b) supercells with periodic boundary conditions (PBCs). In a cluster calculation the group of atoms is "terminated" with (spurious) hydrogen to use up the extra electrons on the outer atoms, but is otherwise surrounded by an infinite amount of empty space. This technique has some strong advantages but the hydrogens and empty space do not behave much like the surroundings of a real macroscopic sample. In a supercell calculation, on the other hand, the cluster of atoms is surrounded by an infinite array of copies of itself. (It is periodically repeated.) If there is no defect present this is perfect for describing real materials. However, if there is a defect present it will interact with an infinite array of spurious "images" of itself seen in the periodic boundaries, and this is a major weakness in the supercell approach. There are two types of unwanted interactions involved: elastic interactions, which can easily be truncated (albeit introducing further errors), and electrostatic interactions which cannot. These result in errors in the calculated formation, binding and migration energies of defects, errors which can be on the same order as the energies themselves. For practical supercell sizes they need not even be negligible for neutral defects, since dipolar and quadrupolar contributions can remain significant.

In recent years various correction schemes have been suggested [1-4] to compensate for these errors. They are usually based upon fits to quasi-classical models and/or multipole expansions of the electrostatic interactions. They have met with varying levels of success but are so far generally considered insufficiently reliable for regular use. There are more direct approaches, however. Probert and Payne [5] recently presented a detailed *ab initio* study of the neutral vacancy in Si, considering all aspects of convergence, from basis set and k-point sampling to size and symmetry of supercells. They demonstrated that the use of 'large' supercells (here meaning 200+ atoms) can be essential for obtaining the correct physical results. We, meanwhile, have been studying the case of InP, with both neutral and charged defects. By examining the dependence on supercell size of both the formation energies and the relaxed structures of various simple defects we have demonstrated the advantages of not only using large supercells but also of finite size scaling the results.

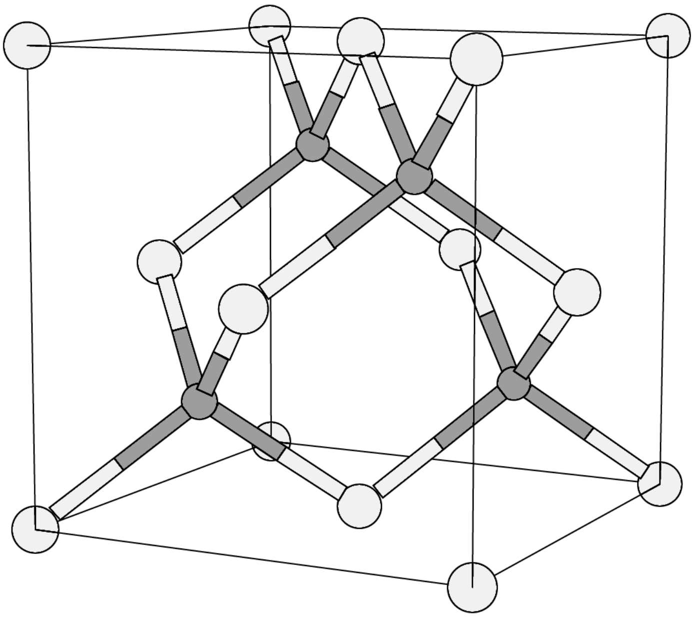


Figure 2
The 8 atom simple cubic supercell of InP, with no defects. (White: P, grey: In.) Larger cells are formed as multiples of this cell: 2x2x2, 3x3x3, 4x4x4 etc. The supercell is then used to mimic an infinite crystal by using periodic boundary conditions. The various defect supercells are formed by adding, removing or exchanging atoms.

The supercell size dependence is itself dependant upon cell symmetry, so we use only simple cubic supercells of 8, 64, 216 and 512 atoms. The 8 atom cell is shown in Figure 2. The other cells are just multiples of it. We allow all atoms to relax, except for those located on the surface of the cell. This restriction truncates the elastic interactions between adjacent supercells, and leaves 1, 5, 7, and 9 shells of atoms free to relax about the defect in the four supercells. No restrictions are placed upon the symmetry of relaxations. Calculations in the 8 and 64 atom supercells can be done easily on local workstations, but for the 216 atoms cell relaxation calculations can take weeks, so most of ours are done at PDC. For the 512 atoms supercell it would be impossible to proceed at all without access to the computing facilities at PDC.

In Figure 3 we show some typical results, taken from a forthcoming papers [6] plotted against $1/L$ where L is the supercell size. It is clear that there is both a linear term and a higher order term and that the signs and magnitudes of the two (both absolute and relative) vary greatly from defect to defect. In some cases the scaling is almost purely linear (Zn_i^{+2} for example) whilst in others (see P_{In}^{+2}) the linear term can be almost zero so that $E_d^C(L)$ first falls and then rises. By careful fitting of various curves to various trial scaling forms [6] we have shown that the higher order term goes primarily as an inverse cube. This is not, in fact, surprising, since it is known [1-4] that the leading electrostatic errors in E_d^C scale as L^{-1} and L^{-3}. Hence we fit our data (solid lines in figures) to

$$E_d^C(L) = E_d^\infty + a_1 L^{-1} + a_3 L^{-3} \tag{2}$$

E_d^∞ is then the finite size scaled formation energy for an infinitely large supercell. To get an idea of the accuracy of the fitting and of the individual values of E_d^C, four more fits (dotted lines) are added for each defect. For each of these lines one of the four data points has been omitted, to give an idea of the error bar on E_d^∞. We almost always get very good fits and rather small error bars. What this shows is that we have indeed managed to account well for the effects of the limited size of the supercells (10s & 100s of atoms rather than 1000s and 10000s). It also shows that the minimum number of fit parameters needed to describe the supercell dependence of E_d^∞ is 3. This in turn means that the minimum number of supercells needed in order to obtain meaningful results is 4, hence this work would simply not be possible without access to the facilities at PDC, which enable us to go up to these large supercells.

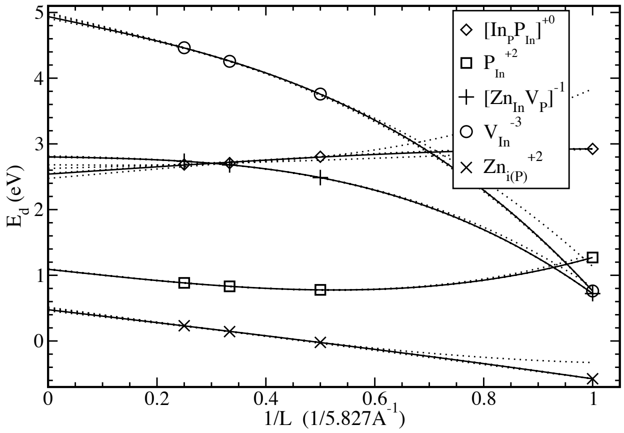


Figure 3
Examples of the scaling of formation energies with supercell size. The horizontal axis is in units of the inverse of the 8 atom supercell length (5.827Å). Hence

1.00 -> 8 atom cell,
0.50 -> 64 atom cell,
0.33 -> 216 atom cell,
0.25 -> 512 atom cell.

(Where we do not get good fittings, for example for V_p^{+0} [7] it is normally due to Jahn-Teller behaviour which changes the local symmetry of defects, and does not fully develop in the smaller cells.)

Many of the results we obtain using the finite size scaling overturn the conclusions previously reached using small cells such as the 64 and 216 atom cells, sometimes even overturning what one might have expected from the 512 atom cell.

For example, most previous calculations to find the most stable charge state of the indium vacancy V_{In} in strongly p-type InP have relied upon the 64 atom supercell, and found the most stable to be -3 (see Figure 4.) This is in fact incorrect. In the 216 atom cell -1 becomes more stable, but we need to go all the way to the 512 atom cell to find that the most stable is really the 0 charge state. A worse case is the relative stability of Zn in its interstitial and substitutional forms. See Figure 5. Zn is a commonly used p-dopant in various III-V semiconductors, and yet we find that in p-type InP the n-doping Zn_i is more stable, even in the 512 atom cell. Only once the results are scaled do the two forms start to become equal in formation energy. Growing the material with a small excess of P over In is then able to tip the balance in favor of Zn_{In} but only in the large L limit is this sufficient. The results in the 512 atom cell alone remain unphysical until scaling is taken into account.

Compensation

Another issue we have looked at recently is the structure of defects in heavily Zn doped p-type InP. This is one of the most common materials used within optoelectronics. The Zn sits substitutionally within the In sublattice (Zn_{In}) where it has a shallow acceptor level. A well known limitation in the use of Zn as a p-dopant in InP is the saturation of the hole concentration in the mid 10^{18} cm^{-3} range. Above this, further increases in Zn concentration do not translate into increases in hole carrier concentration. There are several suspected causes [8] for this, in particular: a) increases in the concentration of interstitial zinc (Zn_i), b) phase separation and c) the formation of other compensating defects, especially phosphorus vacancies (V_P) and complexes of zinc with V_P. The results shown in Figure 5 indicate that the formation of Zn_i is probably quite favourable, but this may well not prove to be the sole cause. Indeed, the formation of Zn-vacancy complexes has been recently observed by two experimental groups using positron annihilation, [8] and [9]. Both groups identify the formation of the same complex, thought to have a binding energy around 0.1-0.4 eV. Both identify the occurrence of two alternative structures for it, associated with the charge states 0 and -1, with the 0/-1 transition occurring when the Fermi level lies 0.2±0.1 eV above the valence band edge. One structure has an effective vacancy volume about the same as that of the free phosphorus vacancy whilst the other has a larger volume. Where the two experimental groups disagree is over the assignment of the charge states. One group labels the larger structure as the 0 charge state, the smaller one as the -1 state. The other group makes the opposite assignment. We recently answered the question computationally [10] specifically identifying the structure of the complex and at the same time determining the extent to which formation of the complex is involved in the problem of carrier compensation. We started with calculations for the most likely candidate complex: namely [Zn_{In}-V_P] – a substitutional Zn atom bound to a nearest neighbor V_P. (It can alternatively be viewed as a V_P with one of its 4 nearest neighbor In atoms replaced by a Zn.)

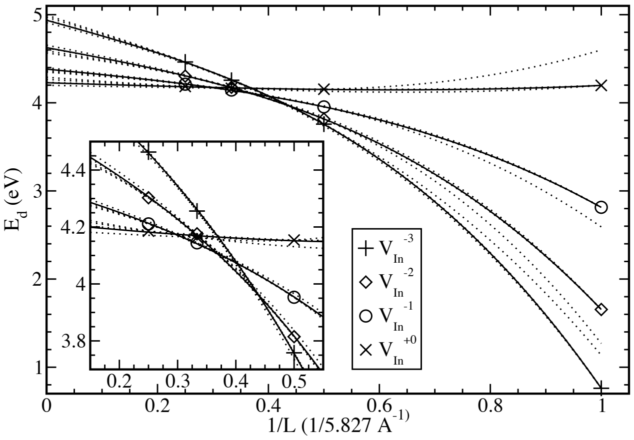


Figure 4
Scaling the formation energy of V_{In} for various charge states in strongly p-type InP (i.e. for $\varepsilon_F = 0$). The most stable state is $Q = 0$ in the large supercell limit, but this does not show up until the 512 atom cell is considered.

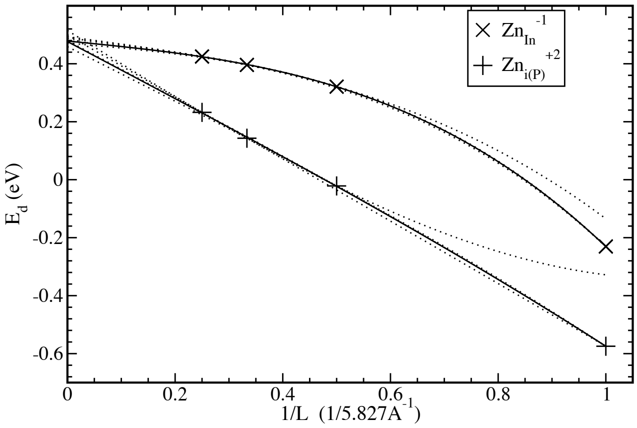


Figure 5
Scaling the formation energy of Zn_{In}^{-1} and $Zn_{i(P)}^{+2}$ in strongly p-type InP. Even in the 512 atom cell Zn remains much more stable in the (experimentally incorrect) interstitial position. The relative stabilities do not become apparent until the results are scaled.

We found that the neutral state of the defect is the most stable form in strongly p-type InP. The structure is shown in Figure 6a: The Zn atom sinks back away from the vacancy into a so called "DX"-like position. This means that the sp^3 bonding at the Zn re-arranges to form three sp^2 bonds with the 3 remaining P neighbors, leaving an empty Zn:$4p_z$ orbital pointing into the vacancy. In Figure 6a we show a section through the "Electron Localization Function" (ELF, see [11]) of the complex. This (roughly speaking) indicates where the electrons are most localized, which in turn corresponds to areas of bonding. In the figure we see bonding between the three In atoms neighboring the V_P, but no bonding at all between Zn and In. This bond re-arrangement effectively increases the volume of the vacancy, exactly as seen in the experiments. In this neutral charge state we found a binding energy of 0.39 eV: right in the range anticipated. We found the 0/-1 transition occurring at 0.50 eV. This is a little higher than expected, but this value is not usually expected to come out of the calculations as accurately as, say, the structure or binding energy. We then examined all of the possible charge states of the complex, right from +2 to -6, and found that the volume changes monotonically across the entire range, shrinking from 14.6 Å^3 at +2, to 7.6 Å^3 at 0, 6.5 Å^3 at -1 and on to 3.1 Å^3 at -6. We find that the free V_P, as previously seen with positrons [10], [12] has a volume of about 5.9 Å^3.

We are thus able to fully reproduce the experimental results, at the same time answering the question of which charge state to ascribe to the observed effective vacancy volumes. This confirms the assignment of the measured Zn-V_P complex specifically to the nearest neighbor [Zn_{In}-V_P] combination. However, we were able to go further, to explain why the shrinkage occurs for example. By studying the difference in charge densities when an electron is added to or removed from each charge state of the complex we found that all electrons added/removed from the complex go to/from the dangling bonds on the 3 remaining In neighbors of the vacancy. See Figure 6b for the case of the 0/-1 transition, for example.

The DX structure and sp^2 bonding at the Zn do not change until the -5 charge state is reached, which occurs well into the conduction band and hence will never actually be seen. As electrons are added to the In atoms the bonding between them strengthens, shortening the In-In distances monotonically. This also increases the amount of negative charge on the In side of the complex. The absence of $4p_Z$ electrons, on the other hand, leaves the positive core of the Zn atom slightly less well screened in the direction pointing towards the In atoms. Hence the Zn-In distances also shrink monotonically with increasing negative charge, but for essentially electrostatic reasons. The experimentally observed volume change is thus due partly to increased bonding, and partly to increased electrostatic interactions across the complex.

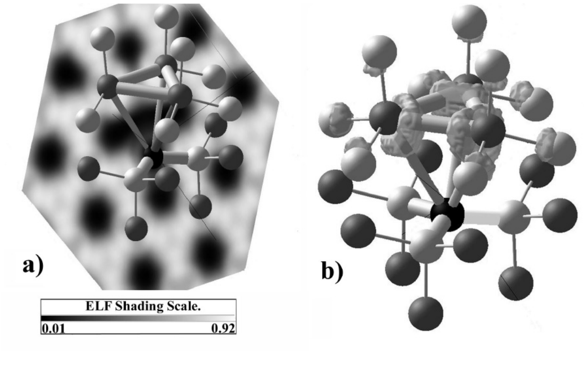


Figure 6
Geometry of the [Zn_{In}-V_P] defect complex in InP. White balls: P, grey balls: In, black ball: Zn. a) ELF density in a plane though the Zn and two of the In atoms. (ELF = 1 means strong bonding, ELF = 0 means no bonding at all.) b) charge density surface showing where the additional electron goes at the 0/-1 transition.

Returning to the issue of compensation, however, we find that the formation of V_P is certainly involved in compensation: in p-type material it forms in a +1 charge state, compensating one carrier per vacancy formed. However, when this V_P^{+1} meets and pairs with an ionized Zn dopant atom (Zn_{In}^{-1}) to form $[Zn_{In}\text{-}V_P]^0$, no further electrons are added or removed from the complex, so no additional compensation takes place. The only way that the existence of the complex could help to explain the compensation issue would be if the binding energy of the complex meant that the effective energy cost for creating additional V_P^{+1} was significantly lower on the P sites neighboring the Zn. However, we find the (scaled) formation energy for free V_P^{+1} to be around 2.5 eV. The binding energy of ~0.4 eV gives an effective formation energy of about 2.1 eV when the V_P neighbors the Zn. This is still far too large to explain the carrier concentration. Hence our current results point towards either phase separation or, (perhaps more likely,) the formation of additional interstitials (particularly Zn_i and P_i) as the principle cause of carrier compensation in heavily Zn doped InP.

References

1. G.Makov & M.C.Payne, Phys Rev B 51 (1995) 4014
2. P.A.Schultz, Phys Rev B 60 (1999) 1551
3. L.N.Kantorovich, Phys Rev B 60 (1999) 15476
4. H.Nozaki & S.Itoh, Phys Rev E 62 (2000) 1390
5. M.I.K.Probert & M.C.Payne, Phys Rev B 67 (2003) 75204
6. C.W.M.Castleton & S.Mirbt, accepted to appear in Phys Rev B (2004) and in preparation.
7. C.W.M.Castleton & S.Mirbt, Physica B 340-342 (2003) 407
8. J. Mahony, P. Mascher and W. Puff, J Appl Phys 80 (1996) 2712
9. J. Slotte et al. Phys Rev B 67 (2003) 115209
10. C.W.M.Castleton & S.Mirbt Phys Rev B 68 (2003) 085203
11. B. Silvi and A. Savin, Nature 371 (1994) 683
12. Alatalo et al. Phys Rev 51 (1995) 4176

Monte Carlo Simulation of Condensed Matter Systems

Mats Wallin, Anders Vestergren, Marios Nikolaou, and Jack Lidmar
Condensed Matter Theory, Royal Institute of Technology (KTH), Stockholm

This is a nontechnical review of our computer simulation project of phase transitions in superconductors and superfluids, which is the research area that was awarded the Nobel prize in physics in 2003. The project focuses on effects created by the presence of strong random disorder. Our simulation results can predict how measurable quantities, for example, the resistance, behave near the superconducting phase transition. The simulation of models of strongly disordered systems is made possible by excessive use of a parallel computer facility. Our simulations often involve thermal averaging as well as averaging over random defects present in the system. This double averaging requires massive computer resources. We have been successful with various very demanding simulations of glass transitions of magnetic flux lines in superconductors.

Phase transitions in condensed matter physics

Condensed matter physics is about understanding properties that arise when a macroscopic number ($N \sim 10^{23}$) of interacting particles come together, that cannot be directly understood from the properties of a single particle. For example, depending on external parameters like temperature and pressure, the same element may be in different phases with completely different properties: solid-liquid-gas, insulating-conducting-superconducting, magnetic-nonmagnetic, transparent-nontransparent, etc. Such emergent behavior underlies the vast richness of different properties in nature, and analyzing such properties is important for understanding how nature works. One main aim of the research is to explain the richness of different phases, and phase transitions between these phases. Such investigations are often very challenging, and computer simulation is often an indispensable tool. Simulations can be used both to understand the properties of a given model, but also to systematically vary system parameters to optimize properties, in a way that would be too slow and expensive in the lab. It also allows very useful visualizations that can lead to valuable insights into the problems studied. An illustration from our simulation is shown in Figure 1.

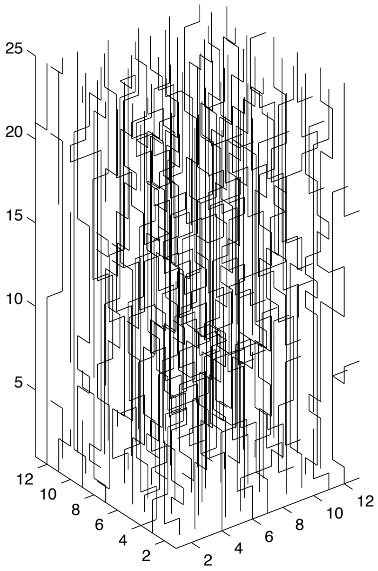


Figure 1
Vortex line configuration generated in our Monte Carlo simulation of the superconducting phase transition in the presence of random disorder.

The superconducting phase transition

An explosion of research followed the discovery of high temperature superconductivity in 1986, but despite massive efforts many important problems are still unsolved. Superconductivity research includes both basic and applied science, and has various important technological applications. The research aims at understanding how superconductivity works in detail, and to use such insights in order to construct superconductors with optimal properties for various applications.

The project mainly considers various so called second order phase transitions in the presence of sample imperfections that play the role of random disorder. Why is it important to understand the effects of disorder? Disorder is always present in real systems, and whereas it often degrades the properties of a material, for superconductors it often has the opposite effect and can actually improve the properties of the superconductor, which is highly desirable.

In our discussion, which is highly simplified and non-technical, we will mainly use the example of superconductivity as a model system. For the other physical systems in this project, the physics is closely related. As the temperature is lowered from the normal metallic phase, through the superconducting transition temperature T_C, into the superconducting phase, the resistance drops from a finite value in the normal phase to zero in the superconducting phase. In high temperature superconductors T_C is typically about 100 Kelvin, which is –173 degrees Celsius. One important research problem is to investigate the details of how the resistance goes to zero at $T = T_C$.

Ohm's law tells us that voltage V is resistance R times current $I, V = RI$. If the resistance is zero, a current can flow without any voltage. A current induced in a superconducting ring will therefore run "forever", without any battery, and is called a supercurrent. On the other hand, if the resistance is nonzero, a voltage source is needed to maintain the current. Therefore resistance is associated with the degradation of supercurrents, or with voltage across the system, so that the power dissipated is $P = VI$. In the superconducting state the quantum mechanical wave function of the superconducting electrons becomes nearly constant across the sample. Superconductivity is a macroscopic quantum effect that can be easily produced, for example in an ordinary piece of lead wire, by just cooling below its superconducting transition temperature.

Voltage can arise by the Faraday induction law, which associates a voltage with a time dependent magnetic field. Small time dependent magnetic field fluctuations are formed in the superconductor as thermal fluctuations, having the form of loops of closed magnetic field lines, or vortices, inside the superconductor. A supercurrent circulates around the magnetic field inside each vortex, that shields the magnetic field away from the interior of the superconductor. A configuration of vortex lines generated in our simulation is shown in Figure 1.

Critical power laws and scaling analysis

Scaling analysis means to relate physical measurements to characteristic scales of length and time. This is very useful both for analyzing experiments and numerical simulation data, and offers ways to connect results from theory, experiments, and simulations.

The typical size of a vortex loop is called the correlation length, ξ . At the phase transition, loops become big and approach the size of the sample. The loop expansion is described by a divergence of the correlation length of the form $\xi = constant \times |T - T_c|^{-\nu}$, where v is called the correlation length exponent. Similarly, typical loop motion becomes slow at the transition, and the associated time scale, called the correlation time τ, diverges like $\tau \sim \xi^z$, where z is the dynamic critical exponent. A main problem in the project is to find such power law relations and calculate the critical exponents.

Critical power laws, like the ones for the correlation length and time, are directly measurable in experiments, and values for the critical exponents can be determined from experimental data. For example, the resistance close to the transition behaves like $R \sim |T-T_c|^{vz}$. This relation follows from a scaling argument, which is much too complicated to go into here in detail [1, 2], but a simplified description goes as follows. To relate the resistance to the correlation length and time, use the fact that the resistance R is proportional to voltage by $V = RI$. Voltage is proportional to the rate of change of the difference of the phase of the superconducting wave function across the system, which has dimension of one over time, i.e., $V \sim 1/\tau$. Hence $R \sim V \sim 1/\tau \sim 1/\xi^z \sim |T - T_c|^{\nu z}$, where the last step used $\xi \sim |T - T_c|^{-\nu}$.

The above argument shows how the resistance behaves close to the superconducting transition at T_c. However, in a simulation of a finite system, the sidelength L plays the role of the characteristic length scale at the transition. For the resistance this gives $R \sim L^{-z}$ at $T=T_c$. Similar so called finite size scaling relations hold for other quantities at $T=T_c$. For example, the density of superconducting electrons is $\rho_s \sim |T - T_c|^{(d-2)\nu} \sim L^{2-d}$. These relations can be used to connect experimental results to finite size scaling properties of the corresponding quantities calculated in a simulation.

Models of superconductors

Here we will briefly describe some of the physical models used in the project. We typically consider models defined on a three-dimensional simple cubic lattice of size $L \times L \times L_z$. The *three dimensional XY model (3DXY)* is defined by the Hamiltonian, or energy function

$$H = -\sum_{\langle i,j \rangle} J_{ij} \cos(\theta_i - \theta_j - A_{ij})$$

Here $\Sigma_{<i,j>}$ denotes the sum over all near neighbor pairs of lattice sites, J_{ij} is a constant, θ_i is the phase of the superconducting wave function at site *i*, and A_{ij} is proportional to the integral of the vector potential from site *i* to *j*. Vortices are permitted by the periodicity of the cosine function, given by plaquettes in the lattice around which the phase winds by $\pm 2\pi$. A vortex line configuration from our simulation is shown in Figure 1. Random disorder can be modeled by using a position dependent coupling constant J_{ij}.

The *lattice loop model* is defined by

$$H = \frac{1}{2} \sum_{ij} q_i V_{ij} q_j + \sum_i g_i q_i^2$$

Here q_i is a line variable on the link *i* in the lattice, which is constrained to form closed loops, possibly closing across the periodic boundaries of the system. V_{ij} is the interaction between parallel line segments in the system, which is a periodic version of a screened Coulomb potential, $V = e^{-r/\lambda}/r$, where the λ is the range of the interaction. The last term models disorder in terms of a position dependent random coupling g_i.

Remarkably, the loop model describes two very different physical systems, that are both of great current interest [3, 4]. First, the loops can describe vortex loops in superconductors, and is then closely related to the model above formulated in terms of phases. Indeed a mathematical transformation exists between the two models in the case without random disorder. Second, the loops can be worldline trajectories of quantum Bose particles. Thus there is a deep and beautiful analogy between quantum worldlines of bosons and vortex loops in superconductors. This analogy is very useful, because the same simulation gives very useful predictions for both systems.

The probability distribution for vortex loop configurations is a Boltzmann distribution, such that the probability for a configuration is $P = e^{-\beta H}/Z$, where $\beta = 1/T$ is inverse temperature, *H* is the energy of the configuration, and *Z* is the normalization factor. The probability distribution for the random disorder is usually taken to be a rectangular distribution.

Monte Carlo methods

In the simulations we calculate thermal and disorder averaged expectation values of physical variables that can be compared to experiments. For example, the average energy is defined as

$$E = \langle H \rangle = \frac{1}{NM} \sum_{sv} H(s, v)$$

where s runs over N generated vortex configurations with a given disorder configuration, and v runs over M disorder configurations, where we typically use $M = 10^3$. The superfluid density is $\rho_s = \partial f^2/\partial\Delta^2$, where f is the free energy density and Δ is a twist in the phase at the boundary of the sample, or, alternatively, proportional to the mean square fluctuation of the vortex loop area [2]. The resistance is obtained from the Kubo formula [1]

$$R = \frac{1}{2T}\sum_{t=-\infty}^{\infty}\langle V(0)V(t)\rangle$$

At a phase transition the dynamics of the system slows down and the correlation time diverges [1]. This means that the simulation becomes very hard to converge, and it becomes highly desirable to use improved MC schemes to enhance the convergence. We use three such methods that we will briefly describe. These methods can improve the convergence of the simulations by several orders of magnitude compared to a naive approach.

The *Wolff algorithm* [5] is a convergence acceleration scheme that applies the 3DXY model in zero magnetic field. Here a new configuration of θ_i is generated by simultaneously updating a whole cluster lattice sites. The cluster is grown from an initial seed site, and the growth is terminated with a probability given by the Boltzmann distribution. This collective update method eliminates the critical slowing down of the dynamics, and vastly enhances the convergence. It is enough to use about 200 update attempts per lattice site to reach sufficient convergence of the simulation.

The *worm algorithm* [6] applies to the loop model. The MC moves are directed random walks of inserting connected line segments on the links of the lattice. The probability for a link to be added to the walk is given by a local probability distribution related to the Boltzmann distribution. A loop move is completed when the random walk terminates at its starting point which is chosen at random. The usefulness is mainly in the case of short range interactions, where convergence of the simulation is dramatically improved.

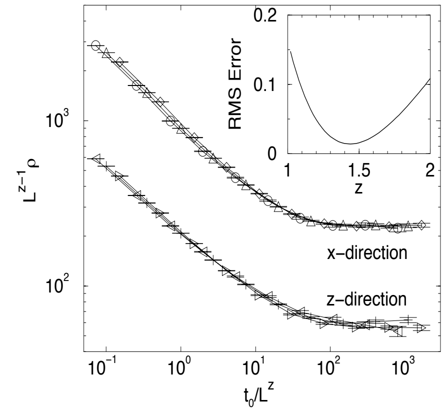
The *temperature exchange algorithm* [7] applies to disordered systems. A set of n copies of the system sharing the same disorder is simulated in parallel, at a set of closely spaced temperatures. In addition to the usual MC update, attempts are made to exchange configurations between the different temperatures. The effect is to repeatedly rise and lower the temperature for each copy of the system, which avoids getting stuck to metastable states of the disorder potential, and thereby greatly improves convergence.

Results for the vortex glass

The vortex glass transition in high temperature superconductors in magnetic fields with random pointlike disorder has attracted enormous attention, both theoretically, experimentally, and from simulations. The vortex glass phase is a superconducting phase at low temperature where the vortex lines are frozen in positions determined by the distribution of the random disorder in the system. Theoretically the existence and properties of the vortex glass have been quite unclear despite several previous simulations. We have done the first simulation that conclusively demonstrates the existence of the vortex glass, and calculated several properties [8]. A vortex glass configuration is shown in Figure 1, which shows a line configuration which is frozen in time or glassy.

Figure 2 shows the scaling properties of the resistivity ρ, which is proportional to the resistance. The figure shows that the resistance goes to zero at the vortex glass transition, which means that the vortex glass phase in our simulation is indeed superconducting. The scaling properties of the data, where simulation results for different system sizes are scaled with suitable powers of space and time to make the data curves fall on top of each other, give information about the critical exponents z,ν that can be used to compare our simulation to experiments. During the grant period we have also studied scaling behavior of a vortex glass model in two dimensions. Here we have for the first time shown that there is no glass phase at any finite temperature from an equilibrium dynamic scaling analysis.

Figure 2
Scaling plot of MC data for the resistance in the vortex glass.



Results for the Bose glass

We have done extensive studies of the so called Bose glass transition [9], which is the vortex glass transition but with linearly correlated disorder, instead of uncorrelated point disorder. This also applies to the localization transition of quantum boson worldlines in (2+1) space-time dimensions. Figures 3 and 4 show MC data for the superfluid density and the resistance at the Bose glass transition, for a model with no screening of the interaction and a finite applied magnetic field, which is studied here for the first time. From a scaling analysis we obtain critical exponents that can be compared to experiments. We have also studied the Bose glass transition in zero field during the grand period.

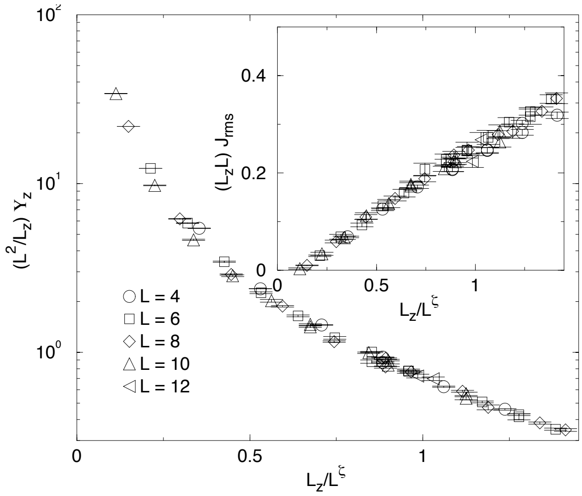


Figure 3
Scaling of MC results for the superfluid density of the Bose glass model.

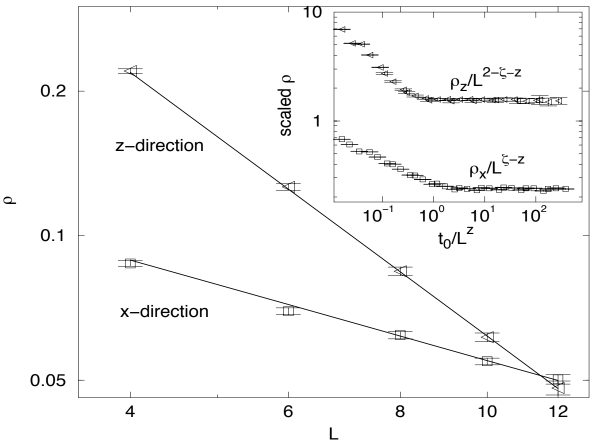


Figure 4
MC data for resistance of the Bose glass.

Significance of the PDC facilities for the project

The simulations during the time 2000-2002 were performed on the Strindberg facility at PDC, KTH. Our simulations of disordered systems are very CPU demanding, despite our sophisticated convergence acceleration algorithms. The reason is that the simulations involve a double average over vortex configurations and over disorder configurations. Since the disorder averaging involves many independent simulations, it is perfectly suited for a cluster computer system. Our project has been made possible to a large extent due to the PDC grant. Our wish for the near future is a huge cluster system where the demands of our simulations can be fully satisfied.

References

1. J. Lidmar, M. Wallin, C. Wengel, S. M. Girvin, and A. P. Young, Phys. Rev. B 58, 2827 (1998).
2. J. Lidmar and M. Wallin, Phys. Rev. B 59, 8451 (1999).
3. M. Wallin, E. S. Sørensen, S. M. Girvin, and A. P. Young, Phys. Rev. B 49, 12115 (1994).
4. J. Lidmar and M. Wallin, Europhys. Lett. 47, 494 (1999).
5. The Wolff algorithm was introduced in: U. Wolff, Phys. Rev. Lett. 62, 361 (1989).
6. The worm algorithm is described in: F. Alet and E. S. Sørensen, Phys. Rev. E 67, 015701(R) (2003).
7. The temperature exchange method was introduced in: K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. 65, 1604 (1996).
8. A. Vestergren, J. Lidmar, and M. Wallin, Phys. Rev. Lett. 88, 11704 (2002).
9. A. Vestergren, J. Lidmar, and M. Wallin, Phys. Rev. B 67, 092501 (2003).

Weak Chemical Bonds

**M. Cavalleri, D. Nordlund, H. Ogasawara, H. Öström, B. Brena,
M. Nyberg, M. Odelius, A. Nilsson and L.G.M. Pettersson**
FYSIKUM, AlbaNova, Stockholm University

In spite of its biological, technological and industrial importance the description of H-bonding in terms of different chemical interactions remains an open issue. Theoretical models based on small clusters of atoms or molecules can be run on Linux PCs, but more extended, periodic models require the resources of PDC, especially for our ab initio simulations of the spectra of liquids where the dynamics becomes important. The large number of structures that must be investigated also allow for a parallelization with e.g. 32 smaller cluster models run simultaneously on 32 processors. Below we will give some examples of what can be learned from the combination of experimental and computational techniques that we use in our studies.

Introduction

The hydrogen bond (H-bond) between water molecules (5-10 kcal/mole) is substantially weaker than normal covalent or ionic bonds (50-200 kcal/mole), but it is still the factor providing the unique properties of liquid and solid water. It is furthermore responsible for *e.g.* base-pairing among the nucleobases in DNA and for the overall ordering observed among amino acids adsorbed on a substrate. In spite of its biological, technological and industrial importance the description of H-bonding in terms of different chemical interactions remains an open issue.

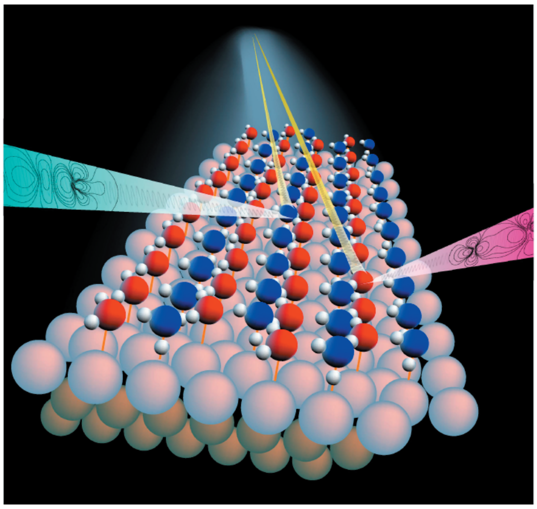


Figure 1
The figure depicts atom-specific probing of the water-surface chemical bond using soft X-ray spectroscopy. When an oxygen atom in water is excited by a soft X-ray, the excited atom emits photons or electrons. The analysis of energy and polarization of absorbed and emitted photons gives insight into the surface chemical bond of waters binding respectively through oxygen and hydrogen (red and blue cones). We can compute the soft X-ray spectra to visualize the molecular orbitals of these species. (Artwork by H. Ogasawara.)

It is still surprising to find that application of new experimental techniques in combination with density functional theory (DFT) calculations and large-scale simulations demonstrate that even the local coordination of liquid water holds surprises: each molecule is H-bonded to only two neighbors rather than close to four as presently believed [1]. Furthermore, when a thin film of ice forms on a metal surface, the structure of the first, or contact, layer on the metal does not resemble normal ice. Instead a new form of ice, flat ice, is formed [2]. Characteristic for this layer is that every other molecule forms an unusual bond to the metal through one of the hydrogens instead of through the oxygen as illustrated in Figure 1. This type of bond is interesting since it may be viewed as representing the initial step in the dissociation of water on the metal, which in turn can lead to the release of hydrogen. In terms of a clean, environmentally innocuous fuel for the future, hydrogen generated from splitting water holds the greatest promise at present.

Another instance where hydrogen interaction with a metal surface is of great importance is in dehydrogenation of hydrocarbons. The saturated alkanes, such as methane, ethane and e.g. *n*-octane, show only weak bonding to metals, that is commonly denoted physisorption with bond strengths comparable to a weak H-bond. As we find for the case of H-bonding it is not correct, however, to conclude that the weak physisorption bond is the result of a weak interaction. We find significant changes in the electronic structure that can only result from stronger interactions, where a balance between attractive and repulsive forces in the end result in only a weak net attraction. Thus, also in the case of *n*-octane on the copper (110) surface we find that the common picture of a physisorbed alkane must be revised [3,4].

The investigations mentioned above have all applied a direct combination of experimental and theoretical techniques. Our group combines both theory and experiment and is furthermore located both at Stockholm University and at the Stanford Synchrotron Radiation Laboratory, California. The experimental techniques use synchrotron radiation to measure the unoccupied electronic levels through X-ray absorption spectroscopy (XAS) and the occupied levels through X-ray emission spectroscopy (XES). Since both processes involve transitions from or to a strongly localized core-level these spectroscopies measure the electronic structure at a specific atom in a molecule in a liquid or adsorbed on a surface. The transitions are of dipole type, which for a 1s core-level implies that only the local 2p character of the molecular orbitals at that site is observed. We have implemented techniques to compute the corresponding spectra theoretically and, having ascertained that experimental and theoretical spectra match, we can provide further information on the electronic structure from the calculations. The theoretical models are based on either small cluster models of the surface that can be run on our Linux PCs or on more extended, periodic models requiring the resources of the PDC. The latter is the case in particular for our ab initio simulations of the spectra of liquids where the dynamics becomes important. Below we will give some examples of what can be learned from this combination of experimental and computational techniques.

Alkanes on metals: the case of octane

Alkanes are saturated hydrocarbons, i.e. contain only single bonds between the carbon atoms, and thus do not have e.g. a low-lying π^* orbital with which to mix in order to form new bonds. Due to this they are rather unreactive and adsorb on metals with a sufficiently low binding energy per carbon for the interaction to be characterized as physisorption.

For a physisorbed molecule one does not expect large changes in the electronic structure and in particular only minor effects on the geometry of the adsorbate. However, this turns out not to be true for one of the classical cases of physisorption, i.e. alkanes on copper and platinum, where on the latter surface even methane (CH_4) shows new electronic states from the interaction with the metal. In the case of octane on Cu(110) the molecules form an ordered overlayer with adsorption on the rows in the surface (Figure 2). This allows projecting the electronic structure on all three directions using symmetry-resolved X-ray absorption spectroscopy (XAS) and X-ray emission spectroscopy (XES) in combination with density functional theory (DFT) spectrum calculations [3,4]. We find new adsorption-induced states in the XES spectra, which we assign to interaction between the bonding CH orbitals and the metal surface. These are not immediately reproduced in the theoretical spectra computed at the optimized geometry; the distance to the surface is too long.

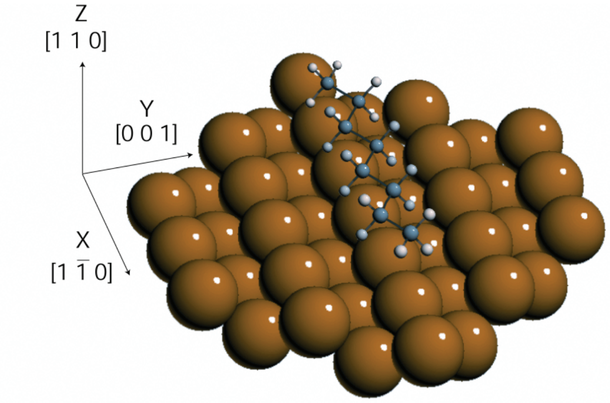


Figure 2
n-octane adsorbed on row of Cu atoms in the (110) surface

The problem turns out to be related to the difficulties in DFT to describe weak interactions and in particular the van der Waals interaction between separated or weakly connected systems. The interaction curve is very shallow and a small error in the energetics can result in a large error in the computed distance over the surface. The spectral features, however, provide a more sensitive measure of the distance than the total energy since they depend on molecular orbital changes as the adsorbate approaches the surface and these in turn depend on the overlap leading to essentially exponential distance dependence. Comparison of measured and computed spectra for different geometries can thus provide an alternative approach to structure determination in the absence of a reliable treatment of van der Waals interactions in DFT.

By performing a systematic computational investigation of the influence of different structural parameters on the XAS and XES spectra we are then forced to conclude that the molecule must be substantially closer to the substrate than given by the optimization and that the molecular geometry is significantly distorted relative to the gas phase structure [4]. The bonding to the surface leads to a strengthening of the carbon-carbon bonds and a weakening of the carbon-hydrogen bonds, consistent with a rehybridization of the carbons from sp^3 to $sp^{2.8}$ [3]. Although the resulting modifications are relatively small in magnitude they indicate that the picture of physisorption and weak bonding to a substrate must be reconsidered; rather than due to a weak interaction we find that the small adsorption energy is due to a delicate balance between significant exchange-repulsion coupled with some covalent bond-formation to the surface. The bonding involves a direct C-H····Cu interaction involving some donation from the metal into the antibonding C-H σ^* orbital resulting in loss of C-H bond-order and, as a consequence, increase in internal C-C bonding. That adsorption of an alkane on copper can result in significant molecular orbital changes is unprecedented and leads to new insights on the mechanism of C-H activation on metal substrates.

Adsorbate stabilization through hydrogen bonding: the amino acid glycine on copper

To address the mechanism of self-assembly of bio-organic molecules on surfaces it is important to analyze the intermolecular interactions driving the structure formation. In many systems the dominating intermolecular interaction takes place via hydrogen bonding (H-bonding). One such example is the simplest amino acid glycine, which has an interesting and rich surface chemistry where several different adsorption structures are obtained during the deposition on a metal surface [5,6]. The deprotonated species that forms the most stable structure is termed glycinate.

In our previous studies the intermolecular interactions were not taken into account in the theoretical model, since it was clear that the dominating binding contribution was to the substrate and the main goal was the local electronic and geometric structure. In the case of glycinate, hydrogen bonding between the molecules is a much weaker interaction (up to a few tenths of an eV) than the adsorbate-surface chemical bond (several eV). Therefore, the local structural properties of the molecule, such as orientation, bond lengths and angles, are largely determined by the interaction of a single molecule with the surface.

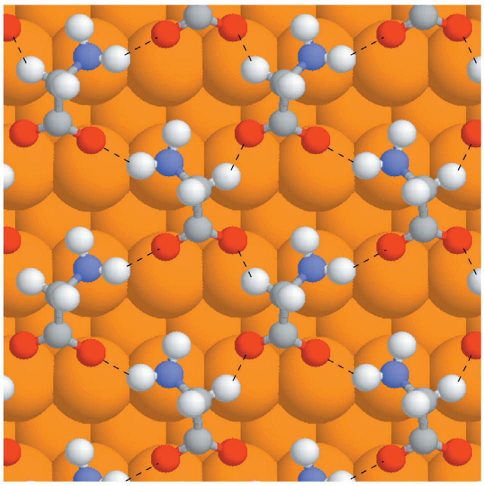


Figure 3
Periodic overlayer structure of glycinate on Cu(110). Color coding: nitrogen blue, oxygen red, carbon gray and hydrogen white.

However, there were certain aspects, in particular in the computed N XAS spectra, where an overall agreement between theory and experiment was missing that could be attributed to intermolecular interactions. Furthermore, the specific (3x2)pg LEED pattern of glycinate on the surface [7] is not easily explained assuming noninteracting glycinate molecules. Instead intermolecular forces between adsorbates should be the underlying reason for the observed LEED pattern for these molecules.

As can be seen in Figure 3, each oxygen is involved in H-bonds, one where one oxygen binds to a NH-group and one where one oxygen binds to a CH-group and to the other proton of the NH-group. The expectation is that the NH-O hydrogen bond is much stronger than the CH-O bond due to the stronger electronegativity of the N atom as compared to the C. As seen from the figure the amino group can participate in two hydrogen bonds, although the distance and the angle are optimal in only one direction.

For the optimization of the structures we used periodic DFT calculations on a 3-layer slab with the CPMD code [8]. The second and third layers of the Cu atoms were held fixed after full optimization of the clean slab. We used the B-LYP functional [9,10] norm-conserving pseudo-potentials with a 90 Ry kinetic energy cut-off for the plane wave expansion for the Kohn-Sham wave-functions and a (221) Monkhorst-Pack k-point mesh [11]. The structure was furthermore optimized with the DaCapo code [12], where we used ultra-soft pseudopotentials [13] with a cutoff of 25 Ry for all the atoms. The structures were optimized using a two-layer slab model of the surface. The calculations were performed at the GGA level [14] using a (441) k-point mesh [11]. The topmost Cu atoms were allowed to move while the second layer was held fixed. The unit cell consisted of 12 Cu atoms (two layers) and two glycinate molecules.

The calculations of the spectra were done using both the periodic models with the CPMD code [8] and on cluster models of the adsorbate system using the StoBe-deMon code [12]. In the latter case the largest model studied consisted of three glycinate on an all-electron Cu_{37} cluster, which is the minimum required in order to describe the different possible H-bonds to one of the molecules. The results in the case of the nitrogen K-shell XAS for models of a single adsorbate, two interacting glycinates and finally the full three glycinates on Cu_{37} are shown in Figure 4 below [15].

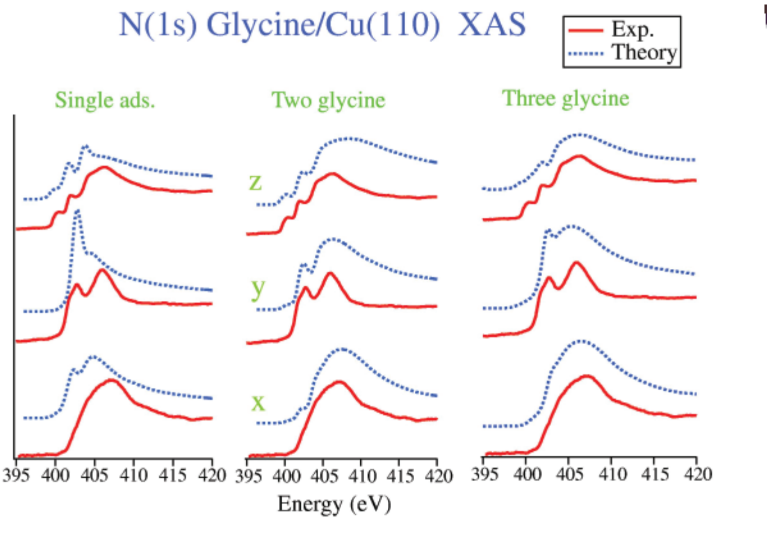


Figure 4
Comparison of computed symmetry-resolved spectra for one, two and three glycinates on cluster models of Cu(110). Red lines are experimental spectra.

The comparison with experiment is improved dramatically just by inclusion of the nearest neighbor molecule (see the middle column which displays the calculated spectra for the model consisting of two hydrogen bonded molecules). The most dramatic effect is seen in the p_y spectrum, where for only one adsorbate the first peak at about 402 eV is dominating whereas the experimental spectrum shows a strong feature at 407 eV. When we include the neighboring molecule providing the NH-O hydrogen bond in the simulation the second peak instead becomes dominant in accordance with experiment. Also the appearance of the p_y and p_z spectra is vastly improved over the single-adsorbate model. It is clear that the main discrepancy of the single-adsorbate model is the lack of hydrogen bonding to neighboring adsorbates. The quality of the computed N1s spectra is not further improved by inclusion of a second H-bond (bonded to the H-bond accepting carboxyl group) as in the third column in the figure. Due to the local character of the utilized spectroscopy, the CH-O bond cannot significantly perturb the N spectra, but has some minor effects on the O and C spectra, since these atoms are directly involved in that particular bond. In conclusion it is clear that including the weak H-bond in the model has profound effects on the electronic structure as measured through XAS and reproduced by the calculations [15]. We can thus use XAS as a sensitive measure also of the presence of H-bonding.

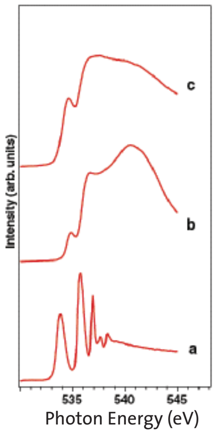


Figure 5
XAS spectra of water. (a) gas, (b) ice, (c) liquid

Hydrogen Bonding in Liquid Water

Having established that H-bonds are detectable using XAS we now turn to the classical case of liquid water. The problems connected with XAS electronic structure measurements on the liquid due to the high vapor pressure in combination with the ultra-high vacuum required for electron detection has recently been overcome [1]. In addition, improvements in the resolution in the hard X-ray regime has enabled X-ray Raman scattering (XRS) to be applied [16], the two techniques have quite different penetration depths into the liquid and provide independent measures of the electronic structure. The surprising result of these measurements is displayed in Figure 5, where we compare the XAS spectra of water in three aggregation states: (a) gas, (b) solid ice and (c) liquid. Considering the local character of XAS it is immediately clear that the local environment in the liquid must be significantly different from the tetrahedral coordination in bulk ice [1]. This is a surprising and controversial conclusion since neutron and X-ray scattering data in combination with molecular dynamics (MD) simulations has lead to the conclusion that the average coordination in the liquid should be similar to that in ice; on the average 3.5 H-bonds albeit distorted compared to the ice. So what is then the coordination in the liquid according to these new data?

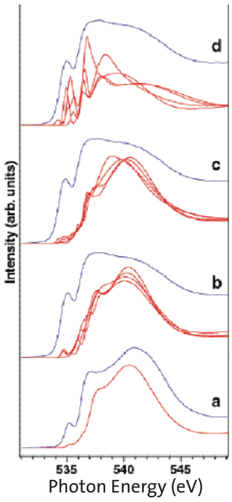


Figure 6
Computed XAS spectra (red) compared with experiment (blue). a) Bulk ice, b) distorted tetrahedral coordination in the liquid, c) broken acceptor bond in the liquid and d) broken donor bond in the liquid.

In order to address this question we have in our earlier work [1] performed classical MD simulations to generate cluster models for specific situations, i.e. as indicated in Figure 6 where we reproduce well the experimental ice spectrum (the features at 535 and 536.5 eV are due to defects and surface contributions) and find similar spectrum contributions for all cases, except when a donating H-bond is broken (d, top). The asymmetry at the free O-H in this case leads to a localization of the first excitation as an antibonding O-H orbital with enhanced p-character [17]. The intensity in the pre-edge region (535-537 eV) is entirely due to molecules in this bonding situation and we immediately see that a majority of the molecules must have a broken donating H-bond.

This interpretation is based on a large number of theoretical simulations, but the conclusions must be checked by independent experiments. The surface of ice provides such a test case since it is known to be terminated with about half the molecules in precisely this situation, i.e. with a free O-H basically pointing out from the surface. Indeed, surface sensitive measurements on these species provide a spectrum very similar to that of the liquid, while the pure bulk contribution is well represented by the computed ice spectrum (a) in Figure 6 [18]. Thus we must conclude that the liquid is much less coordinated than hitherto believed with only 2 strong H-bonds per molecule (one accepting and one donating), which is indicative of chain- or ring-like structures dominating the liquid. This is not reproduced by MD simulations even of the Car-Parrinello ab initio type and this becomes a severe challenge to theory [18].

We have implemented XAS and XES spectroscopies within the CPMD formalism [19], which allows us to follow the development of the spectra during the dynamics in the liquid as well as the time-evolution of the core-excited state; this is valid as a description of the development of specific situations. However, in order to exploit this development fully to compute average spectra that reflect the experimental distribution of coordinations in the liquid, we need to improve the description of the interaction in the CPMD formalism. The forces are computed within the DFT approximation, but without including van der Waals interactions. We have run large simulations where the forces include empirical estimates of van der Waals interactions, but it is clear that an improved description is required. Such developments are performed, e.g. in the group of Lundqvist in Gothenburg. So far the formalism is restricted to specific geometrical cases but may become of significance also for the description of the liquid in a more general formulation.

The case of the first contact-layer of water on platinum discussed below provides an example where this formalism would be applicable [2]. It also demonstrates that DFT needs to be extended for weakly interacting systems; the geometry optimization using DaCapo leads to a too large distance and, similar to the case of *n*-octane above, computed spectra must be used in order to find the correct distance.

H_2O/Pt(111): flat ice

How water binds to metals is one of the important questions in surface chemistry. Apart from the basic science issues it is also a question of major economic importance: corrosion of metals is a multi-billion dollar yearly loss to society. A major ingredient is the presence of water on these surfaces. In view of this it is surprising that the structure of water in contact with a metal had not already been reliably determined.

The experimental XPS (X-ray Photoelectron Spectroscopy) data using surface and bulk core-level shifts at the Pt 4f level showed that all water molecules were in contact with the metal atoms in the surface [2]. The contact-layer must thus be rather flat and not at all the normal ice structure with tetrahedral angles. The task then becomes to determine the orientation of this basically two-dimensional ice layer on the surface: hydrogens up or hydrogens down?

For a two-dimensional hexagonal ice-layer each water molecule participates in three hydrogen bonds within the layer: either two donating and one accepting or two accepting and one donating hydrogen bond. In the latter case one O-H is free to either point out towards vacuum (hydrogens up) or to bond to the metal (hydrogens down).

The existing model of the ice layer in terms of a buckled bi-layer of bulk ice structure was based on the belief that the O-H interaction with the metal would be repulsive. The experimental XAS spectrum did not show any indication of free O-H groups, i.e. the characteristic features of the broken donating H-bonds observed in the liquid in Figure 6d are absent [2]. A theoretical confirmation of the structure through optimization and spectrum calculations for the two alternative structures is needed for firm conclusions. However, this turned out to be difficult due to the weak bonding to the surface.

Optimizing the structure using gradient-corrected DFT calculations for periodically repeated overlayers as well as for cluster models of the adsorbate layer gave very similar binding energies; the difference is within the error bars of the calculations and cannot be trusted [2]. Computing the core-level shifts for the two types (in terms of type of in-layer H-bonding) of oxygens showed a much larger difference than observed experimentally. Furthermore, the peak at 532 eV from the interaction with the Pt 5d band had very low intensity in the computed XAS spectrum and an incorrect energy position. Both the XAS spectrum and the XPS shifts are, however, sensitive to the distance of the layer from the surface.

The 532 eV peak is due to charge-donation from the oxygen lone-pair into the Pt 5d band and is thus dependent on the overlap between these orbitals [2]. Moving the adsorbate layer stepwise closer to the metal surface we find an increased intensity in this peak and an improved energy position. Similarly, the core-level shifts for the oxygens in the two different bonding situations show different behavior: one increasing and the other decreasing when the distance becomes shorter. In a region around 2.3-2.4 Å they are within the error bars of the experiment that shows only one peak. For the same distance the XAS spectrum is also in agreement with experiment, but only for the hydrogen-down layer, which thus must be concluded to be the correct structure.

The exciting and surprising final outcome of this investigation is thus that the layer of water in contact with the metal is rather flat. Half of the molecules in the layer bind through the oxygen as usually assumed, but the other half bind through an unexpected attractive interaction with the hydrogen to the metal [2]. That a hydrogen can be both part of the water molecule and still bind strongly to a metal is very important for understanding the dissociation of water, which is the outstanding issue in corrosion and electro-chemistry as well as for the splitting of water to make hydrogen, a source for future clean energy.

References

1. Myneni, S.; Luo, Y.; Näslund, L. Å.; Cavalleri, M.; Ojamäe, L.; Ogasawara, H.; Pelmenschikov, A.; Wernet, P.; Väterlein, P.; Heske, C.; Hussain, Z.; Pettersson, L. G. M.; Nilsson, A. *J. Phys: Condens. Mat.* **2002**, *14*, L213; Editor's Choice, *Science*, **295**, 1975 (2002)
2. Ogasawara, H.; Brena, B.; Nordlund, D.; Nyberg, M.; Pelmenschikov, A.; Pettersson, L. G. M.; Nilsson, A. *Phys. Rev. Letters* **2002**, *89*, 276102. *Phys. Rev. Focus,* December 31, 2002; Editor's Choice, *Science* **299**, 475 (2003); Highlight *Angewandte Chemie* **42**, 3458-3460 (2003).
3. Öström, H.; Triguero, L.; Weiss, K.; Ogasawara, H.; Garnier, M. G.; Nordlund, D.; Nyberg, M.; Pettersson, L. G. M.; Nilsson, A. *J. Chem. Phys.* **2003**, *118*, 3782.
4. Weiss, K.; Öström, H.; Triguero, L.; Ogasawara, H.; Garnier, M. G.; Pettersson, L. G. M.; Nilsson, A. *J. El. Spec. Rel. Phen.* **2003**, *128*, 179.
5. Barlow, S.; Kitching, K.; Haq, S.; Richardson, N. *Surf. Science* **1998**, *401*, 322.
6. Nyberg, M.; Hasselström, J.; Karis, O.; Wassdahl, N.; Weinelt, M.; Nilsson, A.; Pettersson, L. G. M. *J. Chem. Phys*. **2000**, *112*, 5420.
7. Booth, N. A.; Woodruff, D. P.; Schaff, O.; Giessel, T.; Lindsay, R.; Baumgärtel, P.; Bradshaw, A. M. *Surf. Science* **1998**, *397*, 258.
8. The calculations have been performed with the code CPMD version 3.6, written by Jurg Hutter et al., Max-Planck-Institut fur Festkörperforschung, Stuttgart, IBM Research Laboratory Zurich., 1995-2000.
9. Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
10. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785.
11. Monkhorst, H. J.; Pack, J. D. *Phys. Rev. B.* **1976**, *13*, 5188.
12. deMon-KS StoBe version 1.0, Principal authors: M.E. Casida, C. Daul, A. Goursot, K. Hermann, A. Koester, L.G.M. Pettersson, E. Proynov, A. St-Amant,

and D.R. Salahub. Contributing authors: V. Carravetta, H. Duarte, N. Godbout, J. Guan, C. Jamorski, M. Leboeuf, V. Malkin, O. Malkina, M. Nyberg, L. Pedocchi, F. Sim, L. Triguero, and A. Vela, deMon Software, 2001.
13. Vanderbilt, D. *Phys. Rev. B* **1990***, 41*, 7892.
14. Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244.
15. Nyberg, M.; Odelius, M.; Nilsson, A.; Pettersson, L. G. M. *J. Chem. Phys. 2003, 119, 12577.*
16. Bergmann, U.; Wernet, P.; Glatzel, P.; Cavalleri, M.; Pettersson, L. G. M.; Odelius, M.; Lin, P.-C.; Nilsson, A.; Cramer, S. P. *Phys. Rev. B* **2002**, *66*, 092107.
17. Cavalleri, M.; Ogasawara, H.; Pettersson, L. G. M.; Nilsson, A. *Chem. Phys. Letters* **2002**, *364*, 363.
18. Wernet, P.; Nordlund, D.; Bergmann, U.; Ogasawara, H.; Cavalleri, M.; Näslund, L. Å.; Hirsch, T. K.; Ojamäe, L.; Glatzel, P.; Odelius, M.; Pettersson, L. G. M.; Nilsson, A.; Science 2004, 305, 995.
19. Cavalleri, M.; Odelius, M.; Nilsson, A.; Pettersson, L. G. M.; J. Chem. Phys., in press.

Monte Carlo Simulations of Solutions Containing Charged Macromolecules

Per Linse and Marie Skepö
Division of Physical Chemistry 1, Department of Chemistry and Chemical Engineering, Lund University

In this report, we will briefly account for our recent advances on the attraction between like-charged colloids and initial studies of the complexation between charged colloids and opposite charged polyelectrolytes performed 2000 with the use of resources at PDC.

Introduction

Aqueous solutions of charged macromolecules, such as charged colloids, polyelectrolytes, and mixtures of charged colloids and polyelectrolytes have gained large scientific attention in the last decades for both fundamental and practical reasons [1-3]. Nearly all industrial and biological processes involve solutions of charged macromolecules, e.g., drug delivery systems, detergents, shampoos, cosmetics, papers, and paints, but also many diseases are associated with malfunctions at the colloidal level. The physico-chemical properties of those systems are dominated by electrostatic forces, which are still not fully understood and need to be further characterized.

Charged colloids are spherical or nearly spherical particles with radius ranging from c. 10^{-9} to 10^{-6} m. Examples of colloids include globular proteins, aggregates formed by self-associating surfactants, microemulsions, and silica and latex particles. In aqueous solution, colloids acquire surfaces charges and the nearby solution contains counterions, making the system electroneutral. Normally, two like-charged colloids repel each other. The Derjaguin–Landau–Verwey–Overbeek (DLVO) theory [4, 5], a cornerstone of colloid science, predicts such a purely repulsive electrostatic force among like-charged colloids *under all conditions*. However, recent experimental findings [6] appear to challenge this theory, in particular when multivalent counterions are present.

Polyelectrolytes are charged linear molecules that display a high solubility in water and strong adsorbing capacity to surfaces bearing an opposite charge. An interesting feature is that they can act both as a stabilizing and as a destabilizing agent in particle suspensions [7]. The complexation of polyelectrolytes with oppositely charged macroions is therefore an essential component in many technical and biological systems. Examples include synthetic polyelectrolytes adsorbing on various charge colloids as well as the packaging of DNA in chromosomes through histone-DNA complexes. More recent simulation studies have been made by Wallin [8] and theoretical investigations by Muthukumar with coworkers [9, 10] and Netz with coworkers [11, 12].

Model

Theoretical studies of highly charged colloids and polyelectrolytes possess several challenging aspects. Two of them are the long-range Coulomb interaction and the existence of spatial correlations among the charged species on different length scales. Therefore, a simple model needs to be selected, but the model should still possess relevant components for describing salient features of the systems of interest.

In our investigations, the so-called primitive model has been adopted. In this model the charged colloids (latter referred to as macroions) as well as the small ions (latter referred to as counterions) are explicitly represented by charged hard spheres interacting through a Coulomb potential, whereas the solvent is treated as a dielectric medium. Similarly, polyelectrolytes are represented by chains of charged segments connected by harmonic forces and the stiffness of the chains is regulated by angular force terms. See [13-15] for further details.

Method

The static properties of the model systems of interest were obtained by performing Monte Carlo (MC) simulations using the in-house software MOLSIM. In the investigation of the macroion solution, two special techniques were necessary. First, the Ewald summation was applied to handle the long-range interactions among the charged species [16]. Second, due to a large accumulation of the counterions near the macroions, normal MC simulation with single particle displacements becomes hopelessly ineffective. This is of course a real physical effect, but it still constitutes a major problem of performing such simulations. A few years ago, we introduced the cluster move technique into this area [17], which became an extremely powerful remedy for this difficulty.

Structure in colloidal solutions

In 1999, it was unambiguously demonstrated that a suspension of charged colloids and counterions described within the primitive model could display an effective attraction between macroions leading to a phase instability of the system [18]. This observation led us to make a more extensive investigation of the phase behavior of such systems. For a compact description, it can be shown that such systems without added salt can accurately be described using only three reduced parameters [13]; our choice is the macroion-to-counterion charge ratio, Z_r, the macroion volume fraction, ϕ_M, and a coupling parameter relating the electrostatic interaction between two counterions with the thermal energy, Γ_{II}.

A large part of the parameter space experimental accessible has been investigated by extensive MC simulations [13]. Figure 1 shows the investigated area in the (Z_r, Γ_{II})-plane and the location of different real systems. The directions of increase in (i) counterion charge Z_I, (ii) macroion charge Z_M, and (iii) the combined parameter $(R_M \varepsilon_r T)^{-1}$, where R_M denotes the macroion radius, ε_r the permittivity of the solvent, and T the temperature, are also shown in the figure. Snapshots of three selected systems are also provided in Figure 1.

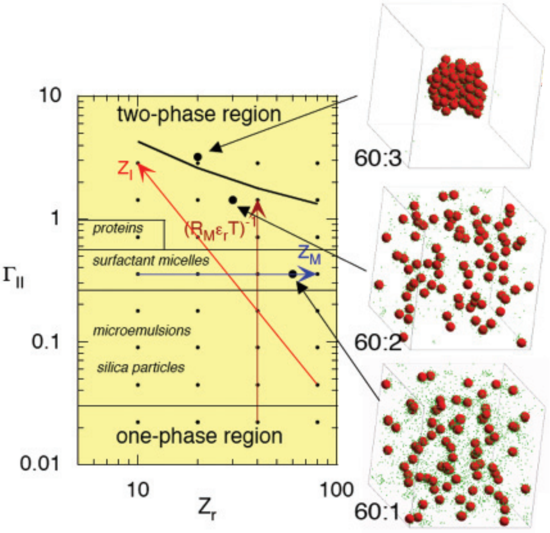


Figure 1
Simulated state-points (dots) in the (Z_I, Γ_{II})-plane for macroion volume fractions ϕ_M = 0.00125 - 0.08 and the locations of solutions of various charged colloids with monovalent counterions in aqueous solution at room temperature. The directions of increase in (i) counterion charge Z_I, (ii) macroion charge Z_M, and (iii) the combined parameter $(R_M \varepsilon_r T)^{-1}$ are shown by arrows. The border between the one-phase region (bottom) and a two-phase region (top part) appearing at $\Gamma_{II} \approx$ 1-5 is indicated (thick solid curve). The locations in the (Z_I, Γ_{II})-plane of the systems characterized by (Z_M, Z_I) = (-60, +1), (-60, +2), and (-60, +3) are given (filled circles) and snapshots of these systems at ϕ_M = 0.016 are also shown with the particles are drawn to scale. Periodic boundary conditions are applied in all three directions to avoid surface artefacts. Data taken from [13, 18].

At small electrostatic coupling ($\Gamma_{II} < 0.1$), the system is characterized by a weak electrostatic interaction and the macroions can come close to each other. Such solutions are stable and a one-phase system appears. At intermediate electrostatic coupling ($0.1 < \Gamma_{II} < 1$), the repulsion among the macroions becomes appreciable. The snapshot labeled “60:1” illustrates the fact that the macroions (red spheres) are well separated. Moreover, the counterions (green dots) are fairly well spread throughout the full solution. Again the solution is thermodynamically stable. At larger electrostatic coupling ($\Gamma_{II} \approx 1$), achieved, e.g. by replacing the monovalent counterions by divalent ones, significant structural changes appear. The snapshot labeled “60:2” shows that most counterions are accumulated near the macroions and that macroions occasionally are near each other forming pairs. The system is just barely thermodynamically stable. At an even larger electrostatic coupling ($\Gamma_{II} > 2$), e.g. realized by having trivalent counterions, the outcome of the simulations is that all macroions and most counterions form a single aggregate, see snapshot labeled “60:3”. Obviously, the system is now thermodynamically unstable and two coexisting phases occur, one rich and one poor in the charged species. At this condition the conventional picture of a charged stabilized colloidal suspension, according to the DLVO theory, does not hold. Briefly, the missing component in the

DLVO theory is a short-ranged attractive force originating form spatial correlations among counterions localized near different macroions [13, 19]. At intermediate electrostatic coupling, the electrostatic mean-field repulsion as described by the DLVO theory is sufficiently long-ranged making the correlation attraction not effective, whereas at large electrostatic coupling the mean-field repulsion is more short-ranged than the correlation attraction making the latter the dominating force. For more details on the extensive mapping of the parameter space including structural and thermodynamic results, see [13].

The difference in the effective interactions between two like-charged macroions for the 60:1 and 60:2 systems are further illustrated in Figure 2 displaying the effective potential acting between the two macroions [20]. For the 60:1 system, only a long-ranged repulsive potential appears, whereas for the 60:2 system the long-ranged repulsion is strongly reduced in extension and magnitude and, moreover, a local potential energy minimum appears at r ≈ 45 Å corresponding to two macroions separated with one layer of counterions. These results are consistent with Figure 1 and the associated discussion.

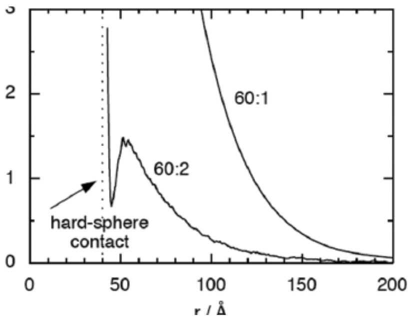


Figure 2
Effective macroion-macroion potential vs. the macroion separation for the 60:1 and 60:2 systems obtained from an inverse Monte Carlo simulation procedure at macroion volume fraction $\phi_M = 0.0168$. The macroion radius R_M is 20 Å. Data taken from [20].

In summary, the presented results unambiguously demonstrate that a solution of like-charged colloidal particles plus counterions interacting solely through repulsive hard-core forces and long-range Coulomb forces (the so-called primitive model) displays conditions where two phases of different electrolyte density are in equilibrium with each other, in contradiction to the celebrated DLVO theory, but consistent with experimental observations [6].

Polyelectrolyte-macroion complexation

The complexation between polyelectrolytes and oppositely charged macroions and the phase behavior of such solutions are very rich owing to the large number of physical parameters characterizing these systems. Important parameters include (i) the macroion size and charge, (ii) the polyelectrolyte length, linear charge density, and stiffness, (iii) the stoichiometric ratio of the two types of macroions, and (iv) the amount and nature of added salt.

In the initial part of the thesis by Marie Skepö (former Jonsson) [21], the polyelectrolyte-macroion complexation was studied. Figure 3 displays typical snapshots of binary complexes involving polyelectrolyte of different flexibility with the macroion-to-polyelectrolyte charge ratio $\beta = 1/4$ (panel (a) and (b)) and for 1:4 complexes with charge ratio $\beta = 1$, i.e. equivalent amount of polyelectrolyte and macroion charges (panel (c) and (d)).

When one macroion is present in the solution both the flexible and the stiff chain complex the macroion. The flexible chain wraps loosely around the macroion with one or two tails stretching away from the macroion surface (Figure 3a), while the stiffer chain remains rod-like even when it has complexed the macroion, and only a few segments are in contact with the macroion surface (Figure 3b). With four macroions being present, all macroions are still complexed independently of the chain flexibility. When the chain is flexible, the macroions form a compact structure (Figure 3c), which opens up as the chain becomes stiffer, and in the presence of the stiff polyelectrolyte chain the macroions are nearly linearly arranged (Figure 3d). We also notice that the separation between the macroions increases with increasing chain stiffness.

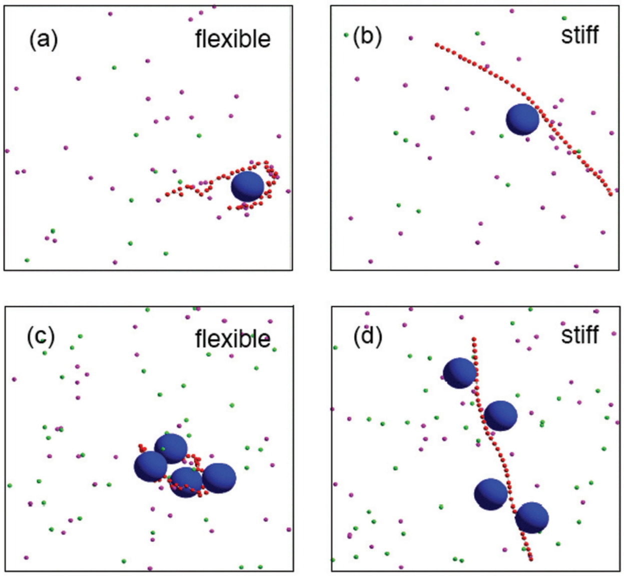


Figure 3
Typical configurations of systems containing one polyelectrolyte with 40 segments (connected red dots) and one or four macroions (blue spheres) and their counterions (unconnected dots). In (a) and (c), the polyelectrolyte is flexible and in (b) and (d) stiff. The box length is 258 Å, and particles are drawn to scale. Periodic boundary conditions are applied in all three directions to avoid surface artefacts. Data from [15].

The maximum amount of macroions able to be complexed to a single polyelectrolyte is of large interest. This limit is governed by an intricate balance between electrostatic and chain bending forces in combination with mixing entropy. In Figure 4 the complex charge ratio β^c is given as a function of the solution charge ratio β. At β up to one, there is a quantitative binding; whereas at larger β the complex charge ratio levels off at c. 1.5, implying that at saturation the absolute charge of the complexed macroions exceeds that of the polyelectrolyte by c. 50%. The net charge of the complex is opposite of the polyelectrolyte, and hence a charge reversal has appeared. Interestingly and counter-intuitively, the charge reversal is strongest for the stiff polyelectrolyte. That observation was confirmed by free energy calculation of successive macroion binding and is now understood as an effect of chain entropy [15].

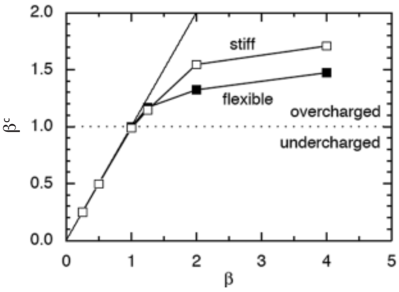


Figure 4
Complex charge ratio β^c vs. the stochiometric charge ratio β for flexible and stiff polyelectrolytes. Quantitative complexation (thin solid line) and border between undercharged and overcharged complexes (dotted line) are also indicated. Data from [15].

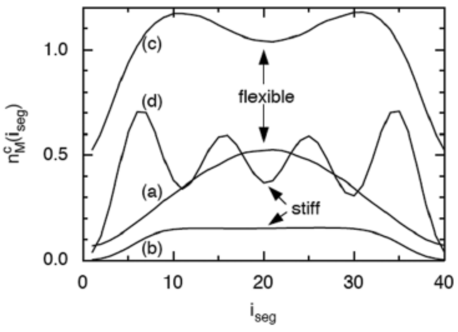


Figure 5
Complexation function n_M^c (i_{seg}), denoting the total number of macroions complexed to a segment, vs. the position i_{seg} of that segment along the chain. The conditions for the curves labelled (a)-(d) are the same as for panels (a)-(d) in Figure 3. Data from [15].

The structure of the polyelectrolyte-macroion complexes has been quantified by various methods [14, 15]. In particular, the introduction of the complexation function n_M^c (i_{seg}) denoting the total number of macroions complexed to the different polyelectrolyte segments have been very useful. Figure 5 shows this function for the four cases discussed in Figure 3. As seen from Figure 5, a single macroion is preferentially complexed to the center of the flexible chain (single maximum of curve (a)) with two significant tails, whereas the binding region of the stiff chain is much more extended (extended flat region of curve (b)). In the case of four macroions complexed to the flexible polyelectrolyte, the macroions are grouped two by two (two maxima appearing for curve (c)). Finally, with four macroions complexed to the stiff chain, the sequential binding along the extended chain gives rise to four clear maxima in the complexation function (curve labeled (d)). In subsequent simulation studies, we have examined the phase stability of such systems and, in particular, the appearance of redissolution have been considered in great detail [21].

The large interest in this area is manifested by e.g. related simulation studies by Chodanowsaki and Stoll [22, 23] and by Akinchina and Linse [24, 25].

Use of PDC facilities

In our work on the stability of solutions of charged colloids, started around 1997, we have greatly benefited from using Strindberg at PDC for our parallel MD and extensive MC simulations. At that time Strindberg provided a vital extension of our local computer capacity. However, in more recent time, much of our computation has been transferred to clusters of workstations and today to PC clusters.

Further outlook

There is presently an intense experimental, theoretical, and simulation activity in the area of solutions containing charged macromolecules. The development is driven by the enhanced capability of tailor such systems for specific applications, as for example in formulation of new drug delivery systems, and by new experimental techniques facilitating experiments on single molecules. In parallel, new theories are developed and the need for more extensive numerical simulation is increasing to understand experimental outcome and to assess the validity of various theories.

References

1. Russel, W.B., D.A. Saville, and W.R. Schowalter, *Colloidal dispersions*. 1989, New York: Cambridge University Press.

2. Arora, A.K. and B.V.R. Tata, eds. *Ordering and phase transitions in charged colloids*. Complex fluids and fluid microstructures, ed. R. Rajagopalan. 1996, VCH Publisher, Inc.: New York.

3. Holm, C., P. Kékicheff, and R. Podgornik, eds. *Electrostatic effects in soft matter and biophysics*. Vol. 56. 2001, Kluwer Academic Press: London.

4. Derjaguin, B.V. and L. Landau, *Theory of the stability of strongly charged lyophobic sols and the adhesion of strongly charged particles in solutions of electrolytes.* Acta Physiochim. (URSS), 1941. **14**: p. 633-662.

5. Verwey, E.J. and J.T.G. Overbeek, *Theory of the stability of lyophobic colloids*. 1948, Amserdam: Elsevier.

6. Jönsson, B. and H. Wennerström, *When ion-ion correlations are important in charged colloidal systems*, in *Electrostatic effects in soft matter and biophysics*, C. Holm, P. Kékicheff, and R. Podgornik, Editors. 2001, Kluwer Academic Press: London. p. 171-202.

7. Napper, D.H., *Polymeric stabilization of colloidal dispersions*. Colloidal Science, ed. R.H. Ottewill and R.L. Rowell. 1983, London: Adademic Press.

8. Wallin, T., *Polyelectrolyte-surfactant interaction in dilute suspensions. A study based on Monte Carlo simulations and mean-field calculations*, in *Physical Chemistry 1, Center for Chemistry and Chemical Engineering*. 1998, Lund University: Lund.

9. von Goeler, F. and M. Muthukumar, *Adsorption of polyelectrolytes onto curved surfaces.* J. Chem. Phys., 1994. **100**: p. 7796-7803.

10. Kong, C.Y. and M. Muthukumar, *Monte Carlo simulations of adsorption of a polyelectrolyte onto charged surfaces.* J. Chem. Phys., 1998. **109**: p. 1522-1527.

11. Netz, R.R. and J.-F. Joanny, *Complexation between a Semifleible Polyelectrolyte and an Oppositely Charged Sphere.* Macromolecules, 1999. **32**: p. 9026-9040.

12. Kunze, K.-K. and R.R. Netz, *Salt-Induced DNA-Histone Complexation.* Phys. Rev. Lett., 2000. **85**: p. 4389-4392.

13. Linse, P., *Structure, phase stability, and thermodynamics in charged colloidal solutions.* J. Chem. Phys., 2000. **113**: p. 4359-4373.

14. Jonsson, M. and P. Linse, *Polyelectrolyte-macroion complexation. I. Effect of linear charge density, chain length, and macroion charge.* J. Chem. Phys., 2001. **115**: p. 3406-3418.

15. Jonsson, M. and P. Linse, *Polyelectrolyte-macroion complexation. II. Effect of chain flexibility.* J. Chem. Phys., 2001. **115**: p. 10975-10985.

16. Allen, M.P. and D.J. Tildelsley, *Computer simulation of liquids*. 1987, New York: Oxford.

17. Linse, P. and V. Lobaskin, *Monte Carlo simulations of solutions of like-charged colloidal particles*, in *Simulation and Visualization on the Grid*, B. Engquist, et al., Editors. 2000, Springer: Berlin. p. 165-180.
18. Linse, P. and V. Lobaskin, *Electrostatic attraction and phase separation in solutions of like-charged colloidal particles.* Phys. Rev. Lett., 1999. **83**: p. 4208-4211.
19. Linse, P., *Mean force between like-charged macroions at high electrostatic coupling.* J. Phys. Condens. Matter, 2002. **14**: p. 13449-13467.
20. Lobaskin, V., A. Lyubartsev, and P. Linse, *Effective macroin-macroion potentials in asymmetric electrolytes.* Phys. Rev. E, 2001. **63**: p. 020401.
21. Skepö, M., *Structure and phase stability of polyelectrolyte-macroion solutions*, in *Physical Chemistry 1, Center for Chemistry and Chemical Engineering*. 2002, Lund Univesity: Lund.
22. Chodanowsaki, P. and S. Stoll, *Polyelectrolyte adsorption on charged particles in the Debye-Hückel approximation. A Monte Carlo study.* Macromolecules, 2001. **34**: p. 2320-2328.
23. Chodanowsaki, P. and S. Stoll, *Polyelectrolyte adsorption on charged particles: Ionic concentration and particle size effects. A Monte Carlo approach.* J. Chem. Phys., 2001. **115**: p. 4951-4960.
24. Akinchina, A. and P. Linse, *Monte Carlo simulations of polyion-macroion complexes. 1. Equal absolute polyion and macroion caharges.* Macromolecules, 2002. **35**: p. 5183-5193.
25. Akinchina, A. and P. Linse, *Monte Carlo simulations of polyion-macroion complexes. 2. Polyion length and charge density dependence.* Macromolecules, 2003. **107**: p. 8011-8021.

Quantum Modeling of Molecular Materials

Pawel Salek and Hans Ågren
Laboratory of Theoretical Chemistry, The Royal Institute of Technology (KTH), Stockholm

A general overview is given of past, present and some future scheduled work within the DALTON program system for molecular and materials modeling with special emphasis on the utilization of the PDC center. Sample applications of a linear scaling modeling routine are briefly described.

Introduction

Materials modeling – that is, mathematical simulations of materials in order to describe their properties, spectra, reactions, and interactions – has become an increasingly viable approach in computational based research, covering a wide range of phenomena of technical as well as fundamental importance. Adhering to the general philosophy of such modeling, the ultimate goal of the DALTON network and computer program {1} is to model and understand materials from a rigorous quantum perspective. This understanding is to be accomplished by transcending the traditional boundaries of rigorous electronic-structure simulations, from the scale where each atom makes a difference to the scale where the bulk properties can be derived. We are in particular concerned with quantum modeling of magnetic and electric properties, of the distribution of nuclei and the electron density, and of any type of force, reaction, or spectrum that can be associated with molecules and materials. A goal is to provide theoretical support to experimental activities such as synthesis and materials characterization. In this review we highlight some aspects of this project with special emphasis of computing at the PDC center.

Background of the DALTON project

The current status of quantum chemistry is such that the properties of molecular systems can be predicted with a high degree of accuracy and reliability when the systems are of small to medium size. Thus, the user of quantum-chemical programs is not restricted to calculating the total energy at a given molecular configuration. Rather, she or he can make predictions of molecular properties – for example, of molecular structure, reaction enthalpies, atomization energies, vibrational energy levels, electronic spectra, dynamic polarizabilities and hyperpolarizabilities, magnetizabilities, NMR and EPR parameters – at different levels of theory, depending on the size of the system at hand and the accuracy required. The ability to carry out such predictions, mostly in a fully automated manner, is an essential feature of modern molecular modeling. In recent years, one of the most important developments in rigorous quantum chemistry (including ab initio wave-function theory and density functional theory (DFT)) has concerned calculations on large systems. However, although methods and programs have already been developed for the optimization of the energy of large systems, the field is still very much in its infancy.

The adaption of second quantization enabled the original authors of the DALTON program to develop a uniform scheme for calculations of properties for a large variety of wave functions, expressed in a compact manner in terms of the algebra of a few elementary operators. In particular, an important analytical transferability between the wave function and its associated properties was established – namely, once the optimal values for the parameters characterizing the wave function have been generated, all its properties follow analytically and automatically, without the need for further approximations or interference by the user. Furthermore, a common formalism – and indeed, to a large extent, also a common code – was developed for treating all kinds of molecular perturbations and their associated properties. These aspects, which were the cornerstones of our visions at the start of the DALTON project twenty years ago, have now been almost fully exploited at the molecular level.

More recent work within DALTON has emphasized the linear-scaling implementation of the Coulomb problem, making the methodology applicable for an order of magnitude larger systems. Furthermore, other recent accomplishments are the completion of the fourth-order toolbox, to integrate fully the technology of DFT, and to take large steps in relativistic implementations. Now, the challenge of electronic-structure theory is to move fully into the nano regime. From having constituted a core of a few Scandinavian groups, the DALTON project now involves many contributors throughout Europe. The efforts have materialized in the DALTON program system, with some 30 authors [1]. DALTON is fully documented and released to the public for research. Collecting more than 1000 user licenses all over the world, DALTON forms also a main vehicle for training of students. This software package thus represents the state-of-the-art for the modeling of molecules and materials.

DALTON research method

The approach behind DALTON is twofold: on the one hand, to develop hierarchical methods that can be converged in a controlled manner, making possible meaningful comparisons with and interpretations of experimental data; on the other hand, to develop less expensive methods, directly applicable to very large systems. The two approaches are not mutually exclusive; rather, both must be pursued to achieve the state goal of *a priori* nano quantum chemistry. The computational methods are subdivided in three subclasses as described in the following.

Wave-function and density-functional methods

The first step in any purely theoretical investigation of molecular systems is to generate some representation of the molecular electronic state. This can either be achieved directly, through the calculation of some approximate wave function, or indirectly, through the calculation of an approximate electron density function, invoking the techniques of density-functional theory. We take the approach that both techniques are needed for an efficient treatment of molecules and molecular materials. Wave-function methods have the advantage of generality and rigor, having the ability to treat any electronic state and any electronic structure in an unbiased and systematic manner, allowing us to approach the exact solution in a hierarchical, rigorous way. Density-functional methods, by contrast, are much more widely applicable, allowing large molecular systems to be treated in a manner that is less expensive but also less rigorous than wave-functional theory.

Analytic response methods

Analytic response theory, which represents a particular formulation of time-dependent perturbation theory in terms of response functions, constitutes a core technology in the DALTON project. Response functions are important in that they provide a uniform, universal representation of the response of a system to external perturbations, applicable to all computational models (density-functional as well as wave-function models) and to all perturbations (dynamic as well as static perturbations). The theory is analytical in the sense that the properties are calculated from analytically derived expressions at the frequency of the experiment, avoiding the unstable and less expensive finite-difference procedures. It is also model transferable in that, once the computational (wave-function or density-functional) model has been defined, all properties are obtained on an equal footing, without further approximations. A main advantage of response theory is that the traditional method of summing over excited states is replaced by the solution of linear equations, without explicit knowledge of the excited states. This enables us to obtain the property of interest in a size-extensive manner, provided the underlying computational model is size-extensive. Over the years, response theory has been extended and diversified in the DALTON framework to encompass a wide variety of properties and electronic-structure models. In particular, it has been implemented in a gauge-invariant property toolbox, applicable to fourth order for all the main reference-state technologies mentioned above – that is, for multiconfigurational self-consistent field, coupled-cluster, and density-functional theories.

Analytic gradient methods

In chemistry, the calculation of molecular gradients and Hessians – that is, the forces and force constants in the Born-Oppenheimer approximation – is of primary importance. An efficient evaluation of these properties – in particular, of the molecular gradient – is essential for the efficient exploration of potential-energy surfaces, the identification of molecular equilibrium structures, reaction transition states, chemical pathways, and so on. In principle, molecular gradients and Hessians can be evaluated using the general methods of response theory. However, since the atomic orbitals used in molecular calculations are perturbation-dependent in that they are fixed to the nuclei and thus move with changes in the molecular geometry, the calculation of gradients and Hessians is much more difficult than the calculation of other properties, where the atomic orbitals are unaffected by the perturbation. In particular, this means that the evaluation of molecular gradients requires us to evaluate also contributions from the two-electron interactions (two-electron integrals), greatly increasing the computational efforts.

Review of current modeling area

Nanoscale quantum modeling. As mentioned in the introduction of this review an essential objective of current developments is to transfer the scale of quantum modeling from the atomic domain into the nano domain. A primary target of DALTON work is therefore to address linear-scaling techniques (i.e. techniques with a cost proportional to the system size) and to extend their applicability range into the nano regime, beyond 10000 atoms.

Large-scale developments of high-level correlation methods. Due to its computational simplicity, the main thrust of the development of quantum methods towards the nanoscale regime takes place within the realm of density functional theory (DFT). Nevertheless, since modern DFT is somewhat semi-empirical in nature, it is essential that a similar development is also undertaken for the hierarchical ab initio methods in order to benchmark and test DFT.

Multi-configurational relativistic DFT. Apart from extending the scale and scope of the applicability range of current electronic-structure techniques into the nano domain, it remains important to study and develop the basic aspects of electronic-structure theory. Two important goals are to treat general spin states in DFT and to develop a general correlation approach that combines the dynamical correlation recovered by DFT with the near-degeneracy correlation recovered by multi-configurational reference expansions.

Modeling of characterizing technologies. In addition to the interaction with external electromagnetic fields, the interaction of the electrons with internal molecular fields such as those generated by spinning nuclei and electrons, although exceedingly weak, provides numerous possibilities for diagnostics of materials by spectroscopy. Within the DALTON project we have therefore involved ourselves a lot in development on modeling of various magnetic resonance processes – NMR, EPR, ODMR, as well as of many properties in the IR, optical and X-ray regions.

Molecular- and nano-electronics. A long-standing goal of molecular electronics has been to replace semiconductors by individual molecules so that electronic devices can be produced orders of magnitude smaller than their current semiconductor counterparts. The goal of our studies in this field is to gain more of the necessary understanding of the physical and chemical processes involved, to develop efficient computational approaches to simulate the electron transportation in molecular devices, and ultimately to use these techniques to improve device performance.

Molecular- and nano-photonics. To fully understand the functionality of various photonics devices such as conductors, wave-guides, switches and displays, it is essential to master and simulate the basic photon-matter interaction. An immediate task of the DALTON network is to merge the quantum-mechanical methodology for electromagnetic properties with the dynamical theory of pulse propagation into a versatile toolbox for simulations of optical properties of materials in general.

Computational aspects – usage of PDC resources

We use the computing resources for two primary purposes: production calculations [4] and code development. While there are several projects being developed, the probably most challenging one is the work on the low-scaling density functional routines. The theoretical side of this task has been under active investigation during last few years but practical implementations are still very rare.

A method that has inherently rather low scaling of N^2 to N^4 is density functional theory (where the exponent depends on actual implementation, used basis set etc). Its Kohn-Sham formulation leads to equations very similar to Hartree-Fock and these in turn have very long tradition in quantum chemistry, with years of research spent on optimizing for performance and many available algorithms and implementations.

The self-consistent Kohn-Sham algorithm has two expensive elements:

1. Evaluation of the Kohn-Sham (KS) matrix. This is the dominant part, scaling like N^2 and with large prefactor because it includes calculation of many two-electron, six-dimensional integrals.
2. Determination of molecular orbitals or alternatively a density matrix corresponding to the given KS matrix. This is usually done by diagonalization of the KS matrix. This process scales like N^3 but the operations involved are rather simple and appropriate subroutines (we use DSPEV routine of LAPACK numerical algebra package) are usually optimized for the hardware and the prefactor is low – this is why this time becomes noticeable with at least several hundreds of basis functions.

We currently focus on removing the first bottleneck [3] but there is also some undergoing effort to address the theoretical foundations of the second problem as well [2].

Computation of the Coulomb matrix via multipole method

In order to construct the KS matrix in linear time, one has to be able to construct all the expensive terms in this regime. The Coulomb term and exchange and correlation (XC) terms are attacked from different angles. Computation of the Coulomb term can be seen as an evaluation of interaction between a set of M charge clouds. This process in principle scales like M^2 since all the pairs need to be taken into account. A method has been developed to express whole subsets of charges with single multipoles. In this way, interactions do not need to be computed pairwise but instead, the entire set is partitioned into groups, the potential generated by the groups is expressed via multipole expansion, and the entire interaction is computed by adding up interactions between groups. One actually constructs a hierarchy of larger and larger charge groups so that the number of subgroups in a group is roughly constant.

Evaluation of the exchange-correlation matrix with sparse matrix formalism

Computation of the XC contribution to KS matrix needs to be approached from another side. The time-consuming part in the evaluation of the XC contribution

$$F_{\mathrm{xc};pq} = \int v_{\mathrm{xc}}(\rho(\vec{r}))\psi_p(\vec{r})\psi_q(\vec{r})d\vec{r}\,,$$

where v_{xc} is exchange-correlation potential and $\psi(\vec{r})$ are basis functions, is the evaluation of the density and the computation of all matrix elements. The integration is done numerically over a set of grid points. For each grid point, the density must be computed and contributions to the matrix elements distributed over F_{xc} matrix. Both these operations can be expressed via matrix-vector operations with $\mathbf{D}$ being the density matrix and v being a vector of basis function values at the given grid point

$$\rho = \mathbf{v}^T\mathbf{D}\mathbf{v} \qquad \Delta F_{\mathrm{xc}} = v_{\mathrm{xc}}\mathbf{v}\mathbf{v}^T$$

The integration time is in principle proportional to N^3 because the number of grid points is proportional to the system size, and the N^2 matrix elements need to be evaluated for each of them. Fortunately, the density matrix $\mathbf{D}$ and vectors $\mathbf{v}$ are sparse for large systems and roughly a constant number of elements of $\mathbf{v}$ are nonvanishing.

The computation of response functions is based on similar building blocks and can be implemented analogously. The difficult part is to organize a calculation of the above multiplications in a memory-bandwidth friendly scheme, and in such a way that one obtain long innermost loops. This can be done by grouping neighbor grid points into batches and processing them in one shot organizing loops in such a way that the inner loop goes over grid points. This yields simple stride-one loops that can easily be optimized by compilers.

Benchmarks

The linear scaling is best demonstrated on simple linear systems where one can arbitrarily extend the system just by adding new building blocks. All the benchmark calculations presented in this article were carried out with STO-3G basis set and with explicit treatment of all electrons.

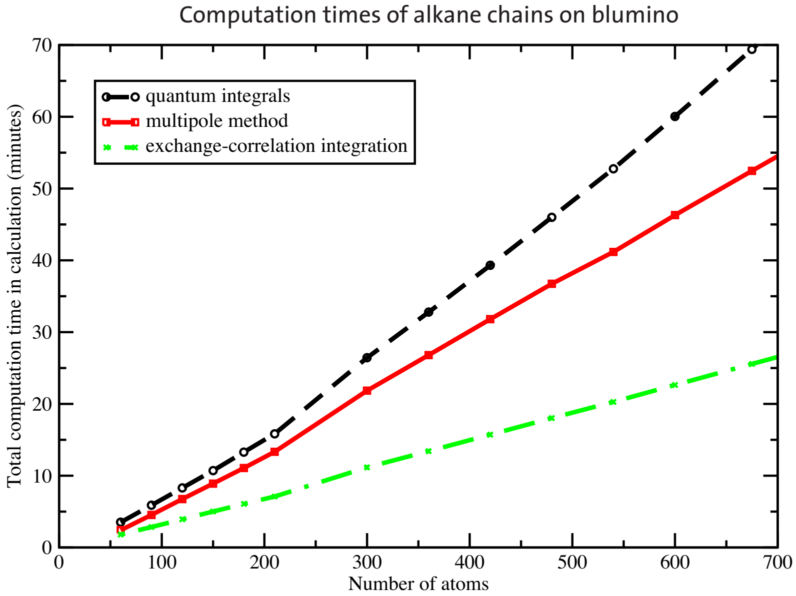
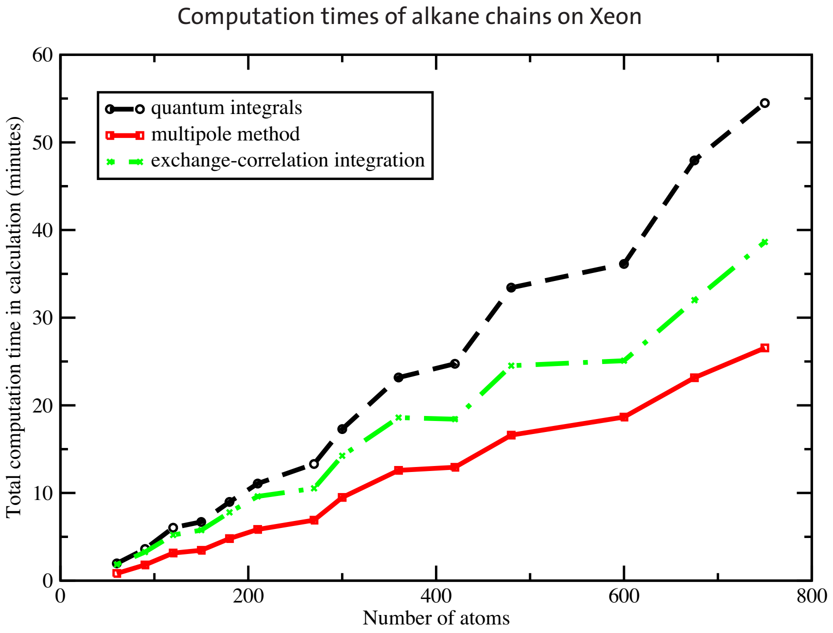


Figure 1
Total time of a self-consistent scheme for a set of alkane chains. The program was run on a single Itanium2 processor with 900MHz clock.

Figure 1 displays the total computation time for a set of alkane chains calculated on the new PDC cluster Lucidor consisting of 90 nodes, each having 2 Itanium2 processors with 900MHz clock. Only single processor on the node was used during the calculations.

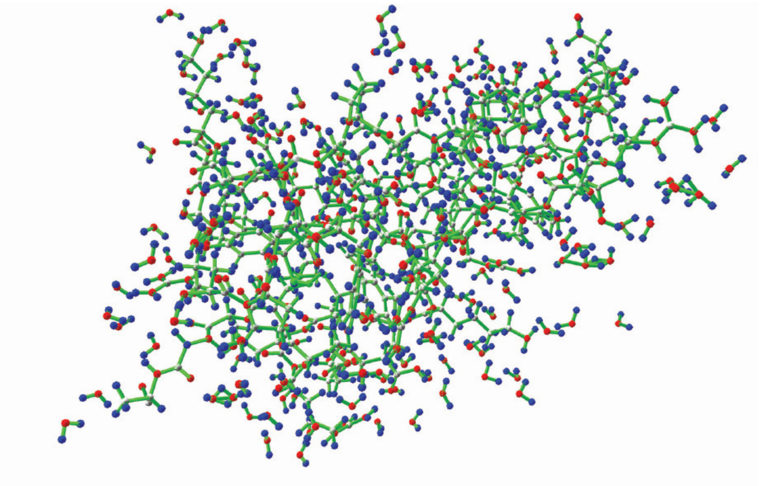
For comparison, we present also in Figure 2 timings of the same code run on a high-end PC with two Pentium Xeon 2.8GHz processors. Again, only one processor was used for calculation. It should though be stressed that such large calculations hit the 2GB memory limit per process on 32-bit machines that limits the range of possible property calculations.

Figure 2
Total time of a self-consistent scheme for a set of alkane chains. The program was run on a single Pentium Xeon processor with 2.8GHz clock.



The current version of the program can compute efficiently the Kohn-Sham matrix but for really large molecules – like the one shown on Figure 3 – other parts of the code consume substantial amount of time. Still, achieving such a milestone marks, in our view, the arrival of a new era in quantum chemistry.

Figure 3
A BPTI protein of salmon consisting of 1396 atoms. Structure obtained from X-ray crystallography experiment.



References

1. T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, J. Olsen, K. Ruud, H. Ågren, T. Andersen, K. L. Bak, V. Bakken, O. Christiansen, P. Dahle, E. K. Dalskov, T. Enevoldsen, H. Heiberg, H. Hettema, D. Jonsson, S. Kirpekar, R. Kobayashi, H. Koch, K. V. Mikkelsen, P. Norman, M. J. Packer, T. Saue, P. R. Taylor, and O. Vahtras. DALTON, *an ab initio electronic structure program, Release 1.0.* See http://www.kjemi.uio.no/software/dalton/dalton.html, 1997.

2. H. Larsen, J. Olsen, P. Jørgensen, and T. Helgaker. Direct optimization of the atomic-orbital density matrix using the conjugate-gradient method with a multi-level preconditioner. *J. of Chem. Phys,* 115(21):9685–9697, 2001.

3. M. Watson, P. Salek, P. Macák and T. Helgaker. Linear scaling method and program for computing molecular properties. *to be published.*

4. P. Salek, O. Vahtras, T. Helgaker, and H. Ågren. Density-functional theory of linear and nonlinear time-dependent molecular properties. *J. Chem. Phys.,* 117: 9630, 2002.

Computational Experiments of Complex Fluid Flow Behavior

Luca Brandt, Geert Brethouwer, Dan S. Henningson, Astrid Herbst, Arne V. Johansson, Erik Lindborg
Department of Mechanics, Royal Institute of Technology (KTH), Stockholm

The Department of Mechanics, KTH, has been one of the leaders in the use of so called direct numerical simulations (DNS) for the understanding of fluid physics. Fast and accurate methods to solve the governing equations of fluid dynamics, a nonlinear set of partial differential equations called Navier-Stokes equations, enable flow experiments to be performed on today's supercomputers. Complex and chaotic phenomena such as turbulence can be directly simulated, even resolving the details of the swirling eddies and irregular fluid motion. The area of direct numerical simulation of fluid flows is now an internationally established avenue of investigation into complex flow physics and represents a new method of obtaining scientific results, in addition to traditional physical laboratory experiments and theoretical investigations. The high accuracy and detailed validation of the DNS methodology has now made this field stand on its own and to reveal new and exciting scientific results.

Many of the simulations performed by investigators using DNS at the Department of Mechanics, have utilized the resources at PDC. This report gives several examples of such results, each pertaining to important application areas and answering questions relating to complex fluid physics.

Transition to turbulence in a boundary layer

Transition to turbulence in a boundary layer subject to high levels of free-stream turbulence is of practical importance in many applications, e.g. electronics cooling, aerodynamics and boundary layers on gas turbine blades, but can still not be reliably predicted. In particular, in gas turbines, where a turbulent free stream is present, the flow inside the boundary layer over the surface of a blade is transitional for $50 - 80\%$ of the chord length. Wall shear stresses and heat transfer rates are increased during transition and a correct design of the thermal and shear loads on the turbine blades must take into account the features of the transitional process. Some parts of the phenomena observed and considered relevant in transition in boundary layer subject to free-stream turbulence have been analyzed in detail in the past using simple flow configurations and known repeatable disturbances. Simulations are performed to study this phenomenon on a flat-plate boundary layer flow.

Direct numerical simulations (DNS) of laminar-turbulent transition in a boundary layer subject to free-stream turbulence have recently been performed, in collaboration with P. Schlatter from Swiss Federal Institute of Technology Zurich (ETH Zürich). The spectral code used has been developed at the Department of Mechanics at KTH and was optimized using portable OpenMP directives for shared memory machines. It solves the Navier-Stokes equation for incompressible flows. It uses Fourier series expansion in the wall-parallel streamwise and spanwise directions and Chebyshev series in the wall-normal direction. The nonlinear terms are computed in physical space (pseudo-spectral treatment). A fringe region is implemented to force periodic boundary condition in the streamwise direction thus enabling the simulation of spatially evolving flows.

Isotropic and homogeneous free-stream turbulence is generated at the inflow of the computational domain. Simulations are performed with free-stream turbulence of intensity from $Tu = 4.7\%$ to $Tu = 1.5\%$ and different characteristic integral length scales. For the largest simulations 1152 × 201 × 192 grid points have been used for a computational domain of 900 × 80 × 90 boundary layer thickness at the inlet $Re_{\delta^*} = 300$.

In particular, it is observed how free-stream disturbances enter and affect a laminar boundary layer. The generation of streamwise elongated perturbations (streamwise streaks), mainly in the streamwise velocity component, is the dominant feature. We also studied the nonlinear evolution of these streaks. In fact, if the streaks reach a threshold amplitude, they become unstable initiating a rapid transition to a turbulent flow. Single streaks are seen to undergo a rapid wiggly motion and then break down into turbulent spots. The formation of streaks and the subsequent break down of the streaks can be observed in Figure 1. The turbulent spots merge giving a fully turbulent flow further downstream. Moreover the influence of the free-stream turbulence intensity and characteristic length scales on the location of transition is investigated.

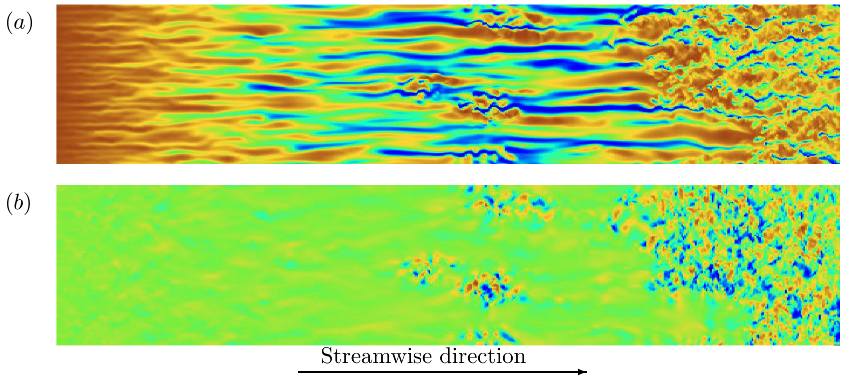


Figure 1
(a) Instantaneous streamwise velocity and (b) spanwise velocity in a plane parallel to the wall inside the boundary layer. Blue colour indicates low streamwise velocity and negative spanwise velocity. The formation of streamwise elongated streaks and turbulent spots can be seen.

Influence of periodic excitation on a turbulent separation bubble

The performance of diffuser-like flows occurring in many technical applications e.g. modern engine air intakes is strongly affected by separation which causes an excessive pressure loss. Therefore flow control is needed to improve the flow quality, to prevent separation and to promote reattachment.

DNS of the turbulent flow over a flat plate with an imposed outer pressure distribution are creating a strong adverse pressure gradient. As a consequence of the adverse pressure gradient, flow close the surface separates and backflow occurs. A region of counterflow can be clearly seen in the instantaneous flow configuration presented in Figure 2 (dark blue areas surrounded by white line).

The effects of time-periodic perturbations on such a separating turbulent boundary layer flow over a flat plate are studied at $Re_{\delta^*} = 400$. Periodic forcing is implemented through a local volume force. The leading parameters affecting the flow are the position where the periodic forcing is applied, its amplitude and the frequency. Periodic forcing at sufficiently high amplitudes in a certain frequency range is shown to be successful in preventing separation, which can be seen from the plot of the instantaneous velocity shown in Figure 2. The closer the forcing is to the separated region the more effective it is.

The calculations were performed on the IBM-SP at PDC, using a MPI version of the code for DNS of boundary layers developed at the Department of Mechanics.

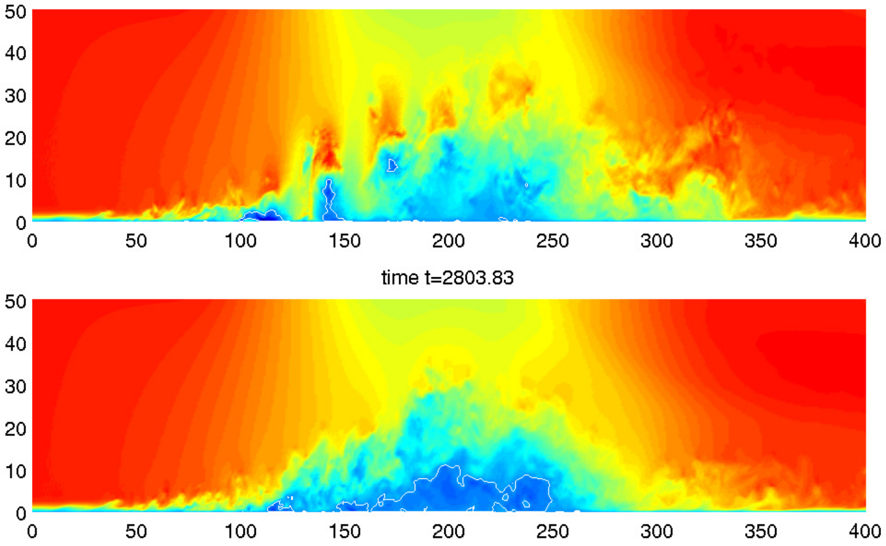


Figure 2
Instantaneous streamwise velocity (negative: dark blue to positive: red, white line: zero) for simulation with forcing (above) and without forcing (below).

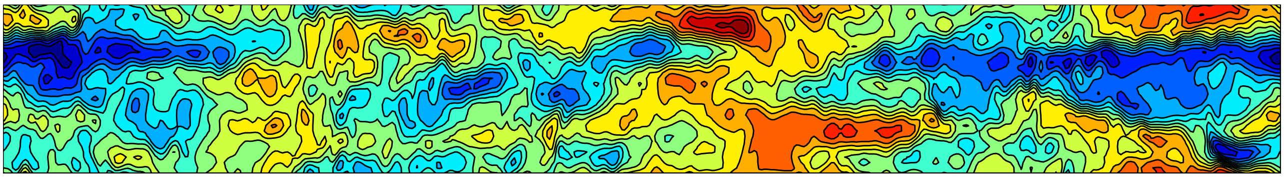
The energy cascade in a strongly stratified fluid

The horizontal mesoscale (wave lengths ~ 1–500 km) wave number spectra of kinetic and potential energy measured in the free atmosphere have been debated for several decades in the scientific community. It is a remarkable feature of these spectra that they exhibit a $k^{-5/3}$ power law, with wave number k, just as the spectrum of three-dimensional isotropic turbulence. However, it is easily argued that these spectra cannot be explained by classical three-dimensional turbulence. Instead, it has been speculated that they can be explained by the existence of an inverse cascade of energy, from small to large scales, a phenomena which can be observed in two-dimensional turbulence. An alternative hypothesis has also been advocated, that the spectra arise from a forward cascade of nonlinearly interacting gravity waves.

Thus we are here faced with an attractive scientific problem: There is an important observation and two alternative and principally different hypotheses have been put forward to explain it. With present experimental techniques it is not easy to determine which, if any, of the two alternative hypotheses is the correct. On the other hand, it is an ideal problem to attack by means of numerical simulations.

The answer to which of the two hypothesis is the correct, has been given by a series of numerical simulations of the Boussinesq equations carried out at the Kallsup computer. The Boussinesq equations are the governing dynamic equations for a strongly stratified fluid. Essentially, they consist of a coupled system of non-linear equations determining the velocity, pressure and temperature fields. The equations have been solved for an ideal system consisting of a highly elongated box, in which the fluid is excited at the largest horizontal scale by a random forcing. The number of grid points is 512 × 512 × 32 with 512 in each horizontal direction. The ratio between the vertical side and each of the horizontal sides of the box has been varied from 1/8 to 1/192. Especially in the most elongated boxes, where we also had the strongest stratification, the results were very clear. After some time it could be observed that layers were developed in the box and a stationary state with forward cascade of energy was obtained. Moreover, the horizontal spectra of kinetic and potential energy were exhibiting the observed $k^{-5/3}$-range. In Figure 3 we see a contour plot, in a vertical plane, of a horizontal velocity component, from the simulation with aspect ratio 1/192. The layered and yet chaotic structure of the flow is striking. In Figure 4 we see the kinetic and potential energy wave number spectra from the same simulation. These plots confirm that the dynamics giving rise to the $k^{-5/3}$ atmospheric spectra is a three-dimensional forward cascade.

Two further series of simulations are planned, one to systematically study the influence of system rotation and the Coriolis force on the motion and one to investigate the diffusion of a passive scalar in the Boussinesq system, a problem with important environmental implications.



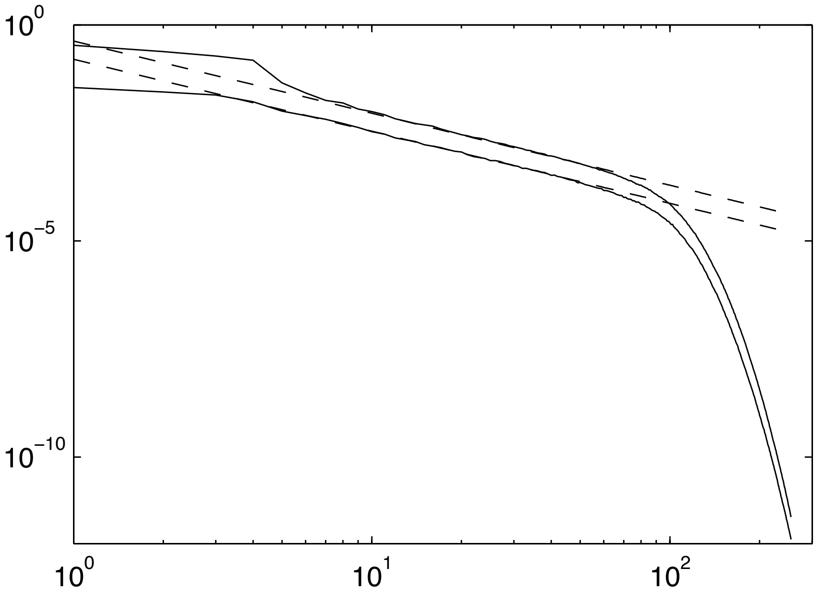


Figure 3
Contour plot in a vertical plane of the horizontal velocity component parallel to the plane, from simulation with vertical to horizontal box ratio 1/192. The vertical side of the box is magnified by a factor 24 in the plot.

Figure 4
Horizontal wave number spectra of kinetic (higher) and potential (lower) energy versus horizontal wave number k. A peak at wave number 4 is seen in the kinetic energy spectrum. This peak is due to the forcing. The dashed lines are of the form $\sim k^{-5/3}$.

Effects of rotation on turbulent flows

Basic physical laws state that the flow of a fluid in a rotating system feels a force as a consequence of the rotation. This force resulting from the rotation of the system is commonly known as the Coriolis force and affects the flow patterns and the unsteady and chaotic turbulent motions in the flow.

Many flows are influenced by the Coriolis force. For instance, atmospheric flows and flows in oceans are significantly affected by the rotation of the earth. Also the flow in industrial apparatus is often influenced by rotation. An example hereof is a gas turbine where the alternation of the turbulence by rotation affects the cooling of the turbine blades and the stability of the combustion in the combustor. In a large number of the aforementioned kind the flow is transporting a scalar, for example, contaminants, heat or fuels and oxidants. It is thus desirable to have a thorough understanding of how rotation influences transport processes. These two questions, how does rotation affect a turbulent flow and how does rotation influence transport processes in turbulent flows, were the subject of the present study.

A major problem is the difficulty of designing an experimental facility that can be used to investigate the effect of rotation on turbulent flows and turbulent transport in detail. An alternative tool is numerical simulation where we easily can take into account the Coriolis force. We decided to simulate the flow in a relatively simple geometry without any walls, but a linear and constant shear is imposed to drive the flow. The geometry is rotating around of the axes and a scalar is present in the flow. The scalar is transported and mixed throughout the fluid as a consequence of the turbulent fluctuating movements in the flow.

The numerical simulations are computationally very demanding and they are not possible without the availability of a supercomputer with many processors. In the present study we used grids with up to 74 million grid points. The memory requirement was more than 12 Gb and about 55 Gb disk space was necessary to store all the data. Even larger scale simulations will be performed in the future which puts higher demands on the computational resources.

Several simulations have been carried out with varying rotation speeds of the geometry. The simulations confirmed our expectation that the rotation has a strong influence on the flow. However, we have now reliable and accurate quantitative information about the change of the turbulence due to rotation and this clearly deepens and enlarges our knowledge about the effects of rotation. Furthermore, the statistics are very useful for the development of engineering models for turbulent flows.

One of the subjects about which we have now a much better understanding is the influence of the rotation on the instantaneous turbulent structures in a flow. Figure 5 presents a visualization of the instantaneous turbulent flow field and shows the regions that are characterized by a rapid swirling motion of the flow, often referred to as vortices. These vortices have a typical elongated tube-like shape and are approximately aligned with the flow direction. However, the vortex tubes are not completely straight according to the visualization but are often curved and are sometimes perpendicular oriented to the flow direction. This is a typical result of the non-linear and chaotic behavior of turbulent flows. The numerical simulations have revealed that the rotation strongly affects these kinds of turbulent structures in a flow.

As the transport and mixing of scalars are strongly allied with turbulent structures we might expect a significant change of the rate and direction of the scalar transport with varying rotation speeds. This is indeed confirmed by the simulations. Figure 6 shows plots of the instantaneous concentration of a scalar in case the flow is not rotating (upper figure) and in case the flow is rotating (lower figure). There is a clear difference between the case with rotation and without rotation. In the case without rotation we observe small plume-like structures in the scalar field which are absent in the case with rotation. It is therefore no surprise that the transport and mixing of the scalar in the flow shows a completely different behavior in these two situations. In the situation with rotation the transport rate of the scalar is much smaller than in the situation without rotation.

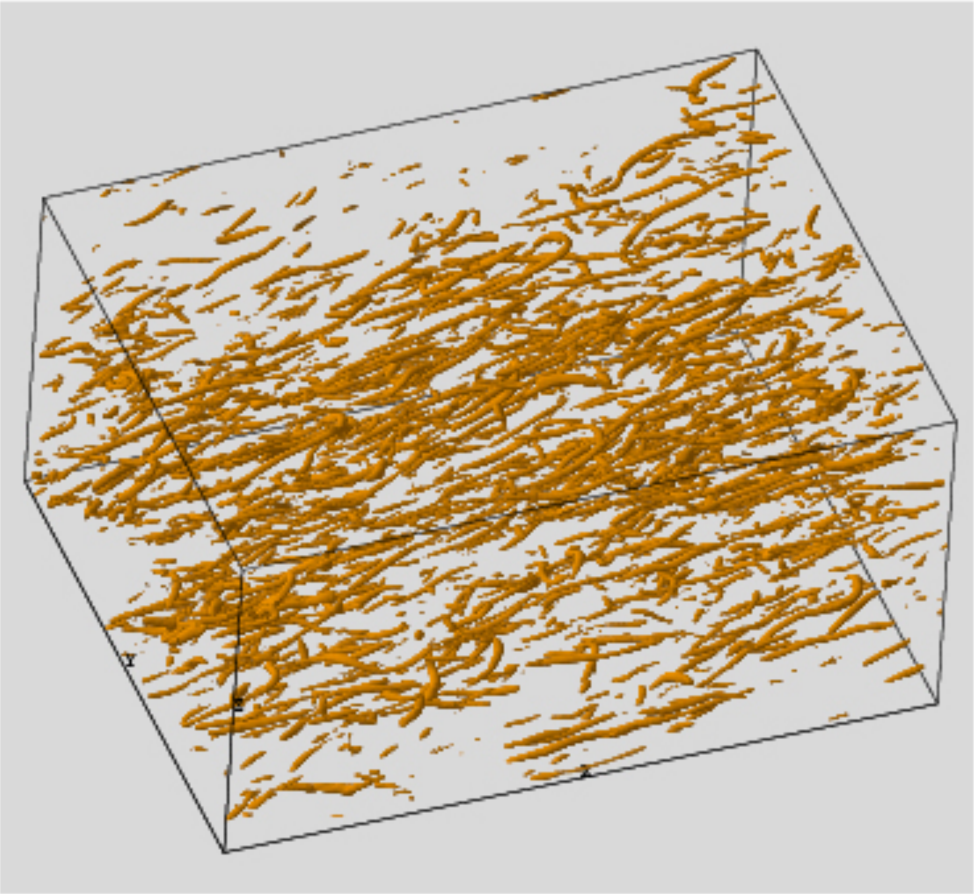
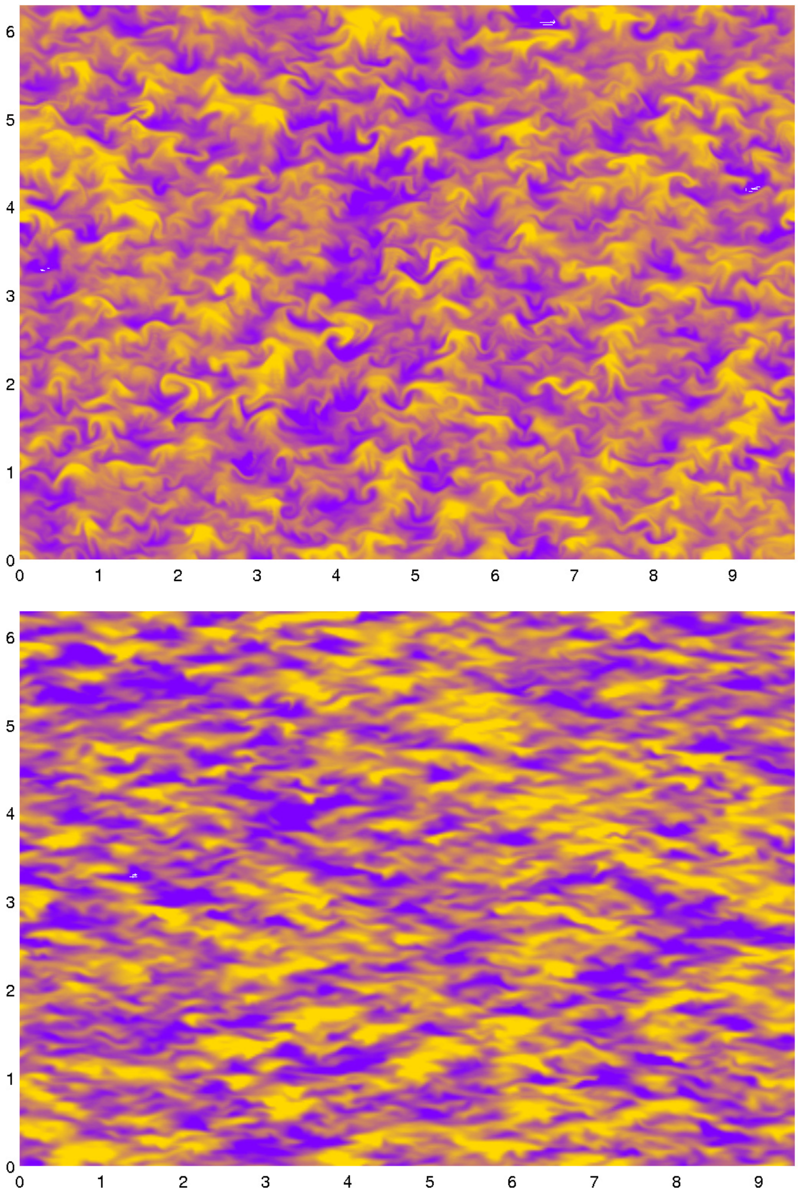


Figure 5
Vortex tubes in a turbulent shear flow.

Figure 6
The instantaneous scalar distribution in a plane perpendicular to the flow direction. The upper figure is for the case without rotation and the lower figure is for the case with rotation.



The projects have been supported by the Swedish Defense Research Agency (FOI), Göran Gustafsson Foundation, the Parallel Scientific Computing Institute (PSCI), the Swedish Energy Agency (STEM) and the Swedish Research Council (Vetenskapsrådet).

References

Brandt L., Henningson D.S. and Schlatter P., *Direct numerical simulations of boundary layer transition under free-stream turbulence*. In Bulletin American Physical Society, 47 (10), 2002, p. 136.

Brandt L., Schlatter P. and Henningson D.S., *Numerical simulations of transition in a boundary layer under free-stream turbulence*. In Advances in turbulence IX, Proc. 9th European Turbulence Conference, Southampton, 2002, p. 17-20.

Brandt L., Schlatter P. and Henningson D.S., *Transition in boundary layers subject to free-stream turbulence*. Accepted for publication in the Journal of Fluid Mechanics.

Brethouwer, G., *The effect of rotation on a rapidly sheared homogeneous turbulent flow with a passive scalar*. In proceedings of the Third International Symposium on Turbulence and Shear Flow Phenomena, Sendai, Japan, 2003, p. 311-316.

Herbst A. H. and Henningson D. S., *Influence of periodic exitation on a turburlent separation bubble.* PSCI report no. 30.07 (2003).

Lindborg, E., *Strongly stratified turbulence: A special type of motion*. In Advances in turbulence IX, Proc. 9th European Turbulence Conference, Southampton, 2002, p. 435-442.