inSiDE

inSiDE • Vol. 5 No.2 • Autumn 2007

Innovatives Supercomputing in Deutschland

Editorial

German supercomputing has advanced at a tremendous speed in the last months. The formation of the Gauss Centre for Supercomputing (GCS) has triggered a closer collaboration among the three national supercomputing centres (HLRS, LRZ, NIC) already visible during the successful ISC' 2007 conference in Dresden. At the same time the formation of the Gauss Centre for Supercomputing has encouraged German mid-range centres to work together more closely towards a future German HPC alliance. With these developments Germany is well prepared to take a leading role in European supercomputing together with its European partners. This is reflected in the EC-funded project PRACE that will soon start operation. This issue of inSiDE presents a first overview of the project from its co-ordinator Prof. Achim Bachem offering a first look at the future path Europe will take in supercomputing.

- Prof. Dr. H.-G. Hegering (LRZ)
- Prof. Dr. Dr. Th. Lippert (NIC)
- Prof. Dr.-Ing. M. M. Resch (HLRS)

After the powerful installations at NIC, HLRS and LRZ over the last three years it was time in Germany for a next big leap forward in 2007, too. The new IBM BlueGene/P system recently installed in Jülich (NIC) continues the tradition of permanent improvement in German HPC. With its peak performance of 222 TFLOP/s it is currently the fastest system in Europe. A description of the hardware and its integration into the existing hardware environment at NIC is given here. With this installation Germany takes again the lead in European HPC with respect to hardware performance in Europe.

Although hardware performance is a good yard stick to measure the quality of German supercomputing it is applications that define the quality of HPC research and form the driving force for further improvements. We therefore have again included a section on applications running on one of the three main German supercomputing systems. In this issue the reader will find eight articles about various fields of applications and how the different architectures have been exploited. The section not only reflects the wide variety of simulation research in Germany, but also nicely shows at the same time how various architectures provide users with the best hardware technology for all of their problems.

Similar to PRACE there are a number of projects that drive research and development in German supercomputing. For this issue of inSiDE we have chosen two such projects. D-Grid is the German Grid initiative aiming at providing German researchers with a sustainable Grid infrastructure. Since 2005 the project has been funded by the German Federal Ministry of Education and Research (BMBF) and has seen two calls for proposals in the last two years. At the core of the project is DGI – the D-Grid Integration project. How it is organized and how it aims at real solutions for German research is described in an article here.

The close integration of industrial and research usage of Grid resources is one of the outstanding features of German supercomputing. BREIN (Business objective driven REliable and Intelligent grids for real busiNess) is a project aiming at bringing the Grid closer into a business-driven environment. The EC-funded project was launched in 2006 and its technical background and goals are laid out in an article here.

As usual, this issue includes information about events in supercomputing in Germany over the last months and gives an outlook of workshops in the field. Readers are invited to participate in these workshops.

Contents

inSiDE

EU-Project PRACE to start in January 2008

Providing computing resources at the highest performance level for computer-based science and engineering in Germany and Europe: this is the common denominator behind current decisions made by German and European supercomputing facilities. First, the three German national supercomputing centres joined up to form the Gauss Centre for Supercomputing (GCS) with combined computing power of more than 120 TFLOP/s located in Stuttgart, Garching, and Jülich. Next, the Gauss Centre and Partners from 14 European countries founded the consortium PRACE – Partnership for Advanced Computing in Europe.

Background

The European Strategy Forum for Research Infrastructures (ESFRI) has identified High Performance Computing as a strategic priority for Europe. Scientists and engineers must be provided with access to capability computers of leadership class in Europe to remain competitive internationally and to maintain or regain leadership. Supercomputers are indispensable tools for solving the most challenging and complex scientific and technological problems through simulations. Following the recommendations detailed in the ESFRI Roadmap the European Commission issued a call in the 7th Framework Programme for a preparatory phase for up to 35 European Research Infrastructures. Based on the initial work of the High Performance Computing in Europe Taskforce (HET), HET members submitted the successful PRACE proposal. In parallel the PRACE consortium signed a Memorandum of Understanding (MoU) on April 17, 2007 in Berlin (inSiDE Vol. 5, No. 1, Spring 2007). Under the MoU work complementing the PRACE project will be undertaken and additional partners are invited to join the MoU.

The Objectives of PRACE

PRACE will prepare the creation of a persistent pan-European HPC service as a single legal entity, consisting of three to five centres, similar to the US HPC infrastructure. PRACE will be the tier-0 level of the European HPC ecosystem. The hosting centres of the planned tier-0 systems will provide the expertise, competency, and the required infrastructure for comprehensive services to meet the challenging demands of excellent users from academia and industry. PRACE will prepare for the implementation of the infrastructure in 2009/2010 by defining and setting up a legal and organizational structure involving HPC centres, national funding agencies, and scientific user communities to ensure adequate funding for the continued operation and periodic renewal of leadership systems, co-ordinated procurements, efficient use and fair access. PRACE will require an initial investment of up to 500 Mio Euro followed by annual funds of 100 Mio Euro for upgrades and renewal in later years. In parallel PRACE will prepare the deployment of Petaflop/s systems in 2009/2010. This includes the procurement of prototype systems for the evaluation of software for managing the distributed infrastructure, the selection, benchmarking, and scaling of libraries and codes from major scientific user communities, the definition of technical requirements and procurement procedures. PRACE will collaborate with the European IT-industry to influence the development of new technologies and components for architectures that are promising for Petaflop/s systems to be procured after 2010.

The European Research Area

The pan-European HPC service will be a part of the European Research Area, in which the Seventh Framework Programme is prepared to invest hundreds of millions of Euros. The PRACE infrastructure will be complemented with network and Grid access, and the services required to enable applications. These include development of parallel application software expertise, packages, and data handling. PRACE will build on the partners' experience and use concepts and services from EC-funded projects such as GÉANT2 and DEISA.

Open to European Researchers

PRACE is designed to meet the most challenging demands for computational resources in Europe. It will be open to all European researchers on a fair basis. Resource allocation will be based on peer review of the scientific quality of the proposal. Scientists from the following scientific communities have already expressed their interest in using PRACE:

- Quantum/theoretical chemistry and molecular dynamics
- Computational fluid dynamics
- Climate, weather, meteorology, earth systems, geophysics
- Physics
- Astronomy
- Astrophysics, gravitational physics
- Health, life and biosciences
- Nano/material science
- Computational engineering

An invitation to use the prototype systems to scale selected applications to exploit Petaflop/s will be issued in 2008.

Co-ordinator/Contact:
Prof. Dr. Achim Bachem
a.bachem@fz-juelich.de

The PRACE partners are:

- France: GENCI – Grand Equipement National pour le Calcul Intensif
- Germany: GCS – Gauss Centre for Supercomputing, represented by:
  - Forschungszentrum Jülich GmbH
  - Universität Stuttgart – HLRS
  - Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften
- The Netherlands: NCF – Netherlands Computing Facilities Foundation
- Spain: BSC – Barcelona Supercomputing Centre – Centro Nacional de Supercomputacion
- UK: EPSRC – Engineering and Physical Sciences Research Council
- Austria: GUP – Institut für Informatik der Johannes Kepler Universität Linz
- Finland: CSC – The Finnish IT Centre for Science
- Greece: GRNET – Greek Research and Technology Network
- Italy: CINECA – Consorzio Interuniversitario per il Calcolo Automatico dell'Italia Nord Orientale
- Norway: UNINETT – Sigma AS
- Poland: PSNC – Poznan Supercomputing and Networking Centre
- Portugal: Universidade de Coimbra
- Sweden: SNIC – Swedish National Infrastructure for Computing
- Switzerland: CSCS – Swiss National Supercomputing Centre

• Achim Bachem

Forschungszentrum Jülich (FZJ)

Turbulent Flow and Acoustic Simulations of a Jet Flow and the Flow around a High-lift Airfoil Configuration

The increasing growth of airline traffic and stricter noise regulations of airports demand acoustic prediction methods, which enable a low noise design of aircrafts already during the development process. For the analysis for both airframe and jet noise a hybrid method has been developed, which is based on a two-step approach using a Large-Eddy Simulation (LES) of the flow field and approximate solutions of Acoustic Perturbation Equations (APE) for the acoustic field.

Although useful theories, experiments, and numerical solutions exist, the understanding of subsonic jet noise mechanisms is far from being comprehensive. In addition, coaxial jets with round nozzles as they occur for aircraft engines can develop flow structures of considerably different topology, compared to single jets, depending on environmental and initial conditions and of the temperature gradient between the inner or core stream and the bypass stream. Not much work has been done on such jet configurations and as such there are still many open questions, for instance, how the mixing process is influenced by the development of the inner and outer shear layers or how the temperature distribution acts on the mixing and on the noise generation mechanisms. The present investigations are carried out in the framework of the European project Cojen to determine the flow and the acoustic field of high Reynolds number coaxial jet configurations taking into account the nozzle geometry and a heated inner stream.

The computational Grid used for the LES has approx. 22 million Grid points. Figure 1 shows instantaneous density contours coloured by the velocity magnitude. Slender torus-shaped structures are generated directly at the nozzle lip. Further downstream, these structures are stretched and become unstable, yielding smaller turbulent structures. The degree of mixing in the shear layers between the inner and outer stream, the so-called primary mixing region, is generally high. This is especially evident in Figure 2 with the growing shear layer instability separating the two streams. Spatially growing vortical structures generated in the outer shear layer seem to influence the inner shear layer instabilities further downstream. This finally leads to the collapse and breakup near the end of the inner core region. Figure 3 shows contours of the perturbation pressure distribution. It is widely accepted that two different noise generation mechanisms exist, one is associated with coherent structures generating noise which is radiated to the downstream direction and the other is related to small scale turbulence structures contributing to high frequency noise normal to the jet axis.

During the landing approach of an aircraft, when the engines are near idle condition, airframe noise is becoming the dominant noise source. Main contributors to airframe noise are the landing gears and high-lift devices, like slats. To identify the flow phenomena generating slat noise, an LES of the flow around an airfoil consisting of a slat and a main wing is performed in the national project FREQUENZ. The computational mesh consists of 55 million Grid points. The flow simulation, performed on the NEC SX-8 at HLRS Stuttgart, is done by a flow solver optimized for vector computers and parallelized by using the Message Passing Interface (MPI). For the here mentioned simulations the maximum obtained floating point operations per second (FLOPS) amounts 6,7 GFLOPS, the average value was 5,9 GFLOPS. An average vectorization ratio of 99,6 % was achieved with a mean vector length of 247. For the statistical analysis and the determination of the acoustical sources samples of the solution are recorded with a total data size of about 7 TB.

The simulation results are in good agreement with the data determined with particle-image velocity measurement carried out at the DLR Braunschweig Windtunnel (AWB). Vortical structures in the flow field are visualized by 2 contours following the work of Jeong and Hussain (1995). Figure 4 reveals areas of turbulent flow located in the boundary layers of the slat and main airfoil, downstream of the slat trailing edge, and in the slat cove region. The transition of the boundary layer from laminar to turbulent flow occurs shortly downstream of the leading edges. The acoustical field shown as pressure contours in Figure 5, show clearly that the slat is the major noise source. At present further analysis of the data is carried out for the determination of the noise generation mechanisms.

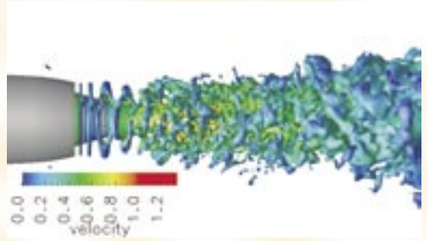


Figure 1: Instantaneous density contours with mapped on velocity field

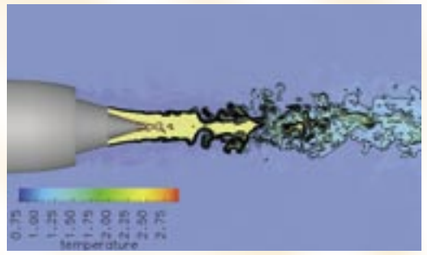


Figure 2: Instantaneous temperature contours

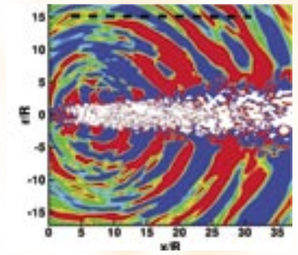


Figure 3: Pressure contours of the single jet

Figure 4: λ_2 contours in the slat region

Figure 5: Pressure contours based on the LES/APE solutions

• Daniel König
• Elmar Gröschel

Aerodynamisches Institut
RWTH Aachen

Autumn 2007 • Vol. 5 No. 2 • inSiDE

Breaking the Load-Balancing Barrier by Block Splitting in the CFD Code NSMB

In a recent comparative benchmarking investigation [1] of the flow simulation code NSMB, we tested its performance on various scalar and vector platforms (Cray XT3 and NEC SX-5 at CSCS, NEC SX-8 at HLRS). In this article we focus on the load-balancing issues which prevent a good parallel scaling on these systems. As a remedy we present the block-splitting technique and study its effect on the parallel performance.

NSMB [2] is a finite-volume solver for compressible flows using structured Grids. It supports domain decomposition into Grid blocks (multi-block approach) and is parallelized using the MPI library on the block level. This means that the individual Grid blocks are distributed to the processors of a parallel computation, but the processing of the blocks on each CPU is performed in a serial manner.

We used one of our typical flow simulations for the benchmarking: the Large-Eddy Simulation (LES) of a jet-in-cross flow configuration related to the film cooling of gas turbine blades [3]. The cooling of turbine blades is often necessary since the temperature of the hot combustor gas typically exceeds the material limits of the blades. As displayed in Figure 1, a coolant jet fed from a large isobaric plenum is issuing through a short inclined nozzle into the hotter cross flow. Thereby it mixes with the hot gas and creates a relatively cold layer that shields the blade surface. In our test case, the computational domain consists of a total of 1,7 million cells, distributed to 34 blocks of largely varying dimensions and cell counts.

Results

Our benchmarking results with a varying number of processors in the parallel computations are summarized in Figure 2. As a performance measure, the number of time steps per wall-clock time is displayed in panel (a), whereas panel (b) shows the parallel efficiency (i.e., the parallel speedup normalized by the number of processors). While for the given test case the processor performance of the NEC vector computers is sufficient for low simulation turnover times even at a low number of processors, the Opteron CPUs of the Cray XT3 are considerably slower. Therefore, correspondingly more CPUs have to be employed on the Cray to compensate.

However, this is usually not easily feasible due to the block-parallel nature of the simulation code in combination with the coarse-grained domain decomposition of our typical flow cases. While the total block count is a strict upper limit for the number of CPUs in a parallel simulation, a severe degradation of the load-balancing (as evident from Figures 2 (a) and (b)) renders parallel simulations with a block-to-CPU number ratio of less than 4 very inefficient (also cf. Figure 3 (a)). On the NEC SX-8, an additional issue is the strongly decreasing average vector length especially on the smallest blocks, see Figure 3 (b).

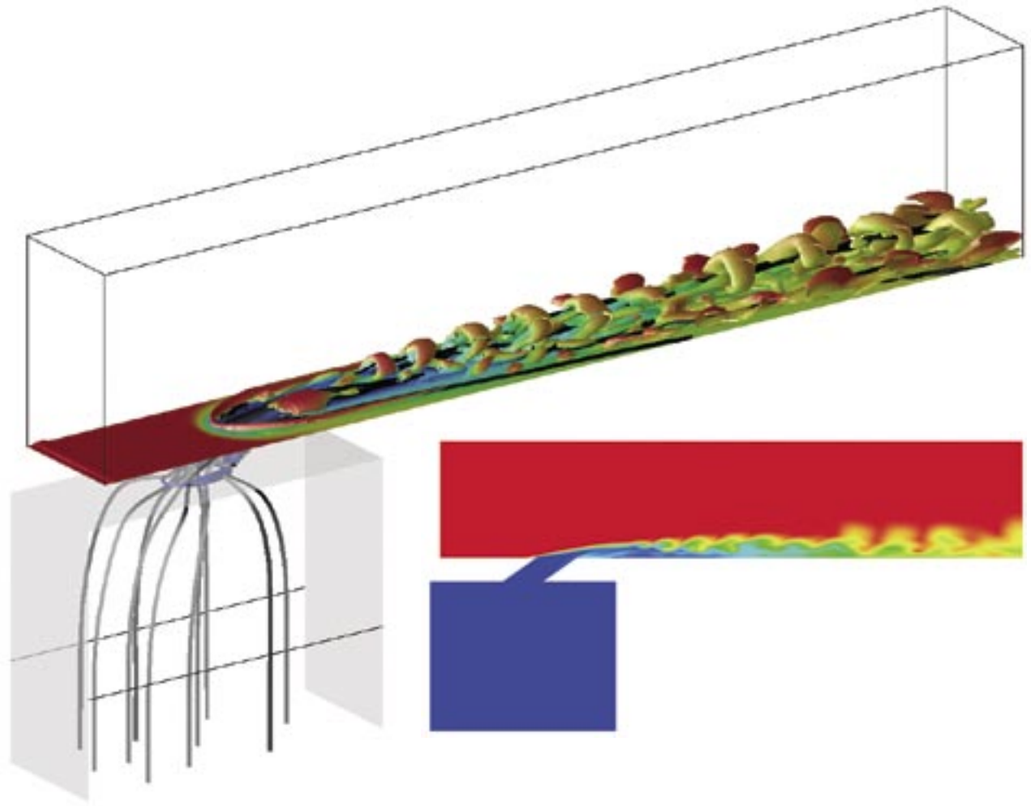
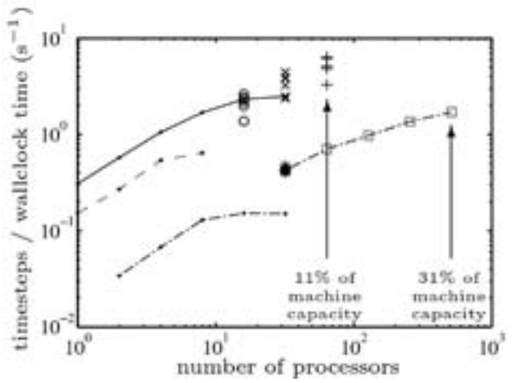


Figure 1 (a): Visualization of flow configuration and instantaneous flow field. The cutting plane displays a colour visualization of the temperature. The colour bar ranges from blue (cold) to red (hot). Additionally shown are vortex structures coloured with temperature and streamlines as thick dark ribbons. Solid walls are shaded in grey, and thin black lines denote the domain boundaries. The small inset picture (b) displays a colour visualization of the instantaneous temperature.

Applications

(a)



(b)

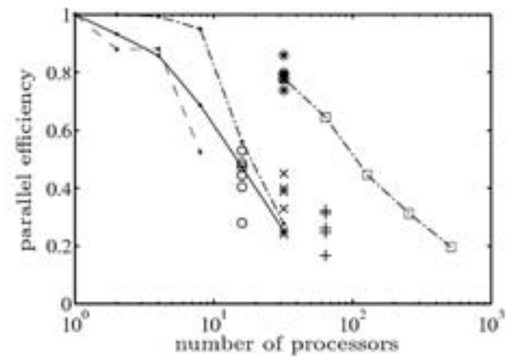


Figure 2: Aggregate performance chart for varying number of processors

(a) Time steps per wall-clock time
(b) parallel efficiency.

——— NEC SX-8 with 34 blocks;
o 16 CPUs
x 32 CPUs
\+ 64 CPUs
on NEC SX-8 with a varying number of blocks ranging from 68 to 340

– – – NEC SX-5 with 34 blocks

–·–·– Cray XT3 with 34 blocks

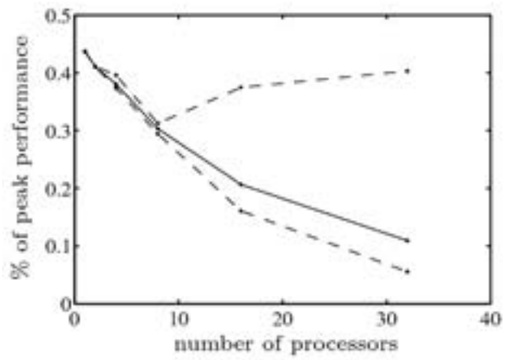
* Cray XT3 with a varying number of blocks ranging from 68 to 680 and 32 CPUs

—⊡— Cray XT3 with 680 blocks

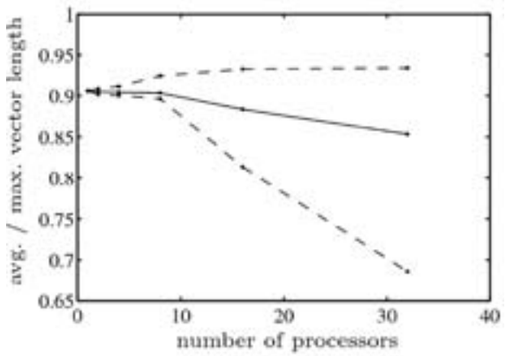
Figure 3: Dependence of (a) average floating-point performance (normalized with the peak performance of 16 GFLOPS per processor) and (b) average vector length (normalized with the length of the vector register of 256) on number of processors on NEC SX-8 for a fixed number of 34 blocks.

——— mean and – – – maximum/minimum values taken over all processors of a simulation.

(a)

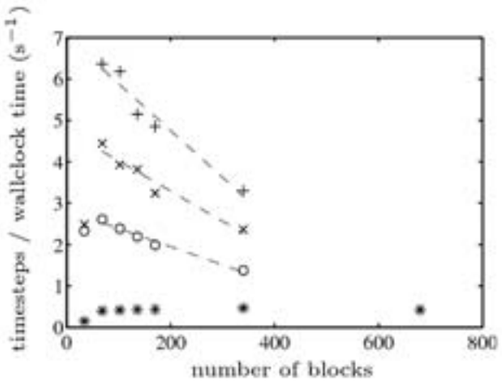


(b)



An alleviation of the load-balancing problem can be found in the block-splitting technique, were the total number of blocks is artificially increased by splitting them with an existing utility programme [4] (at the expense of a memory- and computation-overhead due to an increased number of ghost cells). The finer granularity of the domain subpartitions allows for a more homogeneous distribution of the work to the individual processors, which leads to a more efficient parallelization and a drastically improved performance. While a twofold increase of the number of blocks already caused a more than threefold simulation speedup on the Cray XT3, a further splitting of the blocks did not yield considerable performance improvements, cf. Figure 4. The optimum simulation performance on the Cray XT3 for the given test case was reached for a CPU-to-block number ratio of 10. On the NEC SX-8, the block splitting technique generally also yields favourable results. However, it is advisable to

(a)



(b)

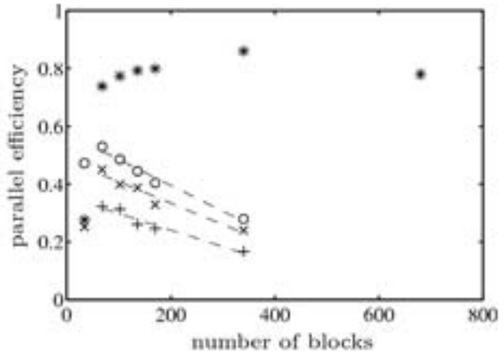


Figure 4: Aggregate performance chart for varying number of blocks

(a) Time steps per wall-clock time
(b) parallel efficiency

o 16 CPUs

◊ 32 CPUs

\+ 64 CPUs on NEC SX-8

– – – linear fit through falling parts of the curves

✶ Cray XT3 with 32 CPUs

keep the total number of blocks on this machine as low as possible to avoid an undue degradation of the average vector length and thus a diminished parallel performance. The most important benefit from block splitting can be seen in the extension of the almost-linear parallel scaling range to a considerably higher number of processors (cf. Figure 2 (b)). Given the low number of blocks in our typical simulations, block splitting seems to be the only possibility allowing for an efficient use of scalar supercomputers with a computational performance of NSMB that is comparable to simulations with a low number of CPUs on vector machines.

However, when employing an increased number of CPUs on the NEC SX-8, corresponding to a similar allocation of the total machine size than on the Cray XT3, the SX-8 performance considerably exceeds the capabilities of the XT3, see Figure 2 (a). Furthermore, an increased number of subpartitions comes along with an overhead in bookkeeping and communication, i.e., an memory- and computation-overhead. Additionally the complexity of the pre- and post-processing rises considerably.

We conclude that a vector supercomputer with a relatively low number of high-performance CPUs such as the NEC SX-8 is far better suited for our typical applications of NSMB than a scalar machine with a large number of relatively low-performance CPUs such as the Cray XT3.

Acknowledgements

Part of this work was carried out under the project HPC-Europa of the European Community. The hospitality of Professor U. Rist (Institute of Aero and Gas Dynamics, University of Stuttgart) is greatly appreciated. We thank U. Küster (HLRS), S. Haberhauer (NEC HPC Europe) and P. Kunszt (CSCS) for fruitful discussions.

References

[1] **Ziefle, J., Obrist, D., Kleiser, L.**
"Performance Assessment and Parallelization Issues of the CFD Code NSMB", High Performance Computing on Vector Systems: Proc. 6th Teraflop Workshop, HLRS Stuttgart, March 26-27, 2007, Springer, to appear

[2] **Vos, J. B., van Kemenade, V., Ytterström, A., Rizzi, A. W.**
"Parallel NSMB: An Industrialized Aerospace Code for Complete Aircraft Simulations", Proc. Parallel CFD Conference 1996, edited by P. Schiano et al., North Holland, 1997

[3] **Ziefle, J., Kleiser, L.**
"Large-Eddy Simulation of Film Cooling: Treatment of Inlet Boundary Conditions for the Isobaric Plenum and Instantaneous Vortex Structures", Proc. Appl. Math. Mech., 2007, submitted

[4] **Ytterström, A.**
"A Tool For Partitioning Structured Multiblock Meshes For Parallel Computational Mechanics", The International Journal of Supercomputer Applications and High Performance Computing, Vol. 11, 1997, pp. 336-343

• Jörg Ziefle
• Dominik Obrist
• Leonhard Kleiser

Institute of Fluid Dynamics, ETH Zurich

Autumn 2007 • Vol. 5 No. 2 • inSiDE 13

Mesoscopi Simulations of Particle-laden Flows

Particle-laden flows are ubiquitous in our daily life and include the cacao drink which keeps separating into its constituents, tooth paste and wall paint which are mixtures of finely ground solid ingredients in fluids or blood which is made up of red and white blood cells suspended in a solvent. Also of great interest for experiments are individual particles solved in a liquid which are moved around by mechanical or optical forces. Long-range fluid-mediated hydrodynamic interactions often dictate the behavior of such systems, but are not taken into account in most analytical and numerical works due to their complexity. Various simulation methods, like Brownian Dynamics and Stokesian Dynamics to name a few, have been developed to simulate particle-fluid mixtures. All of them have their inherent strengths but also some disadvantages. In particular, often even todays most powerful computers are not able to treat experimentally relevant particle numbers. A number of more recent methods attempt to describe the time dependent long-range hydrodynamics properly with the computational effort scaling linearly with the number of particles only. These include recent mesoscopic methods like dissipative particle dynamics, the Lattice-Boltzmann Method, or stochastic rotation dynamics to solve the fluid flow and a molecular dynamics solver for the dynamics of the particles.

In this article we give a short description of the application of such methods to a selected range of applications. These include single particles in very complex environments as well as Brownian and non-Brownian suspensions. In Brownian suspensions, i.e. where thermal fluctuations are of importance, particle diameters are roughly on the micrometer scale or below. In macroscopic systems, thermal fluctuations can be neglected. In order to avoid finite size effects and to gain good statistics, large systems are needed. While in the single particle case, we have to use a very high resolution of the fluid solver in order to obtain the proper flow field around a particle embedded in a complex geometry, in the case of suspensions, also many thousands of solved particles are required. Thus, in both cases we have truly high performance computing applications requiring hundreds of CPUs and terabytes of file storage. An important focus of our work is on modelling an experimental setup as closely as possible. For example, we simulate typical experiments common for microfluidic applications. In particular, we are interested in studying the effect of boundary slippage by modelling a modified atomic force microscope and the application of optical tweezers to colloidal crystals.

Lattice-Boltzmann Studies of Boundary Slippage

The most popular mesoscopic simulation method is the Lattice-Boltzmann Method which has been extended to the simulation of particle-laden flows by Anthony Ladd in the early 90s. One of the main advantages of this method is that it can be easily parallelized and the codes nicely scale on thousands of CPUs. We currently use a number of different implementations which are parallelized using MPI or OpenMP and have been found very efficient on many different machines. The Lattice-Boltzmann Method does not contain thermal fluctuations without further modifications. Therefore, it is most well suited for particle diameters of at least a few tens of micrometers. Currently, we use this method for two different projects. The first one is related to a basic question from the topic of microfluidics. Here, recent experiments have found a so-called boundary slip, i.e. the velocity of a fluid at a surface was found to be larger than the velocity of the surface. This can be seen as a contradiction to the common sense of most scientists working in fluid dynamics. However, on the micro- or nanoscale, effects like dissolved gasses, electrostatic repulsion or impurities become important. These effects often cannot be resolved by the measurement apparatus. A common experiment to quantify boundary slip applies a modified atomic force microscope, where a sphere is oscillated in the vicinity of a surface. From the force acting on the sphere, one can compute the velocity of the fluid close to the wall. We are investigating such experiments by simulating a moving sphere along surface geometries obtained from atomic force microscope measurements of experimental groups and give quantitative estimates of the order of the boundary slip. A typical setup is shown in Figure 1. For simulation of realistic setups, very large lattices are needed since the surface variations are small compared to the sphere radius causing the need of simulation lattices of the order of 512^3 to 1024^3. While our standard Lattice-Boltzmann solver is commonly used on the NEC SX-8 at HLRS in Stuttgart and on the IBM p690 in Jülich [1, 2], our suspension code has been found to perform very well on the new Opteron Cluster at SSC in Karlsruhe.

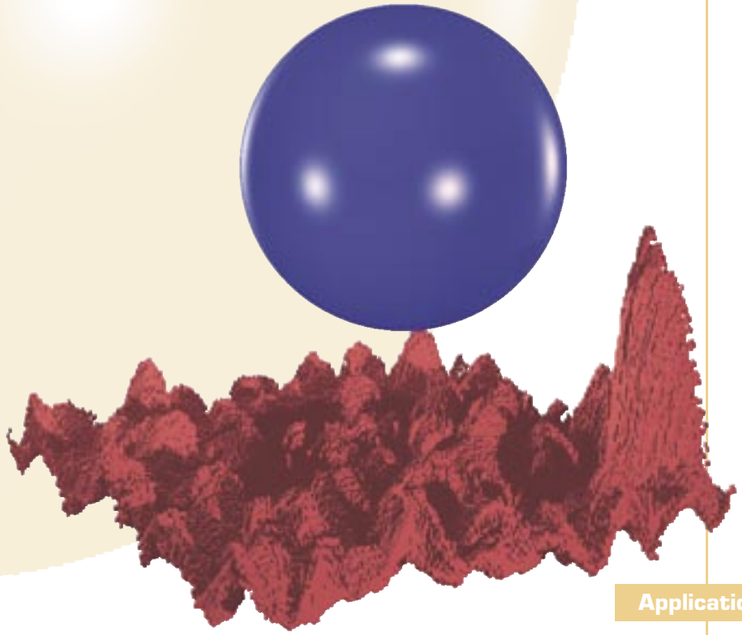


Figure 1: Simulation of a typical slip measurement: a sphere surrounded by a fluid (not shown) is oscillated in the vicinity of a surface. From the measured force on the sphere, information about the boundary conditions can be obtained. (For better visibility, the sphere radius was reduced.)

Simulation of clay-like Colloids

For civil engineering material properties of soils are of central importance. The design of fundaments for large buildings gets its bearings from the building to construct, but from the ground on which it is located as well. Especially clay and silt are very complex materials, since they consist of many sub-micrometer sized particles. The interactions between these particles determine the properties of the soil as a whole. The compressibility of the soil material is one of the aspects of interest, when buildings are constructed, but the behaviour under shear is at least of the same importance if we think of landslides. To gain a deeper understanding of silt and clay our collaborators have performed standard soil mechanics experiments on a model system: alumina powder suspended in water.

The salt concentration and the pH-value can be controlled in the laboratory and the composition, surface properties, and particle size of powder are well defined. A typical experiment is to prepare a suspension of a given volume fraction of suspended particles. This suspension is divided into several vessels and the pH-value is adjusted differently in each of the vessels. After some weeks a sediment is formed and its height is different in each of the vessels. Since the suspension has cleared up, one can conclude that most of the material has settled down, but the sediment has a different density, depending on the pH-value. The dependence of the interactions between the particles leads to different porosities in the sediment. This statement can be tested by measuring the compressibility of the sediment: loose packings are more brittle and can collapse after a certain bearing pressure is exceeded. However, the question remains, why different types of sediments appear and how the differences can be quantified. The starting point here are the interactions among the individual soil particles. Their surface chemistry, the salt concentration of the water contained in the pores of the soil, as well as the pH-value are the most important ingredients for an adequate description. Especially, the microscopic structure, i.e., the local order of the soil particles is one of the features which one has to consider to understand the properties of the soil. By computer simulations we can reproduce the cluster formation process and thus obtain the microscopic structure of the sediment. This process and its dependence on the interactions among the particles plays the key role for the formation of sediments, and essentially determines the mechanical properties of such a sediment. Since we are interested in the aggregation of clusters we need large scale simulations. Not only large numbers of soil particles (order of 10^5) are needed, but also due to low particle concentrations, a large computational effort is needed to simulate the fluid between them (order of 10^7 fluid particles for the stochastic rotation dynamics algorithm). The process of the cluster growth can be characterized by the cluster size distribution and its development in time, depending on the interactions between the particles [3,4].

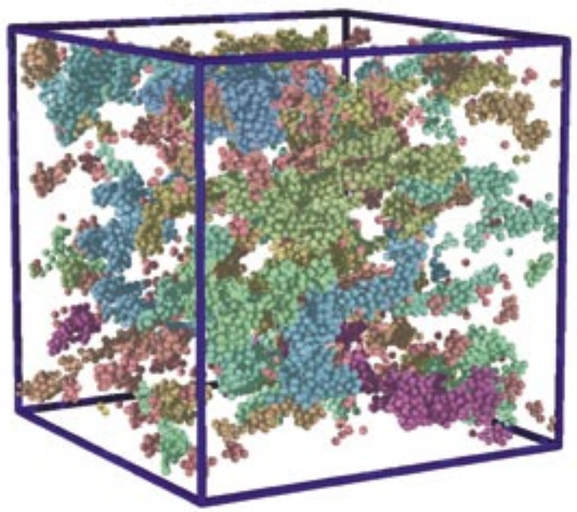


Figure 2: Simulation of cluster formation in a diluted suspension of silt particles: microscopic particles stick to each other and form larger and larger clusters, which settle down on the ground and form the sediment. 10^5 soil particles and 10^7 fluid particles have been simulated on 32 CPUs of the IBM p690 at NIC in Jülich.

Optical Tweezers in colloid Science

We are interested in the reaction of a colloidal crystal to external disturbances. Colloids are an interesting model system for such questions because the typical timescale (seconds) and length scale (micrometers) are far easier accessible then those of atoms and molecules (femtoseconds and angstroems). At the same time the interaction potential between colloids can be tuned in many ways, e.g. by coating them with polymers, adding salt to the suspension etc. Thereby the condition under with a crystal is formed can be changed and examined as desired. We model a colloidal crystal suspended in water by means of stochastic rotation dynamics for the fluid solvent and a molecular dynamics algorithm for the colloidal particles in the same manner as in the previous section. We study the reaction of the crystal to external disturbances by dragging an impurity through the crystal using an optical tweezer. The focus of the optical tweezer is moved with time, thereby pulling the impurity along. The optical tweezer is modelled with a harmonic potential, i.e. as if the impurity were connected with the trap centre by an ideal spring. Optical tweezers trap a colloid (or even an atom) in the focus of a laser beam: this is because a dielectric is always driven towards the strongest field gradient. They are a very important tool in soft condensed matter physics: colloids can not only be trapped, they can also be moved around individually by moving the focus of the laser beam. Thereby the colloidal system can be controlled with an accuracy that would be impossible for an atomic system.

To analyze the effects of the disturbance, we can either measure the distance that the impurity stays behind the focus - and thereby the force required to drag the impurity through the crystal - or examine the reaction of the crystal itself. To get a quantitative measure for the amount of damage done, we count the number of colloids that are no longer six-fold coordinated. In computer simulations the dependence of e.g. the crystal's stiffness on a wide range of parameters can be examined, especially those that are not easily accessible in an experiment. For our simulations we use the same simulation code as for the suspensions described in the previous section. An example of a rather small system is given in Figure 3. For large tweezer velocities, much larger crystals are needed and for studying the relaxation of the crystal, we also have to simulate for at least a few seconds of real time causing these simulations to cost up to a few thousand CPU hours each.

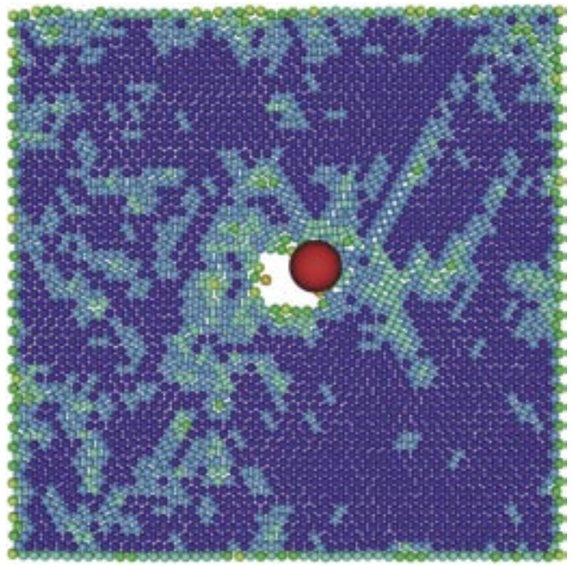


Figure 3: Simulation of a large colloidal particle dragged through a crystal consisting of smaller colloids by means of an optical tweezer. The colouring denotes defects occuring due to the distortion of the crystal.

References

[1] **Harting, J. and Giupponi, G.**
High Performance Computing in Science and Engineering '06, edited by W. Nagel, W. Jäger and M. Resch

[2] **Harting, J., Giupponi, G., Coveney, P. V.**
Physical Review, E 75, 041504, 2007

[3] **Hecht, M., Harting, J., Herrmann, H. J.**
Physical Review, E 75, 051404, 2007

[4] **Hecht, M., Harting, J., Herrmann, H. J.**
International Journal of Modern Physics, C 18, 501, 2007

• Jens Harting
• Martin Hecht
• Christian Kunert
• Rudolf Weeber

Aerodynamisches Institut
RWTH Aachen

Computational Steering: Interactive Flow Simulation in civil Engineering

Large and compute-intensive Computational Fluid Dynamics (CFD) simulations are usually run in non-interactive mode as batch processes on queuing systems of high-performance computers. Non-interactive means that the user cannot modify the layout of the simulation after the job has been initiated. In a first step typically referred to as pre-processing, the model geometry is mapped to a computational Grid. Together with boundary conditions such as surface temperatures or flow velocities a file or a database record is generated which completely describes this setup information. The latter may be submitted to a batch queue as a "computing job". As soon as the required hardware resources are available, the computation starts and continuously saves its output to disk. The user may evaluate the simulation results during a subsequent post-processing step, after the job has been completed. The above described workflow is practical for problems when one is interested in obtaining detailed and accurate information in a fluid flow investigation. For the wide variety of investigations and the explorative character of an integrated environment which is necessary for performing case studies, however, the user would like to directly interact with a running simulation and to immediately visualize the corresponding physical reactions. "To interact with the simulation itself" is the basic idea of Computational Steering (Mulder et al., 1999). To fulfil the requirement of immediate – or at least low latency – responses to user interactions during a running simulation, a fast CFD solver must be installed on a supercomputer or a computing cluster which is coupled to a steering and visualization workstation by a most efficient communication concept. What about the industrial application? Industry sectors with large-scale production such as the automotive industry usually invest considerable amounts of time and sink money in the design phase of new product prototypes. In contrast, the specialty of civil engineering is the construction of "unique" copies. The design phase consequently has to be much shorter and less extensive to be profitable in the building industry. Due to the lack of time for detailed simulations during the planning phase, it is common practise to rely construction and design solely on empirical rules. It is well known that the associated shortcomings and their belated elimination is a cost intensive issue. This leads to the desire for an interactive simulation tool for preliminary investigations, which offers the possibility to run short simulation cycles to proof or to find a basic concept, possibly followed by a few selected s imulation setups for more detailed investigations.

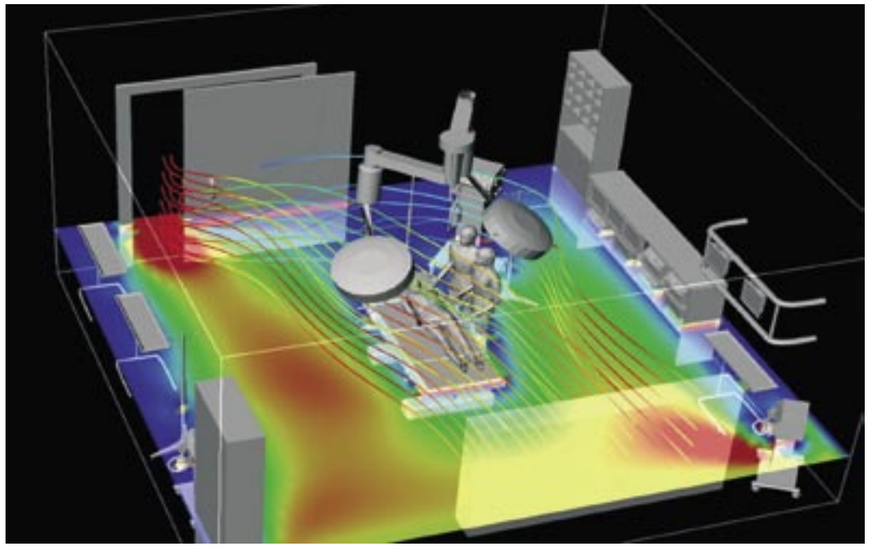


Figure 1: Screenshot of a Computational Steering session while exploring the capabilities of a ventilation system of a surgery room. A user may modify geometry and boundary conditions online during the simulation while results are immediately updated.

With this situation in mind, the Computational Steering application iFluids has been developed at the Chair for Computational Engineering at the Technische Universität München in co-operation with the Leibniz Computing Centre (LRZ) (Wenisch et al., 2005-2007). The latter institute provided special support in using this technology on high-performance supercomputers. As an example, Figure 1 shows the application of the tool for the fluid flow simulation of a surgery room. A sterile air stream is directed towards the patient's wound during an operation in order to prevent infection through bacteria of the room's unfiltered air.

Description of iFluids

iFluids is an application primarily designed as a tool for indoor air flow simulations, but it may easily be extended to support simulation studies in other fields with a focus on geometric setup. Its distinguishing feature as compared to conventional, i.e. non-interactive, Computational Fluid Dynamics applications is its layout as a true Computational Steering framework for high-performance computers. Users are able to visualize current simulation results on-the-fly close to real-time and to interact with the continuously running simulation. Besides basic interactions such as (re-)starting, stopping and pausing the computation, the user can adjust global simulation parameters and, most important, he or she can modify the geometry of the simulated scene as well as its boundary conditions. This is possible either on standard workstation desktop hardware or on high-end virtual reality user interfaces. The interactions comprise adding, deleting, and transforming simulation objects, as well as setting and changing their boundary conditions and configuring their respective parameters. Finally, the computational kernel can be adapted with regard to its numerical model and thus the best optimization available for a particular hardware platform. In an associated project, the institute also investigated the topic of collaborative engineering in this context. In order to support online co-operation between engineers which offers a considerable advantage in large projects, the application has been designed in such a way that multiple visualization and steering clients can be connected to the simulation (Borrmann et al., 2006). Figure 2 sketches the basic software design of iFluids and presents a scheme of how the visualization and steering front-end (VIS) is connected to the simulation kernel. The latter may be

Applications

processed on a supercomputer which may be located elsewhere. At LRZ's setup, the VIS process is usually run on a graphics workstation or visualization cluster, while the parallel simulation code is executed on multiple nodes of the SGI Altix 4700 supercomputer. Within the Altix, all processes communicate via vendor-optimized MPI (Message Passing Interface). Inter-machine communication between VIS and the master node of the simulation (SIM-M) is realized by using either Globus MPICH-G2 (http://www3.niu.edu/mpi) or PACX-MPI (http://www.hlrs.de/organization/pds/projects/pacx-mpi). The most recent pressure, velocity and temperature fields of a simulation are sent to the visualization within short time intervals. A user may analyze the results and, accordingly, may change the scene geometry or its simulation parameters. As mentioned, this is possible online, i.e. throughout the ongoing computation.

Extensions for interactive thermal Comfort Assessment

As the assessment of indoor thermal comfort is becoming increasingly important in the industrial environment, the research group currently also focuses on the integration of a local thermal comfort model into the steering prototype. This includes the development of a fast radiation solver, a numerical thermal manikin modeling the heat exchange between the body and environment with a (passive) system taking physical and physiological properties, the blood circulation and the (active) human thermoregulation system into consideration, furthermore a model for the local and global temperature sensation, and a model for thermal comfort assessment. The idea behind this ongoing work is to directly visualize the local thermal comfort perception on the artificial skin of a "numerical dummy" model. Colours will thereby indicate the level of local (dis)satisfaction. It is referred to (van Treeck et al., 2007) for details.

Summary

The central benefit resulting from Computational Steering is the "interactivity" as a means of intuitive experimentation during and with the simulation by its users. The achievable performance is quite satisfactory if appropriate compute hardware such as a cluster or supercomputer is available (Wenisch et al., 2005). Computational Steering may serve as a helpful tool in the daily engineering practice in the near future. However, a particular requirement for its real application is that these supercomputing resources are available during the engineer's working time. This becomes evident in case of a concurrent collaborative session as resources should be available in time with the appointment. The latter may be realized using an enhanced scheduling system that offers the flexibility to reserve required resources for a certain day and time of an appointment – a feature referred to as, advanced reservation. This feature is understandably not widely used in computing centres and usually not accessible for the public.

When working with large computational Grids and accordingly with large visualization data sets, the communication between computation, visualization and steering front-end is a bottleneck – especially between remote sites. To countervail this situation, a special Grid middleware would be desirable designed for supporting remote visualization on specialized hardware which is efficiently connected to the supercomputer in order to transfer previously computed visualization data to the client, e.g. in terms of a simple video stream. In this sense, also virtual reality environments with special input devices and independent views to simulation scenes as extensions for collaborative engineering should be supported. First products and projects towards this direction are already available, but these tools do not yet sufficiently address all of these requirements.

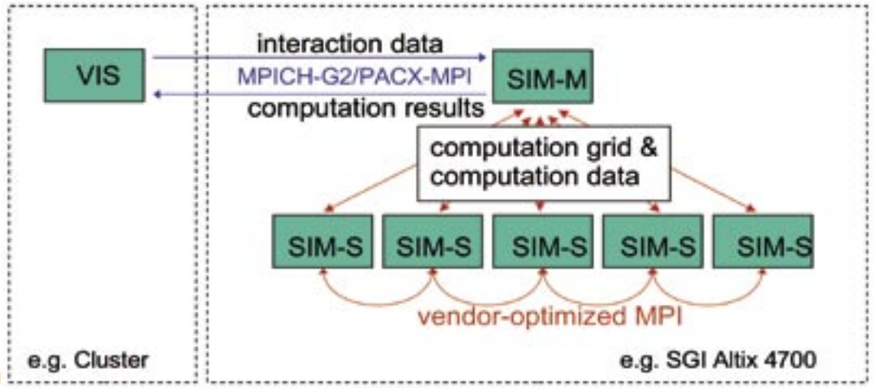


Figure 2: Application design of iFluids: This figure shows the priciple design with the visualization and steering client running on specialized hardware, while the simulation is processed on a supercomputer. The communication between the components is realized via MPICH-G2 or PACX-MPI for intermachine communication and vendor-optimized MPI for internal communications.

Acknowledgements

The authors are grateful to the Bayerische Forschungsstiftung (Germany), KONWIHR (Competence Network for Technical Scientific High-Performance Computing in Bavaria, Germany) and to the SIEMENS AG, Corporate Technology, for their financial support.

References

[1] **Borrmann, A., Wenisch, P., van Treeck, C., Rank, E.**
Collaborative Comp. Steering: Principles and Appl. in HVAC Layout, Integr. Computer-Aided Eng. (ICAE), 13(4): 361-376, 2006

[2] **Mulder, J. D., van Wijk, J., van Liere, R.**
A survey of computational steering environments, Future Gener. Comput. Syst., 15(1): 119-129, ISSN 0167-739X, Elsevier, 1999

[3] **Wenisch, P., Wenisch, O., Rank, E.**
Optimizing an interactive CFD simulation on a supercomputer for comp. steering in a virtual reality env., In: A. Bode and F. Durst (eds.): High Perf. Comp. in Sci. and Eng., pp. 83-93, Springer, 2005

[4] **Wenisch, P., Wenisch, O., Rank, E.**
Harnessing High-Perf. Comp. for Comp. Steering, Lect. Notes in Comp. Sci.: Recent Adv. in Parallel Virtual Machine and Message Passing Interf., 3666: 536-543, Springer, 2005

[5] **Wenisch, P.**
Computational Steering of CFD Simulations on Teraflop-Supercomputers, PhD thesis, Technische Universität München, to be submitted in 2007

[6] **Wenisch, P., van Treeck, C., Borrmann, A., Rank, E., Wenisch, O.**
Computational steering on distributed systems: Indoor comfort simulations as a case study of interactive CFD on supercomputers, Int. J. of Parallel, Emergent and Distr. Syst., 22(4): 275-291, 2007

[7] **van Treeck, C., Rank, E., Krafczyk, M., Toelke, J., Nachtwey, B.**
Ext. of a hybrid thermal LBE scheme for Large-Eddy sim. of turbulent convective flows, Comp. and Fluids, 35: 863-871, 2006

[8] **van Treeck, C., Wenisch, P., Borrmann, A., Pfaffinger, M., Wenisch, O., Rank, E.**
ComfSim – Interaktive Sim. d. therm. Komforts in Innenräumen auf Höchstleistungsrechn., Bauphysik, 29(1): 2-7, Ernst&Sohn, 2007

• Petra Wenisch[1]
• Christoph van Treeck[1]
• Leonhard Scheck[2]
• Ernst Rank[1]

[1] Chair for Computational Engineering in Civil and Environmental Sciences, TU München

[2] Leibniz-Rechenzentrum (LRZ)

Autumn 2007 • Vol. 5 No. 2 • inSiDE 21

Electron Paramagnetic Resonance Parameters of Semiquinone Bioradicals

Electron Paramagnetic Resonance (EPR) spectroscopy is a standard method of detecting and characterising paramagnetic species. In recent years, theoretical analyses of EPR parameters have become an important complement to experimental EPR, as improvements in method and increases in computational power have expanded the ability of theory to predict, analyse and confirm experimental results.

Semiquinone radical anions ($C_6R_4O_2^{\bullet-}$) are an important class of bioradicals which serve as electron transport agents in photosynthesis and respiration in all living things. They have been extensively characteried by EPR studies, and theoretical analyses of their EPR parameters go back decades. While initially concerned with deriving electronic spin density from experimental data, theoretical studies are now able to calculate the electronic orbitals ab initio and predict experimental data from them, with reasonable accuracy when dealing with main-group radicals. Most calculations on semiquinone radical anions have been limited to considering only the immediate environment of the molecule (the first solvation sphere, if in solution), and the effects of thermal motion have been ignored or estimated in a rough manner. These are approximations frequently made in computational chemistry, due to the computational costs involved in doing otherwise. Recently, however, supercomputing power has been used to simulate semiquinone behaviour more realistically: both by including a larger and more realistic solvation environment, and by explicitly calculating the motion of the system over time (a technique termed "Molecular Dynamics", or MD) at a quantum mechanical level of theory (Density Functional Theory, or DFT).

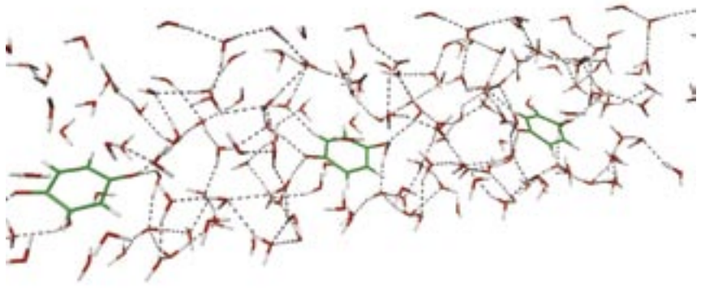


Figure 1: Benzosemiquinone radical anion in aqueous solution (periodic repetition shown in the x-direction)

The system thus treated was benzosemiquinone radical anion ($C_6R_4O_2^{\bullet-}$, or $BQ^{\bullet-}$), the simplest, most prototypical of the semiquinone bioradicals, included with 60 water molecules in a periodically-repeating box (see Figure 1). The interaction of BQ with solvent was closely studied, as both hydrogen bonding and the presence of a dielectric medium strongly affect the EPR parameters. Significantly, hydrogen bonding to $BQ^{\bullet-}$ is more extensive than was believed from smaller calculations. The MD simulations confirmed the theoretical prediction, and tentative experimental indication, of "T-stacking" hydrogen bonding to the C_6 ring, as these were found to occur frequently. The EPR parameters, both electronic g-tensors and hyperfine coupling data (which parameterize the interaction between electronic and nuclear spins), were calculated using DFT-based methods. By varying which water molecules were included in the calculations (e.g. none, or only those H-bonded to BQ, or only those within 4 Å, and so forth) and comparing the results, it was also possible to elucidate the effects of different effects on the EPR parameters. For instance, the effects of T-stacked hydrogen bonding alone, or of hydrogen bonding to oxygen, or both together, could be identified. The effects of the bulk solvent beyond the first solvation sphere were found to be significant, both for g-tensors and for hyperfine data.

Simulating the behaviour of $BQ^{\bullet-}$ over time allowed, for the first time, an assessment of how thermal motion affects the EPR data. One notable result was that several parameters display marked short-range (~30 fs) oscillations (seen in the plot of the x-component of the g-tensor in Figure 2). This arises because of the sensitivity of numerous parameters to the C-O bondlength; these periodic oscillations reflect the C-O bond-stretching vibrational motion. Thermal motion also has a small overall effect on the time-averaged EPR data, which can be seen by comparison with the results of static calculations. Together with the information on the solvation sphere, this should allow more informed use of static calculations using DFT methods to predict and analyse semiquinone EPR in the future. The next step is to extend this methodology to biologically important semiquinones, such as ubi- or plastosemiquinone. These are considerably larger than benzosemiquinone, and have long hydrocarbon side chains which make them lipophilic enough to traverse lipid membranes. Work has begun on simulating ubisemiquinone, a ubiquitous bioradical important in both respiration and photosystem II. Ubisemiquinone radical anion, or $UQ^{\bullet-}$, has an additional level of complexity in its behaviour due to the presence of methoxy side groups: these also attract hydrogen bonds, and their orientation affects the EPR parameters significantly.

These calculations were performed on the SR-8000 machine of the Leibniz-Rechenzentrum. The MD was performed with the CPMD code, a fast and well-parallelized code developed by J. Hutter et al. at IBM Zürich and the Max-Planck-Institute at Stuttgart. Property calculations were carried out using the Turbomole code of R. Ahlrichs et al. at the University of Karlsruhe, and the MAG-ReSpect code of V. Malkin et al. at the University of Würzburg and the Slovak Academy of Science in Bratislava.

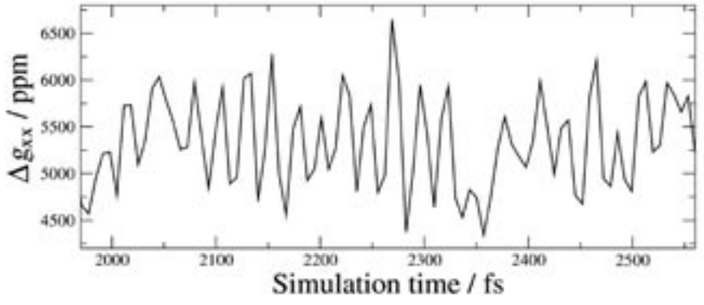


Figure 2: Plot of Δg_{xx} against time

• James Asher

Anorganische Chemie, Julius-Maximilians-Universität Würzburg

Massively Parallel Multilevel Finite Element Solvers on the Altix 4700

In the most recent TOP-500 list, published in June 2007, the HLRB II at the Leibniz Computing Centre of the Bavarian Academy of Sciences is ranked at position 10 for solving a linear system with 1,58 million unknowns at a rate of 56,5 Teraflops in the Linpack benchmark. However, this impressive result is of little direct value for scientific applications. There are few real life problems that could profit from the solution of a general dense system of equations of such a size.

Typical supercomputer applications today fall primarily in two classes. They are either variants of molecular dynamics simulations or they require the solution of sparse linear systems as they e.g. arise in finite element problems for the solution of Partial Differential Equations (PDEs). These two application classes become apparent when one reviews the history of the Gordon Bell Prize, the most prestigious award in high end computing. All Gordon Bell Prizes fall in either of these two categories. It is also interesting to see the correlation between architecture and application. For example, when the Earth Simulator, a classical vector architecture, was leading the TOP-500 list, the Bell Prize was awarded to applications with significant PDE content. More recently, the prize has been awarded for molecular dynamics-based applications, since this is the realm of the IBM/BlueGene systems that have been leading the list in the past few years. This, however, is not an indication that the interest in fast PDE solvers has declined, and therefore we will report here on our results for a massively parallel finite element solver for elliptic PDEs. The HLRB II system is an SGI-Altix that went into operation in September 2006 with 4,096 processors and an aggregate main memory of 17,5 Terabytes ("phase 1"). In April 2007 this system was upgraded to 9,728 cores and 39 Terabytes of main memory ("phase 2"). In particular in terms of available main memory, this is currently one of the largest computers in the world. Though the HLRB II is a general purpose supercomputer, it is especially well suited for finite element problems, since it has a large main memory and a high bandwidth. With this article we would like to demonstrate the extraordinary power of this system for solving finite element problems, but also which algorithmic choices and implementation techniques are necessary to exploit this architecture to its full potential.

The test problem reported in this article is a finite element discretization on tetrahedral 3D finite elements for a linear, scalar, elliptic PDE in 3D, as it could be used as a building block in numerous more advanced applications. We have selected this problem since it has a wide range of applications, and also, because it is an excellent test example for any high performance computer architecture.

Algorithms for very large Scale Systems

In this article we focus on multigrid algorithms [1,2], since these provide mathematically the most efficient solvers for systems originating from elliptic PDEs. Since multigrid algorithms rely on using a hierarchy of coarser Grids, clever data structures must be used and the parallel implementation must be designed carefully so that the communication overhead remains minimal. This is not easy, but our results below will demonstrate excellent performance on solving linear systems with up to 3 x 1,011 unknowns and for up to almost 10,000 processors.

Before we turn to the techniques that make this possible, we would like to comment on why we do not use domain decomposition methods that might be suggested as an alternative to using multigrid. In particular, it may be argued that it is easier to implement parallel domain decomposition efficiently than parallel multigrid, since it avoids the need of a Grid hierarchy. However, the price for the ease of implementation is a deterioration of the convergence rate that makes plain domain decomposition methods unsuitable on massively parallel systems. Of course, they can be improved by using an auxiliary coarse space within each iteration of the algorithm. In our view, however, with a coarse space, domain decomposition methods lose much of their advantage, since they will suffer from essentially the same bottleneck as multigrid.

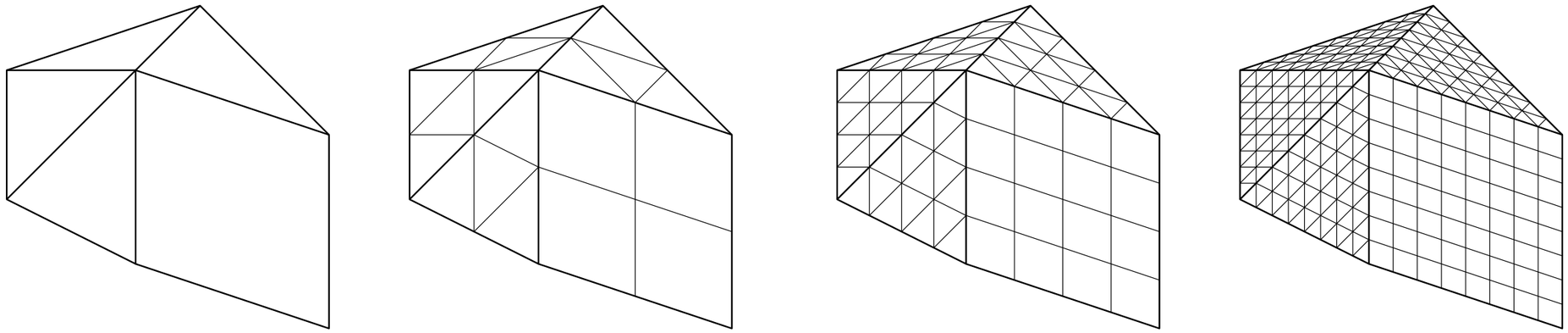
We wish to point out that the need for global communication is a fundamental feature of elliptic PDEs and so there is no hope to get around it by any algorithm. Multigrid methods seem to use the minimal amount of computation and also the minimal amount of communication that is necessary to solve the problem. Using a hierarchical mesh structure is essential not only to keep the operation count optimal, but also to keep for the amount of communication minimal. The difficulty is to organize and implement this efficiently. Using a mesh hierarchy can only be avoided if one is willing to pay by using more iterations. In total this may therefore lead to even more communication. We believe that the computational results below demonstrate that multigrid is the method of choice for solving extremely large PDE problems in parallel.

Hierarchical Hybrid Grids

HHG ("Hierarchical Hybrid Grids") [1,2,3] is a framework for the multigrid solution for finite element (FE) problems. FE methods are often preferred for solving elliptic PDEs, since they permit flexible, unstructured meshes. Among the multigrid methods, algebraic multigrid also supports unstructured Grids automatically. Geometric multigrid, in contrast, relies on a given hierarchy of nested Grids. On the other hand, geometric multigrid achieves a significantly higher performance in terms of unknowns computed per second. HHG is designed to close this gap between FE flexibility and geometric multigrid

Applications Autumn 2007 • Vol. 5 No. 2 • inSiDE

Figure 1: Regular refinement example for a two-dimensional input Grid. Beginning with the input Grid on the left, each successive level of refinement creates a new Grid that has a larger number of interior points with structured couplings.



performance by using a compromise between structured and unstructured Grids: a coarse input FE mesh is organized into the Grid primitives: vertices, edges, faces, and volumes that are then refined in a structured way as indicated in Figure 1. This approach preserves the flexibility of unstructured meshes, while the regular internal structure allows for an efficient implementation on current computer architectures, especially on parallel computers.

Parallelization

To exploit high-end computers, the programs must be parallelized using message passing. The HHG framework is an ideal starting point for this, since the mesh partitioning can be essentially accomplished on the level of the coarse input Grid, that is, with a Grid size that can still be handled efficiently by standard mesh partitioning software like Metis. Figure 2a shows a simple 2D example of such a Grid distribution. Two triangular elements are assigned to the two processes P0 and P1. The unknowns on the edge between the elements are coupled to both elements and are thus needed by both processes. This introduces communication (Figure 2b) and is equivalent to using ghost nodes, as is typical in parallel mesh algorithms. The edge data structure itself can be assigned to any one of the two processors.

In order to avoid excessive latency, the algorithmic details and communication must be designed carefully. The multigrid solver uses a Gauß-Seidel smoother that traverses the Grid points in the order of the primitives of the coarse input mesh: vertices - edges - faces - volumes. During the update of any such group, no parallel communication is necessary, because a vertex, for example, can only be connected to another vertex indirectly via an edge. This means that, rather than sending many small messages, each type of primitive can have its parallel dependencies updated as a single large message, which greatly reduces communication latency. However, a true Gauß-Seidel sweep traversing over the Grid points still requires too many communication steps during each iteration, since neighboring Grid points might belong to different processes. The current HHG implementation ignores a few such dependencies, thus giving the smoother the characteristics of a Jacobi Iteration at the affected points. Numerically, this leads to a slight deterioration of the convergence rate, but the reduction in execution speed more than outweighs this effect.

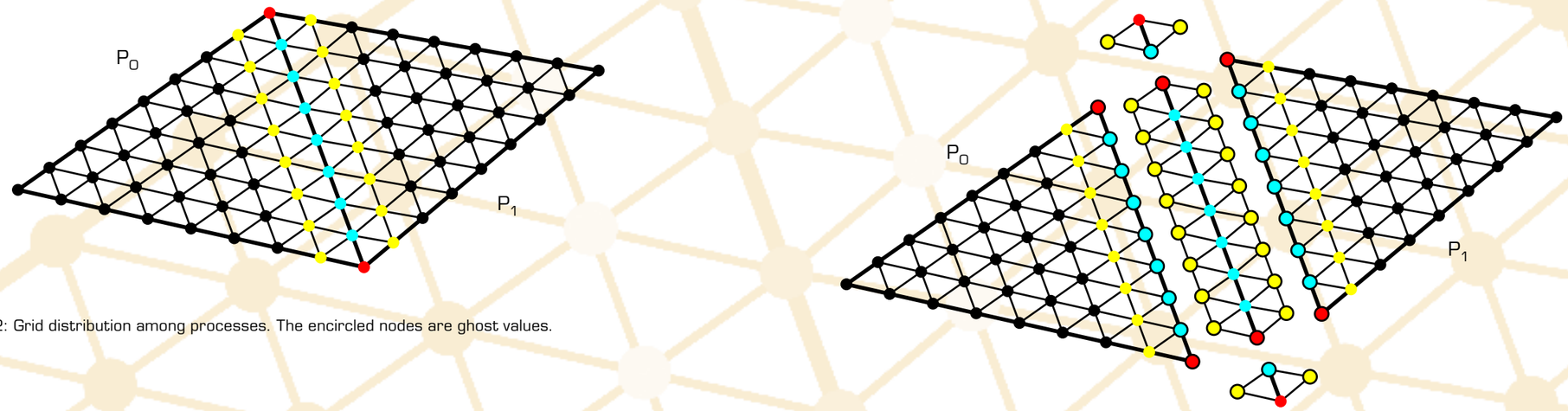


Figure 2: Grid distribution among processes. The encircled nodes are ghost values.

World Record in linear System Solving

In our lagest computation to date, we have used 9,170 cores of HLRB II and HHG to solve a finite element problem with 307 billion unknowns in 93 seconds run time. The problem itself is artificially designed by meshing a cube. This is necessary to ease the construction of problems with varying mesh size for our scalability study. However, the HHG data structures would allow for an arbitrary tetrahedral input mesh.

We believe that this is the largest finite element system that has been solved to date. Additionally, we point out that the absolute times to solution are still fast enough to leave room for using this solver as a building block in e.g. a time stepping scheme.

The results in Table 1 additionally show the results of a scaling experiment from 4 to 9,170 compute cores. The amount of memory per core is kept constant and the problem size is chosen to fill as much of the available memory as possible. If the program were perfectly scalable, the average time per V-cycle would stay constant throughout the table, because the ratio of problem size (i.e. workload) versus number of cores (i.e. compute power) stays constant. Near perfect scaling is seen as measure of the quality of an algorithm and its implementation. For the HLRB II in installation phase 1 the computation time increases only by a factor of 2.2 when scaling from 4 to 3,825 cores. This is still not perfect but in our view acceptable, especially when compared to other algorithms and especially in terms of the absolute compute time. Note that perfect scalability is the more difficult to achieve the faster a code is.

While every core of HLRB II phase 1 still had its own memory and network interface, the new dual-core configuration provides less bandwidth per core since two cores must share an interface. Additionally, a part of the installation is now configured as so-called “high density partitions” where two dual-core processors and thus four cores share one interface. Benchmark results including these high density partitions are marked with an asterisk in table 1. The timings for 64, 504 and 2,040 cores show that the dual-core processors of phase 2 account for approximately 39 % deterioration in runtime compared to phase 1. Scaling on the regular partitions shows a runtime increase from 4.93 s on 64 cores to 6.33 s on 6,120 cores. On the high density partitions, the runtime deteriorates to 7.06 s on just 128 cores but then increases only slightly further to 7.75 s for our largest runs.

Conclusions

The HHG framework and the HLRB II have been used to solve a finite element problem of world-record size, exceeding previous results by more than an order of magnitude, see [2,5]. HHG draws its power from using a multigrid solver that is especially designed and carefully optimized for current, massively parallel high performance architectures. The SGI Altix-architecture is found to be well-suited for large scale iterative FE solvers.

Number of cores	Number of unknowns ($\cdot 10^6$)	Average time per V-cycle (sec)		Time to solution ($\|\|r\|\| < 10^{-6} \|\|r_0\|\|$)	
		Phase 1	Phase 2	Phase 1	Phase 2
4	134.2	3.16	6.38 *	37.9	76.6 *
8	268.4	3.27	6.67 *	39.3	80.0 *
16	536.9	3.35	6.75 *	40.3	81.0 *
32	1,073.7	3.38	6.80 *	40.6	81.6 *
64	2,147.5	3.53	4.93	42.3	59.2
128	4,295.0	3.60	7.06 *	43.2	84.7 *
252	8,455.7	3.87	7.39 *	46.4	88.7 *
504	16,911.4	3.96	5.44	47.6	65.3
2,040	68,451.0	4.92	5.60	59.0	67.2
3,825	128,345.7	6.90		82.8	
4,080	136,902.1		5.68		68.2
6,120	205,353.1		6.33		76.0
8,152	273,535.7		7.43 *		89.2 *
9,170	307,694.1		7.75 *		93.0 *

Table 1: Scaleup results for HHG.
With a convergence rate of 0.3, 12 V-cycles are necessary to reduce the starting residual by a factor of 10–6. The entries marked with * correspond to runs on (or including) high density partitions with reduced memory bandwidth per core.

• Tobias Gradl
• Ulrich Rüde

Chair for System Simulation, University Erlangen-Nuremberg

References

[1] **Bergen, B., Gradl, T., Hülsemann, F., Rüde, U.**
A Massively Parallel Multigrid Method for Finite Elements, Computing in Science and Engineering, 8:56-62, 2006

[2] **Bergen, B., Hülsemann, F., Rüde, U.**
Is 1.7 x 1010 Unknowns the Largest Finite Element System that can be solved today? Proceedings of the ACM/IEEE Proceedings of SC, Seattle, Nov. 12-18, 2005, ISBN 1-59593-061-2

[3] **Hager, G., Bergen, B., Lammers, P., Wellein, G.**
Taming the Bandwidth Behemoth. First Experiences on a Large SGI Altix System. In: inSiDE, 3(2):24, 2005

[4] **Bergen, B., Wellein, G., Hülsemann, F., Rüde, U.**
Hierarchical Hybrid Grids – Achieving TERAFLOP Performance on Large Scale Finite Element Simulation, International Journal of Parallel, Emergent and Distributed Systems, 22(4):311-329, 2007

[5] **Adams, M. F., Bayraktar, H. H., Keaveny, T. M., Papadopoulos, P.**
Ultrascalable implicit finite element analyses in solid mechanics with over a half a billion degrees of freedom, ACM/IEEE Proceedings of SC 2004: High Performance Networking and Computing, 2004

Sustaining Tflop/s in Simulations of Quantum Chromodynamics

The first computer delivering a performance of more than one Tflop/s in the Linpack benchmark appeared on the TOP-500 list in June 1997 [1]. Sustaining a Tflop/s with a real application in everyday production runs is another story. The single CPU performance has to be good, the programme must scale sufficiently well, and it should not use the whole computer. For German quantum chromodynamics researchers sustaining one Tflop/s became reality with the installation of the current generation of supercomputers at LRZ and NIC, the SGI Altix 4700 and the IBM BlueGene/L. A third supercomputer offering comparable performance in single QCD applications is the apeNEXT at NIC/DESY Zeuthen.

Quantum chromodynamics (QCD) is the theory of strongly interacting elementary particles. The theory describes particle properties like masses and decay constants from first principles. The starting point of QCD is an infinite-dimensional integral. To deal with the theory on the computer space-time continuum is replaced by a four-dimensional regular finite lattice with (anti-) periodic boundary conditions. After this discretization the integral is finite-dimensional but rather high-dimensional. The high-dimensional integral is solved by Monte-Carlo Methods.

The basic building blocks of QCD are called quarks (matter particles) and gluons (particles mediating the interaction of quarks). The quark fields cannot be represented directly on a computer. In the computations they appear as large sparse matrices which describe systems of linear equations. QCD programmes spend most of their execution time in solving theses systems of linear equations. One research aim of the lattice QCD community is finding better algorithms by which less systems of equations have to be solved or which improve the convergence of iterative solvers. In any solver and an overall QCD programme the multiplication of the so-called hopping matrix with a vector is the dominant operation.

Sustaining Tflop/s in a QCD programme practically means sustaining Tflop/s in the hopping matrix multiplication. The prerequisite is a parallel computer with an excellent network. For example, in a Fortran/MPI implementation about 30 % of the compute time is needed for communication on the BlueGene/L.

How can one Tflop/s be sustained? At the single CPU level QCD programmes benefit from the fact that the basic operations involve small complex matrices. One can perform at the order of ten floating point operations per memory access. As a rule of thumb the resulting performance is about 20-25 % of peak when programming in Fortran or C. The single CPU performance can be considerably improved by employing low level programming techniques like assembler, multimedia streaming functions, or the BlueGene double hummer routines.

QCD programmes are parallelized by a domain decomposition. If one aims at one Tflop/s the domains become so small that their surface to volume ratio is at the order of one or even larger. This has the effect that a domain completely fits into a large data cache, which is the case on the Altix. On the other hand the data from the large domain surface has to be communicated to eight nearest neighbour processes. On the remote processes that data will not be in the cache but has to be fetched from main memory. Performance benefits from data caches but a substantial fraction of the data is never cached.

At the software level important optimization techniques are prefetching data from memory and overlapping communication and computation. These techniques could be used at a high level on LRZ's previous machine, the Hitachi SR8000-F1. On that machine prefetching instructions were inserted by the compiler which led to a single CPU performance of more than 40 % of peak for the hopping matrix multiplication implemented in Fortran. In a hybrid programming approach, employing Fortran, OpenMP and MPI, one could achieve that communication and computation overlap. The resulting parallel performance was 30-40 % of peak [2].

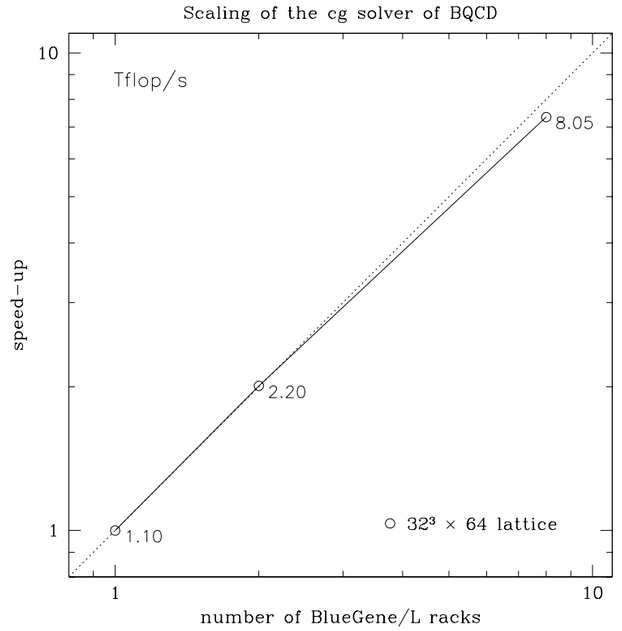


Figure 1: Strong scaling and overall performance of the conjugate gradient kernel with the assembler version of the hopping matrix multiplication on the BlueGene/L. The dotted line indicates linear scaling.

On the BlueGene and the Altix hiding communication is not so straightforward to implement. On the eight-way SMP nodes of the SR8000 one could use one thread for communication while seven threads compute resulting in 12,5 % communication overhead. Using one of the two cores of the BlueGene or Altix processors for communication would produce 50 % communication overhead. However, on both machines lower level techniques are available by which one can try to hide communication overhead.

In our code BQCD (Berlin quantum chromodynamics programme) the hopping matrix multiplication was implemented in assembler for the BlueGene and the Altix [3].

On the BlueGene the network can be directly accessed using special load/store instructions. In the course of the computation each node needs to receive part of the data from the boundary of its neighbouring nodes, and likewise it has to send part of the data from its boundary to neighbouring nodes. In order to hide communication latency the assembler kernel always looks ahead a few iterations and sends data that will be needed by a remote node. When a CPU needs data from another node, it polls for arriving data packets. We can see that communication and computation really overlap by studying strong scaling: when going from one to eight racks the speed-up is 5.3 for the Fortran/MPI implementation but it is 7.3 for the assembler code (see figures). Optimizing computations alone would have decreased the speed-up of the Fortran/MPI programme because the communication part would be unchanged.

Scaling of the cg solver of BQCD

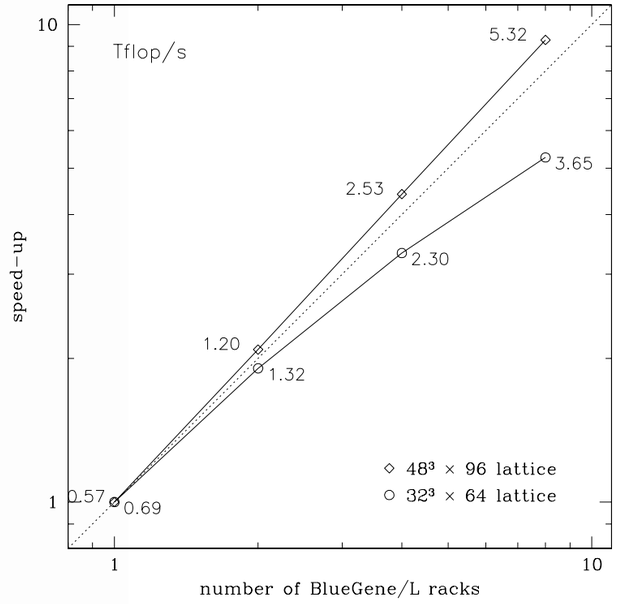


Figure 2: Strong scaling and overall performance of the conjugate gradient kernel with the Fortran/MPI version of the hopping matrix multiplication on the BlueGene/L. The 32^3 x 64 lattice is typically used today. The larger lattice will be interesting in the future. It displays super-linear scaling.

On the Altix a promising method for hiding communication latency is using Altix Shmem-pointers by which one can access remote memory directly without any function calls. The idea is to directly write to remote memory in the course of computations similar to the approach taken on the BlueGene. We used Shmem-pointers in a Fortran/C and an assembler implementation. Unfortunately there was no gain in both cases. Nevertheless, re-writing the hopping matrix multiplication in assembler improved the overall performance by about 30 %.

In production runs typically one BlueGene rack (2,048 cores) is used and a performance of 1,1 Tflop/s or 19 % of peak is sustained. On the Altix almost 1,5 Tflop/s or 23 % of the peak performance are measured when using 1,000 cores. Technically speaking these values were obtained for the whole conjugate gradient solver in a lattice QCD formulation with clover improved Wilson fermions employing even/odd preconditioning. Other groups achieve similar performance figures with their implementations. In other words people sustain one Tflop/s or more on one eighth of the BlueGene/L at NIC or an even smaller part of the Altix at LRZ which makes the Tflop/s available as the normal sustained performance of QCD simulations.

References

[1] www.top500.org

[2] Schierholz, G., Stüben, H.
Optimizing the Hybrid Monte Carlo Algorithm on the Hitachi SR8000, in Wagner, S., Hanke, W., Bode, A., Durst, F. (Eds.), High Performance Computing in Science and Engineering, Munich, 2004, Springer-Verlag

[3] Streuer, T., Stüben, H.
Simulations of QCD in the Era of Sustained Tflop/s Computing, Contribution to Parallel Computing 2007 (ParCo2007), Aachen and Jülich, September 4-7, 2007 (in preparation)

• Hinnerk Stüben

Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB)

Applications

Ab Initio Simulations of functional magnetic Materials

SFB 445
Nanoparticles from the Gasphase:
Formation, Structure, Properties

MSM
MAGNETIC SHAPE MEMORY
A DFG PRIORITY PROGRAMME

The recent evolution of supercomputing power allows for an ab initio treatment of systems in the nanometer size regime. This comes at hand, where the technological progress requires an increasing miniaturization of functional units. Often, atomistic simulations of materials properties on realistic length scales can easily be carried out by classical molecular dynamics simulations using empirical model potentials or hybrid methods which permit simulations in the mesoscopic regime. In some cases, however, a full quantum-mechanical approach is necessary to obtain an accurate description of the electronic structure.

A prominent example is the quest for ultra-high density magnetic recording media. Here, the exponential increase in storage density over time still seems unbroken – current lab demos reach values around 400 GBit/in^2 while the expectations of the manufacturers go up to 10 to 50 TBit/in^2 in the future. For this, it will be necessary to employ self-assembled patterned media of new materials like Fe-Pt and Co-Pt, which are characterized by an extremely large Magnetocrystalline Anisotropy Energy (MAE). This might allow bit sizes of 3-4 nm in diameter in the far end without compromising information stability by thermal fluctuations [1].

Another example of equal technological impact is the search of novel magnetically driven actuator materials, which would allow for, e.g., simple microscopic devices without directly attached power supply. In the prototype Magnetic Shape Memory (MSM) alloy Ni_2MnGa, magnetic field induced strains of up to 10 % have been achieved [2]. Again, the unusually large MAE of this materials plays a crucial role - in connection with an extremely high mobility of martensitic twin boundaries. The physical picture is that a change of the direction of the magnetic field will cause a growth of the martensitic variants possessing an easy axis with proper alignment to the external field at the expense of the others, leading to an overall change of crystal shape. The nature of this process is not fully resolved yet, thus detailed understanding of the origin of the high mobility of the twin boundaries will be a prerequisite for the systematic development of novel actuator materials with improvements concerning their environmental working range and their brittleness which are pivotal for final applications. The ground state properties of the electronic system can be determined from first principles within the density functional theory, by solution of the Kohn-Sham equations which allow a unique description of the total energy by a functional of the electronic charge density distribution (for an introduction, see, e.g., [3]).

The Kohn-Sham equations are usually solved iteratively within a self-consistency approach. The Vienna Ab-Initio Simulation Package (VASP) [4], which we employ for our calculations, uses a plane wave basis set for the description of the electronic wavefunctions of the valence electrons and the projector augmented wave approach for the interaction with the nuclei and the core electrons. In each self-consistency cycle, the eigenvalue problem has to be solved employing the ScaLAPACK eigensolver and Fast Fourier Transforms between real and reciprocal space. From this, the forces acting on the ions can be derived, which are subsequently used within a conjugate gradient energy minimiziation scheme for structure optimization or a dynamical molecular dynamics simulation of the ion motion. The VASP code was easily adapted to the BlueGene/L architecture and is capable of handling large systems.

For example, for a system of 561 transition metal atoms with ~4,500 valence electrons reasonable scalability (~70 % of the ideal performance) was achieved using 1,024 compute-nodes [5] on the BlueGene/L in Jülich (JUBL). For larger systems, the code has been successfully tested on up to 8 racks (8,192 compute nodes), calculating the electronic ground state of a system with 8,000 valence electrons [6], yielding a reasonable performance on up to 4 racks. A typical electronic self-consistency step requires around 16 CPU-h in the case of Fe_{561} nano-particles; several thousands of these steps are needed to obtain a sufficiently optimized geometric configuration of the ions. An estimate of the computational power needed for such a structural optimization is given in Figure 1 for different system sizes.

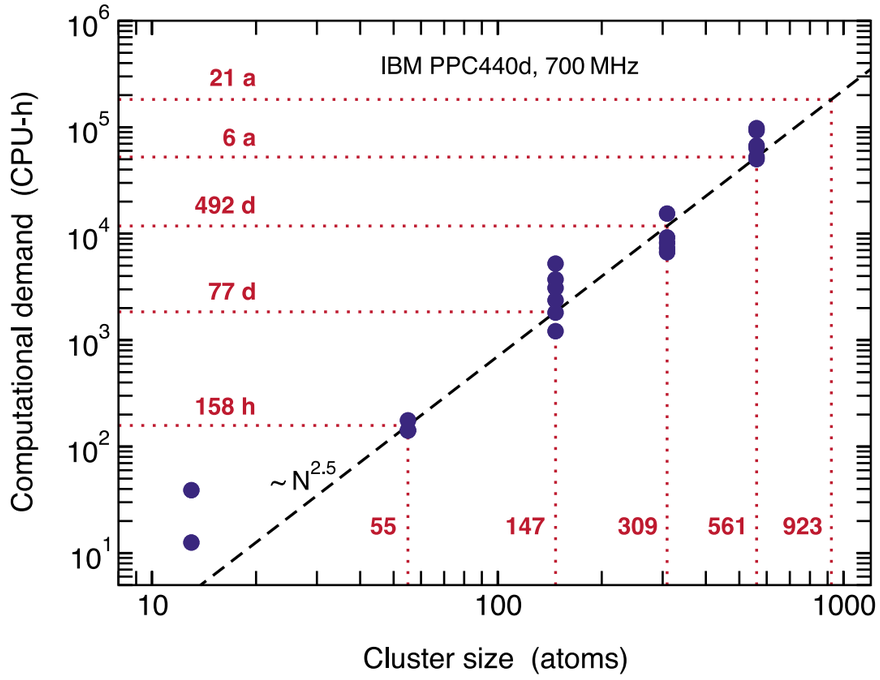


Figure 1: Log-log plot of the computation time per CPU (IBM PPC440d, 700 MHz) needed for a full geometric optimization of Fe-Pt nanoclusters of various sizes on JUBL

A naive fit of a power law yields an increase of the computation time with the system size by an exponent of about 2.5. Since the code appears to work well together with the communication hardware of the BlueGene/L, the most severe limitation is the restricted main memory of the JUBL nodes, which prevents the efficient treatment of much larger systems, which otherwise appears feasible from our previous experience. A partly remedy is that exchange of binary files is possible between the BlueGene/L and the IBM p690 at FZ Jülich (JUMP), so that wavefunction data can be transferred from JUBL to JUMP to continue particularly memory consuming single point calculations, e.g., the determination of non-collinear spin structures or magnetocrystalline anisotropy energies, allowing the exploitation of the advantages of both machines in certain cases.

Our initial calculations on JUBL were dealing with a longstanding problem from fundamental science, i.e., the size dependent evolution of morphologies of elemental iron nanoclusters. Here, we were able to identify a previously unreported structure, possessing a face centreed cubic (fcc) core, while retaining an icosahedral outer shape [7]. For clusters of 55 atoms, this so-called Shellwise Mackay-Transformed (SMT) morphology has the lowest energy while for cluster sizes above around 150 atoms the body centreed cubic (bcc) structure predominates, which is also the ground state of bulk iron. However, the SMT clusters remain, unlike icosahedra and the fcc cuboctahedra, within the range of thermal energies to the bcc isomers and may thus be encountered in the formation of small Fe nanoparticles at finite temperatures. The relative stability of SMT structures can be traced back to a bcc-like co-ordination of the atoms in the outermost shells (cf. Figure 2).

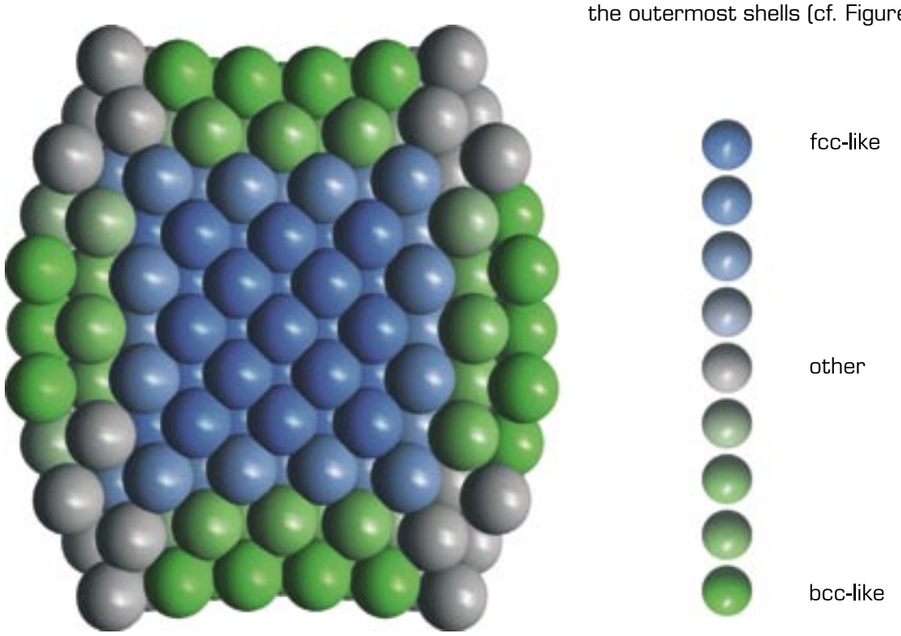


Figure 2: Cross section of a SMT nanoparticle of elemental iron (561 atoms). The color coding describes the local co-ordination of the atoms obtained by a common neighbor analysis.

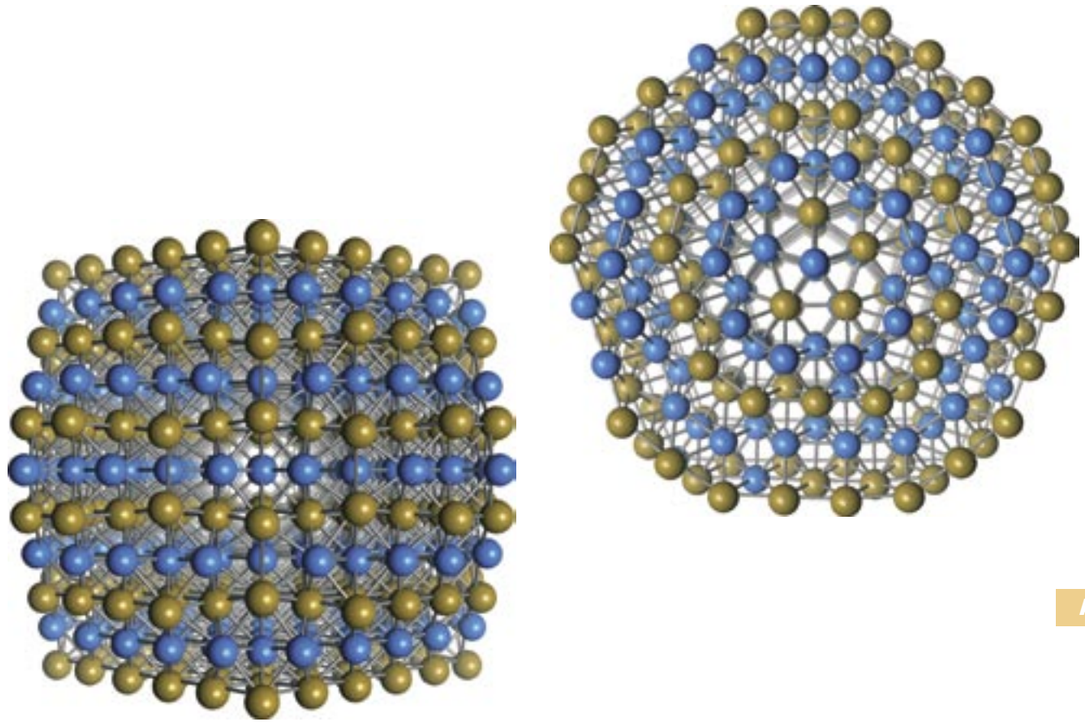


Figure 3: Examples of two morphologies of $Fe_{265}Pt_{296}$ nanoparticles (~2.5 nm in diameter) Left: $L1_0$ ordered cuboctahedron with alternating Fe (blue) and Pt (orange) layers along the [001] direction, which is expected to yield a large MAE. Right: Cross-section of an icosahedral isomer perpendicular to one of its five-fold symmetry axes with radially alternating Fe- and Pt-rich shells, which is the energetically most favorable structure found for this system size.

The computational power provided by JUBL enabled us to compare various morphologies of Fe-Pt nanoparticles of up to 561 atoms corresponding to diameters of about 2.5 nm, which is sufficiently close to the technologically important size range to obtain important trends. Figure 3 shows two different morphologies of $Fe_{265}Pt_{296}$ nanoparticles. On the left is the $L1_0$ ordered cuboctahedron, which provides, due to the alternation of Fe and Pt layers in [001]-direction the largest possible magnetocrystalline anisotropy energy. The isomer on the right is an icosahedron with shellwise alternating Fe- and Pt-rich layers which is about 30 meV/atom lower in energy than the cuboctahedron and is also the lowest energy structure found in our simulations. From our results, multiply twinned morphologies as icosahedra and decahedra are generally more favorable than $L1_0$ cuboctahedra in the investigated size range. Multiply twinned structures, however, do not provide a comparable MAE, even if perfectly ordered, due to the different crystallographic orientations of the individual twins. Our findings correspond well to experimental observations reporting difficulties to obtain Fe-Pt nanoparticles with a sufficiently large magnetocrystalline anisotropy at diameters of 4 nm and below. For an accurate determination of crossover sizes and material specific quantities as surface and twinning energies, calculations of larger particles are projected.

Our calculations concerning the MSM effect are aiming at gaining a microscopic picture of the close interrelation between electronic structure, especially magnetism, and phase stability, which we expected to play a decisive role for the high mobility of the martensitic twin boundaries in Ni-Mn-Ga alloys. As a first step, we studied in a quasistatic approach the shear-induced motion of an ideal twin boundary between two variants of the martensitic $L1_0$-phase. Figure 4 shows five snapshots of a 512 atom system of the off-stoichiometric $Ni_{2.2}Mn_{0.8}Ga$ alloy before and while applying shear on the periodic supercell. With our calculations, we can monitor the energy landscape of this process and changes in the electronic structure which will help us to understand the origin of the high twin boundary mobility occurring in particular martensitic phases at certain compositions. The long term goal is to additionally include lattice defects and to simulate magnetic field induced twin boundary motion. This necessitates larger systems, the handling of non-collinear magnetism and spin-orbit interactions and thus implies considerably increased computation time and memory requirements. With its improved communication system and its extended main memory, the new BlueGene/P installation in Jülich could provide an important step towards its realization.

Acknowledgement

We thank Dr. Pascal Vezolle (IBM) for his help in improving the performance of the VASP code on the BlueGene/L and Inge Gutheil (ZAM, FZ Jülich) for providing an optimized ScaLAPACK library. Financial support was granted by the Deutsche Forschungsgemeinschaft through SPP 1239 (Change of microstructure and shape of solid materials by external magnetic fields) and SFB 445 (Nanoparticles from the gasphase: Formation, structure, properties).

References

[1] **Sun, S., Murray, C. B., Weller, D., Folks, L., Moser, A.**
Monodisperse FePt nanoparticles and ferromagnetic FePt nanocrystal superlattices, Science 287, 1989 (2000)

[2] **Ullakko, K., Huang, J. K., Kantner, C., O'Handley, R. C., Kokorin, V. V.**
Large magnetic-field-induced strains in Ni2MnGa single crystals, Appl. Phys. Lett. 69, 1966 (1996)

[3] **Kohn, W.**
Electronic Structure of Matter – Wave Functions and Density Functionals, Nobel Lecture, January 28, 1999

[4] **Kresse G., Furthmüller, J.**
Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54, 11169 (1996); Kresse, G. and Joubert, D.: From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B 59, 1758 (1999); http://cms.mpi.univie.ac.at/vasp/

[5] **Gruner, M. E., Rollmann, G., Hucht, A., Entel P.**
Massively parallel density functional theory calculations of large transition metal clusters, Lecture Series on Computer and Computational Sciences 7, 173, 2006

[6] **Frings, W., Hermanns, M.A., Mohr, B., Orth, B. (Eds.)**
Report on the BlueGene/L Scaling Workshop 2006, Technical Report FZJ-ZAM-IB-2007-2, 2007

[7] **Rollmann, G. and Gruner, M. E., Hucht, A., Meyer, R., Entel, P., Tiago, M. L., Chelikowsky, J. R.**
Shellwise Mackay transformation in iron nano-clusters, Phys. Rev. Lett. 99, 083402, 2007

• Markus E. Gruner
• Peter Entel

Fachbereich Physik, Universität Duisburg-Essen

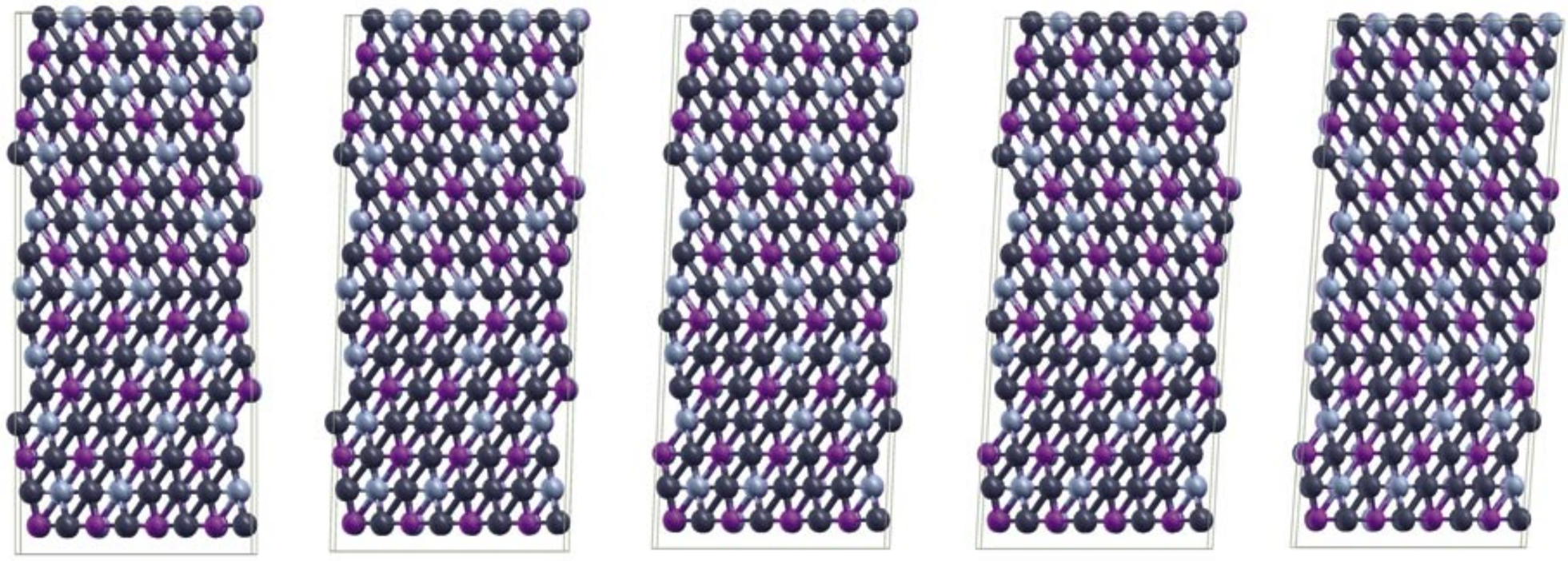


Figure 4: Sequence of snapshots of 512-atom super-cells containing two variants of the martensitic $L1_0$ phase of the off-stoichiometric $Ni_{2.2}Mn_{0.8}Ga$ alloy (Ni: black; Mn: blue; Ga: purple) before (left) and after (right) shear-induced coherent twin boundary motion. This was realized in a quasi-static approach by shearing the simulation cell in small steps and subsequent relaxation of the atomic positions. The twin boundary, which is initially located in the centre of the cell, is moving downwards during this process.

Towards an intelligent Grid for Business – the BREIN Project

BREIN
Business objective driven REliable and Intelligent grids for real busiNess

http://www.Gridsforbusiness.eu

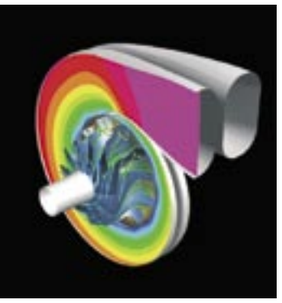
In today's world, enterprises, independent of their size, have to co-operate closely with other companies to keep their competitiveness, as no company is capable of fulfilling all requirements alone. But setting up these collaborations is still difficult and extremely costly, difficult to realize and manage, and often highly risky for all involved parties. With the increasing complexity of modern day market requirements, it becomes more and more difficult for individual entities, particularly small to medium enterprises, to compete cost-efficiently with the major market players. With this in mind, the Grid and web service community developed the notion of so-called "Virtual Organizations": collaborations between organizations that expose their capabilities via standardized interfaces to the internet, thus enriching the individual's capabilities beyond its economic limits.

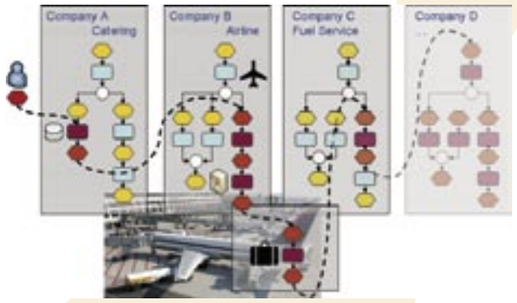
BREIN

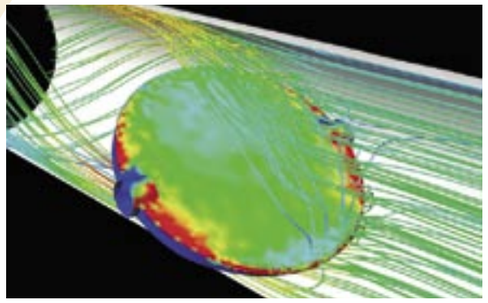
The primary aim of the BREIN integrated project consists in enabling business experts and average users equally to use the enhanced capabilities of Grid technologies to improve their business, be it as a consumer or a provider of services exposed over the internet.

From a service provider perspective, this means specifically to enable business entities to wrap up their capabilities and existing products/solutions in a way that allows their secure, protected exposure. In order to fully support business administrators, the BREIN framework will provide the means to not only technically enable users to expose their capabilities, but also to reduce the general efforts in the tasks involved in setting up and maintaining such an infrastructure. This ranges from interpreting the requirements so as to map them to the configuration of the infrastructure, to actual management of this setup with respect to the business goals and business policies of the provider.

As opposed to this, a customer will require to make use of the capabilities of a resource provider with as little effort as possible, so as to enable him/her to realize complex, multiparty processes without having to have explicit business planning expertise. The BREIN framework will thus bring in the means to simplify this task by providing a framework that can analyse the business goal under the restrictions of the individual participants' business policies and the (available) capabilities of the resource providers.







The full BREIN infrastructure will thus provide a framework that will make it easier for its users to exploit resources across the internet with a maximum of support for optimizing the capabilities whilst maintaining constraints of all parties involved.

In order to achieve the project's ambitious goals, BREIN will combine Grid & Web Service technologies with the capabilities of Multi-Agent Systems, Semantic Web and Artificial Intelligence. As will be elaborated in more detail below, such a combination will result in an enhanced messaging infrastructure that allows for realizing the extended business requirements.

Keeping in mind that such a system will be used by a human user, this cannot be restricted to technical requirements only. The consortium, therefore, analyzes real world business requirements with the exemplary support of two selected test scenarios in the areas of airport management and distributed computing, thus covering a wide area of different applications. The above mentioned technologies are examined in detail with respect to a) their capabilities to meet the requirements and b) in how far they can be combined with each other so as to realize an integrated middleware that is simple to use. The project will thereby make use of existing results of the respective technologies to a maximum degree to avoid unnecessary "reinventions of the wheel".

The main outcome of BREIN will hence consist in a common framework that not only shows how such a middleware can be realized, but also how technologies can be combined in order to mutually enhance the individual area's capabilities. BREIN will furthermore realize this fusion of technologies by providing a reference implementation of the framework. This reference implementation will allow demonstrating the capabilities of the middleware in the context of selected testbed scenarios and will provide the according communities with new technologies in their respective areas.

To guarantee and demonstrate the usability of the BREIN platform in real business, two scenarios, one in the area of commercial oriented High Performance Computing (Virtual Engineering Virtual Organization) and the other one in the area of optimization of logistics in an Airport were chosen to deliver realistic requirements for the design of the framework.

Project Facts

The BREIN project has started in September 2006 and is realized by a consortium of 16 partners across the Europe with an overall budget of M 9,9 €. HLRS is in this project taking over the role of Technical Management.

© Flughafen Stuttgart, 2006

• Bastian Koller

Höchstleistungs-
rechenzentrum
Stuttgart

The D-Grid Integration Project

http://dgi.d-Grid.de

In 2003 some German scientific institutions founded the German Grid Initiative D-Grid with the goal to establish in Germany tools and resources for the new paradigm of Grid computing. Since 2005 the German Federal Ministry of Education and Research (BMBF) is funding the D-Grid initiative generously. In September 2005 the first BMBF-funded Grid projects in Germany were launched. The D-Grid integration project (DGI) belongs to the first group of projects of the D-Grid Initiative, it was started also in September 2005. DGI provides the important infrastructure components that are shared among the various commercially or scientifically oriented disciplines participating in D-Grid. Therefore, it supports those disciplines in establishing their own community-specific Grid infrastructure. In addition, DGI will integrate those infrastructure components that are developed by a community project and are of interest to other community Grids as well. Finally, it is anchor to additional so-called gap and service Grid projects that provide additional services that may be used by the community Grids. DGI is a large project which includes 22 partner institutions from nearly all larger German scientific organizations.

Internally, the project is divided into four work packages that interact with each other:

Work Package 1
D-Grid base-software provides support for various Grid middleware and data management software components. Those components have been selected based on the results of pre-project working groups that included participants from many disciplines. They include the Globus Toolkit V4, UNICORE, and gLite/LCG as middleware tools already in use at many installations, GridSphere as general portal framework, the Grid Application Toolkit (GAT) as a set of generic and flexible APIs for accessing Grid services, SRM/dCache, OGSA/DAI, SRB and Datafinder as components to access data that are organized in various ways. Further, VO management concepts are developed in this work package and have resulted in a separate project IVOM. The partners in this work package support the community Grids by providing tutorials to learn about the components and appropriate installations packages. Further, the components are adapted to the needs of the communities or the developers of the communities are supported in customizing those tools to fit their requirements.

Work Package 2
This Work package builds up a Core-D-Grid for proving the infrastructure. This will be used as a prototype to test the operational functionality of the system. This work package also deals with the challenges of Monitoring, Accounting and Billing of the Grid resources.

Work Package 3
The network infrastructure in D-Grid is based on the DFN Wissenschaftsnetz X-WiN. Work package 3 will provide extensions to the existing network infrastructure according to the needs of Grid middleware used in D-Grid. Further tasks are to build an Authentication and Authorization Infrastructure (AAI) in D-Grid, develop firewall concepts which are essential for building and operating a secure D-Grid network infrastructure and set up Grid specific CERT services by Computer Emergency Response Teams.

Work Package 4
The purpose of this work package is the integration of the deliverables from the Grid integration project and the deliverables from the different community projects in one common D-Grid platform. Work package 4 also deals with the challenge of archiving sustainability in D-Grid and Grid-based e-Science systems generally. The longterm and sustainable usage of a national or international Grid infrastructure for e-Science is a huge and cost-intensive challenge, just like the installation of such an infrastructure. Legal requirements like legal frameworks and their consideration by technical implementations, software licenses, privacy, adequate business models and security are essential for achieving sustainability of emerging Grid infrastructures.

In co-operation with some community projects the DGI has established a reference installation to help institutions to integrate their resources into the Grid. The reference installation covers the three middleware flavours used in D-Grid (Globus Toolkit V4, UNICORE and gLite/LCG) as well as some data access mechanisms like SRM/dCache and OGSA/DAI. In 2006 and again in 2007 significant amounts of additional resources have been funded by the Federal Ministry of Education and Research (BMBF). So it became possible to install and integrate a Germany-wide Grid infrastructure, consisting of thousands of CPUs at 24 locations and more than one Petabyte of distributed disk space. Figure 1 shows the distribution of the D-Grid resources on a map of Germany. These generous resources can now be used in a shared mode by the community projects joined under the roof of D-Grid.

Figure 1: D-Grid resources in Germany

Based on the feedback obtained by the communities, a second phase of the project has been proposed. This second phase will continue and intensify the support for the communities over the next three years, and it will build a powerful Core-Infrastructure that will become an important part of a sustainable D-Grid world consisting of many different Community Grids. This long-term and sustainable D-Grid world should provide Grid services to many Grid communities from science and industry in Germany.

• Klaus-Peter Mickel
(DGI co-ordinator)

Forschungs-
zentrum Karlsruhe

Remote Visualization at the LRZ

http://www.lrz-muenchen.de/services/compute/visualization

The LRZ has acquired a powerful graphics server, which is to be used for the visualization of large data sets produced by simulations that are running on the HLRB II and the Linux cluster. This visualization system is based on a SUN x4600 server with 8 dual-core Opteron CPUs, 128 GB RAM and a local RAID array with a capacity of 3 TB. A 10 Gbit ethernet network interface and a direct connection to the CFXS file system of the HLRB II allow for a fast access on the simulation results. The graphics capabilites are provided by two nVidia Quadroplex units, which are connected via PCIexpress and contain two high-performance Quadro FX5500 graphics cards each, so that in total 4 GB of graphics memory are available.

In contrast to conventional solutions, the high graphics power is also available to users working remotely on the new system. This feature may not appear to be very advanced, as it has been possible for many years to redirect the graphical output of a program running on one machine (the server) over the network to the display connected to another computer (the client). In fact, this "network transparency" is one of the basic features of the X window system used on all Unix-like operating systems. Furthermore, there are other solutions like VNC that provide a similar functionality also for Microsoft Windows and other operating systems. However, in the case of 3D visualization applications enormous amounts of data have to be transferred over the network in this approach. The total size of the stream of graphics commands needed to display one frame can become similar to the size of the data that is visualized, e.g. to produce a volume rendering of a $1{,}024^3$ data set at least one gigabyte of data has to be transferred from the server to the client. For a 100 MBit/s network connection this would take about two minutes per frame, which does not allow for interactive frame rates (and besides, in general the client graphics card will not be able to handle such amounts of data).

Figure 1: The new remote visualization system of the LRZ, a 16 core Sun x4600

A solution for this problem is provided by the Sun Shared Visualization software package, which consists mainly of the open-source projects VirtualGL and TurboVNC. These tools intercept the 3D graphics commands issued by OpenGL applications on the server (normally these commands would be sent to the remote users' machine). The intercepted commands are then executed on the powerful server graphics cards and the resulting images are compressed and sent to the users' computer. Instead of huge amounts of 3D data only compressed 2D images have to be transfered over the network. Therefore a 100 MBit/s ethernet connection is sufficient to allow for a image quality that is nearly indistinguishable from the one experienced by local users. For slower connections down to several MBit/s the image quality is reduced, but still acceptable.

Owing to the fact that the system is based on standard x86_64 technology a broad variety of visualization applications is available. At the moment the visualization applications Amira, Ensight, ParaView and Visit are supported on the graphics server. More specialized application can be installed on user request. The software installed on the Linux cluster will also be available on the visualization server. Users of the HLRB II and the Linux cluster can apply for an account on the remote visualization server using the form on this web page, which contains also the system documentation:

http://www.lrz-muenchen.de/services/compute/visualization

• Leonhard Scheck

Leibniz-Rechenzentrum (LRZ)

Systems

IBM BlueGene/P in Jülich: The Next Step towards Petascale Computing

When in 2004/2005 the IBM BlueGene technology became available, Research Centre Jülich (FZJ) recognized the potential of this architecture as a Leadership-class system for capability computing applications. A key feature of this architecture is its scalability towards PetaFlop Computing based on low power consumption, small footprint and an outstanding price performance ratio.

In early summer 2005 Jülich started testing a single BlueGene/L rack with 2,048 processors (inSiDE Vol. 3, No. 2, p. 18). It soon became obvious that many more applications than expected can be ported to efficiently run on the BlueGene architecture. Due to the fact that the system is well balanced in terms of processor speed, memory latency and network performance, many applications scale excellently up to large numbers of processors. Therefore in January 2006 the system was expanded to 8 racks with 16,384 processors, funded by the Helmholtz Association.

The 8-rack system has successfully been in operation for almost two years now. Today about 30 research projects, which were carefully selected with respect to their scientific quality, run their applications on the system using job sizes between 1,024 and 16,384 processors. During a BlueGene Scaling Workshop at FZJ experts from Argonne National Laboratory, IBM and Jülich helped to further optimize some important applications. It could be shown that all these applications succeeded in efficiently using all 16,384 processors of the machine.

Computational scientists from many research areas took the chance to apply for significant shares of BlueGene/L computer time to tackle unresolved questions which were out of reach before. Because of the large user demand and in line with its strategy to strengthen Leadership-class computing, Research Centre Jülich decided to order a powerful next-generation BlueGene system. In October 2007 a 16-rack BlueGene/P system with 65,536 processors was installed mainly financed by the Helmholtz Association and the State of North Rhine Westphalia. With its peak performance of 222,8 TFlop/s, Jülich's BlueGene/P – dubbed JUGENE – is currently the biggest supercomputer in Europe.

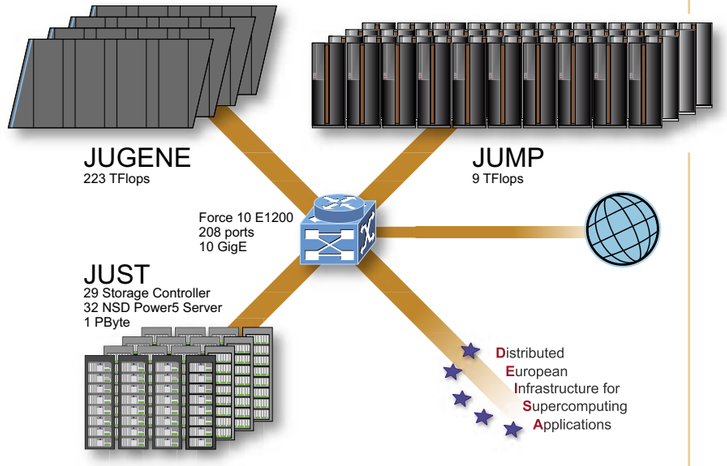
	BlueGene/L	BlueGene/P
Node Properties		
Processor	PowerPC® 440	PowerPC® 450
Processors per node (chip)	2	4
Processor clock speed	700 MHz	850 MHz
Coherency	Software managed	SMP
L1 cache (private)	32 KB per core	32 KB per core
L2 cache (private)	7 stream prefetching 2 line buffers/stream	7 stream prefetching 2 line buffers/stream
L3 cache (shared)	4 MB	8 MB
Physical memory per node	512 MB	2 GB
Main memory bandwidth	5,6 GB/s	13,6 GB/s
Peak performance	5,6 GFlop/s	13,6 GFlop/s
Torus Network		
Bandwidth	2.1 GB/s	5.1 GB/s
Hardware latency (nearest neighbour)	200 ns (32B packet) 1.6 µs (256B packet)	160 ns (32B packet) 1.3 µs (256B packet)
Global collective Network		
Bandwidth	700 MB/s	1,700 MB/s
Hardware latency (round trip worst case)	5.0 µs	3.0 µs

Table 1: BlueGene/L vs. BlueGene/P

The important differences between BlueGene/P and BlueGene/L largely concern the processor and the networks (see Table 1) while the principal build-up of BlueGene/L was kept unchanged. Key features of BlueGene/P are:

- 4 PowerPC® 450 processors are combined in a fully 4-way SMP (node) chip which allows a hybrid programming model with MPI and OpenMP (up to 4 threads per node).
- The network interface is fully DMA (Direct Memory Access) capable which increases the performance while reducing the processor load during message handling.
- The available memory per processor has been doubled.
- The external I/O network has been upgraded from 1 to 10 Gigabit Ethernet.

These improvements are also reflected by the application performance. A code from theoretical elementary particle physics, for example, on BlueGene/P runs at 31,5 % of the peak performance compared to 26,3 % on BlueGene/L. Furthermore, the increased memory of 2 GB per node will let new applications run on BlueGene/P.



JUGENE is part of the dual supercomputer complex in Jülich, embedded in a common storage infrastructure which was also expanded. Key part of this infrastructure is the new Jülich storage cluster (JUST) which was installed in the third quarter of 2007, increasing the online disk capacity by a factor of ten to around one PetaByte. The maximum I/O bandwidth of 20 GB/s is achieved by 29 storage controllers together with 32 IBM Power 5 servers. JUST is connected to the supercomputers via a new switch technology based on 10 Gigabit Ethernet. The system takes on the fileserver function for GPFS (General Parallel File System) and provides service to the clients in Jülich and to the clients within the international DEISA infrastructure as well.

With the upgrade of its supercomputer infrastructure FZJ has taken the next step towards Petascale Computing and has strengthened Germany's position to host one of the future European supercomputer centres.

• Michael Stephan
• Klaus Wolkersdorfer

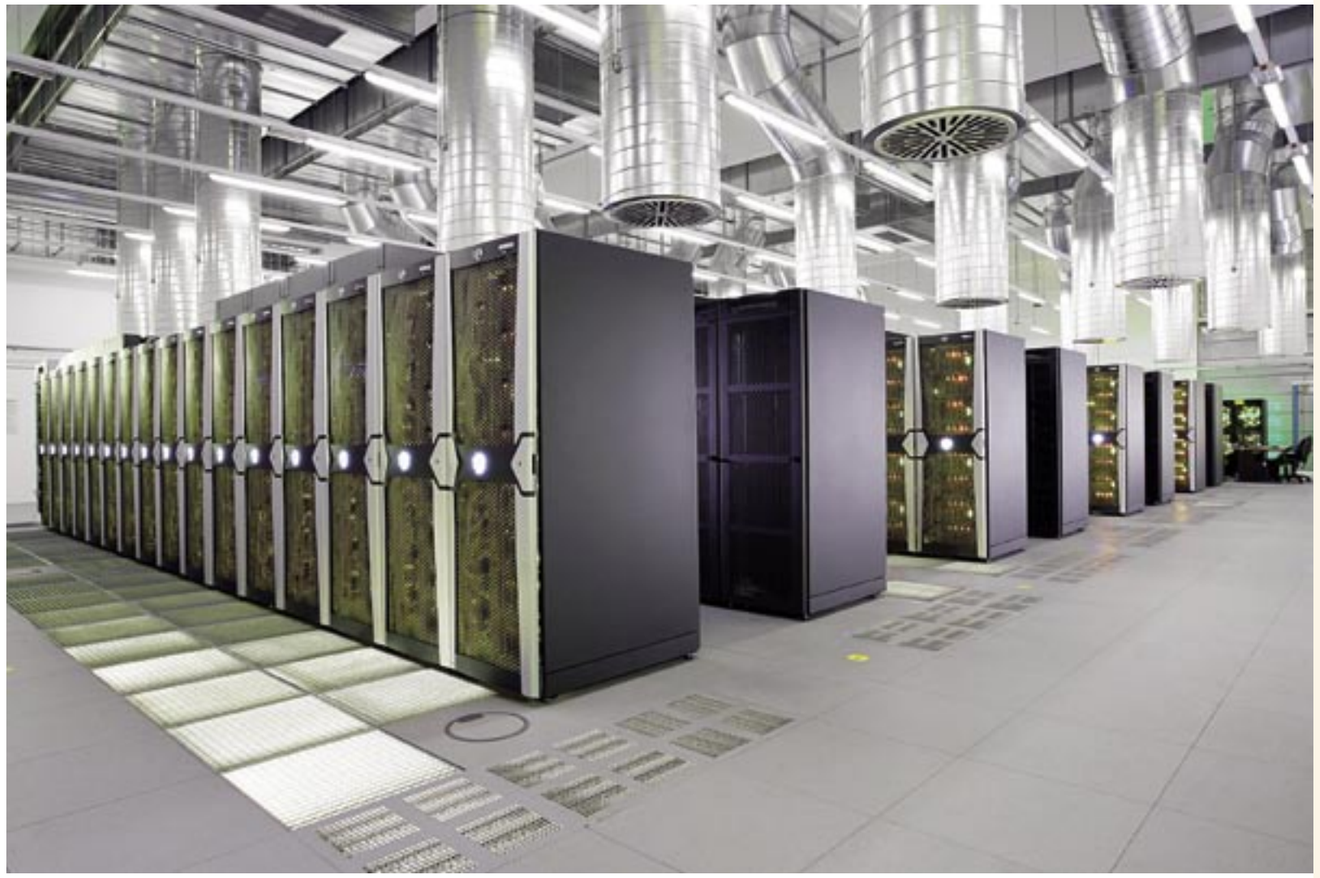
Forschungszentrum Jülich (FZJ)

Leibniz Computing Centre of the Bavarian Academy of Sciences (Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, LRZ) in Munich provides national, regional and local HPC services.

Each platform described below is documented on the LRZ WWW server; please choose the appropriate link from www.lrz.de/services/compute

Contact
Leibniz-Rechenzentrum

Dr. Horst-Dieter Steinhöfer
Boltzmannstraße 1
85748 Garching/München
Germany
Phone +49-89-3 58 31-87 79



View of „Höchstleistungsrechner in Bayern HLRB II“, an SGI Altix 4700
Foto: Kai Hamann, produced by gsiCom

Compute servers currently operated by LRZ are

System	Size	Peak Performance (GFlop/s)	Purpose	User Community
SGI Altix 4700 19 x 512 way	9,728 Cores 39 TByte	62,259	Capability computing	German universities and research institutes
SGI Altix 4700 256 way	256 Cores 1 TByte	1,640	Capability computing	Bavarian universities
SGI Altix 3700 BX2 128-way	128 processors 512 GByte memory	820	Capability computing	Bavarian universities
Linux Cluster Intel IA64 2-way	68 nodes 136 processors 816 GByte memory	870	Capability and capacity computing	Bavarian universities
Linux Cluster Intel IA64 4- and 8-way	19 nodes 84 cores 250 GByte memory	440	Capacity computing	Munich universities
Linux cluster Intel IA32 Intel&AMD EM64T	154 nodes 192 processors 320 GByte memory	850	Capacity computing	Munich universities

A detailed description can be found on LRZ's web pages: www.lrz.de/services/compute

Based on a long tradition in supercomputing at Universität Stuttgart, HLRS was founded in 1995 as a federal Centre for High Performance Computing. HLRS serves researchers at universities and research laboratories in Germany and their external and industrial partners with high-end computing power for engineering and scientific applications.

Operation of its systems is done together with T-Systems, T-Systems sfr, and Porsche in the public-private joint venture hww (Höchstleistungsrechner für Wissenschaft und Wirtschaft). Through this co-operation a variety of systems can be provided to its users.

In order to bundle service resources in the state of Baden-Württemberg HLRS has teamed up with the Computing Centre of the University of Karlsruhe and the Centre for Scientific Computing of the University of Heidelberg in the hkz-bw (Höchstleistungsrechner-Kompetenzzentrum Baden-Württemberg).

Together with its partners HLRS provides the right architecture for the right application and can thus serve a wide range of fields and a variety of user groups.

Contact
Höchstleistungsrechenzentrum
Stuttgart (HLRS)
Universität Stuttgart

Prof. Dr.-Ing. Michael M. Resch
Nobelstraße 19
70500 Stuttgart
Germany
Phone +49-711-685-872 69
resch@hlrs.de
www.hlrs.de

View of the NEC SX-8 at HLRS

Compute servers currently operated by HLRS are

System	Size	Peak Performance (GFlop/s)	Purpose	User Community
NEC SX-8	72 8-way nodes 9,22 TB memory	12,670	Capability computing	German universities, research institutes, and industry
TX-7	32 way node 256 GByte memory	192	Preprocessing	German universities, research institutes, and industry
Intel Nocona Cluster	205 2-way nodes 240 GB memory	2,624	Capability and capacity computing	Research institutes, and industry
Cray Opteron	129 2-way nodes 512 GByte memory	1,024	Capability and capacity computing	Research institutes, and industry
Cray XD1	8 12-way nodes 96 GByte	500	Industrial development	Research institutes, and industry

A detailed description can be found on LRZ's web pages: www.lrz.de/services/compute

The John von Neumann Institute for Computing (NIC) is a joint foundation of Forschungszentrum Jülich, Deutsches Elektronen-Synchrotron DESY, and Gesellschaft für Schwerionenforschung GSI to support supercomputer-aided scientific research and development. Its tasks are:

Provision of supercomputer capacity for projects in science, research and industry in the fields of modelling and computer simulation including their methods. The supercomputers with the required information technology infrastructure (software, data storage, networks) are operated by the Central Institute for Applied Mathematics (ZAM) in Jülich and by the Centre for Parallel Computing at DESY in Zeuthen.

Supercomputer-oriented research and development in selected fields of physics and other natural sciences, especially in elementary-particle physics, by research groups of competence in supercomputing applications.

At present, two research groups exist: the group Elementary Particle Physics, headed by Zoltan Fodor and located at the DESY laboratory in Zeuthen and the group Computational Biology and Biophysics, headed by Ulrich Hansmann at the Research Centre Jülich.

Education and training in the fields of supercomputing by symposia, workshops, school, seminars, courses, and guest programmes.

Contact
John von Neumann -
Institut für Computing (NIC)
Zentralinstitut für
Angewandte Mathematik (ZAM)
Forschungszentrum Jülich

Prof. Dr. Dr. Thomas Lippert
52425 Jülich
Germany
Phone +49-24 61-61-64 02
th.lippert@fz-juelich.de
www.fz-juelich.de/nic
www.fz-juelich.de/zam

The IBM supercomputers "JUBL" (top) and "JUMP" (bottom) in Jülich (Photo: Research Centre Jülich)

Compute servers currently operated by NIC are

System	Size	Peak Performance (GFlop/s)	Purpose	User Community
IBM BlueGene/P "JUGENE"	16 racks 16,384 nodes 65,536 processors PowerPC 450 32 Tbyte memory	222,800	Capability computing	German universities, research institutes and industry
IBM BlueGene/L "JUBL"	8 racks 8,192 nodes 16,384 processors PowerPC 440 4 TByte memory	45,875	Capability computing	German universities, research institutes and industry
IBM pSeries 690 Cluster 1600 "JUMP"	41 SMP nodes 1,312 processors POWER4+ 5,1 TByte memory	9,000	Capability computing	German universities, research institutes and industry
IBM BladeCentre-H "JULI"	2 racks 56 Blades 224 PowerPC 970 MP cores 224 GByte memory	2,240	Capability computing	Selected NIC projects
IBM Cell System "JUICE"	12 Blades 24 Cell processors 12 GByte memory	4,800 (single precision)	Capability computing	Selected NIC projects
AMD Linux Cluster "SoftComp"	66 compute nodes 264 AMD Opteron 2.0 GHz cores 264 GByte memory	1,000	Capability computing	EU SoftComp community
Cray XD1	60 dual SMP nodes 120 AMD Opteron 2.2 GHz processors 264 GByte memory	528	Capacity and capability computing	NIC research group "Comp. Biology and Biophysics"
apeNEXT (special purpose computer)	4 racks 2,048 processors 512 GByte memory	2,500	Capability computing	Lattice gauge theory groups at universities and research institutes
APEmille (special purpose computer)	4 racks 1,024 processors 32 GByte memory	550	Capability computing	Lattice gauge theory groups at universities and research institutes

HLRS Workshop Program

http://www.hlrs.de/news-events/events/

Each year, the HLRS organizes several workshops for scientific knowledge exchange and training in programming on high performance systems. The scientific workshops are dedicated to topics in state-of-the-art teraflop computing, scalable global parallel file systems, and tools for debugging and performance analysis of parallel programs. In the annual Results and

Review Workshop, teraflop applications in the framework of the federal HPC projects at HLRS are presented.

Parallel Programming Workshops

HLRS has set up series of training courses in parallel programming. They are organized at HLRS and also at several other HPC institutions: ZIH (TU Dresden), LRZ Garching, NIC/ZAM (FZ Jülich), CSCS (Manno, CH), and also at the University of Kassel. At the end of most theoretical lectures, hands-on sessions (mainly in C and Fortran) allow participants to immediately test and understand the basic constructs of, e.g., the Message Passing Interface (MPI), the shared memory directives of OpenMP, iterative solvers, and methods in computational fluid dynamics.

2007 – Retrospect
Scientific Workshops
6th Teraflop Workshop (March 26-27)
6th HLRS/hww Workshop on Scalable Global Parallel File Systems (April 16-18)
1st HLRS Parallel Tools Workshop (July 9-10)
The 3rd Russian-German Advanced Research Workshop on Computational Science and High Performance Computing (Novosibirsk, Russia, July 23-27)
German-Ukrainian Workshop on Simulation (Krim, Ukraine, September 9-16)
High Performance Computing in Science and Engineering - The 10th Results and Review Workshop of the HPC Centre Stuttgart (October 4-5)
Parallel Programming Workshops: Training in Parallel Programming and CFD
Parallel Programming and Parallel Tools (TU Dresden, ZIH, February 12-15)
Introduction to Computational Fluid Dynamics (University of Kassel, March 5-9)
Iterative Linear Solvers and Parallelization (HLRS, March 12-16)
NEC SX-8 Usage and Programming (HLRS, March 19-20)
1st HLRS Parallel Tools Workshop (HLRS, July 9-10)
Iterative Linear Solvers and Parallelization (LRZ, Garching, September 17-21)
Message Passing Interface (MPI) for Beginners (HLRS, October 8-9)
Shared Memory Parallelization with OpenMP (HLRS, October 10)
Advanced Topics in Parallel Programming (HLRS, October 11-12)
Introduction to Computational Fluid Dynamics (HLRS, October 15-19)
Parallel Programming with MPI & OpenMP (FZ Jülich, ZAM/NIC, November 26-28)
International Teaching
Summer School on Parallel Computing (Novosibirsk, Russia, July 9-20)
Parallel Programming with MPI & OpenMP (CSCS Manno, CH, August 8-10)
Tutorial on Hybrid MPI and OpenMP Parallel Programming, at SC '07 (Reno/Nevada, USA, November 12)
Training in Programming Languages at HLRS
Fortran for Scientific Computing (February 26 - March 2)
C++ for Scientific Computing (March 26 - April 5)
Fortran for Scientific Computing (October 22-26)
Other Training at HLRS
Beginners' Course: Autodesk 3ds Max (January 15-18)
Beginners' Course: Adobe Photoshop (January 22)
Beginners' Course: Adobe Photoshop (June 29)
Hybrid Grid Methods with GRIDGEN (February 7)
Mathematical Modeling and Simulation with COMSOL Multiphysics (June 11+12)

2008 – Workshop Announcements
Scientific Workshops
10th HLRS Metacomputing Workshop (March 12-14)
8th Teraflop Workshop (April 10-11)
7th HLRS/hww Workshop on Scalable Global Parallel File Systems (April 14-16)
2nd HLRS Parallel Tools Workshop (July 7-9)
High Performance Computing in Science and Engineering - The 11th Results and Review Workshop of the HPC Centre Stuttgart (planned October 2008)
Parallel Programming Workshops: Training in Parallel Programming and CFD
Parallel Programming and Parallel Tools (TU Dresden, ZIH, February 11-14)
Iterative Linear Solvers and Parallelization (HLRS, February 25-29)
Introduction to Computational Fluid Dynamics (University of Kassel, March 3-7)
NEC SX-8 Usage and Programming (HLRS, March 27-28)
2nd HLRS Parallel Tools Workshop (HLRS, July 7-9)
Parallel Programming with MPI & OpenMP (CSCS Manno, CH, August 12-14)
Iterative Linear Solvers and Parallelization (LRZ, Garching, September 15-19)
Introduction to Computational Fluid Dynamics (HLRS, September 22-26)
Message Passing Interface (MPI) for Beginners (HLRS, October 6-7)
Shared Memory Parallelization with OpenMP (HLRS, October 8)
Advanced Topics in Parallel Programming (HLRS, October 9-10)
Parallel Programming with MPI & OpenMP (FZ Jülich, ZAM/NIC, November 26-28)
Training in Programming Languages at HLRS
C++ for Scientific Computing (March 10-20)
Fortran for Scientific Computing (March 31 - April 4)
Fortran for Scientific Computing (October 27-31)
Other Training at HLRS (not yet fixed, planned similar to our courses 2007)
Beginners' Course: Autodesk 3ds Max
Beginners' Course: Adobe Photoshop
Hybrid Grid Methods with GRIDGEN
Mathematical Modeling and Simulation with COMSOL Multiphysics
URLs
http://www.hlrs.de/news-events/events/
http://www.hlrs.de/news-events/events/2008/parallel_prog_spring2008/
http://www.hlrs.de/news-events/events/2008/prog_lang_spring2008/

In the series of programming language courses, HLRS has restarted teaching FORTRAN due to its ongoing relevance in HPC programming.

In some courses, several different HPC aspects are combined. In Iterative Solvers and Parallelization, the focus is on iterative and parallel solvers, the parallel programming models MPI and OpenMP, and the parallel middleware PETSc. Thereby, different modern Krylov Subspace Methods (CG, GMRES, BiCGSTAB ...) as well as highly efficient preconditioning techniques are presented in the context of real life applications.

The course Introduction to Computational Fluid Dynamics deals with current numerical methods for Computational Fluid Dynamics. The emphasis is placed on explicit finite volume methods for the compressible Euler equations. Moreover outlooks on implicit methods, the extension to the Navier-Stokes equations and turbulence modeling are given. Additional topics are classical numerical methods for the solution of the incompressible Navier-Stokes equations, aero-acoustics and high order numerical methods for the solution of systems of partial differential equations. The last day is dedicated to parallelization of explicit and implicit solvers.

The Parallel Tools Workshop combines the aspects of a scientific workshop – the tool developers present their latest results – with aspects of an HPC course – the users of such tools can learn about using modern debugging and profiling and optimization tools.

1st Parallel Tools Workshop at HLRS

In the last issue of inSiDE, the ParMA project has been introduced. The aim of this project is to better integrate various tools for parallel development to better suite the programmer's difficulties of up-coming highly parallel multicore architectures. As a logical step, HLRS organized and hosted on the 9th and 10th of July, the 1st Parallel Tools Workshop with contribution of tool developers from various institutions.

The major aim of this workshop was to bring together the tool developer community and the professional user base. The concept of combining talks with live-demos, practicals and hands-on sessions proved to be successful: the workshop attracted 65 participants, mostly from German universities, research institutions and industry, and altogether 10 participants from the UK, France and the US.

The sessions were split into topical areas currently of interest: from single processor performance optimization, to parallel debugging and various tools for performance analysis and optimization, as well as advanced programming models and integrated development environments for parallel applications.

To enable the participants to get first-hand experience, each tool was made available on a cluster at HLRS, with a single command for setup using Oscar module-files. This allowed easy access to the wide range of tools.

With a session on Parallel Debugging tools the capabilities of DDT were shown, as well as memory debugging for using valgrind integrated into Open MPI. The MPI correctness tool Marmot and the Intel Message Checker were presented. Then, several tools for single and parallel performance analysis were introduced, ranging from Valgrind/Kcachegrind, the Intel Performance Tools Suite, OPT-tool by Allinea, the tools Kojak and TAU, as well as the Trace Analysers Vampir and Paraver and the performance prediction tool Dimemas.

Finally, Intel Threading Tools were introduced, followed by a talk on Intel Cluster OpenMP and the integration of PTP in Eclipse.

Although presenting a dozen tools in a two-day workshop requires a tough schedule, all speakers gave not only an overview of their respective tool, but also were able to show the particular software's strengths.

We hope to continue this successful workshop with a similar broad set of topics and tools and hope to welcome as many participants next year, again.

Welcome to Germany
Land of Ideas

21st of July - HLRS
A Place in the Land of Ideas

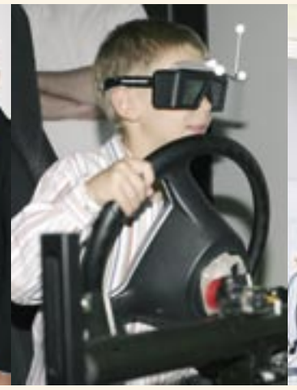
Germany - Land of Ideas - is a common and non-party initiative of the German Government and German economy under the patronage of Federal President Horst Köhler.

Over 1,500 applications for the year 2007 were received. The application of the High Performance Computing Centre convinced the jury. As "Selected Place 2007" the HLRS opened its doors on July 21st for one day and together with its co-operation partners invited the public into the simulated worlds of science and industry.

For the awarding ceremony HLRS welcomed Undersecretary of State Dr. Birk of the Ministry of Science, Research and Art of the State of Baden-Württemberg, the Rector of the University of Stuttgart Prof. Ressel, Mayor Dr. Müller-Trimbusch of the City of Stuttgart and the representative of the German Bank Dr. Rainer Grünenwald. Dr. Birk pointed out the importance of HPC for the state of Baden-Württemberg, mayor Müller-Trimbusch emphasized the relevance of simulation for the city of Stuttgart. Rector Ressel congratulated the HLRS and expressed his wish for the HLRS to thrive and prosper in the future as one of the key institutions of the University of Stuttgart.

After the awarding ceremony the centre was opened to the public. HLRS and its partners showed highlights of HPC and simulation among them a tour through the SX-8 computer room and Augmented Reality demonstrations of companies like Mercedes, RECOM, and Visenso. Institutes of the University of Stuttgart showed the use of HPC and visualization in fields like helicopter flight, water turbine simulation, blow flow simulation, many particle simulation, and ironing simulations. HLRS presented results of European lead projects like ViroLab and Akogrimo.

Besides the scientific presentations about 1,000 visitors enjoyed attractions like the Porsche driving simulator and an archery contest with an inflatable jumper topping the program for the many visiting children.



JARA -
Jülich-Aachen Research Alliance

On August 6, 2007 the Rector of RWTH Aachen University and the Chairman of the Board of Directors at Research Centre Jülich signed an agreement on establishing the Jülich-Aachen Research Alliance (JARA). The two scientific institutions will create a model for an internationally highly respected partnership between university and non-university research. The alliance will initially comprise the following three research sections:

- JARA-BRAIN
  Translational Brain Medicine
- JARA-FIT
  Fundamentals of Future Information Technology
- JARA-SIM
  Simulation Sciences

Other sections will be created in the future with energy research at the top of the list. Within the sections, research objectives will be defined and created in a joint process, the resources and equipment available to research and education will be shared, educational activities will be developed together, and personnel and investment decisions will be taken together. The co-operation is intended to be as unbureaucratic as possible.

JARA-SIM

Based on supercomputing as a key technology in many future-oriented research fields, JARA-SIM will promote the simulation sciences in disciplines like nanoscience, biology, energy and environmental sciences, and medicine. The German Research School for Simulation Sciences, whose mission is to integrate supercomputing into a high-class Master and PhD education for computational scientists and engineers, will be an important part of JARA-SIM.

Within JARA-SIM the Aachen Centre for Computational Engineering Sciences and the future Jülich Institute for Advanced Simulation will also join forces. The Research Centre Jülich, in the European project "Partnership for Advanced Computing in Europe" (PRACE), is co-ordinating the construction of a European high-performance computing and simulation infrastructure. Jülich is currently also leading the Gauss Alliance which unites the three German national supercomputing centres in their quest for a European top-level centre.

JARA | Jülich Aachen Research Alliance

Towards Petascale Computing in Europe – 3rd DEISA Symposium in Munich

Distributed European Infrastructure for Supercomputing Applications

Petascale computing was the theme of the third annual DEISA Symposium, held in Munich on 21st and 22nd of May 2007. DEISA (Distributed European Infrastructure for Supercomputing Applications) is a consortium of leading national supercomputing centres which currently deploys and operates a sustainable, distributed supercomputing infrastructure on a European scope. The DEISA symposium was organized by the Leibniz Computing Centre of the Bavarian Academy of Sciences and Humanties and by the Computing Centre of the Max Planck Society in Garching. It took place in the rooms of the Bavarian Academy of Sciences and Humanities in the Munich Residenz. The symposium focussed on perspectives for petascale computing in Europe, looking at the initiatives and strategies being deployed to enhance the outreach of high performance computing in Europe.

In his welcome address Professor Hegering, Head of the Board of Directors of the Leibniz Computing Centre has referred to the German efforts towards petascale computing: The three national high performance computing centres in Garching, Jülich, and Stuttgart (HLRS) have concentrated their resources in the Gauss Centre for Supercomputing (GCS), a step on the way to a petascale computing system in Germany. Professor Hegering referred to the excellent scientific and economic infrastructure of the Greater Munich area, which, to a large extend, is due to the continous support of the State Government of Bavaria.

Next, the Bavarian State Minister for Science, Research and the Arts, Dr. Thomas Goppel, opened the symposium. In his opening address Minister Goppel pointed out that supercomputers have become integral to the attractiveness and viability of a research hub. "Supercomputers are an absolute priority subject in the USA. But Bavaria has also done its homework: For quite a while the National Supercomputing System I at the Leibniz Computing Centre was the fastest supercomputer world-wide which was being used solely for scientific (non-military) purposes. Last year the National Supercomputing System II was installed at the LRZ. Since its expansion in April 2007 we can once again be proud to have one of the ten fastest supercomputers worldwide located here in Bavaria. We shall continue to ensure that supercomputing interests are properly taken into consideration in the future." Since Europe has fallen behind the USA and Japan in this arena, Dr. Goppel welcomed the efforts of the European Commission to build a supercomputer infrastructure within the EU that encompasses multiple centres with petaflops capabilities. He pointed out that Germany would be superbly well suited to serve as one location of that infrastructure. He further pointed out that it is just as important "that we invest in people", as it has been done in Bavaria for the last six years with the Competence Network for Technical and Scientific High Performance Computing (Kompetenznetzwerk für Technisch-Wissenschaftliches Hoch- und Höchstleistungsrechnen, KONWIHR).

This network has operated as a magnet for the world's top scientists. "The European Supercomputer should also be accompanied by such a program", he emphasized.

The first day of the event focussed on existing HPC initiatives and strategies: DEISA in Europe, NAREGI in Japan, and TeraGrid in the USA. Technology trends for petascale computing were discussed and various scientific cases for petascale computing in Europe were presented. In his speech Dr. Mario Campolargo, Head of Unit "GEANT and e-Infrastructures" at the European Commission, discussed the e-Infrastructure area in the EU's 7th framework programme and the way towards a European Supercomputing Infrastructure. DEISA's role in the new cycle of e-Infrastructures in FP7 is to provide concrete support for the deployment of a European HPC eco-system, including new European petaflops machines. Professor Dr. Victor Alessandrini, co-ordinator of the DEISA consortium, pointed out that with DEISA an environment is provided as well as services. These services facilitate access to high performance computers for European scientists as well as supporting DEISA administrative functions. The established infrastructure will be a good basis for a European eco-system in high performance computing. Professor Dr. Achim Bachem, Head of the Research Centre Jülich and Speaker of the Gauss Centre for Supercomputing, concluded the first day with a presentation of PRACE, the new "Partnership for Advanced Computing in Europe", which shall narrow the gap to the top players in the area of supercomputing. He emphasized that it would be crucial to establish legal structures and a commitment of authoritative payment obligations of the participating countries. The focus of the second day was on scientific results that were achived by using the DEISA infrastructure. As an introduction Professor Bode, full professor in the Department of Informatics of the TU München, presented the technological challenges of petascale computing: The efficent usage of hundred-thousands of processor cores demands new programming models and algorithms. Professor Bode explicitly argued for a multifaceted architecture landscape and against a monoculture in terms of computing technology: "We need competition in this domain, too". The scientific presentations from internationally renowned scientists requiring petascale computers involved a demonstration of the newest achievements in climate simulation to study global warming, the simulation of fusion reactors as a future unlimited energy source, and the simulation of the human heart as embedded in the human physical system. The symposium was concluded by presentations about the drawbacks of astrophysical simulations due to missing details requiring even larger supercomputers and the enormous advancements in the simulation of molecular structures for the development of new materials.

The presentations of the symposium: http://www.deisa.org/news_events/deisa_events/munich_symposium.php

Activities

New Cylindrical Stereoscopic Projection System at Research Centre Jülich

The visualization of data gained from experiments and simulations is a very important tool for the analysis and the presentation of scientific results. Nevertheless, common desktop-based display systems are not sufficient for the visualization of very large datasets with a complex 3D geometry. Also, visualization devices with small screens are not able to present results to a larger audience. For these reasons, large screen displays based on projection techniques are frequently used for a clear and detailed presentation of scientific data.

Such a device – or more precisely – a cylindrical, stereoscopic projection system, was installed in the rotunda of the Central Institute for Applied Mathematics at the Research Centre Jülich in early summer 2007. It is designed as a three channel frontprojection system with a large curved screen (11 m wide and 3 m high). Six projectors featured with Infitec stereo filters are used to produce three stereoscopic images. To guarantee that the three display channels are generating one seamless composite image, the sub images are linked by an overlap area of about 20 % of the image width. An edge-blending technique is used to reduce the added luminance in the overlap area to the regular level. The projectors are connected to six image processing units, which are responsible for enhanced colour correction, edge-blending, and image warping. The image warping is needed to compensate for the image distortion, which originates from the projection of the flat image onto the curved screen. A cluster of four Linux PCs serves as the render system for the display device. Three PCs are used as render nodes, each generating one stereo image pair. One PC is used as the master node on which the visualization application is running. To stream graphical commands from the master to the render PCs, all PCs are connected by a fast internal 1 Gbit/s network. The new visualization system can be used by scientists for data analysis as well as for presentations, typically for an audience of 20-30 people. As the curved screen has no edges at the transitions of the display channels, a homogeneous image can be presented to all viewers.

One application, which was successfully ported to the new projection system, is the visualization of data from the modelling and simulation of fire and smoke spread. Fire simulations using CFD methods become an important tool to control and optimize the safety in buildings, like sports stadiums, train stations or airports. A widely accepted

software package in the fire safety community is the Fire Dynamics Simulator (FDS), an open source project of the National Institute of Standards and Technology (NIST), USA. FDS includes Large Eddy Simulations (LES) as well as Direct Numerical Simulations (DNS). The geometrical resolution necessary to obtain reliable LES results together with the complexity of real-life buildings makes parallel simulations inevitable. An efficient parallelization is mandatory but complicated due to the strong dynamics of the combustion. Currently the parallel version of FDS is evaluated on different high performance systems. To enable a massively parallel usage on Jülich's BlueGene systems JUGENE and JUBL an alternative parallelization concept for FDS is under development.

One of the pictures shows the smoke spread in a subway station (primary fire is located inside a carrier) visualized on the new projection system using the open source software Smokeview (NIST). This investigation was done in co-operation with the safety consulting office insa4, Wuppertal. The corresponding simulations were performed on Jülich's supercomputer JUMP.

High Performance Computing Courses and Tutorials

LRZ
www.lrz.de

Tutorial:
Programming with Fortran 95/2003: Object orientation and design patterns

Date

February 4-6, 2008

Location

LRZ Building, Munich/Garching

Contents

The new Fortran standard 2003 offers new features which provide support for object oriented programming. However, it is indeed possible to implement many important design patterns using Fortran 95 only. This course has the purpose of giving an introduction on how to use the object-oriented features of Fortran at the 95 and 2003 levels without getting in the way of good application performance; furthermore also the following items are discussed:

- C interoperability and exception handling features in Fortran 2003
- Tools for programming and handling the build process (Make, Eclipse)

The participants of the course have the opportunity to experiment with the introduced concepts in hands-on sessions.

Webpage

http://www.lrz.de/services/compute/courses/#OOFortran

Workshop:
Performance Analysis of parallel programs with VAMPIR

Date

February 7, 2008

Location

LRZ Building, Munich/Garching

Contents

The performance of parallel programs is often constrained by bottlenecks in the communication structure which may only become obvious when running many MPI tasks. Isolating these issues requires analysis of very large trace files containing the necessary information about the MPI calls used. The "next-generation" version of VAMPIR and the VAMPIRtrace library allows efficient generation and analysis of these trace files. This workshop, given by specialists from TU Dresden, provides an introduction to using the VAMPIR toolset, as well as a hands-on session for users to apply the tool to their own codes.

Webpage

http://www.lrz.de/services/compute/courses/#Vampir

Course:
Parallel Programming of High Performance Systems

Date

March 17-20, 2008

Location

RRZE Building, Erlangen
LRZ Building, Munich/Garching
(via video conference)

Contents

Day 1

- Introduction to High Performance Computing
- Processor and system architectures
- Programming paradigms and languages
- Memory hierarchy and caches
- Intel Itanium Architecture
- HPC systems at LRZ and in Germany

Day 2

- Intel Compiler, tools and libraries
- SGI Altix architecture and tools
- Basics of optimization and performance evaluation
- Elements of Parallel Programming

Day 3

- Parallel Programming with MPI (incl. hands-on sessions)
- Intel Tracing Tools and MPI correctness checking
- Tuning of I/O

Day 4

- Parallel Programming with OpenMP (incl. hands-on sessions)
- Intel Threading Tools
- Advanced Examples for Optimization

Webpage

http://www.lrz.de/services/compute/courses/#ParallelProgramming

NIC
www.fz-juelich.de/nic

Parallel Programming with MPI, OpenMP, and PETSc

Date

November 26-28, 2007

Location

NIC/ZAM, Research Centre Jülich

Contents

The focus is on programming models MPI, OpenMP, and PETSc. Hands-on sessions (in C and Fortran) will allow users to immediately test and understand the basic constructs of the Message Passing Interface (MPI) and the shared memory directives of OpenMP. This course is organized by NIC/ZAM in collaboration with HLRS.
Presented by Dr. Rolf Rabenseifner, HLRS

Webpage

http://www.fz-juelich.de/zam/neues/termine/mpi-openmp

CECAM Tutorial Programming Parallel Computers

Date

February 11-15, 2008

Location

NIC/ZAM, Research Centre Jülich

Contents

This tutorial provides a thorough introduction to scientific parallel programming. It covers parallel programming with MPI and OpenMP. Lectures will alternate with hands-on exercises.

Webpage

http://www.cecam.fr/indexphp?content=activities/tutorial

Education in Scientific Computing

Date

August 4 - October 10, 2008

Location

NIC/ZAM, Research Centre Jülich

Contents

Guest Students' Programme "Scientific Computing" to support education and training in the fields of supercomputing. Application deadline is April 30, 2008.

Webpage

http://www.fz-juelich.de/zam/gaststudenten

High Performance Computing Courses and Tutorials

HLRS
www.hlrs.de

Parallel Programming with MPI, OpenMP and PETSc

Date

February 11-14, 2008

Location

Dresden, ZHR

Contents

The focus is on programming models MPI, OpenMP, and PETSc. Hands-on sessions (in C and Fortran) will allow users to immediately test and understand the basic constructs of the Message Passing Interface (MPI) and the shared memory directives of OpenMP. The last day is dedicated to tools for debugging and performance analysis of parallel applications. This course is organized by ZIH in collaboration with HLRS.

Webpage

http://www.hlrs.de/news-events/external-events/

Iterative Linear Solvers and Parallelization

Date

February 25-29, 2008

Location

Stuttgart, HLRS

Contents

The focus is on iterative and parallel solvers, the parallel programming models MPI and OpenMP, and the parallel middleware PETSc. Thereby, different modern Krylov Subspace Methods (CG, GMRES, BiCGSTAB ...) as well as highly efficient preconditioning techniques are presented in the context of real life applications. Hands-on sessions (in C and Fortran) will allow users to immediately test and understand the basic constructs of iterative solvers, the Message Passing Interface (MPI) and the shared memory directives of OpenMP. This course is organized by University of Kassel, HLRS, and IAG.

Webpage

http://www.hlrs.de/news-events/events/

Introduction to Computational Fluids Dynamics

Date

March 3-7, 2008

Location

University of Kassel

Contents

Numerical methods to solve the equations of Fluid Dynamics are presented. The main focus is on explicit Finite Volume schemes for the compressible Euler equations. Hands-on sessions will manifest the content of the lectures. Participants will learn to implement the algorithms, but also to apply existing software and to interpret the solutions correctly. Methods and problems of parallelization are discussed. This course is organized by University of Kassel, HLRS, and IAG, and is based on a lecture and practical awarded with the “Landeslehrpreis Baden-Württemberg 2003“ (held at University of Stuttgart).

Webpage

http://www.hlrs.de/news-events/external-events/

C++ for Scientific Computing

Date

March 10-20, 2008

Location

Stuttgart, HLRS

Contents

This introduction to C++ is taught with lectures and hands-on sessions. This course is organized by HLRS and Institute for Computational Physics.

Webpage

http://www.hlrs.de/news-events/events/

NEC SX-8 Usage and Programming

Date

March 27-28, 2008

Location

Stuttgart, HLRS

Contents

The first day is focused on vectorizing and parallelizing on NEC SX-8, the second day is dedicated to parallel I/O.

Webpage

http://www.hlrs.de/news-events/events/

Fortran for Scientific Computing

Date

March 31 - April 4, 2008

Location

Stuttgart, HLRS

Contents

This course is dedicated for scientists and students to learn (sequential) programming scientific applications with Fortran. The course teaches newest Fortran standards. Hands-on sessions will allow users to immediately test and understand the language constructs.

Webpage

http://www.hlrs.de/news-events/events/

2nd HLRS Parallel Tools Workshop

Date

July 7-9, 2008

Location

Stuttgart, HLRS

Contents

Developing for current and future processors will more and more require parallel programming techniques for application and library programmers. This workshop offers to the industrial and scientific user community, as well as the tools developers itself an in-depth workshop on the state-of-the-art of parallel programming tools, ranging from debugging tools, performance analysis and best practices in integrated developing environments for parallel platforms. Participants and tools developers itself will get the chance to see the strengths of the various tools. Therefore, this workshop is focused on persons who already know about parallel programming, e.g. with MPI or OpenMP. Hands-on sessions will give a first touch and allow to test the features of the tools.

Webpage

http://www.hlrs.de/news-events/events/

inSiDE

inSiDE is published two times a year by
The German National Supercomputing
Centres HLRS, LRZ, and NIC

Publishers
Prof. Dr. Heinz-Gerd Hegering, LRZ
Prof. Dr. Dr. Thomas Lippert, NIC
Prof. Dr. Michael M. Resch, HLRS

Editor
F. Rainer Klank, HLRS klank@hlrs.de

Design
Stefanie Pichlmayer sp028@hdm-stuttgart.de

Authors
Dr. James Asher asher@mail.uni-wuerzburg.de
Prof. Dr. Achim Bachem a.bachem@fz-juelich.de
Prof. Dr. Peter Entel entel@thp.uni-duisburg.de
Tobias Gradl tobias.gradl@informatik.uni-erlangen.de
Elmar Gröschel e.groeschel@aia.rwth-aachen.de
Dr. Markus E. Gruner me@thp.uni-duisburg.de
Dr. Jens Harting j.harting@icp.uni-stuttgart.de
Dr. Martin Hecht martin.hecht@icp.uni-stuttgart.de
Prof. Dr. Leonhard Kleiser kleiser@ifd.mavt.ethz.ch
Daniel König d.koenig@aia.rwth-aachen.de
Bastian Koller koller@hlrs.de
Christian Kunert christian.kunert@icp.uni-stuttgart.de
Klaus-Peter Mickel mickel@iwr.fzk.de
Dr. Dominik Obrist obrist@ifd.mavt.ethz.ch
Prof. Dr. Ernst Rank rank@bv.tum.de
Prof. Dr. Ulrich Rüde ruede@informatik.uni-erlangen.de
Leonhard Scheck leonhard.scheck@lrz.de
Dr. Michael Stephan m.stephan@fz-juelich.de
Dr. Hinnerk Stüben stueben@zib.de
Dr. Christoph van Treeck treeck@bv.tum.de
Rudolf Weeber rudolf.weeber@icp.uni-stuttgart.de
Petra Wenisch wenisch@bv.tum.de
Klaus Wolkersdorfer k.wolkersdorfer@fz-juelich.de
Jörg Ziefle ziefle@ifd.mavt.ethz.ch

If you would like to receive inSiDE regularly, send an email with your postal address to klank@hlrs.de