INSIDE • Vol. 13 No.2

Innovatives Supercomputing in Deutschland



is published two times a year by The GAUSS Centre for Supercomputing (HLRS, LRZ, JSC)

Publishers

Prof. Dr. A. Bode | Prof. Dr. Dr. Th. Lippert | Prof. Dr.-Ing. Dr. h.c. Dr. h.c. Hon. Prof. M. M. Resch

Editor

F. Rainer Klank, HLRS

klank@hlrs.de

Design

Pia Heusel, HLRS Linda Weinmann, HLRS

heusel@hlrs.de weinmann@hlrs.de

Fine dust particle distribution in the human lung during inspiration. The particles, tracked by a Lagrangian model, are transported down the trachea and distribute in the left and right main bronchi before they deposit in higher lung generations where they can cause severe damage to the respiratory system potentially leading to lung cancer.

Gonzalo Brito Gadeschi¹, Lennart Schneiders¹, Christoph Siewert^{1, 2} Andreas Lintermann^{1, 3} Matthias Meinke¹, Wolfgang Schröder^{1, 3}

¹ Institute of Aerodynamics, RWTH Aachen University, Germany, ² Laboratoire Lagrange, OCA, CNRS, Nice, France, ² SimLab "Highly Scalable Fluids & Solids Engineering", JARA-HPC, RWTH Aachen University, Germany

Editorial

journal on Innovatives Supercomputing in Germany.

In this issue we introduce two new supercomputer systems of GCS that were installed this year at the Leibniz Supercomputing Center (LRZ) at Garching and the High Performance Computing Center Stuttgart (HLRS) at Stuttgart.

SuperMUC Phase 2 – as the system at LRZ is called is operational since July of this year and already shows outstanding results for the users of GCS.

Hazel Hen - the new system at HLRS passed acceptance tests in September and is fully operational since October of this year. First tests show excellent results. Hazel Hen is expected to be the fastest European system available through PRACE.

The next step in the spiral of innovation that drives German HPC will be a new installation at Jülich Supercomputing Center (JSC) that is expected to be available in 2017.

All GCS systems are targeting high-end users that are able to make best use of hundreds of thousands of cores. The application section in this issue of inside introduces some of the results achieved on our systems.

The large variety of success stories that we have received shows that our systems play a major role not only in traditional fields such as physics, CFD, and Enjoy reading! climate research but also in innovative applications - supporting communities that are driven by the need for ever

Welcome to this new issue of inside, the faster systems but at the same time have to develop new methods to harvest the potential of HPC.

> As a consequence all three GCS partner centers are active in research and development and keep growing these activities. The focus is on improving the quality of applications from an HPC point of view. To harvest the performance of specialized systems codes have to be adapted, special architectural features have to be exploited, and power consumption has to be held at bay.

> It is only logical that all three GCS partners are active in a number of newly created European Centers of Excellence. Building excellent services on top of excellent resources has been a major goal of GCS for years. With the additional European funding this process can be accelerated and new networks can be created across the European research and economic space.

> Another goal of GCS has been to establish an excellent training program. Through its German activities and further supported by PRACE GCS has created the most comprehensive training program in Europe. We provide an overview of all courses for the coming nine months and invite our users - both in Germany and Europe - to join us to learn about HPC and discuss your needs for future programs.

We look forward to your feedback.

- Prof. Dr. A. Bode (LRZ)
- Prof. Dr. Dr. Th. Lippert (JSC)
- Prof. Dr.-Ing. Dr. h.c. Dr. h.c. Hon. Prof. M. M. Resch (HLRS)

З

LRZ: SuperMUC Phase 2 operating

5 PB of data are managed with an aggregated bandwidth of 100 GB/s across the distributed environment. A total main memory of just under 500 Terabyte is now available. "Energy efficiency is a key component of today's computing devices – from smart phones to supercomputers," explains Arndt Bode, Chairman of the LRZ. "With Phase 2 of SuperMUC, LRZ continues to act as



Figure 1: Front view on SuperMUC Phase 2 (without the front doors)

On June 29, 2015, the expansion of SuperMUC was officially put into operation during an opening ceremony at the Leibniz Supercomputing Centre (LRZ) in Garching near Munich. "SuperMUC Phase 2" provides an additional peak performance of 3.6 Petaflops to the previously available 3.2 Petaflops.

The festive inauguration of "SuperMUC Phase 2" was kicked off by a symbolic act: Dr. Ludwig Spaenle, Minister of Science at the Bavarian Parliament, Stefan Müller, Parliamentary State Secretary to the Federal Minister of Education and Research, Prof. Dr. Karl-Heinz Hoffmann, President of the Academy, Prof. Dr. Arndt Bode, Chairman of the LRZ, Martina Koederitz, General Manager of IBM Germany, and Christian Teismann, Vice President and General Manager, Global Account Business Levono, jointly pressed the "Red Button" symbolizing the start-up of the system expansion to SuperMUC.

Significantly reduced Footprint

The extension to SuperMUC, an IBM System X iDataPlex which first became operative in mid 2012, went according to the earlier defined systems roadmap. 86,016 processor cores in 6,144 Intel Xeon E5-2697 v3 processors, based on Intel's latest technology, were added to the previously available 155,656 processor cores, lifting the maximum theoretical computing power to now 6.8 Petaflops. This performance boost comes with surprisingly little space requirements: While more than doubling the overall system performance, Phase 2 only requires one fourth of the footprint of SuperMUC Phase 1.

Users will especially benefit from the fact that – like with SuperMUC Phase 1 – the system expansion refrains from using so called accelerators. Because of this, existing applications can continue to be used without any major adaptations to the software.

The LRZ supercomputing infrastructure now offers additional 7.5 Petabyte of SAN/ DAS storage on GPFS Storage Servers (GSS). By combining IBM's Spectrum Scale technology with Lenovo System x-Servers, LRZ offers a detailed description of all components of SuperMUC on its webpage https://www.lrz.de/services/compute/ supermuc/systemdescription/.

Energy efficient Supercomputing

SuperMUC will continue to be one of the most energy efficient supercomputers in the world. The proven hot water cooling technology, implemented by IBM, was also applied to Phase 2 of the installation: Through a network of micro channels, the cooling system circulates 45 centigrade warm water over active system components, such as processors and memory, to dissipate heat. Thus, no additional chillers are needed. The use of the latest processors, which allow an adaptation of their frequency to the specific needs of the computations, adds to the efforts of reducing the power usage. In combination with the use of energy optimizing operating system software these energy saving means result in an overall reduction of system power usage by approx. 40 per cent thus saving several millions of Euros every year compared to conventional cooling.

a pioneer in this field as we deliver proof that it is possible to significantly lower the energy consumption in data centres, thus drastically reducing the operating costs."

Enabling Big Science

Like SuperMUC Phase 1, the LRZ system expansion has been designed for exceptionally versatile deployment. The more than 150 different applications, which run on SuperMUC on average per year, range from solving problems in physics and fluid dynamics to a wealth of other scientific fields, such as aerospace and automotive engineering, medicine and bioinformatics, astrophysics and geophysics, amongst others. The results of the first two years of research supported by SuperMUC are available as a report (https://www.lrz.de/ services/compute/supermuc/magazinesbooks/). One of the highly recognized research projects that was only possible using simulations on SuperMUC gave new insights into the evolution of insects and of birds. Alexandros Stamatakis' article in this issue explains how Super-MUC supported this breakthrough that led to several publications in "Science", featured on their cover page.



Figure 2: Prof. Dr. Arndt Bode, Chairman of the LRZ, explaining to State Minister Dr. Spaenle and Parliamentary State Secretary Stefan Müller how LRZ saves millions of Euros by cooling SuperMUC with warm water.



Figure 3: Stefan Müller, Parliamentary State Secretary to the Federal Minister of Education and Research, Prof. Dr. Arndt Bode, Chairman of the LRZ, Martina Koederitz, General Manager of IBM Germany, Christian Teismann, Vice President and General Manager, Global Account Business Levono, Dr. Ludwig Spaenle, Bavarian State Minister of Science, and Prof. Dr. Karl-Heinz Hoffmann, President of the Bavarian Academy of Sciences.

• Ludger Palm

Leibniz Supercomputing Centre (LRZ), Germany During the first weeks of operation, scientists using SuperMUC Phase 2 already achieved impressing world records:

 the biggest Fast Fourier Transformation for more than 10,000³ grid points using the CFD code "FLASH"

• the largest QM/MM replica exchange simulation to date with statistics that allow 100 times better accuracy in computing molecular spectra in solution with the molecular dynamics code IPHIGENIE

- reducing the computation time for a seismic problem from 16 hours to 55 seconds
- the biggest simulation of the visible universe with the GADGET code.

The results of the GADGET simulation of the universe were then visualized again using SuperMUC to produce a very impressive movie (http://www.mpa-garching.mpg.de/HydroSims/Magneticum/ Images_Movies/left_hd.avi). Klaus Dolag et al. explain this further in this issue.

Financing of SuperMUC

As with the first installation phase, SuperMUC's system expansion including service expenses and operating costs has been funded through the project PetaGCS with the Federal Ministry of Education and Research (BMBF) and the Bavarian State Ministry of Science, Research and the Arts covering the expenses in equal shares. Like SuperMUC Phase 1, SuperMUC Phase 2 is now available for academia in Germany as part of the Gauss Centre for Supercomputing GCS and for many researchers in Europe via the Partnership for Advanced Computing in Europe, PRACE.

contact: Ludger Palm, ludger.palm@lrz.de

HLRS almost doubles Computing Capacities with new HPC System Hazel Hen

With the installation of HPC system Hazel Hen, an extension to the XC4Osystem Hornet, the High Performance Computing Center Stuttgart (HLRS) has now completed the second and final phase of the PetaGCS project, an activity launched by the Gauss Centre for Supercomputing (GCS) in 2008, which aimed at rolling out petascale High Performance Computing systems at the three GCS locations in Garching, Jülich, and Stuttgart.

Supercomputer Hornet, which was installed at HLRS in 2014 (21 XC40 cabinets, 3,944 compute nodes, main memory of 128 GB per node), had already summed up to a total number of 94,656 cores with a main memory of 493 Terabyte. But this was only the

first step of the configuration in the systems road map. In a second and final step, the HLRS supercomputer has now been extended to a 41 XC40 cabinets system - code named "Hazel Hen" – which provides 7,712 nodes of leading edge technology to academia and industry. Hazel Hen's overall system configuration now consists of 185,088 Intel Haswell E5-2680 v3 cores, a total of 965 TB Main Memory and more than 10 Petabyte of Disk Space. Hazel Hen's peak performance of 7.42 Petaflops (Quadrillion Floating Point Operations per second) almost doubles the performance data delivered by Hornet. The node-node interconnect of the HLRS HPC system continues to be based on the CRAY Aries technology.



Bastian Koller

University of Stuttgart (HLRS), Germany

With this extension, the clear focus on such as the Institute of Aerodynamics sustained performance of real-world applications as they are currently running on the HLRS HPC systems is further underpinned.

Tight Installation Schedule

The installation of the extension and the final consolidation of the two systems (Hornet + extension) into one homogeneous Haswell-System (Hazel Hen) followed a tight time schedule. The additional cabinets were delivered to the HLRS premises in the last week of July 2015 and were installed during the following weeks. On August 17, Hornet was powered down for one entire week and the connections between the systems were installed.

During a short period of around one

week, the system integration and the

system testing were executed and initial

Linpack Benchmarks were run. Addition-

ally, as carried out with system Hornet,

Hazel Hen was made available to users

Science (IAM-CMS) of the Karlsruhe Institute of Technology, for so-called XXL-simulation projects. The goal of this undertaking had been to once again test the endurance of the new system under real life conditions, and the test runs were completed successfully.

and Gas Dynamics of the RWTH Aachen

University and the Institute of Applied

Materials Computational Materials

On October 1, system Hazel Hen was declared fully operational and is now available for general use. With this configuration HLRS currently hosts the largest CRAY installation based on the Haswell-technology worldwide.

contact: Bastian Koller, koller@hlrs.de

P

Preparatory Access to Computing and Support Resources at JSC



Starting this November, JSC is offering a new way of accessing its computing and support resources. Besides submitting a full project proposal via the regular NIC/GCS and JARA-HPC/ VSR calls, users may now apply for Preparatory Access, which includes a limited amount of computing time on JURECA or JUQUEEN for porting and testing purposes as well as support by the JSC Simulation Labs.

Analogously to similar schemes previously introduced within PRACE [1], the JSC Preparatory Access aims to facilitate access to the Jülich supercomputers for researchers with computationally intensive scientific problems, but codes that still need to be made fit for HPC prior to a full proposal. Applications for Preparatory Access to JSC resources may be submitted twice a year before the 1st of November and 1st of May, respectively. These will undergo a technical evaluation by JSC

computingtime.

Reference

access/

contacts: Paul Gibbon, p.gibbon@fz-juelich.de Boris, Orth, b.orth@fz-juelich.de

staff, who will assess the potential of the application to benefit from HPC adaptation and tuning. If approved, users receive a limited computing time budget along with expert assistance from one of the JSC SimLabs for a period of up to four months, to improve the performance of their application and prepare a full project proposal.

Further details on the new JSC Preparatory Access scheme can be found at http://www.fz-juelich.de/ias/jsc/

 Paul Gibbon Boris Orth

> Jülich Supercomputing Centre (JSC), Germany

[1] http://www.prace-ri.eu/prace-preparatory-

Convection permitting Latitude-Belt Simulation

Using the Weather **Research and Forecasting** (WRF) Model

Thanks to the availability of HLRS's petascale HPC system Hornet, researchers at the Institute of Physics and Meteorology of the University of Hohenheim were able to run a highly complex extended-range simulation for a time period long enough to cover various extreme weather events on the Northern hemisphere at a previously unmatched spatial resolution. Deploying the highly scalable Weather Research and Forecasting (WRF) model on 84,000 compute cores of Hornet, the achieved results confirm an extraordinary quality with respect to the simulation of fine scale meteorological processes and extreme events.

Key Facts

- 84,000 compute cores
- 84 machine hours
- 330 TB of Data
- 12,000*1,500 horizontal grid points (resolution 0.03°) with 57 vertical layers
- 535,680 time steps

intense Vb weather situations in Europe are responsible for extreme events like the heat waves in summer 2003 in Central Europe, in August 2010 in Russia (which was associated with flooding of the Odra and in Pakistan) as well as severe flooding events like in summer 2002 in Germany. Blocking Omega and Vb weather situations are caused when long-lived, resonant Rossby waves develop in mid-latitudes. In numerical weather prediction and climate models, at least a resolution of 20 km is required to simulate quasi-stationary Rossby waves. However, to simulate the associated extremes the simulations need to be convection permitting. Further, the high resolution allows the small scale structures to feed back to the large scale systems. Most of the current limited area, high-resolution models apply a domain which is centered over the region of interest and has 2 longitudinal and 2 latitudinal boundaries. Such limited area model applications suffer from a deterioration of synoptic features like low pressure systems due to effects in the boundary relaxation zone. For

Persistent so-called blocking Omega and



Figure 1: Convection permitting model domain with topography as shaded areas. The southern and northern boundaries are at 20°N and 65°N, respectively.

Europe this is mainly caused by the longitudinal boundaries which are located over the North Atlantic at around 20°W and Eastern Europe close to 20°E. Especially the boundary over the North Atlantic can destroy low pressure system features which are responsible for the weather and climate development over Europe. A possibility to overcome this is to run a latitude belt simulation. Here only boundary conditions at the northern and southern boundaries of the model domain are required.

The Weather Research and Forecasting (WRF, http://www.wrf-model.org/) model was applied at 0.03° horizontal resolution for July and August 2013 forcing the model 6-hourly with ECMWF analyses at 20N and 65N. Sea surface and inland lake temperature was updated daily with data from the OSTIA project (http:// ghrsst-pp.metoffice.com/pages/latest_ analysis/ostia.html) of the UK Met Office, which is available at 5km resolution.

Figure 1 shows the domain with the model topography as shaded areas. The model grid consisted of 12,000*1,500 boxes with a horizontal resolution of 0.03°. For vertical discretization, 57 terrain following levels were selected with the model top being at 10hPa (approx. 35km above sea level).

State of the art global numerical weather prediction models have a horizontal resolution of approximately 0.15°. In comparison the latitude belt simulation provides an improved representation of the real topography and thus allows for a better representation of the atmospheric circulation systems. For instance, in the global model of ECMWF, the summits of the Alps and the Himalaya have an altitude of 2,500m and 6,000m which is far away from reality. This can be

crucial e.g. for the simulation of convective precipitation during the summer season. Convection permitting simulations are also beneficial for simulating intense cyclones such as hurricanes over the Gulf of Mexico and typhoons in the Pacific Ocean. Due to the usually





Min SLP 940 hPa

Figure 3: Sea level pressure [hPa] (black contour lines) and 10m wind speeds [m/s] as shaded areas.



Figure 2: Predicted track of typhoon Soulik from the WRF simulation during the period July 11, 2013 and July 13, 2013. The colors show the intensity of the typhoon according to the Saffir-Simpson hurricane wind scale.

Valid: 2013-07-11_20:00:00



coarser resolution of the global model, they would not be able to simulate high intensities as they are observed for e.g. Cat. 4 and 5 storms.

One of the highlights of this study was the simulation of the typhoon Soulik, which developed in the Pacific Ocean in the first 10 days of July 2013. Figure 2 shows the simulated track of the typhoon in 3-hourly intervals from July 11-13, 2013.

It was a category 4 typhoon meaning that average 10m-wind speed exceeded 58m/s. The model was able to simulate the intensity well in accordance with observations, which indicated a central pressure of 925hPa and maximum wind speed of 65m/s.

Figure 3 shows an example of the typhoon's structure with the surface wind speed and sea level pressure during its peak intensity at July 11, 2013. The maximum 10-m wind speed was about 62m/s and the central pressure 940hPa. Also the calm region in the center of the cyclone is clearly visible. The whole simulation experiment was repeated on a 0.12° grid. At this resolution, the track was captured slightly better but the maximum wind speeds were strongly underestimated by about 15m/s compared to the convection permitting resolution simulation with its 0.03° grid. Also the representation of the fine structure of the rain bands was improved in the high-resolution simulation.

The comparison with satellite images reveal that the timing of the typhoon's track in the model was comparable to the observations, although the center of the storm was simulated about 250km further north. The intensity was well captured by the model as shown in Figure 3. In summary, this new configuration of the WRF model demonstrates a significant improvement of the simulation of extreme events.

contact:

Thomas Schwitalla, Thomas.Schwitalla@uni-hohenheim.de

Principal Investigator: Volker Wulfmeyer, Institute of Physics and Meteorology, University of Hohenheim

Prediction of the turbulent Flow Field around a ducted **Axial Fan**

Exploiting the available computing capacities of supercomputer Hornet of the High Performance Computing Center Stuttgart, researchers from the Institute of Aerodynamics (AIA) of the RWTH Aachen University conducted a large-scale simulation run in their efforts to tackle the prediction of the acoustic field of a low pressure axial fan using computational aeroacoustics (CAA) methods. Goal of this project, which scaled to 92,000 compute cores of the HPC system Hornet, was to achieve a better understanding of the development of vortical flow structures and the turbulence intensity in the tip-gap of a ducted axial fan.

Key Facts

- 80 TB of data

Axial fans can be found in many technical applications, from computer CPUs to automotive engines and industrial airconditioning systems. In addition to the performance, noise reduction has become one of the major issues for engineers in recent years. Therefore, efficient numerical methods which are able to accurately predict the acoustic field are required.



Figure 1: Sea level pressure [hPa] (black contour lines) and 10m wind speeds [m/s] as shaded areas.

Thomas Schwitalla

Institute of Physics

and Meteorology,

University of

Hohenheim,

Germany

• 92,000 compute cores 110 machine hours • 1,000 million grid points 320,000 time steps.

Matthias Meinke

Institute of Aerodynamics, **RWTH Aachen** University, Germany

In a joint university-industry project, the prediction of the acoustic field of a low pressure axial fan using computational aeroacoustics (CAA) methods is tackled. The source distribution of the CAA analysis, however, requires highly resolved instantaneous flow field data. Since Reynolds averaged Navier-Stokes (RANS) computations strongly depend on the chosen turbulence model and are not always reliable due to, e.g., the strong streamline curvature dominating the flow structures inside the tip-gap region, the subsonic flow field is predicted by large-eddy simulation (LES). For this purpose, the unstructured flow solver ZFS (Zonal Flow Solver), which solves the Navier-Stokes equations for unsteady and compressible flows in the rotating frame of reference on Cartesian meshes, is used. The overall accuracy of the flow solver in space and time is of second order.

Simulations were performed at a Reynolds number of Re=9.36×10⁵ based on the outer casing diameter and the rotational velocity of the casing wall. The current mesh has approx. 1 billion cells, resolving only a 72° segment of the axial fan, i.e., one out of five blades, to reduce the high computational costs. This high resolution is necessary to accurately resolve the vortical flow structures and their development in the tip-gap region.

Simulations were performed on the CRAY HPC system Hornet of the HLRS. The flow solver has been fully parallelized using the message passing interface (MPI) so that different number of cores can be used for the flow simulation. The minimum number of cores required for a simulation using 1 billion cells is approx. 10,000. However, up to 92,000 cores have already been used.

The required time per simulation, i.e. four full rotations of the rotor using 92,000 cores, is about 110 machine hours.

To obtain accurate statistical data from the turbulent flow field like, e.g., the Reynolds stress tensor and two-point correlations of the velocity components, a large number of samples of the instantaneous field is required. The statistical data requires about 80 TB of disk space.

contact: Matthias Meinke, office@aia.rwth-aachen.de

Principal Investigator: Wolfgang Schröder, Institute of Aerodynamics, **RWTH Aachen University**

A CFD Study of Centrifugal **Compressor with Circum** ferentially non-Uniform Inlet **Distortions using Ansys CFX**

This numerical study is combined with an experimental research previously conducted at the institute on an industrial compressor stage for Flight Propulsion, TU München, on an industrial compressor stage with newly developed IGVs as shown in Figure 1. While the type-A as baseline is a standard NACA profile, the new type-B features a biased camber and type-C is a multi-foil comprised of a rotating tail. Measurement data shows a 40% pressure loss reduction and 2 point efficiency improvement with the new IGV profiles. As a consequence, this CFD investigation attempts to reveal how exactly the IGV interacts with the impeller, and how the inlet distortions induced by IGVs affect the impeller performance. To reproduce the circumferential flow nonuniformities at the impeller inlet section, a full-annulus modelling is required for the numerical study. An example for the inlet conditions created by IGVs is shown in Figure 2. As an alternative, the TBR-FT method using a two-passage model may be useful if the save on computation cost is of prior concern, and thus was tested as well.

The mesh grids applied for the CFD simulation are shown in Figure 3. To quickly represent the speedlines, a singlepassage model with a total of 4.6 million cells was introduced. Ansys TurboGrid was used for generating the impeller mesh with $y_{+} = 1.5$ on the walls and ICEM for the diffuser mesh. The diffuser outlet was artificially contracted to avoid outlet backflow deteriorating

numerical stability. For the transient simulations a full-annulus model with the similar topology was introduced. The mesh in each passage was made much coarser to obtain an overall size of 4.3 million cells. The y+ of the full-annulus impeller walls was 4.5. The numerical simulations were performed by using the commercial CFD solver Ansys CFX v.15.0 based on the finite-volume Reynolds averaged Navier Stokes methods (RANS). The SST model was selected as the turbulence model for this study. A full transient simulation requires 32 to 48 computer cores and 8 computation hours until the result became fully settled upon time progress. The relative small number of cores compared to other large-scale jobs was determined by the mesh size, the HPC licenses and the efficiency of parallelization. For the TBR-FT method, an identical mesh per passage as the full transient case was applied, but only a total of two passages were needed. The 360° inlet conditions



Figure 1: Schematic cross section of compressor stage including the IGV plenum and IGV inside (the red arrows denote the CFD domain); from [1].



Figure 2: Inlet boundary conditions for the full-annulus computation model, shown are total pressure and yaw angle for the three IGV types; from [2].

must be defined as rotational flow boundary conditions onto a locally stationary frame in order to create the time-varied distortions for the rotating impeller. The time step for the full transient and TBR-FT models was set to 1/24 of blade passing period. The measured inlet total pressure, total temperature and yaw angle were set as inlet boundary conditions. Accordingly, the static pressure was set as outlet condition to achieve good stability. The value of static pressure was controlled to result in a correct mass flow rate as the measured one.

Figure 3: Mesh grids at midspan section for the single-passage model (left) and the full-annulus model (right); from [2]; Top right: outlet contraction at diffuser outlet.

The simulation results show that while at O° IGV setting angle the improved homogeneous inflow is responsible for the improved impeller efficiency, at high IGV setting angle the better performance for the new IGV types is mainly attributed to the correct incidence, as illustrated in Figure 4. While the baseline IGV has a large overturned inlet flow direction, the new type-C IGV provides a proper incidence angles due to its strict linearity of yaw angle generated within the working range from O° to 60°. This prevents a pressure side flow separation at the impeller leading edge, which is the case for A-60 by steady calculation. The comparison study between transient and TBR-FT shows that the TBR-FT method is 4.5 times faster, and the result requires 8.5 times less storage memory than the conventional full transient CFD simulation. In addition, a TBR-FT result file contains the Fourier coefficients needed to reconstruct the result back to an arbitrary time step. A good agreement was reached between steady, transient, and TBR-FT methods regarding the prediction of overall performance parameters. However, two discrepancies were found, which are small flow discontinuities across the side boundaries of the two-passage model, and smoothing out the flow unsteadiness non-blade passing frequencies (BPF). These discrepancies can be traced back to the TBR-FT theory in terms of collecting signals based on a phase-shift sampling, and filtering out the non-BPF harmonics. The test result is depicted in Figure 5. It can be observed that while the transient result contains some irregular flow regions at the impeller trailing edge, the TBR-FT computation yields ordered patterns repeating themselves at the same locations. Additionally, some small flow discontinuities can be detected across the side boundaries of each two-passage domain. Despite of discrepancies, the TBR-FT method predicts very similar overall performance (efficiency, work coefficient, head) and circumferentially averaged flow progresses (blade loading, flow profile) as the transient method.

More detailed comparison between the previous test data and the simulation result is being performed to derive useful information from the CFD results on the understanding of impeller flow physics.

The author would like to thank the CFD team at the Leibniz Supercomputing Centre (LRZ der Bayerischen Akademie der Wissenschaft) for their technical support on this project.





Figure 4: Midspan relative velocity contours and vectors at IGV setting angle of 60° for type-A and type-C; from [2].



Figure 5: Midspan total pressure contours, compared between transient and TBR-FT methods; from [3].

References

[2] Chen, N., Erhard, W.

Numerical investigation of a centrifugal compressor stage with IGV induced inlet flow distortions, Conference on Modelling Fluid Flow (CMFF'15), Budapest, Hungary, 2015

[3] Chen, N.

Numerical Investigation of a Centrifugal Compressor with Inlet Distortions Induced by Variable Guide Vanes - A Comparison between Steady, Transient and Transient Blade Row Simulation Methods, ANSYS Conference & 33rd CADFEM Users' Meeting 2015, Bremen, Germany, 2015

contact: Nan Chen, chen@lfa.mw.tum.de



[1] Sezal, I., Chen, N., et al.

Introduction of Circumferentially Non-uniform Variable Guide Vanes in the Inlet Plenum of a Centrifugal Compressor for Minimum Losses and Flow Distortions, Proceedings ASME Turbo EXPO, Montreal, Canada, GT2015-43467, 2015

Nan Chen

Institute for Flight Propulsion, TU München, Germany

Improving the Plasma Simulation Code (PSC)

For medical diagnostics, material sciences, and experiments probing fundamental physics, the development of new sources of radiation and particles created by high intensity laser-plasma interaction has become a major research field. Upcoming laser facilities like the Extreme-Light-Infrastructure (ELI, [1]) have billion Euro budgets. A detailed understanding of the expected effect beforehand can strongly increase and ensure the return on the investment.

The **PSC**

The original Fortran version of the PSC was developed by H. Ruhl in the late 9Os as first full 3D Particle-in-Cell (PIC) code that was released to the public domain (GPL). It is well-tested, widely-recognized and reliable, and has been fundamental to other current codes,

e.g. the EPOCH code. In 2009 H. Ruhl et al. and K. Germaschewski et al. reworked the code into a modular C framework supporting bindings to Fortran and C/CUDA, featuring selectable field and particle pushers [2,3].

Adaptive-Particle-Refinement (APR) for Load-Imbalances and QED Effects

A major challenge for PIC simulations are compact targets like ultra-thin foils or nano balls. For reasonable statistics in later parts of the simulation a very high number of quasiparticles has to be concentrated in the initial target. This implies enormous load imbalances, especially in 3D, as shown in Figure 1. APR is a promising technology to resolve such issues. First measurements indicate a potential speedup by an order of



Figure 1: A test for 3D nano foil simulations. Shown is the electron density (violet = under dense, light blue = over dense) spread after a few cycle high intensity laser pulse interacted with a nano meter foil infinite in x-y plane initially at $z \approx 50$ (= 8 μ m). In the green section the proton background aggregates many electrons. Further the development of layers can be recognized. In the extended violet region a filamentation into the strongly under-dense blue can be observed probably due to a lack of quasiparticles, although the initial quasi particle density exceeded the usual values by far more than an order of magnitude, demonstrating the technical difficulties of compact targets. magnitude. Our findings about quantum electrodynamics (QED) also show that multiple particle weights are required to simulate electron-positron cascades. This can lead to a super-exponential growth of the particle number over time $[O[e^{(e^t)}]$. Such simulations are impossible without APR.

Key technical Features of the **PSC**

- Second and fourth order electromagnetic field solvers, multiple second order particle pushers with varying form factors
- Flexible command line interface exposing a wide range of simulation parameters, solver types, etc.
- Dynamic window: Multiple moving and/or growing simulation areas, that can collide and merge
- Autotools build and automated test system for a wide range of computing platforms
- Modular I/O subsystem including several HDF5/XDMF modules for large scale parallel output
- Dynamic load and memory balancing
 Automated checkpointing and restart facility
- MPI only or hybrid MPI/OpenMP parallelization as well as SSE/AVX(512) micro vectorization
- Support for GPGPU (CUDA) and INTEL XEON PHI (MIC)
- Sophisticated inline performance measurement analysis
- AMR: Adaptive Mesh Refinement (in development).

Ultra-thin Foils

To create novel radiation sources, the interaction of ultra-thin foils with intense lasers has been thoroughly studied in the laser-plasma-community over the last years. The detailed dynamics of the

electron layer, longitudinal and lateral radiation effects, and their spectra are of main interest. At higher field strengths, which will be available in future experiments at ELI, QED effects and self-radiation will play a role. At currently available intensities for few cycle lasers (10^{24} W/m²), the slingshot effect of the electrons is dominant and allows for circularly polarized Attosecond X-ray Pulses (AXP) in the reflected light (see Fig. 2). One focus of our current Gauss supercomputing campaign is the full 3D simulation of nanometer foils with the necessary resolution to cover the frequencies of the AXP. Therefore, up to 512 billion grid cells will be necessary.

Including particles this easily requires up to 200 TB of main memory. Reduced runs with half resolution have been carried out successfullyand the setup for the full resolution run was tested during the "Extreme scale-out of SuperMUC-Phase 2" during May/June 2015.

Node Level-Performance Optimizations

Within an on-going project funded by KONWIHR, the node level performance of the code is improved in collaboration with the LRZ. Special energy configurations on SuperMUC allowed us to do a CPU frequency analysis of the PSC that showed that the field pusher gains almost no speed up from 1.2 up to 2.6 GHz, suggesting it to be most likely memory bandwidth bound. Further measurements confirmed that the field pusher was close to the memory controller limit. Therefore, optimizations efforts have been concentrated on the particle pusher that gained close to perfect linear speed up. Together with analyzing software expertise from LRZ, a core-level vectorization was imple-



mented that indicates a potential double precision performance gain of 60% on SuperMUC Phase 1 (AVX) and 95% on SuperMUC Phase 2 (AVX2).

Strong Scaling

The average exponent for strong scaling varies between -0.96 and -0.99, over 6-10 powers of two of the core number for homogeneous problems. The particle throughput from 1-86k cores increases almost linearly with a scaling exponent of +0.98.

I/O Measurements

On the new GSS file system, performance tests were carried out for the task local output-module of the PSC. To

avoid file system blocking, a directory hierarchy for the threads was created using a serialized rank slide. For a full machine run this is possible in less than three minutes. During a 32.5 TB checkpoint, I/O throughput data from the file system monitoring showed a peak saturation at about 140 GB/s but a mean utilization of around 62 GB/s hinting at a congestion problem. Therefore a serialization for the writers was introduced giving the best throughput with four tasks per node. This increased the average I/O-bandwidth to almost 92 GB/s. The exact reason for the value of four will be investigated further.



Full Machine Scaling on SuperMUC

During the extreme scaling workshop at LRZ in 2014, several problems were identified that hindered scaling above 32k cores and limited it to 65k cores. These problems could be resolved, and we could demonstrate in May 2015 that the improved PSC scales up to the full machine size of 6 islands with 3,072 nodes and 86,016 cores on SuperMUC Phase 2. The results suggest that also 16 or even 18 islands on Phase 1 will be possible.

References

- 1 http://www.eli-laser.eu

contact:

Karl-Ulrich Bamberg, Karl-Ulrich.Bamberg@physik. uni-muenchen.de

Acknowledgements

The authors acknowledge the support of the Arnold Sommerfeld Center for Theoretical Physics at the Ludwig Maximilians University as well as useful editorial suggestions from Patrick Böhl. This work was supported by Grant No. DFG, FOR1048, RU633/1-1, by SFB TR18 project B12 and by the Cluster-of-Excellence Munich-Centre for Advanced Photonics (MAP). For computing resources the Gauss SuperMUC grant pr84me and the financial support of the KONWIHR-III program of the state Bavaria is acknowledged. For technical expertise in running and optimizing the code for large scale application we thank the Astrolab at the Leibniz Supercomputing Centre, LRZ.

Figure 2: Shown is a 3D visualization of a preliminary simulation of a thin foil. The circularly polarized plane wave laser (1024 W/m2) passed the foil, almost to the end of the box. The ion density is illustrated in white-black, electron density in yellow-green and the reflected electro magnetic intensity in blue-red. In the reflected light the longitudinal oscillation of electron sheets created a multi-layerstructured modulation giving raise to very short pulses of high frequency circularly polarized laser light (AXP).



[2] http://www.plasma-simulation-code.net/

[3] The Plasma Simulation Code: A modern particle-in-cell code with load-balancing and GPU support (arXiv:1310.7866)

- Karl-Ulrich Bamberg
- Fabian Deutschmann
- Constantin Klier
- Nils Moschuering
- Hartmut Ruhl

Computational and Plasma Physics, LMU Munich, Germany

Disentangling Evolution on the SuperMUC

In the following we describe how we addressed and solved some major computational challenges to develop software that allowed us to infer the largest (w.r.t. input data size) evolutionary trees of birds and insects published to date. Since the biological results have been broadly addressed in the popular press, we will exclusively focus on aspects related to parallel processing and computer science. Inference of phylogenetic trees, that is, the evolutionary history of distinct species has a plethora of applications in medical and biological research.

Reconstructing such trees represents a challenge because of the immense number of distinct evolutionary hypotheses (unrooted binary trees) that need to be assessed to identify the most plausible tree according to a given stochastic model of evolution. In fact, finding the tree that best explains the data at hand - the molecular sequences of the species under study - is known to be NP-hard. In other words, no efficient algorithm is known for finding 'the' most plausible evolutionary scenario. As a consequence, all widely used software tools for this purpose deploy so-called ad hoc heuristics to find 'good' trees without any guaranteed bounds.

The most commonly used method for determining how well the data is explained by a given evolutionary tree is the phylogenetic likelihood model [1]. To infer trees, the phylogenetic likelihood function is either used to calculate maximum likelihood trees or to sample the posterior probability distribution of trees based upon Bayesian statistics and Markov-Chain Monte Carlo sampling.

Irrespective of the specific probabilistic framework deployed, that is, Bayesian or maximum likelihood based inference, the underlying computational challenges are the same. The major computational bottleneck is the evaluation of the phylogenetic likelihood function on distinct alternative tree topologies and other model parameters. We will outline why supercomputers are required to conduct research in evolutionary biology in the subsequent section.

Motivation

Over the last years the field of molecular biology has witnessed a revolution with respect to molecular data acquisition. The emergence of so-called Next Generation Sequencing (NGS) technologies has dramatically decreased the time as well as cost that is required to extract molecular data from biological samples. In fact, sequencing a genome is currently getting cheaper at a faster pace than computing power cost decreases according to Moore's law. The entire field of evolutionary biology is thus facing a paradigm shift from a hypothesis-driven toward a data-driven science. Unlike a decade ago, now the challenge for biologists is not to obtain the data - albeit catching a tiny animal that moves fast is still not trivial- but to analyze the data.

Nonetheless, one should keep in mind that a mere matter-of-fact transition toward a data-driven science is dangerous, without critically thinking about the psychological pitfalls induced by desperately trying to detect something 'publishable' in an enormous pile of data.

Given the ability to generate huge molecular datasets, biologists increasingly wish to infer evolutionary histories on datasets that comprise the entire transcriptomes or genomes of the organisms under study. The main reason for using more data is the hope that this will allow for resolving longlasting debates with respect to the position of major clades in the tree of life. From the theoretical point of view, we know that the phylogenetic likelihood model is consistent when the sequence length goes to infinity. Thus, using sequences that are as long as they can get (i.e., whole-genome sequences) is reasonable. Nonetheless, biological phenomena such as, for instance, lateral gene transfer between species as well as gene duplication yield the underlying biological processes more complex. As a consequence, gene trees must not always be congruent with the corresponding species tree. The proper modeling of these phenomena currently represents a hot research topic. Putting aside all the simplifying assumptions we do make in our models, one of the major technical challenges when dealing with whole-genome datasets is to efficiently orchestrate phylogenetic likelihood calculations on supercomputers.

In the following we will describe the technical challenges we addressed and solved in the context of two large-scale phylogenetic analysis projects [2,3] that both made it to the cover page of Science in late 2014. In Figure 1 we show the evolutionary tree of insects [2] that was computed on Super-MUC. However, we will not focus on biological results, but on the computer science aspects that finally enabled us and our project partners to conduct these inferences.

Results

To parallelize likelihood calculations we distribute different parts of the genomes (e.g. distinct genes) to processors. All processors then calculate the likelihood score for their part of the data on the same underlying tree and subsequently communicate with each other to compute the overall likelihood score (across all genes) for that tree.

The first step in increasing the parallel efficiency of our codes consisted in a complete re-design of the parallelization scheme [4], since our original master-worker approach [5] proved to be inefficient for large datasets with hundreds to thousands of genes. This lead to improvements in parallel efficiency by up to a factor of three.



Applications

Figure 1: Phylogenetic tree of insects calibrated on an absolute time scale. From Misoff et al., SCIENCE 346:763 (2014). Reprinted with permission from AAAS.



Subsequently, we addressed the problem of data and hence, load distribution to processors. In current likelihood-based analyses each gene (or other subdivision of the genome) is assumed to evolve under a distinct evolutionary model because genes are exposed to different evolutionary pressures. For this reason, genes should ideally be assigned entirely to one processor, while at the same time each processor should perform calculations on the same amount of DNA characters. Since gene lengths differ, this represents a non-trivial scheduling task that is known from theoretical computer science as the NP-hard multi-processor scheduling problem. We first showed this analogy in [6] and were able to demonstrate that parallel execution times can be improved by up to one order of magnitude via common heuristic strategies for distributing genes to processors. However, this does not guarantee that all processors will operate on the same amount of genetic data. Thus, in a second step we developed an algorithm that allows for splitting up genes among processors, balances the amount of genetic data per processor, and, at the same time, minimizes the number of genes that are split up across different CPUs. We showed that, finding the optimal solution for this refined scheduling problem formulation is NP-hard as well, but we also developed an approximation algorithm that is guaranteed to find near-optimal results [7]. We then implemented this algorithm in our codes and showed that it consistently yielded the shortest run times, regardless of the amount of genes and number of processors.

We also focused on improving I/O performance for reading in the molecular data sets. To this end, we re-designed the file input formats as well as the concurrent file accesses carried out by

each processor [8]. These optimizations yielded a performance improvement for the I/O-intensive part of the code by a factor of 15. In addition, we also ported the code to the Intel Xeon PHI [9], albeit this version of the code was not used for the two Science papers mentioned above

Thus, apart from the two biological data analyses, the major outcome of our project is the development and release of the ExaML (Exascale Maximum Likelihood) and ExaBayes (Exascale Bayesian Inference [10]) open-source codes for phylogenetic analyses on supercomputers. Our focus was to explicitly develop codes for x86-based systems like the SuperMUC, since codes that run on such standard systems will be of greatest use to the global evolutionary biology research community.

Another major outcome is that we have developed generally applicable methods (parallelization scheme, I/O scheme, load distribution algorithm) for improving the scalability of codes that rely on computing the phylogenetic likelihood function. These techniques can thus be leveraged by every program for largescale phylogenetic analysis and are not specific to our own search strategies.

The project on insect evolution (www.1kite.org, [2]) is still on-going and we expect to analyze a dataset that contains an order of magnitude more species (1,500 insects) in 2015 and 2016.

Since we can now compute such large trees and because of the computational expansion space we created via our optimizations, we will focus on improving our simplistic models of evolution and replace them by more complex, adequate, and compute-intensive ones.

One major challenge is to develop and implement new models that will allow for reconstructing trees of populations of species and not just representative individuals of species as in the two aforementioned studies. With the continuous progress of next generation sequencing technologies, sampling multiple individuals per species of interest is now technically and economically feasible. We will thus need to develop models that can simultaneously capture population genetic (within species) as well as phylogenetic (among species) processes.

Another major challenge will be to further optimize the core phylogenetic likelihood kernel implementation. Over the years the community has developed an entire zoo of algorithmic tricks for speeding up likelihood calculations. However, none of them globally works best, since most of these shortcuts are dataset-dependent or require setting appropriate threshold parameters. We will thus develop a next generation phylogenetic inference code that implements all of these tricks and automatically selects the most efficient implementation for the dataset at hand using an appropriate calibration procedure. Finally, we plan to integrate all techniques developed for ExaML and ExaBayes into the desktop and server versions of our codes.

References

[1] Felsenstein, J.

Evolutionary trees from DNA sequences: a maximum likelihood approach, Journal of molecular evolution, 17(6), 368-376, 1981.

[2] Misof, B., Liu, S., Meusemann, K., Peters, et al. Phylogenomics resolves the timing and

pattern of insect evolution, Science, 346(6210), 763-767, 2014.

[3] Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., et al. Whole-genome analyses resolve early branches in the tree of life of modern birds,

[4] Stamatakis, A., Aberer, A. J. Novel parallelization schemes for large-scale likelihood-based phylogenetic inference, In Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on (pp. 1195-1204), IEEE, 2013.

- Izquierdo-Carrasco, F. 2066, 2012.
- [6] Zhang, J., Stamatakis, A.
- Stamatakis. A.
- [8] Kozlov, A. M., Aberer, A. J., Stamatakis, A. btv184, 2015.

[9] Kozlov, A. M., Goll, C., Stamatakis, A. Efficient computation of the phylogenetic

likelihood function on the Intel MIC architecture, In Parallel & Distributed Processing Symposium Workshops (IPDPSW), 2014 IEEE International (pp. 518-527), IEEE, 2014.

contact:

Alexandros Stamatakis, Alexandros.Stamatakis@h-its.org Applications

Science, 346(6215), 1320-1331, 2014.

[5] Stamatakis, A., Aberer, A. J., Goll, C., Smith, S. A., Berger, S. A.,

RAxML-Light: a tool for computing terabyte phylogenies, Bioinformatics, 28(15), 2064-

The multi-processor scheduling problem in phylogenetics, In Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International (pp. 691-698), IEEE, 2012.

[7] Kobert, K., Flouri, T., Aberer, A.,

The divisible load balance problem and its application to phylogenetic inference, In Algorithms in Bioinformatics (pp. 204-216), Springer Berlin Heidelberg, 2014.

ExaML version 3: a tool for phylogenomic analyses on supercomputers, Bioinformatics,

[10] Aberer, A. J., Kobert, K., Stamatakis, A.

ExaBayes: massively parallel Bayesian tree inference for the whole-genome era, Molecular biology and evolution, 31(10), 2553-2556, 2014.

- Alexandros Stamatakis^{1,2}
- Alexey Kozlov¹
- Andre J. Aberer¹
- ¹ Heidelberg Institute for Theoretical Studies, Germany
- ² Karlsruhe Institute of Technology, Germany

SuperMUC Phase 2: Challenges of Peta-scale Applications

In spring 2015, the Leibniz Supercomputing Centre (Leibniz-Rechenzentrum, LRZ), installed its new Peta- Scale System SuperMUC Phase 2, consisting of 6 islands with a total number of 86,016 Intel Xeon E5-2697v3 cores (Haswell), with a theoretical peak performance of 3.6 PFlop/s and a sustained LINPACK performance of 2.81 PFlop/s. Selected user groups had the opportunity to use the new system for 28 days in May and June, 2015. During this friendly user phase, the users could run jobs up to the full system size. This article presents results and the lessons learned from the operational point of view.

The friendly user phase had a daytime operation during office hours when users could submit jobs up to the full system size. In the evening, the special queue was replaced by the general queue, allowing jobs of up to 3 islands (43,008 cores) and 2 hours wallclock limit. Selected users had the additional opportunity of running their applications on the whole system for up to 48 hours on weekends. In total, 41 scientists with 14 different applications were participating. 63.4 million core-hours were available, out of which 43.8 million core-hours could be used by the applications. A total number of 6,751 jobs with

Software	Application Science	Project Investigator	Details
SeisSol	Seismology	S. Rettenberger (TUM)	IBM MPI, pHDF5, OMP
SL-Hydro	Astro Physics	P. Edelmann (Uni Würzburg)	IBM MPI, OMP
PSC	Plasma Physics	K. Bamberg (LMU München)	IBM MPI, local I/O
Musubi	Lattice Boltzmann	H. Klimach (U Siegen)	IBM MPI
BQCD	QCD	M. Allalen (LRZ)	IBM MPI, Intel MPI, OMP
CIAO	CFD, Combustion	T. Falkenstein (RWTH Aachen)	IBM MPI, MPI I/O
LS1-Mardyn	Material Science	N. Tchipev (TUM)	IBM MPI, IntelMPI, OMP
FLASH	Astro CFD	C. Federrath (ANU)	IBM MPI, pHDF5
waLBerla	Lattice Boltzmann	F. Schornbaum (FAU Erlangen)	IBM MPI, OMP
Gadget	Cosmology	K. Dolag (USM)	IBM MPI, FFTW3, OMP
ILBDC	Lattice Boltzmann	M. Wittmann (FAU Erlangen)	IBM MPI, IntelMPI
Vertex3D	Stellar Astrophysics	T. Melson (MPA)	IBM MPI, pHDF5, MKL, OMP
Iphigenie	Molecular Dynamics	G. Mathias (LMU München)	IBM MPI, IntelMPI
GPI-2 / GASPI	GASPI Library	M. Kühn (Fraunhofer ITWM)	IBM MPI

Figure 1: Overview of the applications that were running on SuperMUC Phase 2 during the extreme

Results

The performance and system monitoring tools were already active for the extreme scale-out workshop. Some statistics are presented in the Table. a runtime of more than 1 minute were processed by the system. 2,054 jobs had a runtime of more than 10 minutes. The compute nodes of SuperMUC Phase 2 contain 2 sockets with 28 physical with hyper threading active. Only one software package used 56 threads for testing purposes. Most of the applications were using 28 tasks per node. However, a significant amount of projects was using a hybrid approach with varying numbers of OpenMP threads on each node. The two main cases for mixed OpenMP/MPI usage was 2 MPI tasks and 14 OpenMP threads, or 4 MPI tasks and 7 OpenMP threads. For some codes, the processor architecture of SuperMUC Phase 2 with core numbers that are multiples of 7 instead of powers of 2 were a challenge and required some re-adjustment. The power consumption of the system scaled nearly linear with the job size, with more than 1.2 MW for the most power hungry job. It is very crucial for the power consumption of large jobs that the node level performance of the application is optimized. The dif-

Watts/	D		MDI to alte
Node	Percent IDP	cores	IVIPI TASKS
348	97%	86,016	3,072
254	71%	86,016	86,016
249	70%	86,016	86,016
244	68%	86,016	86,016
241	67%	86,016	86,016
237	66%	86,016	86,016
234	65%	86,016	172,032
229	64%	86,016	86,016
223	62%	86,016	172,032
173	48%	86,016	6,144
166	46%	86,016	86,016
159	44%	86,016	6,144
108	30%	84,000	84,000
98	27%	86,016	57,344

scale-out workshop.

ference in power consumption for the scale-out jobs that use all 86,016 cores is quite dramatic, ranging from 120 kW to 1.2 MW. This makes it quite difficult to design the electrical infrastructure,

cores and up to 56 tasks per node and also influences the power bill. For with hyper threading active. Only one software package used 56 threads for testing purposes. Most of the applications were using 28 tasks per node. and also influences the power bill. For example, the compute bound simulation code SeisSol used several times more energy than the memory bound Lattice Boltzmann Code ILBDC.

Lessons learned

The extreme scale-out workshop at LRZ again showed that preparation of a simulation campaign is crucial for the success of the project. Users have to address all the relevant issues, such as scaling tests, choice of OpenMP/MPI balance, intervals for checkpoint or restart files, well prepared input files, I/O strategy, and risk management. Under these conditions, it was possible to use a brand new system like SuperMUC Phase 2 directly after installation and obtain scientific results right from the start. One big advantage of the extreme scale-out workshop was that only one code was running at a time and this code was filling up the whole system. Thus, hardware bugs were much easier to detect and resolve. One especially hard to find bug was a combination of two MPI timeouts that coincided with a hardware problem. During normal user operation, this error would have been almost impossible to detect because of the low probability of three errors occurring simultaneously for smaller jobs. Another important observation is that standard MPI is at its limits. The stack size of the MPI stack is growing on each node and for systems with >80,000 cores it occupies a significant amount of memory. The startup time can exceed several minutes and become a significant part of the overall run time. One way to overcome this bottleneck is the use of hybrid OpenMP/MPI programming models. However, this implies very deep system knowledge on the user side, since process pinning and the

choice of the OpenMP/MPI balance has to be evaluated and decided by the user. Furthermore, I/O strategies have to be developed and tested before the complete system can be used. Using modern parallel I/O libraries is crucial.

Even for hybrid OpenMP/MPI set-ups with a single MPI-task per node, problems arise due to internal limits of the MPI send/receive buffer. This limit is caused by the Integer*4 Byte implementation of the MPI index values. Such problems can be overcome through internal buffering in the applications.

The close interaction between scientists and HPC experts, as well as the thorough preparation by the users and the LRZ, was crucial for the success of this kind of workshop. There are now 25 applications that scale up to the full system size of SuperMUC (147,456 cores of Phase 1 and 86,016 cores of Phase 2, respectively). Two applications showed a sustained performance of more than one PFlop/s for more than 20 hours. The next extreme scale workshop is scheduled for January 2016.

- Nicolay Hammer
- Helmut Satzger
- Momme Allalen
- Anupam Karmakar
- Luigi lapichino
- Ferdinand Jamitzky

Leibniz Supercomputing Centre (LRZ). Germany

The authors would like to thank their LRZ colleagues Matthias Brehm, Reinhold Bader, Alexander Block, Markus M. Müller, Antonio Ragagnin, Vasilios Karakasis, Dieter Kranzlmüller, Arndt Bode, and Herbert Huber. Special thanks to the participants of the extreme scaling workshop: Martin Kühn, Rui Machado, Daniel Grünewald, Philipp V. F. Edelmann, Friedrich K. Röpke, Markus Wittmann, Thomas Zeiser, Gerhard Wellein, Gerald Mathias, Magnus Schwörer, Konstantin Lorenzen, Christoph Federrath, Ralf Klessen, Karl-Ulrich Bamberg, Hartmut Ruhl, Florian Schornbaum, Martin Bauer, Anand Nikhil, Jiaxing Qi, Harald Klimach, Hinnerk Stüben, Abhishek Deshmukh, Tobias Falkenstein, Klaus Dolag, and Margarita Petkova.

contact: Ferdinand Jamitzky, Ferdinand.Jamitzky@lrz.de

Visualizing 10¹¹ Particles from **Cosmological Simulations**

Modern cosmological simulations can contain more than 100 billion particles and simulate the evolution of the Universe over several billion years. Besides numerical and statistical analysis of the raw data, visualizations allow cosmologists to literally fly through this simulated Universe to find remarkable structures within the simulation and to study their dynamic behavior. Moreover, such movies allow the general public to gain insights into the structure formation of the Universe. Here, we present the making of a visualization showcase presented at the Supercomputing Conference SC15 in Austin, Texas: a movie of a fly-through a cosmological simulation of 61 billion particles (320 TByte raw data), rendered on 1,280 CPU-cores, using 90,000 corehours. For the URL of the showcase video and additional information, please see [1]. We also present a new version of the rendering software, realized with GPI-Space, that reduced the time to solution by a factor of 10.

P-Gadaet3-XXL

The cosmological simulation code P-Gadget3-XXL is a cosmological, highly optimized and fully OpenMP/MPI parallelized TreePM-MHD-SPH code that enables large-volume, high-resolution cosmological simulations which follow in detail the various physical processes that are needed for a comparison with the experimental data coming from current and forthcoming astronomical surveys and instruments like PLANCK, SPT, DES and eROSITA. For more details, please see [2,3].

The Data-Set

The data-set originates from the Gauss Centre for Supercomputing (GCS) Large-Scale-Project "Magneticum" [1], which was carried out by a collaboration of scientist from the University Observatory Munich (LMU), the Excellence Cluster Origin and Structure of the Universe, and the Leibniz Supercomputing Center (LRZ).





Figure 2: GANTT-chart of GPI-Splotch: Loading of data, rendering and saving the final images are performed in parallel, leading to an overall performance increase of 10x compared to the hybrid OpenMP/MPI version of Splotch (running on 2.048 cores. Intel Westmere).

Figure 1: Workflow of the parallelized Splotch algorithm. For more details see [6].

All simulations of this project were computed on the SuperMUC system operated at LRZ. The total simulation campaign consumed more than 60 million core-hours of computing time on up to 131,000 cores. It consisted of several medium to large-sized cosmological hydrodynamics simulations, spanning the range from medium to extremely high numerical resolution. Various parameter setups defined in five different types of cosmological boxes were used. The last of these simulations finished in July 2015, data analysis has just started and first scientific publications from this project are expected to be published by the end of 2015. The Magneticum project includes one of the largest cosmological hydrodynamics simulations to date (including sub-resolution models for baryonic matter at galactic scales), evolving 180 billion particles over 100,000 time steps in a box of (2.7 Gpc/h)³ size with a scientific output of more than 300 TByte, using 25 million core-hours on 86,016 cores of SuperMUC Phase 2 (Intel Haswell). With respect to the number of particles, this simulation is 20 times larger than the previous generation of such simulations and corresponds roughly to a size of about 10% of the visible Universe. Only simulations of this size contain enough statistics to enable the analysis of the properties of the largest structures of the universe. For the movie presented here, a box with smaller size but higher resolution (compared to the largest box mentioned above)



Figure 3: Screenshot from the movie.

was used, as it shows finer details, containing 61 billion particles. Rendering the movie using the Splotch visualization tool ran on 1,280 cores on the fat node island of SuperMUC Phase 1 (Intel Westmere), using 32 TByte of main memory for about 72 hours (i.e. about 90,000 core-hours). The movie had to be rendered several times to find the ideal settings. The movie shows a flight through a small part of the evolving Universe (0.25% of the visible Universe), covering the evolution from redshift z = 9 to z = 0, i.e. a time span from around 560 million years after the Big Bang until today. Besides numerical and statistical analysis of the raw data, these visualizations allow cosmologists to find remarkable structures within the simulations and to study their dynamic behavior. Moreover, such movies allow the general public to gain insights into the structure formation of the Universe.

The Renderer Splotch

Splotch is a light-weight, fast, and publicly available rendering software for exploration and visual discovery in particle-based datasets coming from astronomical observations or numerical simulations [5]. The rendering algorithm is designed in order to deal with point-like data, optimizing the ray-tracing calculation by ordering the particles as a function of their "depth", defined as a function of one of the coordinates or other associated parameters. Realistic three-dimensional impressions are reached through a composition of the final color in each pixel, properly calculating emission and absorption of individual volume elements. The strengths of the approach are production of high-quality imagery and support for very large-scale data sets through an effective mix of the OpenMP and MPI parallel programming paradigms (a version using GPGPUs is also available). The movie presented here was rendered using the OpenMP/MPI version of Splotch.

The new and improved GPI-Splotch

While creating the visualization showcase presented here, we also tested an integration of Splotch into GPI-Space (called GPI-Splotch). The development of GPI-Splotch was motivated by the observation that the OpenMP/MPI version of Splotch had a parallel efficiency of only 26% on 32 nodes (512 cores) of SuperMUC, which was determined by a timing analysis of OpenMP/MPI-Splotch. The integration into the GPI-Space framework took about three months and was done in order to reorganize the data-flow in Splotch, thus allowing an overlay of the rendering part with the non-parallel parts of the algorithm like file I/O or the aggregation of partial images, as well as introducing a dynamic load balancing scheme (see Fig. 2). GPI-Splotch uses an interface with only three different functions, which can be developed, tested and compiled without GPI-Space, and can then be connected later-on with GPI-Space through a dynamic library. The virtual memory created by GPI-Space to store partial results in memory is based on GPI-2 and profits from the low latencies and high bandwidth that GPI-2 allows [4]. As a small benchmark, we chose a dataset with 5 billion particles (167 GByte) to render 3,425 frames (fly-through the data-set at a fixed time step), generating 161 GByte of picture data. We rendered this benchmark using OpenMP/ MPI-Splotch, running on 32 nodes (512 cores). This was the minimal possible number of nodes due to memory requirements. The total rendering time for this benchmark was 14.5 hours. A timing analysis of this benchmark revealed that the compute intensive and parallelizable rendering part took 26% of the total time (20% for rendering, 4% for 3D transformations, and 2% for particle coloring), while the rest was taken by load imbalances and the non-

parallel post-processing and writing of the images. Reading the input file was negligible with only 0.3% of the total time. GPI-Splotch needed less memory, which allowed to run the benchmark on 16 nodes, where it already showed a 3.5x performance increase compared to OpenMP/MPI-Splotch on 32 nodes. Moreover, GPI-Splotch scales up to at least 128 nodes (2,048 cores - the largest set-up tested with this small benchmark). On 128 nodes, OpenMP/ MPI-Splotch had basically the same runtime as on 32 nodes, whereas GPI-Splotch was 10x faster here. The next target for GPI-Splotch is to re-write the algorithm that interpolates between time-steps, which currently scales as O((MPI-ranks)²), which currently limits the scalability of Splotch.

References

- visualization showcase: http://www.mpa-garching.mpg.de/ HydroSims/Magneticum/Images_Movies/ left_hd.avi More Information and additional videos: http://magneticum.org [2] Springel, V.
- The cosmological simulation code GAD-GET 2, MNRAS 364(2005)110, 2005
- [3] Bazin, G., Dolag, K., Hammer, N. Gadget3: Numerical Simulation of Structure Formation in the Universe, inSiDE 11(2013)2. 2013
- [4] GPI-Space: http://www.gpi-space.com GASPI Forum: http://www.gaspi-forum.org
- [5] Jin, Z., Krokos, M., Rivi, M., Gheller, C., Dolag, K., Reinecke, M. High-performance astrophysical visualization using splotch. Procedia Computer Science 1(1), 1775-1784 (2010), arXiv:1004.1302, 2010

contact: Helmut Satzger, helmut.satzger@lrz.de

[1] Direct URL to the movie of the

- K. Dolag¹ M. Reinecke² N. Hammer³ F. Jamitzky³ L. lapichino³ • H. Satzger³
- M. Kühn⁴
- M. Rahn⁴
- ¹ Universitäts-Sternwarte München, Germany
- ² Max Planck Institut für Astrophysik, Germany
- ³ Leibniz Supercomputing Centre (LRZ), Germany
- ⁴ Fraunhofer Institute for Industrial **Mathematics** ITWM, Germany

Analysis of 3D Point Clouds using a Parallel DBSCAN Clustering Algorithm

For decades now, scientists have collected huge amounts of data to be analyzed. Machine learning algorithms, which find important information in the data, have become universal tools in data science today. Still, analyzing large and high-dimensional data collections exceeds the capability of default machine leaning implementations on standard computers. The High Productivity Data Processing Research Group at the Forschungszentrum Jülich (FZJ) works on parallel and scalable machine leaning software. This enables a data analysis that is able to leverage the powerful capabilities of modern High Performance Computing (HPC) environments. Driven by the needs of scientific users, their newest parallel implementation of a clustering algorithm, named HPDBSCAN, has reached the state-of-the-art performance in terms of memory usage and speed up.



Figure 1: Example of points clustered by DBSCAN. The algorithm found two clusters (blue and green) and 4 noise points (red).



Figure 2: Speed up of HPDBSCAN, tested on different data sets: - b (Bremen data set), t (Twitter data set).

DBSCAN

DBSCAN - or density based spatial clustering for applications with noise - is the original serial clustering algorithm formulated in 1996 by [1] et al. at the University of Munich. Over the years it became, according to Microsoft Research, the most cited machine learning algorithm [2]. Its core idea is rather simple. While iterating through a dataset, the algorithm tries to find dense areas, which are defined based on the number of neighboring points. These form cluster cores, which, through recursive expansion, are enlarged in the process. Points that cannot be assigned in that fashion are considered to be noise within the dataset.

Parallelization Strategy

The parallelization strategy of the algorithm entails a divide-and-conquer approach. This means that each parallel processor locally clusters a subset of the data, which it then merges with its spatial neighbors. The biggest chal-



Figure 3: Point cloud of Bremen's old town.



Figure 4: Denoised and clustered point cloud of the old city of Bremen. Each color represents another cluster.

lenges in the parallelization required the spatial decomposition, the load balancing of skewed datasets, as well as the lock-free and communicationoptimized merging. Using these techniques, the group was able to achieve scalable performance (see Fig. 2) outperforming previous parallelization attempts of DBSCAN by an order of magnitude in terms of computation time and memory consumption using different datasets [2]. This Highly Parallel HPDBSCAN is implemented as Hybrid MPI/OpenMP application and uses HDF5 files for parallel I/O.

Point Cloud Analysis

One of the application domains where DBSCAN can be used is point cloud analysis. A point cloud is set of threedimensional points that represent an object or environment. These point clouds are taken by specialized cameras and are comparable to an image in 3D.

Figure 3 shows an example using the Bremen data [3]. Using these point clouds, engineers can reconstruct and model the scanned objects and subsequently use them to e.g. search for leaks, deformations and so forth in the object. This is used in, e.g., industry plant safety monitoring, automatic map creations and large-scale archeological excavations. HPDBSCAN can be used for two tasks supporting research questions. First, it can be used to denoise the point clouds from, e.g., false readings, especially those from small objects moving through the scenery. In a second step, it can be used to cluster together these points in order to identify individual objects and distinguish them from others in an automated fashion. These segmented objects can then be classified either using human experts or other machine learning algorithms. An example of such a point cloud analysis step, based on data shown in Figure 3, is presented below in Figure 4 where the old town of Bremen has been scanned for thermal leakage.

Summary and Outlook

HPDBSCAN is a highly scalable implementation of a widely used clustering algorithm called DBSCAN. It is open source and be obtained through source code repositories and compiled. Currently, HPDBSCAN is being deployed on XSEDE resources and evaluated for its permanent installation value. In the future, we will be using HPDBSCAN is foreseen to be used from partners from the Netherlands, the UK and France in order to do archeological data analysis of roman ruins. HPDB-SCAN is also actively used in another research project that is worked on in collaboration with the University of Gothenburg, where we automatically detect water mixing events in the Koljoefjords in Sweden. A wider collection of other scalable machine learning algorithms, by the name of JuML, is currently under development, and is going to be available soon as an opensource package.

References

[1] Ester, M. e. a.

A density-based algorithm for discovering clusters in large spatial databases with noise, Kdd. Vol. 96. No. 34, 1996

- [2] Patwary, M., Palsetia, D., Agrawal, A., Liao, W.-k., Manne, F., Choudhary, A. A new scalable parallel DBSCAN algorithm using the disjoin-set data structure, in High-Performance Computing, Networking, Storage and Analysis (SC), 2012
- [3] Borrmann D., Nüchter, A.

Robotic 3D Scan Repository, 18 03 2015. [Online]. Available: http://kos.infor matik.uni-osnabrueck.de/3Dscans

contact:

Christian Bodenstein, c.bodenstein@fz-juelich.de Markus Götz, m.goetz@fz-juelich.de Morris Riedel, m.riedel@fz-juelich.de

Directing the Morphology of amphiphilic Molecules

Amphiphilic molecules contain at least two structural units that thermodynamically repel each other. Since the two incompatible blocks are covalently bonded into a single molecule, they cannot macroscopically phase separated but, instead, self-assemble into spatially modulated structures whose characteristic length scale is dictated by the molecular extension. Typical examples include the self-assembly of lipid molecules, which are comprised of a hydrophilic, polar head and a hydrophobic tail, into bilayer membranes or synthetic block copolymers, which are comprised of two incompatible flexible chain molecules that are joined at their ends and self-assemble into periodic microphases. Despite the differences in the chemical nature of the constituents and the type of interactions, both - biologically relevant lipids as well as synthetic block copolymers - spontaneously form similar structures (e.g., lamellar sheets or wormlike micelles). The structure formation is dictated by the universal competition between the free-energy cost of the interface between the incompatible components and the entropy loss of arranging these molecules uniformly in space.

This delicate balance gives rise to minuscule free-energy differences between different morphologies (on the order of a fraction of the thermal energy scale kT per molecule), and there exist many competing metastable structures (alternate periodic arrangements, defect structures like dislocations, or localized structures like hydrophobic bridges between lipid membranes). This feature is corroborated by the

protracted annealing times required to observe well-ordered morphologies in block copolymers or the requirement of specialized proteins that provide the free energy required to overcome the barrier in e.g., pore formation, fusion and fission of membranes. In fact, the complex, rugged free-energy landscape of self-assembling amphiphiles has been likened to that of glass-forming materials. The morphology often does not reach the thermodynamically stable state of lowest free energy but, instead, becomes trapped in a metastable state. By exploring these metastable states and the free-energy barriers that separate them, one can either reproducibly trap the system in desired non-equilibrium morphologies [1] or accelerate equilibration of block copolymer structures [2] or control collective changes of membrane topology involved in cellular transport processes [3,4].

Morphological transformations of amphiphilic, soft-matter systems involve the cooperative rearrangement of many molecules on time and length scales ranging from milliseconds to minutes and nanometers to micrometers for lipids and polymers, respectively. These scales are challengingly small for experimental imaging techniques yet too large for atomistic modeling. Since the transformations often involve highly bent interfaces or strongly stretched molecular conformations, also phenomenological continuum models cannot accurately capture them. In turn, coarse-grained models that only incorporate the relevant degrees of freedom are well suited to explore the universal behavior of amphiphilic structure formation and provide direct insights into

Christian Bodenstein

Supercomputing

Centre (JSC).

Germany

Markus Götz

Morris Riedel

Jülich

the kinetics as well as the free-energy landscape. The following two examples illustrate recent progress.

Process-directed Self-assembly of Block Copolymers on chemically guiding Patterns

Block copolymer lithography directs the self-assembly of block copolymers in thin films by sparse, lithographically fabricated, chemical or topographical substrate patterns into dense nanostructures with a critical dimension of a few nanometers. Applications in microelectronic industry require an extraordinarily low defect density of less than 1 defect per 100cm². Computer simulation and self-consistent field theory demonstrated that the excess free energy of a defect is several 100 kT, making the probability that thermal fluctuations generate defects in an initially ordered structure vanishingly small. However, since defects are ob-



Figure 1: top left: Minimum free-energy path from a tight, metastable dislocation dipole in a symmetric diblock copolymer thin film to its equilibrium lamellar structure. The two curves correspond to a laterally homogeneous surface, $\Lambda N=0$, and a stripe-patterned surface. The lower left inset depicts the defect morphology, whereas the upper right inset presents the detailed structure of the first saddle point that constitutes the highest barrier. top right: Dependence of the defect free energy (diamonds) and the highest barrier along the defect annihilation path (squares for laterally homogeneous surface and triangles for stripe patterned surface) on the incompatibility xN. Free energies are scaled by the film thickness, the invariant degree of polymerization and the thermal energy scale kT. (adapted from Ref [5] © American Physical Society 2014). bottom: Kinetics of self-assembly of a symmetric copolymer on a stripe pattern. The stripe spacing L_s=1.83R_a (with R_a being the molecule's end-to-end distance) is larger than the equilibrium spacing 1.68R, and 1.58R, at xN=30 and 20, respectively. Quenching the system from the disordered phase directly to a large incompatibility $\chi N=30$ does not lead to defect-free assembly even at protracted annealing times, 111τ . A shallow quench to $\chi N=20$, however, results in defect-free assembly at 3.3T although the mismatch between the equilibrium spacing and the stripe pattern is larger. The reason is the vanishing of the defect-annihilation barrier around $\chi N^*=26$ (see top-right panel). If the ordered structure is quenched a second time to $\chi N=30$ the stripe pattern remains free of defects, only the segregation between the domains increases. (adapted from Ref. [7] © Materials Research Society 2015).

served in experiments, they must arise during the self-assembly process and, thus, it is important to understand the kinetics of self-assembly and defect annihilation [5,6].

Typically the kinetics of self-assembly can be divided in two stages (see Fig. 1 bottom row) [2]: In the initial stage, the homogeneous structure after solvent evaporation is instable [6], and local domains of incompatible blocks form. This spinodal microphase separation is directed by the lithographic substrate pattern, which imparts overall long-range orientation and registration onto the copolymer morphology, and it lasts the time, τ , a copolymer needs to diffuse its molecular size. The bottom left snapshots of Figure 1 reveal that the morphologies in this early stage are riddled with defects. In the second stage [2,5,7] - defect annihilation and grain formation - defects move in response to long-range strain fields, they collide and may annihilate.

Computer simulations provide insights into defect motion and annihilation mechanisms. Using computer simulations of a highly coarse-grained soft polymer model and self-consistent field theory, we have studied dislocation defects in lamella-forming diblock copolymers [5,7]. Dislocations can easily move along the stripe directions (climb motion) and dislocations with opposite Burgers vectors attract each other. This attraction – Peach-Koehler force - is dominated by near-field and boundary effects, and we find that defects attract each other with a distance-independent force. The dislocation motion critically depends on the distance between the cores perpendicular to the stripes. Upon collision they may spontaneously annihilate

(Fig. 1 bottom, lower row $\chi N=20$) or form long-lived tight dislocation pairs or disclinations (Fig. 1 bottom, upper row $\chi N=30$).

The removal of a metastable tight dislocation pair is a thermally activated process, and we have studied the detailed mechanism by the string method [5]. This computational technique allows us to determine the pathway of defect annihilation and the concomitant free-energy barriers without prior assumption of a reaction coordinate. The top left graph of Figure 1 depicts the free-energy change along the path, α , for different strengths, AN, of the chemical guiding pattern. It consists of sequential breaking and re-joining of connections and, although the starting and ending morphologies are quasitwo-dimensional, a proper description of the transition state requires threedimensional calculations. The top right graph presents the excess free energy of defects and the largest barrier along the defect annihilation path as a function of incompatibility, xN. The defect free energy increases approximately linearly with xN and extrapolates to zero around the order-disorder transition, χN_{0DT} =10.5. The barrier of defect annihilation also linearly depends on χN but it vanishes around $\chi N^*=18$ on an unpatterned surface, $\Lambda N=0$, and we can shift this boundary to larger incompatibilities, $\chi N^*=24$, by increasing the selectivity of the chemical guiding pattern. Thus there is an optimal window of incompatibility, $10.5 < \chi N < \chi N^*$, where defects spontaneously annihilate yet the probability to create defects by thermal fluctuations is vanishingly small. Indeed, the bottom of Figure 1 demonstrates that reducing the incompatibility below XN* for a short initial period can dramatically improve the directed self-assembly [7].

Two Stages of Dynamin-mediated Fission of Membrane Tubes

The structuring of living organisms into cells and sub-cellular compartments is maintained by lipid bilayers undergoing frequent but carefully regulated topological changes [8,9]. During membrane fission a continuous hourglass-shaped membrane tube is divided into two separate bilayers facing each other. Like its reverse, membrane fusion, the intermediate stages of this process involve energetically unfavorable, highly bent morphologies and dynamin proteins provide free energy to substantially deform the lipid bilayer. In the course of fission, dynamin proteins form helical assemblies around the membrane tube and undergo a conformational change upon addition of GTP (Guanosine triphosphate) causing simultaneous constriction, elongation, and twisting [8,9].



Figure 2: Coarse-grained approach to modeling localized constriction of a cylindrical membrane tube by dynamin proteins. (A) Local constriction of a membrane tube by a ring of hexagonal disks representing PHD of dynamin. The disks (blowup) consist of polar (red) and apolar particles (green), which are connected by a network of stiff bonds (grey). For better visibility of the bonds, some particles have been reduced in size. The snapshots in (A) illustrate the transient closure of the tube's interior ("flickering") (C) The disks are evenly distributed over rings encircling the tube, the center of mass of each disk fixed on the ring (blue points). The overall position of the disk was not fixed so the normal to the disk surface (orange arrows) has an arbitrary direction. The two equal rings of radius r separated by the distance d were used for the localized constriction of pre-equilibrated membrane cylinder. (D) To amplify the local membrane constriction, two double-ring systems were used. By representing each rung of the dynamin complex by two rings separated by a small distance Δx and assigning a larger radius, characterized by Δr , to the outer ring, one effectively creates an effective tilt angle, a. The colored circles represent the anchor points, to which the peptide disks are restrained. Note that for better visibility a too large value of Δx in relation to the distance d between the double rings is shown in the diagram. (B) The two double-ring system induces a topological transition from membrane tube to the hemifission structure that is depicted in the snapshot. The constraints, however, fail to induce complete fission during the simulation runs. (adapted from Ref [8] © Nature Publishing Group 2015).

Computer simulations of coarse-grained models can explore the role of the conformation changes of the fission protein and yield insights about the sequence of morphological changes. As a first step we represented the PH (Pleckstrin Homology) domain of the large dynamin protein as amphiphilic disk that shallowly inserts into the membrane constricting the membrane tube and locally inducing positive curvature (see Fig. 2) [8,9].

The fission process is divided into two distinct stages: First constriction and curvature gives rise to flickering states (Fig. 2 left), where the lumen of the membrane tube reversibly opens and closes, and eventually induces the transition from a membrane tube to a metastable hemi-fission intermediate – a worm-like micelle that is the analog of the stalk in membrane fusion. Tailoring the position of the PH domains we can induce tilt and curvature, thereby reducing the required constriction and facilitating the formation of the hemifission intermediate [8].

The hemifission intermediate (Fig. 2 right) is surprisingly (meta)stable [9]. Radial constriction of the dynamin scaffold alone is not sufficient for the hemifission intermediate to spontaneously rupture and complete the fission process. Our simulations indicate that, instead, disassembly of the dynamin scaffold and axial tension may facilitate the final severance of the hourglassshaped bridge. The detailed pathway from hemifission to fission and the role of dynamin, however, remains yet to be explored.

These examples illustrate that the combination of meaningful coarse-grained models for soft matter, efficient simu-

lation techniques, and computational resources provided by supercomputing centers now enable us to explore the complex free-energy landscape of collective processes in self-assembling materials. Ideally these insights will allows us to design processes [1] i.e., time protocols of thermodynamic control parameters (pressure or solvent properties) or localized stimuli imparted by functional molecules that reproducibly direct the collective kinetics into desired morphologies. Such a process-directed self-assembly will allow access to a plethora of nonequilibrium, metastable structures. This investigation clearly is in its infancy. Whereas the examples indicate its usefulness in two different contexts, there remain many open questions related inter alia to a proper choice of the relevant collective order parameter(s), the relation between the molecular dynamics of individual molecules and the collective kinetics of the morphology, as well as the type of the timedependent external control or internal stimuli-response.

I have benefitted from stimulated and enjoyable collaborations with Israel Barragan Vidal, Juan de Pablo, Fabien Léonforte, Vadim Frolov, Marc Fuhrmans, Su-Mi Hur, Weihua Li, Paul Nealey, Juan Carlos Orozco Rey, Sandra Schmid, Yuliya Smirnova, Dewen Sun, Ulrich Welling, and Guojie Zhang. Financial support has been provided by the DFG under grants Mu1674/12, Mu1674/14 and SFBs 803, 937, as well as the FP7 project CoLiSA.MMP and the Volkswagen foundation. Computing time at the Neumann Institute for Computing, Jülich, as well as the HLRN Hannover/Berlin and the GWDG Göttingen is gratefully acknowledged.

References

[1] Müller, M., Sun, D.W.

Directing the self-assembly of block copolymers into a metastable complex network phase via a deep and rapid quench, Phys. Rev. Lett. 111, 267801, 2013

[2] Li, W.H., Müller, M.

Defects in the self-assembly of block copolymers and their relevance for directed self-assembly, Annu. Rev. Chem. Biomol. Eng. 6, 187, 2015

[3] Smirnova, Y.G., Fuhrmans, M., Barragan Vidal, I.A., Müller, M.

Free-energy calculation methods for collective phenomena in membranes, J. Phys. D: Appl. Phys. 48, 343001, 2015

[4] Fuhrmans, M., Marelli, G., Smirnova, Y.G., Müller, M.

Mechanics of Membrane Fusion / Pore Formation, Chem. Phys. Lipids 185, 109, 2015

[5] Li, W.H., Nealey, P.F., de Pablo, J.J., Müller, M

Defect removal in the course of directed self-assembly is facilitated in the vicinity of the order-disorder transition, Phys. Rev. Lett. 113, 168301, 2014

[6] Hur, S.M., Khaira, G., Ramirez-Hernandez, A., Müller, M., Nealey, P.F., de Pablo, J.J. Simulation of defect reduction in block copolymer thin films by solvent annealing, ACS Macro Letters 4, 11, 2015

[7] Müller, M., Li, W.H., Orozco Rey, J.C., Welling, U.

Defect annihilation in chemo-epitaxial directed assembly: Computer simulation and self-consistent field theory, MRS Proceedings 175, mrsf14-1750-kk03-05 (2015)

[8] Fuhrmans, M., Müller, M.

Coarse-grained simulation of dynamin-mediated fission, Soft Matter 11, 1464, 2015

[9] Mattila, J.-P., Shnyrova, A.V., Sundborger, A.C., Rodriguez Hortelano, E., Fuhrmans, M., Neumann, S., Müller, M., Hinshaw, J.E., Schmid, S.L., Frolov, V.A. A hemi-fission intermediate links two mechanistically distinct stages of membrane fission, Nature 524, 109, 2015

contact: Marcus Müller, mmueller@theorie.physik. uni-goettingen.de

First Steps towards predicting the Evolution of Catchment-Scale terrestrial Systems

Model simulations of water and energy fluxes in the terrestrial system, which encompasses the subsurface, the land surface, and the atmosphere, are the backbone of climate and weather prediction, flood and drought forecasting, water resources management, agriculture, and water quality control. In these simulations, data assimilation is key in achieving better predictions of the complex terrestrial system by optimally merging the model with observations to obtain initial model states and estimate prediction uncertainty. In essence, data assimilation is a formalized process in which the uncertainty in a model forecast is balanced by integrating observations into a computer model of a real system. Different philosophies and methods however exist concerning the structure of coupled modelling systems and their integration into data assimilation frameworks. In order to improve predictions of terrestrial systems, it is the central goal of the DFG-funded Research Unit FOR2131 (http://www. for2131.de) to develop a unified data assimilation framework for terrestrial systems from aquifers across the land surface into the atmosphere.

While the atmosphere is relatively accessible for in situ and remotely sensed observations, suitable information about the land surface and subsurface is much harder to get, because of their opacity and extreme spatial and temporal variability. The latter also complicates the evaluation of model predictions as a key ingredient to the development of data assimilation methods and also the ability of terrestrial system models to realistically simulate all relevant processes. In particular, this concerns models used in data assimilation frameworks, which usually require large model ensemble runs and still challenge even the world's most powerful supercomputing facilities today.

Both, the data scarcity issue in the real world and the limitation of computationally-efficient terrestrial system models



Figure 1: Monthly averaged 2m air temperature for January 2007 of the first VR (atmosphere-landsurface simulations). The high 1.1km resolution used here enables the model to capture fine details such as some of the smaller river valleys and other small structures like the hotspots near mountains (most likely localized Föhn effects). Such features are not usually captured in atmospheric models as the coarser resolution typically used is not capable of producing small scale wind effects like these hotspots, which are important for the appropriate representation of exchange processes between the surface and the atmosphere. The boxed area indicates the modeled domain of the second subsurface-landsurface VR shown in Figure 2.

Marcus Müller

Institut für

August

Universität,

Göttingen,

Germany

Theoretische

Physik, Georg-

Average 2m Temperature [°C] of January 2007



for data assimilation, can be mitigated and at least evaluated, by a Virtual Reality (VR) simulation as an approximate representation of the best of our current physical and technical knowledge of a real catchment. A VR simulation invests all available compute power in a single sufficiently long-term simulation with the most sophisticated and most highly resolved model setup deployable on the available supercomputing facility. Suitable forward operators realistically mimic the real measurement process of e.g. instruments and satellites, and are employed to extract virtual observations, which provide the information for a data assimilation system in addition to a suitable compute efficient terrestrial system model. Predictions of system states and fluxes resulting from the assimilation of measurements for state estimation are then compared to their "true" values within VR. This validation will enable us to assess the data assimilation procedures and uncertainty of model predictions in a reliable manner. Types of predictions, which are improved



Figure 2: Surface soil saturation field in relative units for the top 2cm for the second VR (subsurface-landsurface simulations) on April 30, 2007. This VR is modelled at a high resolution of 100m which enables us to capture the river network in detail as well as different zones and features created by topography. soil, and rock properties. Atmospheric forcing data for this VR is obtained from the first VR model run shown in Figure 1.

by integrated models, can be identified as well as observations reducing model uncertainty and improving predictions.

The VR is generated using the Terrestrial System Modeling platform, TerrSysMP [1,3,5], developed within the Transregional Collaborative Research Centre SFB/TR32 (www.tr32.de, [4]). TerrSysMP simulates coupled subsurface, land surface and atmospheric processes by closing the terrestrial water and energy cycles from the bedrock to the atmosphere. These simulations are performed via a JSC/GSC large scale compute time project (project ID HBN29) as High Performance Computing is needed to create the VR at the cutting edge of today's modelling capabilities in terms of spatial and temporal resolution, and model physics. A data assimilation framework has also been developed in combination with the subsurface and land surface components of TerrSysMP [2].

The model uses the Neckar catchment (SW Germany) as a reference, which features topographic, land-use and weather patterns typical for the mid-latitudes. In a first step, two VRs are created, the first by coupled atmosphere-landsurface simulations and the second by landsurfacesubsurface simulations. The first VR uses the atmosphere-landsurface mode of TerrSysMP, which couples the COSMO atmosphere model and the Community Land Model (CLM) to generate virtual observations from satellites, precipitation radars, and meteorological stations for the years 2007-2013 (Fig. 1). This VR provides the atmospheric forcing for the second VR, which uses the surfacesubsurface mode of TerrSysMP (the variably saturated groundwater flow model ParFlow coupled to CLM) to better resolve soil and groundwater processes of the Upper Neckar subcatchment (Fig. 2).

The next step is a VR resulting from the fully coupled TerrSysMP, which will then include the feedbacks of subsurface properties on the atmosphere, such as the effects of soil moisture on atmospheric parameters like near-surface temperatures and precipitation. This will improve the physical consistency of the VR as more system processes and feedbacks are considered. The data assimilation for the fully coupled model will exploit cross-compartmental correlations of state variables and fluxes.

References

[1] Gasper, F., Goergen, K., Kollet, S., Shrestha, P., Sulis, M., Rihani, J., Geimer, M.

Implementation and scaling of the fully coupled Terrestrial Systems Modeling Platform (TerrSysMP) in a massively parallel supercomputing environment-a case study on JUQUEEN (IBM Blue Gene/Q). Geoscientific Model Development Discussions, 7, 3545-3573, 2014

[2] Kurtz, W., He, G., Kollet, S., Vereecken, H.. Hendricks Franssen, H.J.

TerrSysMP-PDAF (version 1.0): A modular high-performance data assimilation framework for an integrated land surface-subsurface model. Submitted to Geoscientific Model Development, 2015

[3] Shrestha, P., Sulis, M., Masbou, M., Kollet, S., Simmer, C.

A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and Parflow. Monthly Weather Review 142(9), 3466-3483, 2014

contact: Clemens Simmer, csimmer@uni-bonn.de

[4]

S. J.

Evaluating the influence of plant-specific physiological parameterizations on the partitioning of land surface energy fluxes. Journal of Hydrometeorology, doi: 10.1175/ JHM-D-14-0153.1, 2014

Simmer, C., Thiele-Eich, I., Masbou, M., Amelung, W., Crewell, S., Diekkrueger, B., Ewert, F., Hendricks Franssen, H.-J., Huisman, A. J., Kemna, A., Klitzsch, N., Kollet, S., Langensiepen, M., Loehnert, U., Rahman, M., Rascher, U., Schneider, K., Schween, J., Shao, Y., Shrestha, P., Stiebler, M., Sulis, M., Vanderborght, J., Vereecken, H., van der Kruk, J., Zerenner, T., Waldhoff, G.

Monitoring and Modeling the Terrestrial System from Pores to Catchments - the Transregional Collaborative Research Center on Patterns in the Soil-Vegetation-Atmosphere System. Bulletin of the American Meteorological Society. doi: http://dx.doi. org/10.1175/BAMS-D-13-00134.1, 2014

[5] Sulis, M., Langensiepen, M., Shrestha, P., Schickling, A., Simmer, C., Kollet,

- Felix Ament¹
- Sabine Attinger²
- Gabriele Baroni²
- Olaf Cirpka³
- Matthias Dursch⁴
- Daniel Erdal³
- Gernot Geppert¹
- Barbara Haese^{5,6}
- Xujun Han⁷
- Harrie-Jan Hendricks Franssen⁷
- Stefan Kollet^{7,8}
- Harald Kunstmann^{5,6}
- Javier Lechuga⁸
- Insa Neuweiler⁹
- Jehan Rihani⁸
- Pablo Saavedra Garfias⁸
- Bernd Schalge⁸
- Clemens Simmer⁸
- Harry Vereecken⁷
- ¹ University of Hamburg, Germany
- ² Helmholtz-Zentrum für Umweltforschung (UFZ) Leipzig, Germany
- ³ University of Tübingen, Germany
- ⁴ European Space Agency
- ⁵ Karlsruher Institut für Technologie, Germany
- ⁶ University of Augsburg, Germany
- ⁷ Forschungszentrum Jülich, Germany
- ⁸ University of
- Bonn, Germany ⁹ Hannover University, Germany

CoeGSS: Center of Excellence for **Global Systems Science**



Centre of excellence

The globalisation of humanity's social and industrial activities, as observed over the past decades, has caused a growing need to address the global risks and opportunities involved. Some of these prominent challenges include:

 The global health risks – from diabetes to pandemics - involved in the spread of unhealthy social habits as well as the opportunity to achieve major global health improvements through healthy behaviour.

 The global diffusion of green growth initiatives, including policy initiatives, business strategies and lifestyle changes for successful as well as inefficient pathways.

 The challenges of global urbanisation, with special focus on the impact of infrastructure decisions regarding indicators like congestion, real estate prices and greenhouse gas emissions.

Approaches that address the abovementioned challenges are investigated by a newly-emerging research area: the Global Systems Science (GSS). However, with these transdisciplinary problems the demand for compute performance due to data and time constraints increases drastically so that the assistance of High Performance Computing is necessary. With respect to the problem statements above, the main topics within the CoeGSS project are the development of the technology and the environment for successful collaboration between the stakeholders dealing with global challenges on the one hand and the High Performance Computing institutions that provide the mandatory capabilities to address those complex challenges at the required scale on the other.

So far, the use of HPC in GSS studies for processing, simulating, analysing, and visualizing large and complex data is very limited due to a lack of tailored HPC-enabled tools and technologies. Whereas typical GSS applications are data-bound, the traditional HPC tools and libraries are optimized to solve the computationally bound problems and thus, are of limited use in this area. The main difference between typical HPC applications and the envisioned GSS ones lies in the data sources and outputs as well as the used algorithms. Whereas lots of traditional high performance application codes, like those of computational fluid dynamics, require massive parallelism and high computational power, GSS applications demand additional capabilities, for instance huge and varying data or in a generic manner, data-centric computation.

Looking for a trade-off between the data-centric programming models of the Cloud infrastructures and the highly efficient and scalable HPC technologies is therefore one of the key challenges for the success of the CoeGSS project.

Use Cases

The CoeGSS project will be based on three complementary use cases, covering different GSS domains with a variety of requirements for processing and data analytics in High Performance Computing environments. Thus, these use cases and their implementations will demonstrate the functionality and flexibility of the developed solutions and the center, which inherits them in its service portfolio.

All use cases build on a single approach, the so called synthetic population, which represents a valid, spatially distributed population that encompasses social habits as well as daily population transitions and hence, forms a perfect basis for global problems for all kinds of domains.

Having those three use cases s baseline for developments, the project will foster inclusion of other application fields, like financial economies with their manifold markets or pandemic simulations for different regions all over the world.

CoeGSS

Role of USTUTT in CoeGSS

Within the CoeGSS project, USTUTT undertakes various roles and responsibilities. In particular, USTUTT is acting as technical coordinator of the overall project. Therefore, all technical developments for GSS and HPC and in particular, the interaction of both domains will be managed and steered into a sustainable direction. Furthermore, USTUTT is in charge of the Center of Excellence operation and ensures the seamless operation of all infrastructures, development and test environments as well as all remaining services.

Besides those two important roles, USTUTT is engaged in various tasks as contributor and task leader, such as the Co-Design task that focuses hardware and software improvements in conjunction with relevant vendors.

Key Facts

CoeGSS is a large-scale project funded by the European Commission under the Horizon 2020 Research and Innovation Programme. It has started at October 1, 2015 and will run until September 30, 2018. The consortium brings together



high	Perf	ormance
low	Sca	alability
high	Data	Locality
	Data S	structuring
Programming	Data Management	Resource Allocat



11 partners, providing comprehensive expertise in social, technical and business domains for both, HPC and GSS.

High Performance Computing Center
 Stuttgart, Universität Stuttgart
 (USTUTT), Germany

• Universität Potsdam (UP), Germany

Technical realizations related to High Performance Computing are managed by the partners USTUTT, PSNC and ATOS, social research including risk assessment is performed by DIA, IMT and GCF. Furthermore, Global Systems Sciences related topics are conducted by GCF, UP, CHALMERS, ISI, IMT and



• Michael Gienger

University of Stuttgart (HLRS), Germany

- Global Climate Forum EV (GCF), Germany
- Instytut Chemii Bioorganicznej
 Polskiej Akademii Nauk (PSNC), Poland
- Fondazione Istituto Per L'Interscambio Scientifico (ISI), Italy
- Scuola IMT Alti Studi Di Lucca (IMT), Italy
- Consorzio TOP-IX (TOP-IX), Italy
- Chalmers Tekniska Hoegskola AB (CHALMERS), Sweden
- ATOS Spain (ATOS), Spain
- The COSMO Company SAS (COSMO), France

• Dialogik Gemeinnuetzige Gesellschaft für Kommunikations- und Kooperationsforschung (DIA), Germany COSMO and a strong business context for exploiting and creating significant services and results is provided by ATOS, TOP-IX and GCF. Especially this in general complementary structure, which still enables cooperation on research and business topics amongst the consortium partners, guarantees a sustainable operation and uptake of HPC and GSS communities.

contact: Michael Gienger, gienger@hlrs.de, Bastian Koller,

koller@hlrs.de

POP: Performance Optimization and Productivity



A Centre of Excellence in Computing Applications

Inaugurated October 1, 2015, the new EU H2O2O Center of Excellence (CoE) for Performance Optimisation and Productivity (POP) provides performance optimisation and productivity services for academic and industrial codes. European's leading experts from the High Performance Computing field will help application developers getting a precise understanding of application and system behaviour.

Established codes, but especially codes never undergone any analysis or performance tuning, may profit from the expertise of the POP services which use latest state-of-the-art tools to detect and locate bottlenecks in applications, suggest possible code improvements, and may even help by Proof-of-Concept experiments and mock-up test for customer codes on their own platforms.

Today's complexity of high performance computer systems and codes makes it increasingly harder to get applications running fast and efficient on the latest hardware. Often expert knowledge and a good amount of experience is needed to figure out the most productive direction of code refactoring. Domain experts in many research areas and industry use computer simulations but lack this knowledge. Thus, their codes are often far away from using the hardware in an efficient way, using much more compute time than needed. By this, they either waste energy, require superfluously oversized and expensive hardware, or just miss research potential as their codes can only handle smaller or less complex problems in the available compute time.

The dominant practise in analysing and presenting the performance of applications and how they scale with increased core counts is to report speed-ups derived from execution timing or domainspecific performance metrics (e.g. simulated time/days). These are global observations but often do not give sufficient insight into the actual inefficiencies inside the simulation programs. In many cases the reference execution is a parallel run with a relatively small core count but can already this experience inefficiencies (e.g., load imbalance), thus masking the effect of such behaviour in the scaling study.

Beyond the global speed-up study, profiles may be collected that point to actual routines that dominate the execution, but often do not give real insight into the fundamental behaviour of the application. As computing applications become more and more complex, their performance is below the optimal levels for a number of reasons:

• Load imbalance that causes waiting on slower processes - This lack of balance can be caused by a different amount of work per process (computa-

tional imbalance) or different performance (hardware-related imbalance) or a combination of both.

• Serialisation – dependencies between code regions that cause chains of delay propagated along different processes.

• Data transfer – time cost of nonoverlapped data transfer of data between processes.

• Instructions Per Cycle (IPC) – actual performance of the sequential computation. This can be significantly below the core peak performance due to issues with the memory hierarchy, instruction mix, non-pipelined instructions, or dependencies.

• Amount of instructions and type of instruction – is the algorithm optimal for the given problem in terms of computational and instruction complexity and code balance?

• I/O, storage – frequency of or the time spent on I/O operations and how it affects computational efficiency.



Figure 1: Performance Analysis Tools of the Score-P Ecosystem.

To overcome this situation, the POP CoE brings two tightly coupled disciplines to the user as a service, which are crucial for the efficient use of parallel computers in the future: First, powerful performance analysis tools, methodologies, and expertise needed to precisely understand and gain real insight into the actual application and system behaviour; Second, deep understanding of programming models and best practice guidance needed to express algorithms in the most flexible, maintainable and portable way, while still being able to maximise the performance achieved.

POP-Technologies

The German project partners will provide their powerful and well-established performance analysis tools centered around the community-developed instrumentation and measurement infrastructure Score-P. Core developers are (among others) the Gauss Alliance members JSC, RWTH Aachen University, Technical University Darmstadt, Technical University Dresden, and Technical University Munich. Score-P can instrument and measure the performance of typical HPC applications written in Fortran or C/C++ and based on the MPI, SHMEM, OpenMP, OmpSs, CUDA, OpenCL and Pthread programming paradigms. It can collect flat and callpath profiles or detailed execution traces in the open CUBE4 or OTF2 formats. Score-P analysis tools include the event trace analyser Scalasca (developed by JSC) which provides very scalable waitstate, delay or root-cause analysis, the event trace analyser and visualisation tool Vampir (developed by TU Dresden), and the online performance analysis tool Periscope developed by GCS partner TU Munich. CUBE4 profiles can be analysed with the help of the Cube browser from JSC or TAU ParaProf (developed by the Univeristy of Oregon).

In addition to the Score-P tool universe, project partner BSC will provide their instrumentation and measurement package Extrae and their event visualiser Paraver. Finally, the POP experts will of course also use project-external tools like vendor products installed on the POP customer target platforms.

POP-Services

The POP CoE team consists out of six partners which are experts in High Performance Computing with long-standing experience in performance tools and tuning as well as researchers in the field of programming models and programming practices. All partners come with a research and development background and proven commitment in application of their knowhow to real academic and industrial use cases. The POP CoE will provide three kind of service levels to its customers – depending on their background, knowledge, and demands:

? Application Performance Audit

This is the primary service of the POP CoE and the starting point for any further work. Applications undergoing this service will be analyzed by the POP experts after an initial discussion with respect to their best practices and provide a first impression of the code status. Within the Performance Audit, performance issues of customer code will be identified at the customer site. It will serve as a starting point for further analysis or initial code refactoring. The duration for a Performance Audit is expected to be around one month and a successful Performance Audit may be seen as a code quality certificate in HPC.

! Application Performance Plan

The Performance Plan Service follows the Performance Audit if the customer needs more detailed knowledge where and how to address specific issues in the code. POP experts together with the customer will develop a plan how and with which tools to analyse the issues under investigation. The POP experts will then analyse the code in detail and give quantified hints to overcome the problems so that they can be fixed by the customer. The duration for a Performance Plan is very problem-specific, but will in general take between one and three months including a closer look into the source code.

✓ Proof-of-Concept

If requested, Proof-of-Concept studies will be performed. This includes experiments and mock-up tests for customer codes. The details of the proof-of-concept study will be decided in very close collaboration with the customer and may include kernel extraction from the application, parallelisation or mini-apps experiments to show effects of the proposed optimisations of the POP experts. As this very complex task goes into deep detail, Proof-of-Concept work should be expected to require about six months.

Besides the above three key services, the POP CoE will also provide a variety of training activities in the field of performance analysis and optimisation based on the user's needs to improve their basic high performance programming knowledge and increase the awareness of performance issues and potentials in general.

POP-Impact

areas of research and indu
a real transversal activity:
It provides access to concation expertise that enable

The POP CoE with its service and training activities will have a wide impact within all areas of research and industry, making it a real transversal activity:

 It provides access to computing application expertise that enables researchers and industry to be more productive, leading to scientific and industrial excellence.



Figure 2: The European HPC Eco-system.

- It improves competitiveness for the Centre's customers by generating a tangible Return on Investment (ROI) in terms of savings, elimination of waste, errors, and delays by making their applications leaner and issue-free.
- As the Centre represents the European world-class expertise in this area, its deployment will strengthen Europe's leading position in the development and use of applications that address societal challenges or are important for industrial applications through better code performance and better code maintenance and availability; POP will drive the cultural shift towards focusing on the health of applications.

The Centre's services will include

training on the use of computational

methods and optimisation of applications.

The Centre will build a repository of

user cases (i.e., the computing applica-

tion issues resolved) which will serve as

a basis for further research in the field.

According to an IDC Report, the global

HPC applications market will grow by 8%

between 2013 and 2018. According to

the same report, HPC is a proven ac-

- Christoph Niethammerr
- José Gracia¹
- Bernd Mohr²
- Brian J. N. Wylie²
- ¹ University of Stuttgart (HLRS), Germany
- ² Jülich
 Supercomputing
 Centre (JSC),
 Germany

celerator of economic competitiveness. With high-end supercomputers now costing \$200-500 million, their ROI can be a scientific advance or corporate profit, revenues, new jobs, or retaining jobs. Also, ROI arguments will become increasingly important for funding systems, which makes any measures improving the system's cost/performance ratio increasingly appealing. The POP CoE directly addresses this issue.

Partners

Barcelona Supercomputing Centre (BSC), High Performance Computing Center Stuttgart of the University of Stuttgart (HLRS), Jülich Supercomputing Centre (JSC), Numerical Algorithm Group (NAG), Rheinisch-Westfälische Technische Hochschule Aachen (RWTH), TERATEC (TERATEC).

Timeframe

October 2015 - March 2018

POP Coordination

Prof. Jesus Labarta, Judit Gimenez Barcelona Supercomputing Center (BSC) Email: pop@bsc.es

http://www.pop-coe.eu

Acknowledgements

This project is supported by the European Commission under H2O2O Grant Agreement No. 676553

contact:

Christoph Niethammer, niethammer@hlrs.de

ESSEX - Equipping Sparse Solvers for Exascale

The Priority Programme 1648 "Software for Exascale Computing" (SPPEXA) of the German Research Foundation is approaching the end of its third of six years. 13 projects started in January 2013 to address various challenges of exascale computing. In this issue, we present project ESSEX.

The major objective of the ESSEX project is to develop an Exascale Sparse Solver Repository (ESSR) for large sparse eigenvalue problems motivated by quantum physics research and apply it, exemplarily, to graphene-based structures and topological insulators. To this end various rather general aspects of sparse eigenvalue problems have to be addressed: Computation of (i) the minimal and the maximal eigenvalue, (ii) a block of eigenpairs at the lower end or in the middle of the spectrum, and (iii) high quality approximations to the complete eigenvalue spectrum. Classic and novel numerical schemes are implemented and optimized for efficient use of heterogeneous supercomputers. Complementing related sparse solver library developments (e.g. Trilinos [1] or sparse MAGMA [2]), the ESSEX project pursues a coherent co-design of all software layers where a holistic performance engineering (PE) process guides code development across the classic boundaries of application, numerical method and basic kernel library (see Fig. 1). The ESSR will finally cover a wide range of quantum physics problems and provide blueprints of sparse solvers adapted and optimized for the exascale challenges of heterogeneity, extreme parallelism, energy/code efficiency and fault tolerance (FT).

Partners from all three software layers are actively involved in ESSEX: The application layer is represented by the group of PI Fehske (Physics, University of Greifswald), expertise on sparse eigensolvers is contributed by PI Basermann (Simulation&Software, DLR) and PI Lang (Applied Mathematics, University of Wuppertal), and the basic building block development (including the project-wide PE and FT activities) is pursued by PI Hager (Erlangen Regional Computing Center) and PI Wellein



Figure 2: Weak and strong scaling performance of the full KPM solver for a topological insulator application with different geometries ("Square" vs. "Bar") on CRAY XC3O. Each node comprises one Intel Xeon E5-2670 and one NVIDIA K2OX card. The weak scaling runs achieve approx. 11% of LINPACK performance.



Figure 3: Runtime of the basic JADA operations for a representative ESSEX matrix with different vector block sizes on one Intel Xeon E5-2670 processor. The ESSR implementation (left panel) outperforms the Trilinos basic building block library (right panel) by up to 3x. See [8] for details.

Projects



Figure 1: Overall ESSEX project structure.

(Computer Science) at the University of Erlangen-Nuremberg. In the course of the project the holistic PE process has been successfully established across these groups. This holistic, layer-crossing concept of ESSEX was key to many achievements in the 2013/14 time frame. After a brief project overview we will present selected results.

The ESSEX application layer leverages ESSR solutions to investigate quantum effects in graphene and topological insulators as topical examples with major public recognition. To determine static and dynamic properties of these quantum systems the above mentioned aspects (i)-(iii) of large scale eigenvalue problems have to be solved for extremely sparse matrices of dimensions up to 10¹⁴ and may involve subspaces of 10² - 10³ eigenvectors. Specific problems which are tackled at the application layer include magnetic edge states in graphene nanoribbons and nanostructured systems of layered topological insulators/superconductors.



Figure 4: Superlattice of gate-defined quantum dots on a graphene sheet.



Figure 5: Single particle spectral function for a disordered graphene guantum dot superlattice calculated with KPM. Flat bands appear due to the dot superlattice and are washed out by disorder.

The numerical methods layer implements and advances state-of-the-art and experimental numerical schemes to determine blocks of eigenpairs including Jacobi-Davidson (JADA) with relevant preconditioners and the FEAST [3] algorithm. These standard methods are complemented by the kernel polynomial method (KPM) [4] which is a widely used approach in quantum physics/chemistry to compute the matrix density of states and can also be easily extended towards a polynomial filtering method to compute blocks of eigenvalues. Related expansion schemes are deployed to determine excitation spectra and dynamical properties of time dependent quantum systems. Though the choice of numerical methods is motivated by the ESSEX application scenario, most of the solvers can directly be applied to many other application areas or serve as blueprints for related methods. Thus enabling the ESSEX applications for exascale is of broad interest.

The basic building block layer establishes an "MPI+X" programming approach which is consistently used across the project and accounts for the potentially strongly heterogeneous and highly parallel node structure of modern supercomputers. It provides a collection of all relevant basic operations (including sparse matrix (multiple) vector multiplication and relevant sparse data structures) and efficient FT strategies, all tailored to the needs of the other two layers. Major design goals for these building blocks are: (i) "Optimal" performance and thus energy efficiency of all relevant operations, (ii) Minimum impact of FT overhead on time to solution. To achieve these goals the basic building block layer drives a structured, model-based performance engineering process across all three layers.

The strong collaboration between the software layers and within the project resulted in many notable contributions in the first 30 months of the ESSEX project. We have selected a few representative results to demonstrate the wide scope and impact of our work:

Introduction of SELL-C-sigma as an architecture independent storage format for sparse matrices [5]. The SELL-C-sigma format achieves high performance on a wide range of matrix classes for all modern HPC architectures and has already been adapted by, e.g. the MAGMA library, Vienna CL or the ExaDUNE project.

- A fully heterogeneous KPM computation for topological insulators has been performed on up to 4,096 nodes of the CRAY XC3O at CSCS Lugano (Switzerland). Exploiting layer crossing optimization the application performance has been boosted by a factor of 3x-5x, resulting in 0.5 PF/s of sustained performance (see Fig. 2) for this extremely sparse matrix application [6]. · Exploiting tailored sparse matrix operations on vector blocks (see Fig. 3),
- it has been demonstrated for the first time that a blocked Jacobi-Davidson variant can outperform classic variants despite its higher numerical effort [7,8].
- Based on the software developed in ESSEX several application problems have already been successfully investigated, e.g., the functionalization of graphene quantum dot lattices [9] (see Fig. 4/5). A detailed overview of the ESSEX project activities is available at http://blogs.fau. de/essex. The ESSR software and blueprints are available at https://bitbucket. org/essex.

We gratefully acknowledge the financial support of the Priority Research Initiative 1648 "Software for Exascale Computing", funded by the German Research Foundation.

References

- [1] Trilinos: https://trilinos.org/
- [2] MAGMA: http://icl.cs.utk.edu/
 - [3] Polizzi, E. (115112), 2009
 - Fehske, H.
 - Fehske, H., Bishop, A. R. C401-C423, 2014

 - Fehske, H.
 - 2014
 - Fehske, H.
 - Fehske, H.

contact: Gerhard Wellein, gerhard.wellein@fau.de

projectsfiles/magma/doxygen/index.html

Density-matrix-based algorithm for solving eigenvalue problems, Phys. Rev. B 79

[4] Weiße, A., Wellein, G., Alvermann, A.,

The kernel polynomial method, Rev. Mod. Phys. 78, pp. 275-306, 2006

[5] Kreutzer, M., Hager, G., Wellein, G.,

A unified sparse matrix data format for modern processors with wide SIMD units, SIAM Journal on Scientfic Computing 36(5), pp.

[6] Kreutzer, M., Hager, G., Wellein, G., Pieper, A., Alvermann, A., Fehske, H.

Performance Engineering of the Kernel Polynomial Method on Large-Scale CPU-GPU Systems, Proceedings of the 29th IEEE International Parallel & Distributed Processing Symposium (IPDPS 2015), 2015

[7] Röhrig-Zöllner, M., Thies, J., Kreutzer, M., Alvermann, A., Pieper, A., Basermann, A., Hager, G., Wellein, G.,

Performance of Block Jacobi-Davidson Eigensolvers, Poster at the International Conference for High Performance Computing, Networking, Storage and Analysis (SC14),

[8] Röhrig-Zöllner, M., Thies, J., Kreutzer, M., Alvermann, A., Pieper, A., Basermann, A., Hager, G., Wellein, G.,

> Increasing the performance of the Jacobi-Davidson method by blocking, submitted

[9] Pieper, A., Heinisch, R. L., Wellein, G.,

Dot-bound and dispersive states in graphene quantum dot superlattices, Physical Review B 89 (165121), 2014

- Achim Basermann¹
- Holger Fehske²
- Georg Hager³
- Bruno Lang⁴
- Gerhard Wellein³
- ¹ German Aerospace Center
- ² Ernst-Moritz-Arndt-University Greifswald, Germany
- ³ Friedrich-Alexander-University Erlangen-Nuremberg, Germany
- ⁴ University Wuppertal: Germany

The Mont-Blanc Project: First Phase successfully finished

With the ever-increasing energy demands and prices and the need for uninterrupted services, data centre operators must find solutions to increase energy efficiency and reduce costs. Running from October 2011 to June 2015, the aim of the European project Mont-Blanc [1] has been to address this issue by developing an approach to Exascale computing based on embedded powerefficient technology. The main goals of the project were to i) build an HPC prototype using currently available energyefficient embedded technology, ii) design a Next Generation system to overcome the limitations of the built prototype and iii) port a set of representative Exascale applications to the system. This project was coordinated by the Barcelona Supercomputing Centre and had a budget of over 14 million EUR, including over 8 million EUR funded by the European Commission.

Improving the energy efficiency of future supercomputers is one of LRZ's main research goals. Hardware prototyping of novel architectures has proven to be successful for the technology watch preceding the selection of large supercomputers. Yet, advances in hardware are only justified if the need for programmability and thus the productivity of application development is still satisfied. Therefore, LRZ's contribution to the Mont-Blanc project was twofold: on the software side, application experts successfully ported two applications to the new system and analysed their performance and productivity. On the hard-

ware side, LRZ's computer architecture experts worked on a system monitoring solution for acquisition and storage of Mont-Blanc sensor data, with particular emphasis on node energy consumption.

Contributions from JSC were also twofold: teams from the Jülich Simulation Laboratories ported their applications to the new system and to the OmpSs programming model [2], and the crosssectional team Performance analysis ported their well-known performance analysis tool components Score-P [3] and Scalasca [4] to the ARM and ARM64 architecture and adapted them to work with the OmpSs programming model [1, Deliverable D5.9].

One of the highlights of the project was the installation of the Mont-Blanc system at the Barcelona Supercomputing Centre. Figure 1 shows the final prototype, consisting of 2 racks, each of which containing 4 standard BullX chassis. Every chassis fits 9 blades, hosting 15 compute nodes each. Every Mont-Blanc compute node comes with a Samsung Exynos 5 Dual System on a Chip, which includes an ARM Cortex-A15 @ 1.7 GHz dual core CPU and an ARM Mali T-604 GPU. In total, the Mont-Blanc system offers 2160 ARM CPU cores and 1080 ARM GPUs, making it a "firstof-a-kind" cluster of this size based on ARM architecture.

The performance of the system was thoroughly tested and evaluated with many real-life application codes: LRZ was involved with two different codes: the lattice Quantum Chromodynamics code BQCD as a representative of a real-world application and the Himeno benchmark. Berlin Quantum Chromodynamics (BQCD) [5] is a Hybrid Monte-Carlo program for simulating lattice QCD with dynamical Wilson fermions. The most important part of the program is a standard conjugate gradient solver. The Himeno benchmark (http://accc. riken.jp/2444.htm] is the kernel of an incompressible fluid analysis code and focuses on the solution of a 3D Poisson equation. It is highly memory intensive and bound by the memory bandwidth. Both codes were successfully ported to the ARM architecture and tested on the Mont-Blanc prototype system. Beyond evaluating the original MPI-only and hybrid MPI+OpenMP versions on the prototype, LRZ also successfully ported BQCD to the OmpSs data flow programming model developed at BSC and investigated the scaling behaviour of code versions combining OmpSs with OpenCL and/or MPI.

JSC ported, tested and evaluated four applications codes: MP2C [6] implements the multi-particle collision dynamics method, which is a particle based description of hydrodynamics taking into account thermal fluctuations and making it possible to simulate flow phenomena on a mesoscopic level. PEPC [7] is a tree code for solving the N-body problem. It is not restricted to Coulomb systems but also handles gravitation and hydrodynamics using the vortex method as well as smooth particle hydrodynamics (SPH). PEPC is a non-recursive version of the Barnes-Hut algorithm with a level-by-level approach to both tree construction and traversals. PROFASI [8] is a Monte Carlo simulation package for protein folding and aggregation simulations. It implements an all-atom protein model, an implicit solvent interaction potential and several modern Monte Carlo methods for simulation of systems with rough energy landscapes. Finally, SMMP (https://trac.version.fz-juelich. de/SMMP) provides advanced Monte Carlo algorithms and several force fields



Figure 1: The new Mont-Blanc prototype system at BSC, Barcelona (1), the Mont-Blanc blade (2) and the Mont-Blanc compute card (3).

to simulate the thermodynamics of single proteins and assemblies of peptides. To port them to the Mont-Blanc prototype, MPI+OmpSs or MPI+OpenCL versions of the codes were developed and evaluated. Figure 2 shows the power profile of a benchmark run of SMMP.

More details on the conducted performance analysis by LRZ and JSC can be found in [1, Deliverable D4.4]. Performance is the most important goal of HPC. However, programming languages should not be judged by the performance that can be reached alone, but also by the ease-of-use, i.e. the programmability. The combination of performance and **Projects**

programmability is commonly referred to as "productivity". To assess the productivity, various software engineering metrics like lines of code and time to solution can be used. For the productivity analysis within the Mont-Blanc project LRZ's software experts concentrated on the number of source lines of code to assess the ease-of-use of parallel programming languages. The number of source lines gives a rough estimate of the time necessary to program the code, as well as of the readability and



Figure 2: Power profile of SMMP. The run with OpenCL uses the GPU in addition to the CPU. The numbers in parenthesis are the energies to solution (1 Ws = 1 J).



Figure 3: Number of total lines of code for various versions of the Himeno benchmark.

maintainability of the code. The productivity analysis has been performed for various code versions of BQCD and the Himeno benchmark using several combinations of parallel programming languages like OpenMP, MPI, OpenACC, CUDA, OpenCL and OmpSs. To show an example of the productivity analysis, Figure 3 presents the number of total lines of codes for different versions of the Himeno benchmark. Further productivity related results are discussed in [1, Deliverable D3.6].

Finally, one of the key requirements for achieving energy efficiency is the capability of retrieving and storing detailed information on the power consumption of the HPC system. Working in close collaboration with Bull, LRZ researchers developed a holistic monitoring solution for the Mont-Blanc prototype, allowing not only to keep track of the power consumed by the system, but also and especially exposing this information to the system user. The monitoring tool was customised to cope with the characteristics of the Mont-Blanc system and features a low overhead transport messaging protocol and a scalable database for storing the monitored data [1, Deliverable D5.8]. In addition to standard cluster monitoring features and following the integration with the SLURM workload management system, the tool can also be used for implementing energy-aware job scheduling. Additionally, this integration will offer the opportunity of defining strategies for energy-aware user accounting. These last research topics alongside with the provision of performance and debugging analysis tools will be further addressed respectively by LRZ and JSC within the second phase of the project, running from 2013 until 2016. The second phase will complement the efforts of the first phase by targeting the system programmability, performance analysis and resiliency. In this phase we will also monitor the evolution of upcoming ARM-based devices and we will define the next Mont-Blanc Exascale architecture, investigating hardware design alternatives and their implications to the current system.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the Mont-Blanc project [1], grant agreement n° 288777 and n° 610402. We would like to thank Dr. Hinnerk Stüben (University of Hamburg) for his continuous collaboration.

References

- [1] The Mont-Blanc project, http://www.montblanc-project.eu/
- [2] The OmpSs programming model, https://pm.bsc.es/ompss

[3] an Mey, D., et al.,

Score-P: A Unified Performance Measurement System for Petascale Applications. Competence in High Performance Computing 2010 (CiHPC), 85-97, 2012

[4] Geimer, M., et al.,

The Scalasca performance toolset architecture. Concurrency and Computation: Practice and Experience, 22 (6), 702-719, 2010

[5] Allalen, M., Brehm, M., Stüben, H.

Performance of quantum chromodynamics (QCD) simulations on the SGI Altix 4700. Computational Methods in Science and Technology CMST 14 (2) 2008

[6] MP2C webpage, http://www.fz-juelich.de/ias/jsc/EN/ Expertise/High-Q-Club/MP2C/_node.html

[8] PROFASI webpage,

contact:

Momme Allalen, Momme.Allalen@lrz.de, David Brayford, David.Brayford@lrz.de, Daniele Tafani, Daniele.Tafani@lrz.de, Volker Weinberg, Volker.Weinberg@lrz.de, Bernd Mohr, B.Mohr@fz-juelich.de, Dirk Brömmel, D.Broemmel@fz-juelich.de, Rene Halver, R.Halver@fz-juelich.de, Jan Meinke, J.Meinke@fz-juelich.de, Sandipan Mohanty, S.Mohanty@fz-juelich.de

Projects

[7] Winkel, M., et al.

A massively parallel, multi-disciplinary Barnes-Hut tree code for extreme-scale Nbody simulations. Computer Physics Communications, Vol. 183 (4), 2012

https://apps.fz-juelich.de/jsc/slbio/ PROFASI_Gallery/index.html

- Momme Allalen¹
- David Brayford¹
- Daniele Tafani¹
- Volker Weinberg¹
- Bernd Mohr²
- Dirk Brömmel²
- Rene Halver²
- Jan Meinke²
- Sandipan Mohanty²
- ¹ Leibniz Supercomputing Centre (LRZ), Germany
- ² Jülich Supercomputing Centre (JSC). Germany

EU Project "AutoTune" finished

The 3 years long European Union project "Automatic Online Tuning" (AutoTune) [1], with the goal of automatically optimizing applications in the area of High Performance Computing, has successfully finished in May 2015.

The project consortium was integrated by six partners, the Technische Universität München (TUM, Munich), the Leibniz Supercomputing Centre (LRZ, Munich) of the Bavarian Academy of Sciences, the CAPS Enterprise (France), the Universitat Autònoma de Barcelona (UAB, Barcelona), the Centre for High-End Computing (ICHEC, Galway) at the University of Galway and the University of Vienna (UNIVIE, Vienna), as well as the associated partner IBM. The partners provided significant technical expertise in the areas of scientific and High Performance Computing that contributed to the successful outcome for the project. As an instance, the LRZ contribution was to provide a plugin for the Periscope Tuning Framework (PTF) [2] - core of the AutoTune project - to optimize the energy consumption of parallel applications running on the petascale system SuperMUC.

The Energy Optimization plugin

The Energy Optimization Plugin (aka DVFS Plugin) for the AutoTune project is based on an energy prediction model developed by IBM for energy aware scheduling optimizations [3] which predicts the best CPU frequency for a parallel application in order to optimize energy to solution. It accurately predicts an optimal CPU frequency with less than

3% deviation from the measured values for the tested applications. This model was modified for the AutoTune Project so that it can be used to search the optimal parameters at the level of the phase or main region. The new version of the model does not rely on node-global data like the memory bandwidth but on corelocal data to allow tuning of individual cores in future versions of the tool.

Figure 1 shows the model's predicted normalized time and normalized energy compared to the normalized values of time and energy of one run of the geophysical application SeisSol. The minimum energy consumption was found at 1.5 GHz. However, visually it is possible to deduce that the best frequencies are between 1.3 GHz and 1.7 GHz. The exact minimum will depend on the actual power leakage and the variability in the measurement. The model shows its minimum at 1.7 GHz.

The DVFS plugin was applied to a wide range of applications with several tuning objectives. As much as 20 automatic tunings were performed that optimized the six applications for energy, power capping, Total Cost of Ownership (TCO), and/or Energy Delay Product (EDP). This high number of tunings performed (see Fig. 2) in a short time is only possible since the tunings are conducted automatically, making this one of the most valued features of AutoTune. The energy and power related costs dominate the expenditures needed to run HPC systems.

While energy savings have a direct impact on these costs, power capping strives to keep power demand at a constant level to avoid penalties due to power spikes. Other metrics like EDP and TCO take into account the costs associated with time and thus consider the productivity of HPC centers. All these tuning objectives plus other policies which allow a certain performance degradation that tune for energy are available in the DVFS plugin. Therefore, the DVFS plugin adapts to the needs of the end user providing a variety of metrics aimed at reducing energy and power related costs.

Additionally, an in-depth documenta-

tion of the AutoTune project has been

provided in the book "Automatic Tuning

of HPC Applications - The Periscope

Tuning Framework (PTF)" [4] edited and

published by the partners. The book cov-

ers different tuning concepts as well as

various tuning plugins like for instance

The European Commission (EC) funded

project in the field of High Performance

Computing (HPC) prolongs its optimiza-

tion services after the official project

ended in April 2015 with a dedicated

centre hosted at the LRZ, the AutoTune

Demonstration Centre (ADC) [5]. The

ADC provides resources to the Parties

for the execution of their exploitation ac-

tivities including hardware and software

as well as a website and organizational

support. As a joint community effort,

the ADC initially consists of the current

partners of the AutoTune project, but

the ADC is open to affiliate also other

Parties with suitable HPC experience

and working activities particularly spin-

offs from the AutoTune project.

AutoTune Demonstration

AutoTune Book

the DVFS Plugin.

Centre (ADC)



software SeisSol

The main Goals of the ADC are:

• To further spread the use of the AutoTune software in the academic as well as commercial HPC user community by dissemination and demonstration of the advantages in the efficient usage of HPC computing resources, accessible through the AutoTune software framework.

 To further exploit the capabilities of energy saving and power steering for user applications.

 To disseminate best practice guides for the use of the tools developed in the AutoTune project, e.g. the Periscope Tuning Framework (PTF), or the Energy Optimization Library (enopt). To organize and to provide trainings

to the users.

• To provide developers with the autotuning tools and to create a platform for interested vendors to test and validate further improvements of the AutoTune tools.

 To support the exchange of information and experience.

customers.



Figure 1: Model comparison of the geophysical simulation

• To further extend the Periscope

Tuning Framework developer community.

To give the potential Spin-off a

platform to demonstrate the tools to



Figure 2: A saving of up to 40% percent for energy consumption was achieved with SeisSol. Optimizing TCO yielded a saving of 0.3143€ per iteration per node on this application.

- To collaborate with other organizations, particularly the Virtual Institute for High Productivity Supercomputing (VI-HPS).
- To broaden the technical expertise of the project partners with respect to tools for HPC in general.

The AutoTune Demonstration Centre offers the following Services:

- Online documentation and best practice guides
- Discussion forums
- AutoTune training events for academic and commercial HPC users
- Individual support for specific applications, including code analysis and code tuning.
- Education and training for bachelor and master students

 The AutoTune web site as a central hub to present the AutoTune project results and activities of the AutoTune Demonstration Centre.

Service Desk and Issue Tracking

References

[1] Miceli, R., Civario, G., Sikora, A., César, E., Gerndt, M., Haitof, H., Navarrete, C., Benker, S., Sandrieser, M., Morin, L., Bodin, F.

> AutoTune: A plugin-driven approach to the automatic tunning of parallel applications. Lecture Notes in Computer Science (LNCS): Applied Parallel and Scientific Computing, 2013

[2] Gerndt, M., Fuerlinger, K.

Automatic Performance Analysis with Periscope. Journal: Concurrency and Computation: Practice and Experience. Wiley InterScience. John Wiley, 2009

[3] Auweter, A., Bode, A., Brehm, M., Brochard, L., Hammer, N., Huber, H., Panda, R., Thomas, F., Wilde, T. (2014) A Case Study of Energy Aware Scheduling on SuperMUC, International Supercomputing Conference (ISC) Proceedings 2014, Sons, Ltd. 2009

[4] Gerndt, M., et al.

Automatic Tuning of HPC Applications - The Periscope Tuning Framework (PTF), ISBN 978-3-8440-3517-9, Shaker Verlag, 2015

[5] http://www.autotune-project.eu/adc_main

contact: Wolfram Hesse, wolfram.hesse@lrz.de

Automated Analysis of the Performance of Jobs on Super-MUC with the PerSyst Tool

The detection of performance bottlenecks is a difficult task on HPC systems like SuperMUC with its more than 25,000 application runs per month. Ideally, all applications are monitored automatically, on-the-fly, without userinteraction and without increasing the runtime of jobs. For the last two generations of HPC systems (HLRB II and SuperMUC) the Leibniz Supercomputing Centre has deployed the PerSyst Tool to achieve these goals (for more details see [1]).

PerSyst is a scalable system-wide monitoring and analysis tool with negligible overhead (< 0.1% increase of the walltime) on application runs. It provides application-level data and on-the-fly analyses which are performed with codified expert knowledge and the assessment of the severity of bottlenecks. These analyses are referred to as Strategy Maps and are designed to reveal bottlenecks. In this article, the I/O bottleneck detection strategy of the Per-Syst Tool is presented as one specific example. Several other strategies to identify bottlenecks are available, such as: network interconnect usage, memory usage, memory bandwidth, single core performance, flops and flop vectorization analysis, and load imbalance.

PerSyst measures every 10 minutes for 10 seconds on SuperMUC Phase 1, a 3.2 petaflop system based on Sandy Bridge-EP Intel® processors and on SuperMUC Phase 2, a 3.6 petaflop system based on Haswell-EP Intel® processors [2]. SuperMUC is connected to a 12 PB GPFS file system. The tool uses the mmpmon command [3] of the GPFS to access the I/O measurements. The measurements are processed on-the-fly according to analysis items, also called properties, in the Strategy Map and are aggregated into 11 percentiles (including maximum and minimum) and the average value.

Disk I/O is extremely expensive and tens of thousand times slower compared to computation cycles [4]. Even if I/O operations remain unavoidable, the forms of accessing a file system can be optimized in diverse ways. The I/O Strategy Map is shown in Figure 1 with properties leading to other refined properties or leading to recommendations. Recommendations are shown in red. Every property shown in the Strategy Map is explained in detail.

A job using 1,280 cores (80 nodes) in SuperMUC Phase 1 with high severities for the I/O properties is selected as an example. The application and details



• Carmen Navarrete

Carla Guillen

Leibniz

• Wolfram Hesse

David Brayford

Matthias Brehm

Supercomputing

Centre (LRZ).

Germany

Figure 1: Strategy Map for I/O analysis.



Figure 2: PerSyst Screenshot: I/O Imbalance. Bandwidth in GB/s for writing data to the GPFS. The x-axis shows time of day.

about it are kept anonymous. The exposure of its bottlenecks via the analyzed properties is discussed to illustrate the detection of inefficiencies with the tool.

I/O imbalance Property: I/O operations in large parallel jobs are often performed with I/O masters. These are tasks assigned to collect and distribute I/O data on a sub-group of tasks or threads. I/O imbalance refers to an unequal amount of requests and/or request sizes (I/O work) among the I/O masters which results in uneven time spent when doing I/O operations. I/O imbalance may arise from a suboptimal ratio of I/O masters to all tasks (for instance, when one master performs all of the I/O in a parallel application). In this case, the distribution of I/O work is performed sub-optimally even if the I/O masters perform an equal amount of requests and request sizes due to an inappropriate distribution of the I/O work on all the tasks. Under the assumption that the ratio of I/O masters is optimal, and not all tasks are I/O masters, tasks which are exempt from performing I/O operations can't be considered imbalanced with respect to the tasks which do.

Figure 2 shows the write bandwidth of the sample job in a timeline. Ten percent of the nodes or less are doing I/O

operations according to the analyzed data. Typically, one core per node (or at least one core every two nodes) should be set as the I/O master task, which means that the application exhibits I/O imbalance.

I/O bandwidth Property: I/O bandwidth gives an insight into the volume requested or sent to the file system. If the bandwidth is small this might indicate an inefficiency of the application due to small sized I/O requests. Small and big are predefined with a threshold within PerSyst.

Figure 3 shows the severities of the properties for the same job as in Figure 2. While Figure 2 shows the bandwidth for writing, Figure 3 shows that its severity (coloured in green) is not significant. Whereas the severity for the read bandwidth is high (in red).

I/O requests size Property: Relatively frequent but small I/O requests are an indication of suboptimal use; the best approach is to try to bundle the I/O data to make larger I/O requests. For the job in consideration, the IO_WrittenBytes has almost no severity but the IO_Bytes-WrittenPerWriteOperation (i.e. request sizes) show a high severity.



Figure 3: PerSyst Screenshot: I/O Severities (x-axis on the top is a timeline).

File metadata request rate Property: File metadata is requested when a file is opened. However, sending metadata has an associated time overhead.

Figure 3 shows the severities due to IO_Opens. In this case, the tool measured more than 1,000 IO_Opens per second on this job (not shown in Fig. 3), which are potentially avoidable.

To the application developers of this job, two main advices can be given. Firstly, to fix the I/O imbalance by allowing more I/O master tasks to write on the disk. Secondly, the users are advised to perform less open and closes by writing more data between an open and a close operation. Alternatively, users can be advised to consolidate files in order to perform less open/close operations.

The PerSyst workflow supports the performance optimization cycle at an HPC centre. This workflow starts with the visualization tool provided by the PerSyst WebAPI. Once the HPC centre detects the jobs with highest severities, the users are contacted and the HPC centre works closely with the application developers to optimize the applications.

All screenshots are from the PerSyst WebAPI. The latest PerSyst Tool de-

References

- [1] Guillen, C. de?id=1237547, 2015
- [2] SuperMUC System,
- Programming Reference, 1 Edition. 2014
- Accessed July 2015

contact: Carla Guillen, Carla.guillen@lrz.de

velopments are currently funded in the FEPA project (Flexible Framework for Energy and Performance Analysis, BMBF Grant: 01|H13009).

Knowledge-based Performance Monitoring for Large Scale HPC Architectures, Dissertation http://mediatum.ub.tum.

https://www.lrz.de/services/compute/ supermuc/systemdescription/

[3] IBM Corporation. GPFS: Administration and http://publibz.boulder.ibm.com/epubs/ pdf/a2314520.pdf, Version 4 Release

[4] Latency Numbers Every Programmer Should Know. Internet Website: http://www.eecs.berkeley.edu/~rcs/ research/interactive_latency.html,

- Carla Guillen
- Wolfram Hesse
- Carmen Navarrete
- Matthias Brehm

Leibniz Supercomputing Centre (LRZ), Germany

JSC's Horizon 2020

Projects for Designing Future HPC Technologies



In 2012 the European Commission (EC) adopted a dedicated HPC strategy [1] which formulates as one objective the independent access to HPC technologies and systems for the EU (the pillars of the strategy implementation are shown in Fig. 1). Within the new program for research and innovation, Horizon 2020, the EC started funding 19 new projects. The call for these projects had been formulated on the basis of the first Strategic Research Agenda [2] of the European Technology Platform for HPC (ETP4HPC). JSC successfully joined the efforts of two consortia which aim for the development of future HPC core technologies and architectures, namely ExaNoDe (European Exascale Processor & Memory Node Design) and SAGE (Percipient StorAGe for Exascale data centric computing), which are coordinated by CEA and Seagate, respectively. The goal of the ExaNoDe project is to design a highperformance, heterogeneous compute element based on the chiplet concept and Unimem memory architecture previously explored in the EUROSERVER project [3]. This memory architecture aims for an elastic allocation of memory resources to different coherence islands by routing load/store operations between differ-

Dirk Pleiter

Jülich Supercomputing Centre (JSC), Germany

SAGE has the objective of providing a next-generation multi-tiered data storage that integrates computing capabilities. The project addresses two important exascale challenges: Today's disk-based storage architectures, which are highly cost-efficient for providing large storage capacity, will not be able to scale bandwidth as compute performance increases. Hierarchical storage architectures comprising high-bandwidth non-volatile memory devices will allow to mitigate this problem. The second challenge is the need for minimizing data movement as it is expensive in terms of energy consumption. SAGE's approach to this challenge is to integrate compute capabilities into the storage hierarchy, i.e. move data processing capabilities to where the data is. These new exascale projects are meant to be a first step within Horizon 2020 towards a European ecosystem for HPC capable of providing exascale class solutions. In a few years the results of these projects are expected to be integrated in extreme-scale demonstrator systems. This will be the litmus test for these development projects as they will have to proof that they can deliver technology which is ready for addressing large-scale computational challenges.

ent chiplets. While ExaNoDe focuses

on the design of future compute nodes,

References

- [1] http://eur-lex.europa.eu/LexUriServ/ LexUriServ.do?uri=COM:2012:0045: FINIFNIPDE
- [2] http://www.etp4hpc.eu/strategy/ strategic-research-agenda/
- [3] http://dx.doi.org/10.1109/DSD.2014.15

contact: Dirk Pleiter, d.pleiter@fz-juelich.de

Future Supercomputers for Brain Research: Pre-**Commercial Procurement** entered final Phase



By developing and expanding the use of

information technology, the Human Brain

Project (HBP) [1] wants to open new op-

portunities for brain research. The goal

of this European project is to enable a

multi-level, integrated understanding of

brain structure and function. Particular

challenging is the enablement of large-



Figure 1: NEST [5], an interactive analysis tool for neural activity data, being used in the aixCAVE at RWTH Aachen University (Source: Virtual Reality Group, RWTH Aachen University).

scale simulations of brain models as todays HPC architectures do not meet their requirements. This includes both, the need for extremely large memory footprint and interactive supercomputing. For realistic network sizes, e.g., the amount of data generated during a simulation becomes too large for being written to an external storage system and thus new memory technologies have to be integrated. Furthermore, the complexity of the simulations requires interactive steering. To ensure that suitable solutions for realizing HBP's future High-Performance Analytics and Computing Platform will exist, the project published in April 2014

• Dirk Pleiter

Jülich Supercomputing Centre (JSC), Germany

References

a tender for a pre-commercial procure-

ment (PCP). PCP [2] is an instrument

promoted by the European Commission

(EC) to foster innovation through public

procurement. It allows for procuring

research and development services to

enable development of new solutions

which would otherwise likely not be

available. A comparison with the already

announced pre-exascale systems in

the US, like Summit at ORNL [3] and

Aurora at ANL [4], confirm additional

research and development efforts will

be required to realize the planned HBP

Platform. This concerns in particular the integration of dense memory technolo-

gies, scalable visualization (see Fig. 1

for a visualization use case) as well as dynamic management of resources

required for interactive access to the systems. By design a PCP is organized

as a multi-phase, competitive process.

During Phase I the suppliers had the

task for sketching-out a design meeting

the different challenges, which then had

been refined in Phase II. At the end of

that phase, three competitors, namely

Cray, a consortium comprising Dell and

the German SMEs Extoll and ParTec as

well as a consortium consisting of IBM

and NVIDIA, presented their design

specifications. An expert committee

evaluated these solutions in July 2015

and recommended awarding Cray and

IBM-NVIDIA with contract for Phase III.

These contractors now have the task

of implementing the proposed solutions

and demonstrate their technological

readiness on pilot systems that will be

installed in 2016. A PCP is a still a new

instrument which needs to be carefully

designed to balance goals, timing and available budget. But the efforts within

the HBP demonstrate that PCP can be

a suitable instrument to drive develop-

ment of future supercomputers.

- [1] https://www.humanbrainproject.eu
- [2] http://ec.europa.eu/digital-agenda/en/ pre-commercial-procurement

[3] https://www.olcf.ornl.gov/summit

- [4] http://aurora.alcf.anl.gov
- [5] http://dx.doi.org/10.1109/
 BioVis.2013.6664348

contact: Dirk Pleiter, d.pleiter@fz-juelich.de

The Vector Computer NEC SX-ACE

In the eighties of the last century High Performance Computing was a synonym for the usage of so-called vector supercomputers, machines that used pipelining in the functional units as well as in data path for the acceleration of numerical codes. Memory bandwidth was high in comparison to the peak floating-point performance. These machines from Cray (later split in Cray Research, Cray Computer Corporation, Super Computer Systems Incorporated), CDC, IBM, Fujitsu, Hitachi and NEC had low chip integration density, but a very sophisticated packaging, which made these technologies expensive. Shared memory parallelism was early integrated and led to OpenMP as standardized parallel model. In the nineties large scale parallel distributed memory computers based on relative inexpensive integrated processors gained traction, the so called "killer micros", together with message passing as parallel model. Indeed these machines replaced the vector supercomputers, but not as fast as expected. One reason for surviving was that vector machines delivered predictable reliable performance for vectorized codes. Vectorization as parallel paradigm can be handled easyer automatically by compilers in contrast to other kinds of parallelization. The other reason was that they came up as parallel machines (Earth Simulator in 2002, HLRS in 2005). Nevertheless, vector machines seem not to play a major/any role today. But is that true? A closer look to modern processors

shows that all of these have SIMD support and are to be used as vector machines. Suppressing vectorization in the code by inappropriate programming (many procedure calls, usage of array of structures instead of structure of arrays, recursive loops, short



Systems



loops, deep information hiding, code flexibility where not needed) implies clear performance losses. SIMD hardware helps for performance increase as well as for reducing energy needed per operation.

Even more, we see with the NEC SX-ACE machine a new vector processor combining the traditional advantage of high memory bandwidth with the cost effectiveness of modern integration technology. The NEC SX-ACE processor has 4 cores each equipped with an Assignable Data Buffer (ADB) serving as a vector cache. Bandwidth to the ADB and to the crossbar connecting the 4 cores to the memory system is 256 GB/s as high as the crossbar bandwidth to memory. The crossbar is directly communicating to the network via the on chip Remote Access Control Unit (RCU). A special scalar unit operates non-vectorizable code and issues vector instructions.

The core peak performance is 64 GFLOPs at a comparable moderate frequency of 1.0 GHz. The memory capacity is 64 GB. A node consists of 1 processor and 16 memory DIMMs. Depending on the IXS-switch, a parallel machine may have up to 512 nodes with 2,048 cores. The HLRS machine has 64 nodes.

The effective performance of a node with 4 cores for vectorizing code can be compared to an Intel Haswell node with two 2.5 GHz processors with 2x12 cores; for appropriate code the SX-ACE core is 6 times faster than a Haswell core. The needs for parallel scaling for comparable performance might be less. The aggregated scalar performance of 4 cores with 1 GHz is definitely less than the aggregated scalar performance of 2x12 cores with 2.5 GHz. The peak performance of the node is not impressive in contrast to its efficiency. The machine has not



Memory(DDR3)

CORE	
Architecture	Vector
Clock Frequency	1.0GHz
SPU decode rate	4 instructions
VPU Performance	64GFlops
ADB size	1MB
ADB bandwidth	256GB/s
Memory bandwidth	64GB/s~ 256GB/s
Core Byte/Flop	1.0 ~ 4.0
CPU	
Cores	4
Performance	256GFlops
Memory bandwidth	256GB/s
CPU Byte/Flop	1.0
Memory capacity	64GB

been build for Linpack Performance. Much more interesting is the efficiency of the HPCG benchmark of 10%.

The machine uses a front-end computer for compilation, scheduling and common access to the ScaTeFS filesystem. There are two Fortran-crossThe SX-ACE will have a successor named Aurora. HLRS will look to the emerging product. Surely, we expect a vector machine. How many cores, how large the bandwidth? If code developers appreciate the potential of implementing vectorizing algorithms, this system will give a new perspective



The machine is quite energy efficient. Our colleagues from the computing center of the Christian-Albrechts-Universität zu Kiel where surprised by their small power bill. They estimated 250 W per node. The Cray XC40 at HLRS takes 370 W per node under load. Tests at HLRS showed 16.5 KW during a linpack run for all 64 nodes.

for High Performance Computing. The SIMD part of the OpenMP 4.0 standard will help as well ideas primarily developed for the vector part of processors as Intel Skylake and Intel Phi.

If interested, we may offer the reader access to test the HLRS SX-ACE.

contact:

Uwe Küster, kuester@hlrs.de Systems

• Uwe Küster

University of Stuttgart (HLRS), Germany



The Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities (Leibniz-Rechenzentrum, LRZ) provides comprehensive services to scientific and academic communities by:

- giving general IT services to more than 100,000 university customers in Munich and for the Bavarian Academy of Sciences
- running and managing the powerful communication infrastructure of the Munich Scientific Network (MWN)
- acting as a competence centre for data communication networks
- being a centre for large-scale archiving and backup, and by
- providing High Performance Computing resources, training and support on the local, regional, national and international level.

Research in HPC is carried out in collaboration with the distributed, statewide Competence Network for Technical and Scientific High Performance Computing in Bavaria (KONWIHR).

Contact:

Leibniz Supercomputing Centre

Prof. Dr. Arndt Bode Boltzmannstr. 1 85748 Garching near Munich Germany

Phone +49-89-358-31-80 00 bode@lrz.de www.lrz.de



Picture of the Petascale system SuperMUC at the Leibniz Supercomputing Centre.

Compute servers currently operated by LRZ are given in the following table

	System	Size	Peak Performance (TFlop/s)	Purpose	User Community
		18 thin node islands IBM iDataPlex 512 nodes, 147,456 cores Sandy Bridge EP 288 TByte, FDR 10 IB	3,185	Capability Computing	German universi- ties and research institutes, PRACE (Tier-O System)
	IBM/Lenovo System x Cluster "SuperMUC Phase 1"	1 fat node island with 205 nodes, 8,200 cores IBM Bladecenter HX5 Westmere EX 52 TByte, QDR IB	78	Capability Computing	
		32 accelerated nodes Intel Ivy Bridge EP, Intel Xeon Phi 76 GByte, Dual-Rail FDR 14 IB	100	Prototype System	
	Lenovo/IBM Nextscale Cluster "SuperMUC Phase 2"	6 medium node islands Lenovo/IBM Nextscale 512 nodes, 86,016 cores Haswell EP 197 TByte, FDR 14 IB	3,580	Capability computing	German universi- ties and research institutes, PRACE (Tier-O System)
	Lenovo/IBM Nextscale Cluster "CooLMUC2"	252 nodes, 7,056 cores Haswell EP 16.1 TByte FDR 14 IB	270	Capability computing	Bavarian Universities
	Megware IB-Cluster "CoolMUC"	178 nodes, 2,848 Cores AMD Magny Cours 2.8 TByte GDR IB	22.7	Capability Computing, PRACE Prototype	Bavarian Universities
	SGI Altix Ultraviolet	2 partitions, 2,080 cores Intel Westmere EX 6.0 TByte main memory	20.0	Capability Computing	Bavarian Universities
	Linux-Cluster	510 nodes, 2,030 cores Intel Xeon EM64T/ AMD Opteron 2-, 4-, 8-, 16-, 32-way 4 7 TByte main memory	13.2	Capability Computing	Bavarian and Munich Universities, LCG Grid

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

Centres

HLR S

First German National Center

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

Service for Industry

Service provisioning for industry 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 industry always has acces to the most recent HPC technology.

Bundling Competencies

In order to bundle service resources in the state of Baden-Württemberg HLRS has teamed up with the Steinbuch Center for Computing of the Karlsruhe Institute of Technology. This collaboration has been implemented in the nonprofit organization SICOS BW GmbH.

World Class Research

As one of the largest research centers for HPC HLRS takes a leading role in research. Participation in the German national initiative of excellence makes HLRS an outstanding place in the field.

Contact:

Höchstleistungsrechenzentrum Stuttgart (HLRS) Universität Stuttgart

Prof. Dr.-Ing. Dr. h.c. Dr. h.c. Hon. Prof. Michael M. Resch Nobelstraße 19 70569 Stuttgart Germany

Phone +49-711-685-8 72 69 resch@hlrs.de / www.hlrs.de

Compute servers currently operated by HLRS

	System	Size	Peak Performance (TFlop/s)	Purpose	User Community
	Cray XC40 "Hazel Hen"	7,712 dual socket nodes with 185,088 Intel Haswell cores	7,420	Capability Computing	European (PRACE) and German Research Organizations and Industry
	NEC Cluster (Laki, Laki2) heterogenous compunting platform of 2 independent clusters	23 TB memory 9,988 cores 911 nodes	170	Laki: 120,5 TFlops Laki2: 47,2 TFlops	German Universities, Research Institutes and Industry
	NEC SX-ACE	64 compute nodes with 256 cores 4 TB memory	16	Vector Computing	German Universities, Research Institutes and Industry

A detailed description can be found on HLRS's web pages: www.hlrs.de/systems



View of the HLRS Cray XC40 "Hazel Hen"

Centres

73



The Jülich Supercomputing Centre (JSC) at Forschungszentrum Jülich is committed to enabling scientists and engineers explore some of the most complex grand challenges facing science and society. Our research is performed through collaborative infrastructures exploiting extreme-scale supercomputing and federated data services.

Provision of supercomputer resources

of the highest performance class for projects in science, research and industry in the fields of modeling and computer simulation including their methods. The selection of the projects is performed by international peer-review procedures implemented by the John von Neumann Institute for Computing (NIC), GCS, and PRACE.

Supercomputer-oriented research and development in selected fields of physics and other natural sciences by research groups and in technology, e.g. by doing co-design together with leading HPC companies.

Implementation of strategic support infrastructures including communityoriented simulation laboratories and cross-sectional teams, e.g. on mathematical methods and algorithms and parallel performance tools, enabling the effective usage of the supercomputer resources.

Higher education for master and doctoral students in close cooperation with neighbouring universities.

Contact:

Jülich Supercomputing Centre (JSC) 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/jsc



JSC's supercomputer "JUQUEEN", an IBM Blue Gene/Q system.

Compute servers currently operated by JSC

	System	Size	Peak Performance (TFlop/s)	Purpose	User Community
	IBM Blue Gene/Q "JUQUEEN"	28 racks 28,672 nodes 458,752 processors IBM PowerPC® A2 448 Tbyte main memory	5,872	Capability Computing	European Universities and Research Institutes, PRACE
	Intel Linux CLuster "JURECA"	1,884 SMT nodes with 2 Intel Haswell 12-core 2.5 GHz processors each 150 graphics processors (NVIDIA K80) 281 TByte memory	2,245	Capacity and Capability Computing	European Universities, Research Institutes and Industry
Y V V X	IBM Cell System "QPACE"	1,024 PowerXCell 8i processors 4 TByte memory	100	Capability Computing	QCD Applications SFB TR55, PRACE

Centres

14th HLRS/hww Workshop on Scalable Global Parallel File Systems



From April 27 to April 29, 2015 representatives from science and industry working in the field of Global Parallel File Systems and High Performance Storage did meet at HLRS for the fourteenth annual HLRS/hww Workshop on Scalable Global Parallel File Systems "The Non-Volatile Challenge". About 75 participants did follow a total of 22 presentations which have been on the workshop agenda.

Prof. Michael Resch, Director of HLRS, opened the workshop with an opening address on Monday morning. In the keynote talk, Eric Barton, CTO of Intel's High Performance Data Division, discussed "A new storage paradigm for NVRAM and integrated fabrics". He explained emerging trends and upcoming technologies which might significantly change the storage landscape. Peter Braam, Braam Research and University of Cambridge, explained the exa-scale data requirements and issues of the square kilometre array which is currently under development. In the first presentation of the Monday afternoon session, Torben Kling-Petersen, Seagate, discussed the Lustre Storage Enterprise HPC technol-

ogy of Seagate which is enabling energy efficient, extreme performance storage solutions. Following new approaches for large HPC systems, Wilfried Oed, Cray, explained the Cray Data Warp solution which can already be deployed in todays Cray XC3O and XC4O systems. Afterwards, Alexander Menck, NEC, gave an overview of the NEC storage portfolio. In the second afternoon session, James Coomer, DDN, introduced the IME Burst Buffer technology and he showed, how real world applications can profit from its usage. Franz-Josef Pfreundt, FhG -ITWM, showed BeeOND, which is BeeGFS (formerly known as Fraunhofer File System) on demand. He explained how a file system can be setup and provided automatically on nodes which have been reserved for a user job, e.g by a batch system. In the last presentation of the day, Mellanox' Oren Duer gave an overview on Mellanox efforts in the storage field including the different technologies they are providing in the field. In the first presentation on Tuesday morning, Akram Hassan, EMC, provided EMCs view to elastic Cloud and object storage. He showed how the object storage solution fits the needs of todays applications. The following talk have been more research oriented. Tim Süß, University Mainz presented results of his studies on the potential of data deduplication in checkpoint data of scientific simulations. The second session started with a presentation of IBM, given by Olaf Weiser, explaining new developments in GPFS, especially GPFS native RAID. For half a year, Lenovo is now a new



and active player on the HPC market. Michael Hennecke provided an overview of Lenovos HPC storage solutions. In the last talk of the morning sessions, Thomas Uhl introduced a new highly performant HSM solution provided by GrauData which is especially working together with Lustre. The second half of the day is traditionally reserved for network oriented presentations. This year a focus have been on the opportunities of Software Defined Networking. Yaron Ekshtein, presented in detail the Pica8 approach for a switch operating system providing an open networking approach for improving cost/performance and scalability of compute clusters. This was followed by Edge Core's Lukasz Lukowski, who showed its bare metal switches as underlying hardware for SDN and how they can accelerate open networking. Klaus Grobe, Adva Optical, went even more down to the hardware and explained solutions for inter-datacenter 400-Gb/s WDM transport.As a preview for the Wednesday session, Radu Tudoran, Huawei, introduced the three storage musketeers of the Big Data era: Reliable, Integrated and Intelligent and Technology-Convergent. The other talks touching also the big data

arena followed on Wednesday morning. Mario Vosschmidt, NetApp, discussed the benefits of declustered RAID technologies and especially covered Hadoop file systems with this. Alexey Cheptsov, HLRS, introduced the Juniper project and how HPC can be used for dataintensive applications with Hadoop and OpenMPI. The last session has been again more future and research oriented. The Seagate perspective to the future of HPC Storage has been given by Torben Kling-Petersen. Michael Kuhn, Hamburg University, followed with a new approach to make future storage systems aware of I/O semantics and finally Thomas Bönisch, HLRS, showed the results of the project SIOX, Scalable IO Extensions. HLRS appreciates the great interest it has received once again from the participants of this workshop and gratefully acknowledges the encouragement and support of the sponsors who have made this event possible.

• Thomas Bönisch

University of Stuttgart, HLRS, Germany

Workshop Sparse Solvers for Exascale: From Building **Blocks to Applications**

A central challenge for exascale computing in science and engineering is the integration of algorithm and software development, crossing the traditional gap between low-level library implementations and high-level application packages. Meeting this challenge requires strong interaction between researchers and practitioners from programming, algorithm development, and the scientific application. From March 23 to 25, 2015 the workshop "Sparse Solvers for Exascale: From Building Blocks to Applications" was held at Alfried-Krupp-Wissenschaftskolleg Greifswald with the dedicated goal of letting people from all three fields exchange their ideas and insights in an informal setting. The workshop was jointly organized by Hans-Joachim Bungartz (Munich), Holger Fehske (Greifswald), and Gerhard Wellein (Erlangen), and supported by DFG through the Prority Programme 1648 "Software for Exascale Computing" (SPPEXA) together with the Stiftung Alfried Krupp Kolleg Greifswald. The workshop program was arranged around five keynote talks. On Monday, Edmond Chow (Georgia Tech) discussed how massive parallelism can be achieved for quantum chemistry computations on large-scale heterogeneous clusters. Yousef Saad (University



Figure 1: Participants of the Sparse Solver Workshop.

Andreas Alvermann

University of

Greifswald,

Germany

Minnesota) explained the ideas behind divide-and-conquer algorithms for large Hermitian eigenvalue problems, addressing the different levels of parallelism in the algorithms. On Tuesday, Marlis Hochbruck (Karlsruhe) presented general time integration schemes for evolution equations from the algorithmic and applications point of view. The day ended with the public evening lecture by Horst Simon (Berkeley Lab), who gave a thought-provoking account of the capabilities and limitations of present and future supercomputer architectures that lead into an engaged discussion with the non-expert audience. On Wednesday, Satoshi Matsuoka (TiTech) addressed communication reducing algorithms that can cope with bandwidth limitations in deep memory hierarchies, while Horst Simon detailed the ideas on usable exascale from his public evening lecture for the professional workshop audience. The program was complemented by thirty-two contributed talks and posters, which further expanded on the variety of topics covered in the keynote talks. The 57 participants (see Fig. 1) from ten different countries were immediately involved in multiple discussions, in spite of the fact that sometimes a common language had to be established first. The consensus that constant exchange between researchers and practitioners from different fields is vital for the future of high-performance computing permeated all discussions. In this way the contacts established during the workshop will likely lead into new fruitful collaborations on exascale applications.

contact: Gerhard Wellein, gerhard.wellein@fau.de

2nd Intel MIC & GPU **Programming Workshop** at LRZ

For the 2nd time the Leibniz Supercomputing Centre as a PRACE Advanced Training Centre (PATC) organized a three day Intel MIC & GPU programming workshop at Garching next to Munich, dated April 27-29, 2015. Around 25 people registered for the workshop and were able to gain experience on high-end GPGPU and Intel Xeon Phi coprocessor based systems. The materials for the presentations and hands-on sessions of the workshop were largely extended to reflect recent developments in parallel programming languages and latest experience with heterogeneous accelerator based systems at LRZ. The workshop covered various GPU and Xeon Phi programming models and optimization techniques. While the first day focused more on the fundamentals of parallel programming with GPUs using CUDA, OpenACC, Python, R and MATLAB, the second day was devoted to various Intel Xeon Phi programming models like native mode vs. offload mode, parallelization approaches like OpenMP, MPI, OpenCL and Intel Cilk Plus, as well as libraries like Intel MKL. On the last day invited speakers Dr.-Ing. Jan Eitzinger from the Regional Computing Centre Erlangen (RRZE) and Dr.-Ing. Michael Klemm from Intel gave lectures about advanced Intel Xeon Phi programming using low-level techniques like intrinsics or assembly language, advanced tuning methodologies and the new offload features from OpenMP 4.0. During many hands-on sessions the participants were able to gain experience on different GPU clus-



Xeon Phi coprocessor.

ters and on the Intel Xeon Phi based system SuperMIC at LRZ (see inSiDE Vol. 12, No. 2, p. 76ff for a description of the system). In addition, the participants also had the opportunity to discuss optimization techniques and test their own codes. Future Intel Xeon Phi trainings are planned by LRZ for 2016, taking place not only at LRZ, but also e.g. in Hagenberg, Austria, during the PRACE Autumn School 2016 and in the Czech Republic, where currently the largest Intel Xeon Phi based system in Europe ("SALOMON") is prepared for production use at IT4Innovations / VŠB - Technical University of Ostrava.

contact: Volker Weinberg, Volker.Weinberg@lrz.de Activities

Figure 1: Participants of the 2nd Intel Xeon Phi & GPU programming workshop at LRZ. The main lecturers Dr. M. Allalen, Dr. V. Weinberg, Dr. D. Brayford (LRZ), Dr.-Ing. M. Klemm (Intel) and Dr.-Ing. J. Eitzinger (RRZE) are each holding an Intel

> Volker Weinberg Momme Allalen

ZENTRU

Leibniz Supercomputing Centre (LRZ), Germany

Workshop Recent Advances in Parallel Programming Languages at LRZ

Since the standards of parallel programming languages are becoming more and more complex and extensive, it can be hard to stay up to date with recent developments. LRZ has thus invited leading a host to an accelerator device using HPC experts to give updates of recent advances in parallel programming languages during a 1-day workshop on June 8, 2015 at LRZ. The workshop attracted more than 50 participants. Languages covered during the workshop were MPI, OpenMP, OpenACC and Coarray Fortran. The workshop started with a talk about the planned extensions to the parallel syntax and semantics of Coarray Fortran as specified in ISO/IEC TS 18508 by Dr. Reinhold Bader, group leader of the HPC group at LRZ and member of the Fortran standardization working group WG5. Dr.-Ing. Michael Klemm, Intel representative in the OpenMP Language Committee, revisited the history of OpenMP, which dates back to 1997. He further presented features of the current OpenMP version 4.0 like offloading and SIMD support and provided an outlook on OpenMP 4.1 and OpenMP 5.0.

Dr. Mandes Schönherr (Cray Inc.) gave a short introduction into OpenACC, a language that provides an efficient way to offload intensive calculations from directives. Finally, Dr. Rolf Rabenseifner from HLRS, who has been a member of the MPI-2 Forum since 1996 and is also in the steering committee of the MPI-3 Forum, gave an overview over the new MPI shared memory programming model and new methods within the MPI standards 3.0 and 3.1. The latter had been approved by the MPI Forum just 4 days before the workshop. The slides of the workshop are available at http://www.lrz. de/services/compute/courses/x_lecturenotes/Parallel_Programming_ Languages_Workshop/. Based on the very positive feedback during the workshop LRZ intends to organise a similar event in 2017.

contact: Volker Weinberg, Volker.Weinberg@lrz.de



Figure 1: Participants of the workshop Recent Advances in Parallel Programming Languages at LRZ. The speakers Dr. Mandes Schönherr (Cray Inc.), Dr. Rolf Rabenseifner (HLRS), Dr.-Ing. Michael Klemm (Intel Corp.) and Dr. Reinhold Bader (LRZ) are standing in the first row.

Workshop on Computational Solar and Astrophysics Modelling



This 5-day summer school took place from 14-18, September and introduced young researchers (advanced master students, PhDs, and junior postdoctoral researchers) to modern open-source numerical astrophysics models with a heavy emphasis on hands-on tutorial sessions. After introductory morning lectures, participants worked with three different open-source software packages to learn about their typical applications and evaluate their performance aspects on the FZ-Jülich supercomputer systems. Participants made use of the MPI-AMRVAC framework covering plasma dynamics in the solar atmosphere and into the heliosphere. For fully kinetic plasma modelling of e.g. the Earth's magnetosphere, the code iPic3D was made available, a C++, MPI-parallelized, implicit moment Particle-In-Cell solver. Participants were also given an introduction to the open source Swift cosmological

particle hydrodynamics code, which has implementations of both SPH and the weighted particle hydrodynamics scheme Gizmo, together with a gravity implementation based on both fastmultipoles and a tree-code. The school was restricted to 28 highly motivated early-career scientists, supported by a Belgian Belspo-funded Inter-university Attraction Pole CHARM (http://wis. kuleuven.be/CHARM) connecting heliospheric to astrophysical communities.

Further information: http://www.fz-juelich.de/ias/jsc/ csam-2015

contact: Paul Gibbon, p.gibbon@fz-juelich.de

Volker Weinberg

Supercomputing

Centre (LRZ),

Germany

Leibniz

Activities

• Paul Gibbon¹ Rony Keppens²

Forschungszentrum Jülich, Germany

² KU Leuven

JSC Guest Student Programme 2015 - GSP rocks JUQUEEN

providing HPC expertise for computational scientists at European universities, research institutions, and industry. A variety of training and educational activities are organised by JSC on a regular basis. One of these activities is the annual Guest Student Programme (GSP) lasting for ten weeks each summer. The participants receive extensive training on cutting edge hardware as well as HPC-related software and algorithms. The acquired theoretical knowledge is turned into hands-on skills by coached work on novel and challenging scientific projects. For many students, the programme has been the foundation of a career in HPC and the basis of fruitful long-term collaborations with their advisers. Some students even return to JSC as PhD candidates focusing on highly parallel applications. Over the past 15 years, 157 students participated in the GSP and this year another 12 got the opportunity to join researchers at JSC. During the highly competitive selection procedure 76 applicants tried to obtain one of the limited number of guest student positions. The selection committee received applications from 24 countries spanning a wide range of scientific domains, e.g. physics, chemistry, computer science, and mathematics.

The Jülich Supercomputing Centre (JSC)

is one of Europe's leading HPC centres



This year's GSP took place from August 3 to October 9. It was supported by CECAM (Centre Europeen de Calcul Atomique et Moleculaire) and sponsored within the IBM University programme. The first two weeks were dedicated to various courses on parallel programming up to advanced level. The lectured HPC techniques range from the usage of MPI on distributed-memory cluster systems to GPGPU programming with CUDA as well as threading via OpenMP. They were complemented by crash courses on LaTeX and revision control techniques with GIT. Equipped with this vital knowledge the participants were ready to focus on the scientific part of the GSP. The range of scientific projects was as diverse as the user community on the hosted supercomputers, covering neuroscience, fluid and molecular dynamics, and safety research. Also represented was fundamental research in elementary particle physics, and mathematical algorithms. The main platforms for code development and simulation were the CPU/ GPU system JURECA and the leadership Blue Gene/Q system JUQUEEN. Next year's GSP will start on August 1, 2016. It will be officially announced in January 2016 and is open to students from natural sciences, engineering, computer science, mathematics and the computer science related branches of neuroscience. For applicants it is mandatory, to have received the Bachelor but not yet the Master degree. The application deadline is March 31, 2016. Additional information of the previous years is available online at www.fz-juelich.de/jsc/gsp.

contact: Ivo Kabadshow, jsc-gsp@fz-juelich.de Supercomputing for Neuroscientists: How HPC can help your **Neuroscience Projects**



Computational neuroscience is developing an interest in problems of increasing complexity and scale, leading to the evolution of projects such as the "Human Brain Project" (HBP) [1] and the "1000 Brains Study" [2] which will include computationally intensive simulations and the analysis of huge data sets. However adapting current software developed for local clusters to the supercomputing environment is often a challenge for the originating labs. The Simulation Laboratory Neuroscience [3] at the Jülich Supercomputing Centre (JSC) aims to bridge the gap between these two environments with a regular series of workwith computational neuroscientists.

Last year, the SimLab Neuroscience began this process with the "Bernstein Network - Simulation Lab Neuroscience" HPC workshop [4], which led to the

porting of several projects to JSC HPC systems. On November 3 of this year, the SimLab presented Supercomputing for Neuroscientists: How High Performance Computing can help your neuroscience projects, a workshop directed towards the German and European neuroscience communities covering:

- Introduction to HPC
- supercomputers
- Participants' projects.

shops bringing together HPC experts Experts in HPC from the SimLab gave presentations on supercomputer architectures, scaled algorithms and massively parallel algorithms. Other experts from the JSC and Jülich's Institute of Neuroscience and Medicine showcased projects that have already

Ivo Kabadshow

Supercomputing

Centre (JSC),

Germany

Jülich

Activities



Current neuroscience projects on supercomputers • Application process and access to

- Anne Do Lam-**Ruschewski²**
- Steffen Graber¹ Anna Lührs¹
- Abigail Morrison¹
- Boris Orth²
- Alexander Peyser¹
- ¹ Simulation Lab Neuroscience -**Bernstein Facility** for Simulation and Database Technology, Institute for Advanced Simulation, Jülich Aachen Research Alliance, Forschungszentrum Jülich, Germany
- ² Division High Performance Computing in Neuroscience, Jülich Supercomputing Centre (JSC), Institute for Advanced Simulation. Jülich Aachen **Research Alliance** Forschungszentrum Jülich, Germany

made use of JSC resources, including large-scale neuronal network simulations such as NEST [5] on the JUQUEEN supercomputer, "Big Data" approaches to experimental electrophysiological analyses, and massive neuroanatomical maps such as the BigBrain [6] using PLI data [7], highlighting issues and opportunities facing neuroscientists as they scale projects up. The computing time application process was explained, the new "preparatory access" model available at the JSC was described, and the HBP HPC Platform as part of the available large-scale neuroscience research infrastructure was introduced. Additionally, participants had an opportunity to show posters and give spotlight talks to catalyze collaboration with HPC experts.

Further details are available at www.fz-juelich.de/ias/jsc/scn.

References

[1] Human Brain Project. URL http://www.humanbrainproject.eu/

- [2] Institute of Neuroscience and Medicine 1, 2, 4 and 8, Forschungszentrum Jülich and University Duisburg-Essen. 1000BRAINS. URL http://www.fz-juelich. de/inm/inm-1/EN/Forschung/1000 Gehirne_Studie/1000_Gehirne_Studie_ node.html
- [3] Simulation Laboratory Neuroscience -Bernstein Facility Simulation and Database Technology. URL http://www.fz-juelich.de/ ias/jsc/EN/Expertise/SimLab/ slns/ node.html

[4] Do Lam-Ruschewski, A., Lührs, A., Morrison, A., Orth, B., Peyser, A., Schenck, W.

Bernstein Network - Simulation Lab Neuroscience, 4-5 June 2014. URL http:// www.fz-juelich.de/SharedDocs/Termine/ IAS/JSC/EN/events/2014/bernsteinhpc-2014.html

NEST Initiative. NEST (Neural Simulation [5] Tool), 2014.

URL https://github.com/nest/nest-simulator

- Institute of Neuroscience and Medicine 1, **[6]** Forschungszentrum Jülich and Montreal Neurological Institute (MNI). BigBrain. URL http://www.fz-juelich.de/inm/ inm-1/EN/ Forschung/_docs/BigBrain/bigbrain_node.html
- Institute of Neuroscience and Medicine 1, Forschungszentrum Jülich. Polarized Light Imaging.
- URL http://www.fz-juelich.de/inm/inm-1/ EN/Forschung/ docs/ PLI/PLI node.html

contact:

Anne Do Lam-Ruschewski, a.dolam@fz-juelich.de, Boris Orth, b.orth@fz-juelich.de

JSC at the 3rd JLESC Workshop

From June 29 to July 1, the 3rd JLESC workshop took place in Barcelona, organized by Barcelona Supercomputing Center. This event was the first one in 2015 of the biannual meetings of the Joint Laboratory on Extreme Scale Computing (JLESC) and the first one with Jülich Supercomputing Centre (JSC) as full partner of JLESC. The Joint Laboratory brings together researchers from the Institut National de Recherche en Informatique et en Automatique (Inria, France), the National Center for Supercomputing Applications (NCSA, USA), Argonne National Laboratory (ANL, USA), Barcelona Supercomputing Center (BSC, Spain), and, since the beginning of this year, RIKEN AICS (Japan) and JSC. The key objective of JLESC is to foster international collaborations on state-of-the-art research related to computational and data focused simulation and analytics at extreme scales. Within JLESC, scientists from many different disciplines as well as from industry address the most critical issues in advancing from petascale to extreme scale computing. The collaborative work is organized in projects between two or more partners. This includes mutual research visits, joint publications and software releases. Every six months, all JLESC partners meet during a workshop to discuss the most recent results and to exchange ideas for further collaborations. With more than 100 scientists and students from the six JLESC partners, the meeting in Barcelona covered a broad range of topics crucial for today's and tomorrow's supercomputing. Together with the other participants, 19 staff members from JSC could catch up on cutting-edge research from the fields



of resilience, I/O and programming models as well as numerical methods, applications, data analytics and performance tools. 8 scientists and students from JSC and German partner universities presented their research and results during contributed talks and Prof. Thomas Lippert highlighted central HPC aspects of the Human Brain Project in his keynote. Prof. Morris Riedel and his group at JSC and University of Iceland gave insight into their research on data analytics during the associated JLESC summer school "Storage, IO and Data Analytics". From December 2 to 4, 2015 the 4th JLESC workshop will be organized by JSC at the Gustav-Stresemann Institut e.V. in Bonn, continuing this successful series of internationally recognized and valued meetings. For more information visit the official JLESC website under http:// publish.illinois.edu/jointlab-esc.

contact: Robert Speck, r.speck@fz-juelich.de Activities



Robert Speck

Jülich

Supercomputing Centre (JSC), Germany

Stefan Krieg¹

Rainer Sommer²

Hubert Simma²

Stefan Schäfer²

• Giannis Koutsou³

Supercomputing Centre (JSC),

Neumann Institute

for Computing

(NIC), DESY

³ Computation-

based Science

and Technology

Research Center

(CaSToRC), Cyprus

Zeuthen

Dirk Pleiter¹

Karl Jansen²

Constantia

¹ Jülich

Germany

² John von

Alexandrou³

Lattice Practices 2015 @ JSC

The 6th training workshop "Lattice Practices" was held at JSC October 14 to 16, 2015. The scope of the Lattice Practices workshops is to provide training in state-of-the-art numerical techniques and the use of information technologies for research in lattice QCD (LQCD). Geared towards PhD students, young researches, and other interested LQCD practitioners, the workshops feature lectures on technical topics accompanied by hands on exercises, with strong emphasis on practical training. Furthermore, a few very recent scientific developments are covered, in order to expose the young researchers and students to potential areas of future research.

This year's workshop was organized by the Joint SimLab "Nuclear and Particle Physics" of Cyprus Institute, DESY, and JSC. Speakers from the SimLab partners and other European institutions gave technical lectures and hands-on tutorials on topics commonly dealt with in their field of research. The topics covered ranged from data analysis and numerical techniques over optimization strategies and computer architecture to "hot" LQCD, with accompanying hands on sessions. Here, the participants were given examples on basic techniques such as binning and error and autocorrelation analysis, but also given typical physics tasks they will likely encounter during their own research. A particular emphasis was put on optimal programming, when the course of lectures and exercises went on to introduce the attendees to code optimization techniques

and HPC architectures in general. This was completed by an introduction to numerical linear solver techniques and deepened in the accompanying exercises for both topics. Completing this year's course of lectures were two talks discussing new simulation techniques and LQCD at finite temperature.

This year's participants came from institutions all over Europe, from Italy to Ireland, but also from as far away as India. This interest demonstrates the need for this series of educational workshops, which was initiated in 2006. A next workshop is planned for spring 2017. The slides of the talks and material of the hands on sessions can be found on the web at: http://www.fz-juelich.de/ ias/jsc/lap15

contact: Stefan Krieg, s.krieg@fz-juelich.de

Smart Data Innovation Lab

Smart Data Innovation Lab

The significantly growing data economy is driven by slogans like "data is the oil of the 21st century" or "the data speaks for itself". But in order to achieve "big insights from data", important research effort still needs to be made, e.g. in terms of parallel, scalable, and even real-time processing of large data quantities ("big data"). Structuring "big data" results in information (called "smart data") which in turn leads to knowledge advantages which can be used to answer important research questions or that contributed to better decision-making processes.

In order to be able to make fast use of this competitive edge for Europe, partners from industry and research have established the Smart Data Innovation Lab (SDIL). The close cooperation between industry and science is intended to improve the conditions for cuttingedge research in the area of data engineering, parallel and scalable machine learning, data mining, and smart data processing. Figure 1 illustrates the conceptual organization of the SDIL initiative.

Besides several important supporting activities with respect to data curation, law, and security, the core benefit of the SDIL initiative is to offer interested communities an SDIL data analytics platform with three cuttingedge industry hardware and software stacks. At the time of writing, the SAP HANA In-Memory database is available on 4 nodes with each 80 CPU-cores, 1 TB RAM, and 20 TB storage. This installation includes software packages like SAP Hana Studio, Client, Smart Data Streaming, Live Tools, and the Predictive Analysis Library. The Software AG Terracotta Big Memory MAX software is available on 8 cores with 64 GB RAM running in a virtual machine. The IBM Watson Foundations is also available with IBM InfoSphere BigInsighs on 6 nodes with each 20 cores, 0.5 TB RAM, and over 300 TB space. The Model-based Predictive Analytics system with IBM SPSS Modeler is provided on 1 node with 20 cores and 1 TB of RAM. The SDIL platform thus offers powerful analytics systems without license issues for SDIL users interested in performing research on modern technologies.



Figure 1: Conceptual organization of the Smart Data Innovation Lab (SDIL).

Figure 2: SDIL combines data sources with cutting-edge systems, analytics, and research.



Figure 3: Brain tissue cutted with 80 micrometer into ~700 layers available as block face images.

As shown in Figure 1, currently four data innovation communities (DICs) are using the platform in different topical areas that offer domain-specific data sources for distinct research projects. Interested organizations from industry and academia are welcome to participate in one of the following four topical areas but are particularly encouraged to participate with ready available data, good analytics algorithms, or interesting research questions.

The DIC Energy is headed by KIT and EnBW and explores important datadriven aspects in the area of energy, such as the demand-driven fine-tuning of consumption rate models based on smart metre generated data. The DIC Smart Cities is headed by Fraunhofer IAIS and Siemens explores data-driven aspects of urban life, such as traffic control, but also waste disposal or disaster control. The DIC Industry 4.0 is headed by Bosch and DFKI and explores important data-driven aspects of the fourth industrial revoluation (towards Smart Factories), such as proactive

service and maintenance of production resources or finding anomalies in production processes. In one of the projects of this particular DIC, Trumpf is working with SAP and KIT on conditionbased monitoring of production systems while Trumpf also is starting to work with IBM, KIT and the Jülich Supercomputing Centre (JSC) on optimizations and classification problems for automatically detecting good or bad welding processes of materials.

The DIC Medicine is headed by Forschungszentrum Jülich and Bayer and works on three different research projects. The JSC and Jülich Institute of Neuroscience and Medicine (INM) closely work together with IBM on a machine learning approach for background segmentation of 3D image volumes of a brain tissue block as shown in Figure 3. The University of Düsseldorf and IBM collaborate on a project about predicting optimal treatment procedures for spinal cord injury patients. In the third project the Ludwig Maximilian University of Munich (LMU) works with JSC and IBM on using machine learning techniques for better supporting the decisionmaking of doctors when picking patientspecific human eye therapies for patients suffering from eye illnesses such as agerelated macular degeneration.

Reference

[1] Smart Data Innovation Lab Website, Online: http://www.sdil.de/de/

contact: Morris Riedel, m.riedel@fz-juelich.de

JSC to participate in four Horizon2020 Centres of Excellence

On May 8, the results of a keenly contested call for new "Centres of Excellence" within the EU Horizon2020 E-INFRASTRUC-TURES Programme were announced [1]. These new funding instruments are intended to harness computational science and big data expertise in HPC in the promotion of scientific discovery and industrial competitiveness. Out of the 20 submitted proposals, 8 projects were approved for initial funding and 4 of these will include active JSC participation - see Table 1 and Ref. [2] for an overview. These four are: POP - Performance Optimisation and Productivity; MaX - Materials Design at the Exascale; E-CAM - an E-infrastructure for Software, Training and Consultancy in Simulation and Modelling; and EoCoE an Energy-oriented Centre of Excellence. All projects - subject to final approval by all participants - plan to start in the autumn of this year.

EoCoE

EoCoE ("echo"), coordinated by the Maison de la Simulation at CEA, France, received the highest grade of the evaluation, and aims to exploit the prodigious potential offered by the maturing HPC infrastructure to foster and accelerate the European transition to a reliable, low-carbon energy supply. EoCoE will achieve its goal via targeted support of four distinct renewable energy pillars: Meteorology (Wind), Materials (Earth), Hydrology (Water) and Fusion (Fire), each of which boasts activities with a heavy reliance on numerical modelling. From the project outset these four pillars will be anchored within a strong

		1	
Coordinator	Partners	Acronym	Proposal Title
CEA	13	EoCoE	Energy oriented Centre of Excellence for computer applications
CNR	12	MaX	Materials design at the eXascale
UCD	18	E-CAM	An e-infrastructure for software, training and consultancy in simulation and modelling
BSC	6	POP	Performance Optimisation and Productivity
DKRZ	16	ESiWACE	Excellence in Simulation of Weather and Climate in Europe
POTSDAM	10	COEGSS	Center of Excellence for Global Systems Science
ктн	11	BioExcel	Centre of Excellence for Biomolecular Research
MPG	11	NoMaD	The Novel Materials Discovery Laboratory

Table 1: Summary of approved EU Centres of Excellence.

transversal multidisciplinary basis providing high-end expertise in applied mathematics and supercomputing science. EoCoE is structured around a central Franco-German hub coordinating a pan-European network, gathering a total of 8 countries and 23 teams, including 5 separate FZJ units from JSC, IEK and IBG. Its partners are strongly engaged in both the HPC and energy fields; a prerequisite for the long-term sustainability of EoCoE and also ensuring that it is deeply integrated in the overall European strategy for HPC.

References

contact: Paul Gibbon, p.gibbon@fz-juelich.de

- Morris Riedel¹
- Christian Bodenstein¹
- Timo Dickscheid²
- Stefan Köhnen²
- Michael Beigl³
- ¹ Jülich Supercomputing Centre (JSC), Germany
- ² Institute of Neuroscience and Medicine, Forschungszentrum Jülich, Germany
- ³ SDIL-Coordinator, Karlsruhe Institute of Technology, Germany

Activities

Paul Gibbon

Forschungszentrum Jülich, Germany

[1] http://ec.europa.eu/research/ participants/portal/doc/call/h2O2O/ h2020-einfra-2015-1/1660102-h2020einfra-2015-1_flash_call_info_en.pdf

[2] http://www.prace-ri.eu (needs update)

Overview Courses GCS

Date(s)	Course Title	Location
2015, Nov 26-27	Introduction to the Programming and Usage of the Supercomputer Resources at Jülich	Jülich
2015, Nov 30 - Dec 1	Advanced Parallel Programming with MPI and OpenMP	Jülich
2015, Dec 7-11	Fortran for Scientific Computing	Stuttgart
2015, Dec 10-11	Node-Level Performance Engineering (PATC course)	Garching
2016, Jan 14	Introduction to hybrid Programming in HPC (PATC course)	Garching
2016, Feb 8-12	Parallel Programming & Parallel Tools	Dresden
2016, Feb 8-12	Programming with Fortran	Garching
2016, Feb 22-26 & Sep 5-9	Iterative Linear Solvers and Parallelization	Stuttgart & Garching
2016, Feb 29-Mar 4 & Sep 12-16	Computational Fluids Dynamics	Siegen & Stuttgart
2016, Mar 7-11 & Sep 19-23	CFD with OpenFOAM®	Stuttgart
2016, Mar 7-11	Parallel Programming of High Performance Systems	Erlangen
2016, Mar 14-16	Parallel I/O and Portable Data Formats (PATC course)	Jülich
2016, Mar 14-16 (not yet fixed)	GPU Programming using CUDA	Stuttgart
2016, Mar 17-18	OpenMP GPU Directives for Parallel Accelerated Supercomputers (PATC course)	Stuttgart
2016, Mar 17-18	Vectorisation and Portable Programming Using OpenCL	Jülich
2016, Mar 21-23	Parallelization with MPI and OpenMP	Mainz
2016, Mar (1-2 days)	Modeling in Computer Science	Stuttgart
2016, Apr (4 days)	VI-HPS Tuning Workshop (PATC course)	Garching

Date(s)		Course Title	Location	-
2016, Apr (1 day)		Introduction to descriptive and parametric statistic with R	Jülich	
2016, Apr 4-7		Advanced Topics in High Performance Computing (PATC course)	Garching	
2016, Apr 12-15 & (Oct 25-28	Cray XC4O Workshop and Parallel I/O Courses	Stuttgart	
2016, Apr 18		Introduction to Parallel In-Situ Visualization with Vislt	Jülich	
2016, Apr 18-19 & N	Nov 3-4	Scientific Visualization	Stuttgart	
2016, Apr 20-21		NEC SX-ACE - Vectorization and Optimization	Stuttgart	
2016, Apr 25-27		Introduction to GPU Programming Using CUDA (PATC course)	Jülich	
2016, Apr (3 days)		Intel MIC&GPU Programming Workshop (PATC course)	Garching	
2016, May 23-24 &	Nov 24-25	Introduction to the Programming and Usage of the Supercomputer Resources at Jülich	Jülich	
2016, May 30 - Jun	2	Programming in C++	Jülich	
2016, Jun 2-3		Introduction to Unified Parallel C (UPC) and Co-Array Fortran (CAF) (PATC course)	Stuttgart	
2016, Jun 6-7		Efficient Parallel Programming with GASPI	Stuttgart	
2016, Jun 13		Introduction to hybrid programming in HPC	Stuttgart	
2016, Jun 13-14		High Performance Computing with Python (PATC course)	Jülich	
2016, Jun 15-16		Cluster Workshop	Stuttgart	
2016, Jun 27-28		High Performance scientific computing in C++	Jülich	
2016, Jun 27-28		Node-Level Performance Engineering (PATC course)	Stuttgart	
2016, Aug 1 - Oct 7		Guest Student Programme at JSC	Jülich	

Courses and Tutorials

Introduction to the **Programming and Usage of** the Supercomputer **Resources at Jülich**

Date & Location November 26-27, 2015 JSC, Forschungszentrum Jülich

Contents

This course gives an overview of the supercomputers at Jülich. Especially new users will learn how to program and use these systems efficiently. Topics discussed are: system architecture, usage model, compilers, tools, monitoring, MPI, OpenMP, performance optimization, mathematical software, and application software.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/sc-nov

Advanced Parallel **Programming with MPI and OpenMP**

Date & Location

November 30 - December 1, 2015 JSC, Forschungszentrum Jülich

Contents

The focus is on advanced programming with MPI and OpenMP. Hands-on sessions (in C and Fortran) will allow users to immediately test and understand the taught constructs of the Message Passing Interface (MPI) and the shared memory directives of OpenMP. Course

language is German. This course is organized by JSC in collaboration with HLRS. Presented by Dr. Rolf Rabenseifner, HI BS.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/mpi

Fortran for Scientific Computing

Date & Location December 7-11, 2015 HLRS, Stuttgart

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/training/2015/FTN2 **Node-Level Performance** Engineering (PATC course)

Date & Location December 10-11, 2015 LRZ, Garching near Munich

Contents

This course teaches performance engineering approaches on the compute node level. "Performance engineering" as we define it is more than employing tools to identify hotspots and bottlenecks. It is about developing a thorough understanding of the interactions between software and hardware. This process must start at the core, socket, and node level, where the code gets executed that does the actual computational work. Once the architectural requirements of a code are understood and correlated with performance measurements, the potential benefit of optimisations can often be predicted. We introduce a "holistic" node-level performance engineering strategy, apply it to different algorithms from computational science, and also show how an awareness of the performance features of an application may lead to notable reductions in power consumption.

Prerequisites

Participants must have basic knowledge in programming with Fortran or C.

Webpage

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

Introduction to hybrid Programming in HPC (PATC course)

Date & Location January 14, 2016 LRZ, Garching near Munich

Contents

Most HPC systems are clusters of shared memory nodes. Such SMP nodes can be small multi-core CPUs up to large many-core CPUs. Parallel programming may combine the distributed memory parallelization on the node interconnect (e.g., with MPI) with the shared memory parallelization inside of each node (e.g., with OpenMP or MPI-3.0 shared memory). This course analyses the strengths and weaknesses of several parallel programming models on clusters of SMP nodes. Tools for hybrid programming such as thread/ process placement support and performance analysis are presented in a "how-to" section. Hands-on exercises give attendees the opportunity to try the new MPI shared memory interface and explore some pitfalls of hybrid MPI+OpenMP programming.

Prerequisites

Basic MPI and OpenMP knowledge as presented in the LRZ course "Parallel programming of High Performance Systems".

Webpage

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

Parallel Programming & Parallel Tools

Date & Location February 8-12, 2016

Dresden, ZIH

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/training/2016/ZIH

Programming with Fortran

Date & Location February 8-12, 2016 LRZ, Garching near Munich

Contents

This course is targeted at scientists with little or no knowledge of the Fortran programming language, but need it for participation in projects using a Fortran code base, for development of their own codes, and for getting acquainted with additional tools like debugger and syntax checker as well as handling of compilers and libraries. The language is for the most part treated at the level of the Fortran 95 standard; features from Fortran 2003 are limited to improvements on the elementary level. Advanced Fortran features like object-oriented programming or Coarrays will be covered in a follow-on course in autumn 2016.

To consolidate the lecture material, each day's approximately 4 hours of lectures are complemented by 3 hours of hands-on sessions.

Prerequisites

Course participants should have basic UNIX/Linux knowledge (login with secure shell, shell commands, basic programming, vi or emacs editors)

Webpage

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

Iterative Linear Solvers and Parallelization

Date & Location

February 22-26, 2016 HLRS, Stuttgart and September 5-9, 2016 LRZ, Garching near Munich

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/training/2016/ITER-S

Computational Fluids Dynamics

Dates & Location Feb 29 - Mar 4, 2016 STS, University Siegen and September 12-16, 2016 HLRS, Stuttgart

Contents

The course deals with current numerical methods for Computational Fluid Dynamics in the context of High Performance Computing. An emphasis is placed on explicit methods for compressible flows, but classical numerical methods for incompressible Navier-Stokes equations are also covered. A brief introduction to turbulence modelling is also provided by the course. Additional topics are high order numerical methods for the solution of systems of partial differential equations. The last day is dedicated to parallelization. Hands-on sessions will manifest the contents of the lectures. In most of these sessions, the application Framework APES will be used. They cover grid generation using Seeder, visualization with ParaView and the usage of the parallel CFD solver Ateles on the local

HPC system.

The course is organized by STS, ZIMT (University of Siegen) and HLRS, IAG (University of Stuttgart).

Webpage

http://www.hlrs.de/training/2016/ CFD-SIEGEN

CFD with OpenFOAM®

Date & Location March 7-11, and Sep. 19-23, 2016 HLRS, Stuttgart

Contents

OpenFOAM® is a widely-used opensource code and a powerful framework for solving a variety of problems mainly in the field of CFD. The five-day workshop gives an introduction to Open-FOAM® applied on CFD phenomena and is intended for beginners and people with CFD background knowledge. The user will learn about case setup, meshing tools and available OpenFOAM® utilities that can be used for pre- and postprocessing tasks as well as job monitoring during runtime.

Additionally, basic solvers and major aspects of code structure are highlighted. Lectures and hands-on session with examples will guide through this course.

Webpage

http://www.hlrs.de/training/2016/ OpenFOAM1

Courses and Tutorials

Parallel Programming of High Performance Systems

Date & Location

March 7-11, 2016 RRZE building, University campus Erlangen, via video conference at LRZ if there is sufficient interest.

Contents

This course, a collaboration of Erlangen Regional Computing Centre (RRZE) and LRZ, is targeted at students and scientists with interest in programming modern HPC hardware, specifically the large-scale parallel computing systems available in Jülich, Stuttgart and Munich. Each day is comprised of approximately 4 hours of lectures and 3 hours of hands-on sessions.

Tentative Schedule

Day 1

- Introduction to High Performance Computing
- Secure shell
- Source code versioning with SVN
- Handling console and GUI-based
- interfaces
- Building programs with GNU MAKE

Day 2

- Basic parallel programming models:
- elements of MPI and OpenMP
- Processor architectures

Webpage http://www.lrz.de/services/compute/

courses/

Fortran 95, C or C++.

• Intel VTune Amplifier

• Principles of code optimization: unrolling, blocking, dependencies, C++ issues. bandwidth issues. performance projections

Advanced MPI programming

Day 4

Day 5

correctness

Prerequisites

Day 3

- Advanced OpenMP programming
- Parallel architectures: multi-core, multi-socket, ccNUMA, cache coherence and affinity, tools for handling memory affinity
- domain decomposition, task parallelism, masterworker, granularity, load balancing, scalability models

• Parallel algorithms: data parallelism,

• Intel tools: MPI tracing and Checking • Intel tools: OpenMP performance and

• Performance analysis with Scalasca

Good working knowledge of at least one of the standard HPC languages:

Parallel I/O and Portable **Data Formats**

(PATC course)

Date & Location

March 14-16, 2016 JSC, Forschungszentrum Jülich

Contents

This course will introduce MPI parallel I/O and portable, self-describing data formats, such as HDF5 and NetCDF. Participants should have experience in parallel programming in general, and either C/C++ or Fortran in particular. This course is a PATC course (PRACE Advanced Training Centres).

Webpage

http://www.fz-juelich.de/ias/jsc/ events/parallelio

Courses and Tutorials

GPU Programming using CUDA

Date & Location March 14-16, 2016 HLRS, Stuttgart

Contents

The course provides an introduction to the programming language CUDA, which is used to write fast numeric algorithms for NVIDIA graphics processors (GPUs). Focus is on the basic usage of the language, the exploitation of the most important features of the device (massive parallel computation, shared memory, texture memory) and efficient usage of the hardware to maximize performance. An overview of the available development tools and the advanced features of the language is given.

Webpage

http://www.hlrs.de/training/2016/CUDA1

OpenMP GPU Directives for Parallel Accelerated Supercomputers

(PATC course)

Date & Location March 17-18, 2016 HLRS, Stuttgart

Contents

This workshop will cover the programming environment of Cray hybrid supercomputer, which combines multicore CPUs with GPU accelerators. Attendees will learn about the directive-based OpenMP programming model whose multivendor support allows users to portably develop applications for parallel accelerated supercomputers.

The workshop will also demonstrate how to use the Cray Programming Environment tools to identify CPU application bottlenecks, facilitate the OpenMP porting, provide accelerated perfor-

mance feedback and to tune the ported applications. The Cray scientific libraries for accelerators will be presented, and interoperability of OpenMP directives with these and with CUDA will be demonstrated. Through application case studies and tutorials, users will gain direct experience of using OpenMP directives in realistic applications. Users may also bring their own codes to discuss with Cray specialists or begin porting.

Webpage

http://www.hlrs.de/training/2016/XK

Vectorisation and Portable Programming Using OpenCL

Date & Location March 17-18, 2016 JSC, Forschungszentrum Jülich

Contents

OpenCL provides an open, portable C-based programming model for highly parallel processors. In contrast to NVIDIA's proprietary programming API CUDA, a primary goal of OpenCL is portability across a diverse set of computing devices including CPUs, GPUs, and other accelerators.

Topics covered will include an introduction to vectorisation, the programming model of OpenCL, datatypes and OpenCL vectorisation features, the tuning for architectures like CPUs, accelerators (GPUs), and co-processors (Xeon Phi), and heterogeneous multi-device programming.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/opencl

Parallelization with MPI and **OpenMP**

Date & Location March 21-23, 2016 Mainz University, ZDV

Contents

The focus is on the programming models MPI and OpenMP. 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 the University of Mainz in collaboration with HLRS. Presented by Dr. Rolf Rabenseifner, HLRS.

Webpage

http://www.fz-juelich.de

Modeling in Computer Science

Date & Location

March 2016 (1-2 days) HLRS, Stuttgart

Contents

This course deals with translations between domain-oriented, mathematical, numerical and algorithmic models in the context of the so-called Simulation Pipeline. Texts treated in this course will stem mainly from the philosophy of computer simulation. Topics include: • Backward propagation of semantics

- in translations
- The changing role of mathematical models
- Verification, validation and evaluation.

Webpage

http://www.hlrs.de/training/2016/PHIL

VI-HPS Tuning Workshop

(PATC course)

Date & Location

April 2016 (4 days) LRZ, Garching near Munich

Contents

This workshop organized by VI-HPS for the German Prace Advanced Training Centre will

- give an overview of the VI-HPS programming tools suite,
- explain the functionality of individual tools, and how to use them effectively, and
- offer hands-on experience and expert assistance using the tools.

Presentations and hands-on sessions are planned on the following topics:

- Setting up, welcome and introduction
- LWM2 lightweight monitoring
- Score-P instrumentation and
- measurement
- Scalasca automated trace analysis
- Vampir interactive trace analysis
- Periscope automated performance analysis
- Paraver trace analysis and
- performance prediction
- MAQAO performance analysis & optimization
- MUST runtime error detection for MPI
- MAP profiling and performance reports.

Webpage

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

Courses and Tutorials

Introduction to descriptive and parametric statistic with R

Date & Location

April 2016 (1 day) JSC, Forschungszentrum Jülich

Contents

Some basic statistical tools are presented and investigated using R freeware and data from pedestrian trajectories. The objective is to propose useful methods allowing to analyze data or to calibrate a model, as well as to learn how to use R. The course is organized in three sessions of two hours:

• The first session is about basic descriptive tools for univariate and bivariate data (such as histogram, kernel density, measures of location/variability, boxplot, covariance, correlation, simple linear regression, qqplot, Bootstrap methods)

• The second session concerns descriptive analysis of multivariate data: Principal component analysis and Multiple linear regression / Estimation by least-square.

• The last session is about parametric statistic: Estimator (definition and property); Punctual estimation of parameters (by maximal likelihood and Bayesian approach); Information criteria (AIC, BIC, likelihood ratio...); Precision of the estimations (confidence and credible intervals); Tests of hypothesis.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/r

Advanced Topics in High Performance Computing

(PATC course)

Date & Location April 4-7, 2016 LRZ, Garching near Munich

Contents

In this add-on course to the parallel programming course special topics are treated in more depth, in particular processor-specific optimisation, I/O and PGAS concepts. The course is provided in collaboration of Erlangen Regional Computing Centre (RRZE) and LRZ. Each day is comprised of approximately 5 hours of lectures and 2 hours of hands-on sessions.

Tentative Schedule

Day 1 • Processor-specific optimisation

Day 2

- Parallel I/O with MPI IO
- Tuning I/O on LRZ's HPC systems

Day 3

• Introduction into Intel Xeon Phi programming and OpenMP 4.0

• Portability of I/O: NetCDF, PnetCDF and HDF5

SIONlib: Scalable I/O library

Day 4

• PGAS programming with Coarray Fortran and Unified Parallel C

Prerequisites

Good MPI and OpenMP knowledge as presented in the course "Parallel Programming of High Performance Systems".

Webpage

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

Fortran for Scientific Computing

Date & Location

April 4-8, 2016 (PATC course) and December 5-9, 2016 HLRS, Stuttgart

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/training/2016/FTN1

Introduction to Python

Date & Location

April 11-13, 2016 JSC, Forschungszentrum Jülich

Contents

This course gives an introduction to the programming language Python. Topics are: data types, control structures, object-oriented programming, module usage. Additionally, Python's standard library and the GUI programming with wxWidgets will be explained.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/python

Cray XC40 Workshop and Parallel I/O Courses

Date & Location

April 12-15, and October 25-28, 2016 HLRS, Stuttgart

Contents

This course consist of two parts. Efficient Parallel I/O (PATC course): An increasing issue in parallel computing is the efficient handling of reading and writing files. The one day workshop Efficient parallel I/O offers users the ability to gain insight on how to optimize the I/O of an application. It shows methods and tools for parallel I/O including MPI-IO, HDF5, NetCDF. Thus the time spent in reading and writing files can be significantly reduced, increasing the total efficiency of an application.

The workshop does include some Cray specific information about tools but is still useful for I/O on other systems as well.

Optimization on Cray XC4O: The 2nd-4th day is mainly intended as a "Bring Your Own Code" (BYOC) workshop, and will cover the Cray Programming Environment, scientific libraries and profiling tools. Specialists from Cray will assist profiling and optimizing your programs.

Target audience are users running applications which are only scaling to a lower number of cores then expected or needed

Prerequisites: In order to attend the BYOC workshop, you will have to bring your own code, and you will be required to do some work on your application before the workshop. For these last three days, it is required that your have already an account on HAZELHEN and your application is already running on the system.

Webpage

http://www.hlrs.de/training/2016/ XC40-1

Autumn 2015 • Vol. 13 No. 2 • inSiDE

Introduction to Parallel In-Situ Visualization with VisIt

Date & Location

April 18, 2016 JSC, Forschungszentrum Jülich

Contents

Vislt is a distributed, parallel visualization and graphical analysis tool for data defined on two- and three-dimensional (2D and 3D) meshes. It lets you instrument your simulation code for in-situ visualization and analysis - visualization capabilities are added inside the simulation so it can visualize the data using the same level of resources as being used to calculate the data.

The course will cover an introduction and the basic aspects of Vislt as a visualization tool in general. In the second half we will focus on the use of Vislt in massive parallel environments. Especially the integration into existing simulation codes for in-situ visualization will be discussed. In a hands-on session the integration of Vislt will be demonstrated on an small parallel application.

Prerequisites

Knowledge about programming of parallel applications in C/C++ or Fortran.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/visit

Scientific Visualization

Date & Location

April 18-19, and November 3-4, 2016 HLRS, Stuttgart

Contents

This two day course is targeted at researchers with basic knowledge in numerical simulation, who would like to learn how to visualize their simulation results on the desktop but also in Augmented Reality and Virtual Environments. It will start with a short overview of scientific visualization, following a hands-on introduction to 3D desktop visualization with COVISE. On the second day, we will discuss how to build interactive 3D Models for Virtual Environments and how to set up an Augmented Reality visualization.

Prerequisites

Some knowledge about Linux, e.g. make, command line editor, Linux shell, experience in C.

Webpage

http://www.hlrs.de/training/2016/VIS1

NEC SX-ACE - Vectorization and **Optimization**

Date & Location April 20-21, 2016 HLRS, Stuttgart

Contents

In spring 2015, HLRS installed a next generation vector computer, a NEC SX-ACE. The participants learn about the configuration of the NEC SX-ACE system at HLRS and how to use this cluster of vectorizing shared memory nodes. One focus is an introduction in vectorization. More experienced users can learn how to optimize programs based on performance measurements. Additional topics are I/O and the optimization of application programs. The first day presents an introduction to the NEC SX architecture and vectorization. The second day morning is focused on usage aspects and the differences between the predecessor system (NEC

SX-9) and the new NEC SX-ACE. The afternoon is dedicated to hands-on sessions. Separate registration of individual days is possible. Participants are encouraged to prepare their own applications for use in the hands-on session. Participants familiar with SX vector computers may wish to register only for Day 2.

Prerequisites

C or Fortran, some knowledge about parallel programming.

Webpage

http://www.hlrs.de/training/2016/NEC

Introduction to GPU **Programming Using CUDA** (PATC course)

Data & Location April 25-27, 2016 JSC, Forschungszentrum Jülich

Contents

GPU-accelerated computing drives current scientific research. Writing fast numeric algorithms for GPUs offers high application performance by offloading compute-intensive portions of the code to an NVIDIA GPU. The course will cover basic aspects of GPU architectures and programming. Focus is on the usage of the parallel programming language CUDA-C which allows maximum control of NVIDIA GPU hardware. Examples of increasing complexity will be used to demonstrate optimization and tuning of scientific applications. This course is a PATC course (PRACE Advanced Training Centres).

Webpage

http://www.fz-juelich.de/ias/jsc/ events/cuda

Courses and Tutorials

Intel MIC&GPU Programming Workshop (PATC course)

Date & Location April 2016 (3 days) LRZ, Garching near Munich

Contents

With the rapidly growing demand for computing power new accelerator based architectures have entered the world of High Performance Computing since around 5 years. Particularly GPGPUs have recently become very popular, however programming GPG-PUs using programming languages like CUDA or OpenCL is cumbersome and error-prone. Recently, Intel developed their own Many Integrated Core (MIC) architecture which can be programmed using standard parallel programming techniques like OpenMP and MPI.

Since 2014, LRZ is running a new Intel Xeon Phi based system SuperMIC. This workshop is especially recommended for users who want to port their applications to SuperMIC or other Xeon Phi based systems.

The course discusses various programming and optimization techniques for Intel Xeon Phi in comparison with GPGPU programming techniques and includes several hands-on sessions on the new SuperMIC system at LRZ and on a GPU cluster. The course is developed in collaboration with the Erlangen Regional Computing Centre (RRZE) within KONWIHR.

Prerequisites

Good working knowledge of at least one of the standard HPC languages: Fortran 95, C or C++. Basic OpenMP and MPI knowledge useful.

Webpage

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

Introduction to the Programming and Usage of the Supercomputer **Resources at Jülich**

Data & Location

May 23-24, and November 24-25, 2016 JSC, Forschungszentrum Jülich

Contents

This course gives an overview of the supercomputers at Jülich. Especially new users will learn how to program and use these systems efficiently. Topics discussed are: system architecture, usage model, compilers, tools, monitoring, MPI, OpenMP, performance optimization, mathematical software, and application software.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/sc-may

Programming in C++

Data & Location May 30 - June 2, 2016 JSC, Forschungszentrum Jülich

Contents

C++ is a multi-paradigm programming language supporting procedural, object-oriented, generic and functional programming styles. This course will provide a practical introduction to the current standard of the language, C++14.

Through a number of simple but instructive exercises, the participants will learn the C++ syntax and familiarise themselves with C++ abstraction mechanisms such as classes and class hierarchies, and templates. The Standard Template Library for C++14 will be introduced in sufficient detail to be useful

This course assumes that the participants have previous programming experience in languages such as C, Python, Java and Fortran, and will introduce the current standard of C++. It is not meant to be a beginners' introduction to programming.

Prerequisites

Good working knowledge of C, Python, Java or Fortran.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/cplusplus

Introduction to Unified Parallel C (UPC) and **Co-Array Fortran (CAF)** (PATC course)

Data & Location June 2-3, 2016 HLRS, Stuttgart

Contents

Partitioned Global Address Space (PGAS) is a new model for parallel programming. Unified Parallel C (UPC) and Co-Array Fortran (CAF) are PGAS extensions to C and Fortran. PGAS languages allow any processor to directly address memory/data on any other processors. Parallelism can be expressed more easily compared to library based approaches as MPI. Hands-on sessions (in UPC and/or CAF) will allow users to immediately test and understand the basic constructs of PGAS languages.

Webpage

http://www.hlrs.de/training/2016/UPC1

Efficient Parallel Programming with GASPI

Data & Location June 6-7, 2016 HLRS, Stuttgart

Contents

In this tutorial we present an asynchronous data flow programming model for Partitioned Global Address Spaces (PGAS) as an alternative to the programming model of MPI. GASPI, which stands for Global Address Space Programming Interface, is a partitioned global address space (PGAS) API. The GASPI API is designed as a C/C++/Fortran library and focused on three key objectives: scalability, flexibility and fault tolerance. In order to achieve its much improved scaling behaviour GASPI aims at asynchronous dataflow with remote completion, rather than bulk-synchronous message

exchanges. GASPI follows a single/multiple program multiple data (SPMD/ MPMD) approach and offers a small, vet powerful API (see also http://www. gaspi.de and http://www.gpi-site.com). GASPI is successfully used in academic and industrial simulation applications. Hands-on sessions (in C and Fortran) will allow users to immediately test and understand the basic constructs of

Webpage

GASPI.

http://www.hlrs.de/training/2016/GASPI

Introduction to hybrid Programming in HPC

Data & Location June 13, 2016 HLRS, Stuttgart

Contents

Most HPC systems are clusters of shared memory nodes. Such SMP nodes can be small multi-core CPUs up to large many-core CPUs. Parallel programming may combine the distributed memory parallelization on the node interconnect (e.g., with MPI) with the shared memory parallelization inside of each node (e.g., with OpenMP or MPI-3.0 shared memory). This course analyses the strengths and weaknesses of several parallel programming models on clusters of SMP nodes. Tools for hybrid programming such as thread/ process placement support and performance analysis are presented in a "how-to" section. Hands-on exercises give attendees the opportunity to try the new MPI shared memory interface and explore some pitfalls of hybrid MPI+OpenMP programming.

Prerequisites

Basic MPI and OpenMP knowledge as presented in the LRZ course "Parallel programming of High Performance Systems".

Webpage http://www.hlrs.de/training/2016/HY-S

Courses and Tutorials

High Performance Computing with Python (PATC course)

Data & Location June 13-14, 2016 JSC, Forschungszentrum Jülich

Contents

Python is being increasingly used in High Performance Computing projects such as GPAW. It can be used either as a high-level interface to existing HPC applications, as embedded interpreter, or directly. This course combines lectures and hands-on session. We will show how Python can be used on parallel architectures and how performance critical parts of the kernel can be optimized using various tools.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/hpc-python

Cluster Workshop

Data & Location June 15-16, 2016 HLRS, Stuttgart

Contents

Nowadays, the IT infrastructure of many companies and institutions includes modern compute clusters. The procurement, operation and the efficient usage of such parallel systems introduces new and complex requirements.

To address these emerging requirements, the High Performance Computing Center of the University of Stuttgart (HLRS) offers a vendor-independent workshop. Topics span from the design of compute clusters to details on different hardware components, operating systems, file systems and modes of operation, touching on specific software solutions. Further, typical problems and strategies for their solution will be discussed

HLRS has been running compute clusters for many years, providing compute resources for scientific and industrial simulations. HLRS is in constant dialoque with users and hardware providers and has accumulated a large knowledge base in cluster computing.

Webpage

http://www.hlrs.de/training/2016/ CLUSTER

High Performance scientific Computing in C++

Data & Location

June 27-28, 2016 JSC, Forschungszentrum Jülich

Contents

Modern C++ with its support for procedural, objected oriented, generic and functional programming styles, offers many powerful abstraction mechanisms to express complexity at a high level while remaining very efficient. It is therefore the language of choice for many scientific projects. However, achieving high performance on contemporary computer hardware, with many levels of parallelism, requires understanding C++ code from a more performance centric viewpoint. In this course the participants will learn how to write programs in C++ that better utilize the modern processors. The participants will learn how to identify and avoid performance degrading characteristics, such as unnecessary memory operations, branch mispredictions, and unintentionally strong ordering assumptions. Two powerful open source libraries to help write structured multi-threaded applications will also be introduced: Intel® Threading Building Blocks and NVIDIA Thrust.

Prerequisites

Good working knowledge of C++, especially the C++11 standard.

Webpage

http://www.fz-juelich.de/ias/jsc/ events/hpc-cplusplus

Courses and Tutorials

Node-Level Performance Engineering

(PATC course)

Data & Location June 27-28, 2016

HLRS, Stuttgart

Contents

This course teaches performance engineering approaches on the compute node level. "Performance engineering" as we define it is more than employing tools to identify hotspots and bottlenecks. It is about developing a thorough understanding of the interactions between software and hardware. This process must start at the core, socket, and node level, where the code gets executed that does the actual computational work. Once the architectural requirements of a code are understood and correlated with performance measurements, the potential benefit of optimisations can often be predicted. We introduce a "holistic" node-level performance engineering strategy, apply it to different algorithms from computational science, and also show how an awareness of the performance features of an application may lead to notable reductions in power consumption.

- Introduction to computer
- architecture - x86 architectures
- Pipelining, out of order processing - Cache hierarchy
- SIMD, SMT, ccNUMA
- LIKWID tools: topology, pin, perfctr

• Microbenchmarking for architectural

- exploration - Scalable and bottlenecked resources
- OpenMP overhead
- Chip-level performance modeling - Roofline model
- Example: vector summation
- Case study: The Jacobi solver
- Modeling with the Roofline model - Model-guided optimizations
- Case study: sparse
- matrix-vector multiply
- Phenomenology
- CRS storage format
- Roofline modeling
- Making best use of parallelism - Simultaneous multi-threading (SMT) - ccNUMA
- Single Instruction multiple Data (SIMD)
- Examples
- Further case studies (optional)
- Energy & Parallel Scalability - Energy consumption of modern processors
- The energy-to-solution metric - Performance engineering == power engineering
- Practical performance engineering with patterns

Prerequisites

Participants must have basic knowledge in programming with Fortran or C.

Webpage http://www.hlrs.de/training/2016/NLP

User-Guided Optimization in High-Level Languages

Data & Location

June 29, 2016 (not yet fixed) HLRS, Stuttgart

Contents

While writing code in High-level languages, HPC-programmers often have a notion of gainful optimization strategies, that should be applied to their code. Nevertheless, nowadays compilers generally use heuristics to decide on optimization strategies. To lower this gap, we will present new ideas and tools to raise the programmers control over compiler optimizations, i.e. 1) Noise - trigger optimizations by annotations

- 2) Sierra SIMD computations with compound data types
- 3) AnyDSL create DSLs with associated optimization strategies

All of these tools will also be applied in hands-on sessions. In addition, an overview of available optimization strategies will be given by compiler experts.

Noise and Sierra are limited to C/C++, even though the underlying concepts ought to be applicable also to other imperative languages, particularly Fortran. AnyDSL uses a dialect of the Rust programming language.

Attendees are highly invited to also join the "Node-level Performance Engineering" - course held on June 27-28 at HLRS that targets related topics!

Webpage

http://www.hlrs.de/training/2016/UG0

Guest Student Programme at JSC

Data & Location

August 1 - October 7, 2016 JSC, Forschungszentrum Jülich

Contents

Guest Student Programme "Scientific Computing" to support education and training in the field of supercomputing. Students of Computational Sciences, Computer Science and Mathematics can work 10 weeks in close collaboration with a local scientific host on a subject in their field. Application deadline is March 31, 2016.

Webpage

http://www.fz-juelich.de/ias/jsc/gsp





Contents

News	
LRZ: SuperMUC Phase 2 operating	4
HLRS almost doubles Computing Capacities with new HPC System Hazel Hen	7
Preparatory Access to Computing and Support Resources at JSC	9
Applications	
Convection permitting Latitude-Belt Simulation using the Weather	
Research and Forecasting (WRF) Model	10
Prediction of the turbulent Flow Field around a ducted Axial Fan	13
A CFD Study of Centrifugal Compressor with Circum-	
ferentially non-Uniform Inlet Distortions using Ansys CFX	15
Improving the Plasma Simulation Code (PSC)	18
Disentangling Evolution on the SuperMUC	22
SuperMUC Phase 2: Challenges of Peta-scale Applications	26
Visualizing 10 ¹¹ Particles from Cosmological Simulations	29
Analysis of 3D Point Clouds using a Parallel DBSCAN Clustering Algorithm	32
Directing the Morphology of amphiphilic Molecules	35
First Steps towards predicting the Evolution of Catchment-Scale terrestrial Systems	41
Projects	
CoeGSS: Center of Excellence for Global Systems Science	44
POP: Performance Optimization and Productivity – A Centre of Excellence in	
Computing Applications	47
ESSEX – Equipping Sparse Solvers for Exascale	51
The Mont-Blanc Project: First Phase successfully finished	54
EU Project "AutoTune" finished	58
Automated Analysis of the Performance of Jobs on SuperMUC with the PerSyst Tool	61
JSC's Horizon 2020 Projects for Designing Future HPC Technologies	64
Future Supercomputers for Brain Research: Pre-Commercial Procurement	
entered final Phase	65
Systems	67
Centres	70
Activities	76
Courses	90