



## *Special Feature: European-US Summer School on HPC Challenges in Computational Sciences*

The first European-US Summer School on HPC Challenges in Computational Sciences organized jointly by the two projects DEISA and TeraGrid took place in Acireale near Catania in Sicily, Oct 3-7, 2010.

The primary objective was to advance computational sciences by enabling and stimulating future international collaboration, innovation, and discovery through the most effective use of HPC, following a suggestion of leading computational scientists from both continents.

Sixty Students and post-docs in total were among the participants, selected from more than 100 applications: 25 from US and 35 from EU universities and research institutions. Females were represented with a fraction of 20%.

25 high level speakers were covering major fields of computational sciences, with 9 being from the US and 16 from Europe.

### The Keynote

Keynote speaker was Barry Schneider, TeraGrid program director of the National Science Foundation (NSF) in the US. He presented the "NSF Vision on HPC Challenges: Creating a New CyberScience (CS) and CyberInfrastructure (CI) Framework for 21st Century Science & Engineering (CF21)". Starting with the profound transformation of science, from the days of Galileo and Newton, to the early work of Stephen Hawking, science breakthroughs were mostly achieved by individuals, whereas recently this situation changed dramatically, with multidisciplinary teams

using powerful (computing) tools to solve complex problems. He then went on to describe the five biggest challenges: computing technology; data, provenance, and visualization; software; organization for multidisciplinary computational science; and education. These challenges result in the need for a CS-CI Ecosystem, and he described the respective cyberinfrastructure framework CF21, an organizing fabric and foundation for science, engineering and education, and its positive impact on the different challenges.

### The HPC Ecosystem: DEISA, TeraGrid, and PRACE

The DEISA overview was given by Hermann Lederer from the Max-Planck-Institute for Plasmaphysics (IPP) in Garching near Munich. His talk "The DEISA Approach to Advance Computational Sciences in Europe" started with describing the 15 partners, the DEISA supercomputers, and the evolution of DEISA from an early testbed in 2004 to a sustainable grid/data-based e-Infrastructure with a professional service. He then went on to summarize the architecture of DEISA, including the 15 different computing environments, the 10 Gb/s network and the continental-scale file system, the Common Production Environment, and the secure login with a single sign-on. The last part of the talk was dedicated to DECI, the DEISA Extreme Computing Initiative which, during the last six years, supported over 250 research projects with grand challenge application simulations.



**DEISA @ SC10**

DEISA will organize a BoF session at SC10 together with PRACE, Partnership for Advanced Computing in Europe. The session called 'PRACE – The European HPC infrastructure created' will take place on Wednesday 17 November from 12:15 to 13:30. The BoF will present the latest news about the PRACE RI. It covers the current status and the future plans; the results to be expected during the EC funded implementation projects, the integration of services currently provided by DEISA within the European HPC ecosystem, and collaboration opportunities for academia and industry.

Visit DEISA-PRACE booth 4021

The “TeraGrid Overview” was presented by John Towns who is the Director of Persistent Infrastructure at the Illinois National Centre for Supercomputing Applications (NCSA), and the TeraGrid Forum Chair. TeraGrid’s mission is to support the most advanced computational science in multiple domains, empower new communities of users, and provide resources and services that can be extended to a broader cyberinfrastructure. TeraGrid is an advanced, nationally distributed, open cyberinfrastructure comprised of supercomputing, storage, and visualization systems, data collections, and science gateways, integrated by software services and high bandwidth networks, coordinated through common policies and operations, and supported by computing and technology experts, that enables and supports leading-edge scientific discovery and promotes science and technology education. It consists of over a dozen organizations and NSF awards working together to provide collective services that go beyond what can be provided by individual institutions.

The last speaker of the introductory session was Giovanni Erbacci from the CINECA Supercomputer Center in Bologna, Italy, with an overview on “PRACE, the Partnership for Advanced Computing in Europe”. After several years of intensive collaboration of European countries, the PRACE preparatory project started in 2008 during which six different HPC prototype architectures have been identified and installed, while the PRACE implementation and operation phase began in July 2010, as an international not-for-profit association under Belgian law with its seat in Brussels, with 20 European partners, and with funding from the EU FP7 program and the PRACE partner states. PRACE aims at creating a European HPC ecosystem by providing a world-leading persistent high-end HPC infrastructure at the top of an HPC provisioning pyramid, managed as a single legal entity. The focus is on deploying systems of the highest performance level (Tier-0), ensuring a diversity of architectures to meet the needs of European user communities, collaborating with vendors and ISVs on strategic HPC technologies, and providing support and training. The first tier-0 system, an IBM Blue Gene/P system with almost 300000 cores, is now available for early access, for European scientists and their peers around the world, through common access



*Especially the sessions on programming, languages, libraries, data, visualization, etc. were of broader interest, because almost everybody has to deal with them.*

methods and user environment at all centres, built on earlier work in the DEISA project.

### Challenges in Materials

Thomas Schulthess from ETH Zurich spoke about “Petascale Computing in Nano- & Materials Science”. His main question was on how challenging it is to turn electronic structure problems into routine calculations. He then looked at different methods for free energy calculations in magnetic nanoparticles and superconductivity and model of high Tc superconductors, and discussed implications on application development.

Axel Arnold from the University of Stuttgart looked at “Advanced Algorithms for Long-Range Coulomb Interactions”, especially the P3C, ELC, and MEMD methods. The Coulomb potential has two problems: first it is singular at each particle position, and second, it is very slowly decaying. He discussed the Ewald method (and its challenges and constraints) which separates the two problems (splits the potential) into smooth potential k-space, and short-ranged potential real space. Finally, he presented the ESPResSo code, an Extensible Simulation Package for Research on Soft matter.

P. Giannozzi from the University of Udine discussed “Simulations at the Nanoscale Using Quantum Espresso”, looking at phenomena happening on a scale of lengths up to a

few tens of nm. The basic theoretical tools are Density-Functional Theory (DFT), Pseudopotentials, Car-Parrinello and other iterative techniques, together sometimes referred to as The Standard Model of Materials Science. Quantum ESPRESSO is a Quantum open-Source Package for Research in Electronic Structure, Simulation, and Optimization. It is a software for atomistic calculations based on electronic structure, using density-functional theory, a plane-wave basis set, and pseudopotentials. The main goals of quantum ESPRESSO are innovation in methods and algorithms, and efficiency on modern computer architectures. A great effort is also devoted to user friendliness and to the formation of a users and developers community. The code is able to do structural modeling (equilibrium structures of molecules, crystal, surfaces), linear response functions (vibrational and dielectric properties), plus some non-linear ones (third-order force constants and dielectric response, non-resonant Raman), chemical reactivity and transition-path sampling (NEB), dynamical modeling (ab initio molecular dynamics), computational microscopy (STM), quantum (ballistic) transport, among others. He closed his presentation with an ab-initio simulation of Protein-Surface Interactions mediated by water, which ran on a DEISA BlueGene/P with 65K cores for 686 atoms.

Marcella Iannuzzi from the University of Zurich focused on the CP2K code for ab-initio Molecular Dynamics. One of the main chal-



*Students valued the fact that they got good information about the latest trends, which one does not yet find in textbooks.*

Challenges in molecular dynamics are the huge differences in time (picoseconds to seconds) and in space scales (a few atoms, to nanometers, to meters), from electronic structures to continuum. A multi-scaling strategy implemented in the CP2K code is to extend the scope by reducing the complexity to propose accurate and realistic models aiming at bridging over time and length scales. She closed her talk with presenting a large number of simulations examples.

### Challenges in Life Sciences

David van der Spoel from Uppsala University in Sweden started the Life Science Session with an overview on "Gromacs: The Road Ahead". He started with a background on Molecular Dynamics, solving Newton's equations of particle motion and their (Lennard-Jones) interaction in a system. He went on to describe the implementation challenges with Gromacs (which is currently used by about 10K academic and industrial users) on large parallel computers, including Jaguar at ORNL, and highlighted especially the scalability challenges for very large problems (e.g. 3.3 Mio particles) and large data sets. Recently, Gromacs has scaled to 150K cores for up to 100 Mio atoms.

Thomas E. Cheatham from the University of Utah provided an "Insight into bio-molecular structure from simulation: promise, peril and barriers to moving beyond the (tera) scale". From the wide field of pharma, such as bio-molecular simulation, molecular docking, lead optimization, drug design, ADMET prediction, and structure prediction, he

mainly looked at simulations of bio-molecules, with a focus on the AMBER code, e.g. for protein folding, structure prediction, computer-aided drug design, calculation of materials properties, and elucidating reaction mechanisms. He addressed the challenges for the computational biologist when moving to what he calls tera- and peta-scale science.

### Challenges in Plasma Physics

Frank Jenko from the Max-Planck-Institute for Plasma Physics in Garching talked

about the HPC Challenges in Plasma Physics: Fusion Energy Research. Main focus was on the efforts to create fusion energy in the laboratory, e.g. through magnetic confinement of plasmas. He discussed the main challenges in fusion physics: small-scale instabilities: turbulent transport; plasma-wall interaction and materials; large-scale instabilities: magneto-hydrodynamics; and the huge range of spatial and temporal scales. He then described GENE, Gyrokinetic Electromagnetic Numerical Experiment, a code which is physically comprehensive and computationally efficient with applications in fusion research and astrophysics, applied for a deeper understanding of fundamental physics issues and direct comparisons with experiments/observations. He further explained these challenges with the example of tackling turbulence with extreme computing in support of ITER, the international joint effort for a 500 MW demonstration fusion power plant. "The successful operation of ITER will depend on our capability to reliably predict, interpret, and optimize ITER discharges by means of petascale (and exascale) simulations – helping to bring the sun on the earth."

Roman Hatzky from Max-Planck-Institute for Plasma Physics in Garching presented

Particle-In-Cell (PIC) Methods which are widely used in Plasma Physics, e.g. in gyrokinetics. In detail, he described the Vlasov-Maxwell equations, the PIC method and its discretization, Monte-Carlo evaluation of the integrals, the gyrokinetic Vlasov equation, and the control variates as variance reduction method. PIC simulations can consist of more than 10<sup>10</sup> markers which makes even 3-dimensional simulation domains with full velocity space possible. They do not evolve "numerical diffusion", and the PIC method is easy to parallelize for parallel computers.

### Challenges in Quantum Chromo Dynamics

Richard Brower from Boston University talked about QCD: Exploring Quarks and Beyond, and described what he called the AAA Challenge: Application (QCD Quantum Fields), Algorithms (Multigrid), and Architecture (GPU computing). He briefly introduced Quantum Field theory and how QCD relates to it, the computer architecture revolution, from Thinking Machine CM-2 to Nvidia GeForce GTX, and MultiGrid algorithms. His lessons: the future is full of hard problems; the trend is toward huge floating point performance, but relatively anemic memory bandwidth; and, inter-node communication is an even greater challenge.

### Challenges in Astro Sciences

HPC Challenges in Astro Sciences, with a special focus on Computational Astrophysics and Cosmology have been presented by Michael Norman from the San Diego Super-



computer Center and the Department of Physics at the University of California. He introduced the students to the wide area of simulation of astrophysical phenomena to better understand interplay of various physical processes at play, and computational cosmology and the simulation of the formation and evolution of gravitationally bound systems in the expanding universe under the influence of cold dark matter, and the formation and evolution of galaxies, quasars, and so on. Then he went on to describe Enzo: an adaptive mesh-refinement,

grid-based hybrid (hydro + N-body) community code for computational astrophysics and cosmology, which is designed to do simulations of cosmological structure formation. The code is able to simulate dark matter, gravity, gas dynamics, magnetic fields, radiation transport, multispecies chemistry, inertia tracer, source, and sink particles, and more. This code has been the basis for the examples presented by Michael Norman: among others galactic star formation and turbulence; galaxies, groups, and clusters; hierarchy of structures from 100 to 109 light years; and the evolution of the Universe (Big Bang expansion). The last part of his presentation dealt with the formation of the first stars and galaxies in the Universe, ab initio simulations spanning 1010 in spatial scales on up to 32K cores of the Blue Waters supercomputer.

## Challenges in Climate Research

Patrick Joeckel from the German Center of Aerodynamics and Aeronautics (DLR) in Oberpfaffenhofen talked about Climate Research and, more general, about Earth System Science. His focus was on the challenges from atmospheric and ocean circulation, and the coupling of atmosphere, land, ocean, ice, and biosphere. He then went on with climate modeling for understand the terrestrial climate system (as a whole), based on physical principles, and the climate variability of the past and the present. Thus, climate modeling serves as a tool to simulate projections for the climate of the future. He described in detail the construction of climate models, from conceptual model (and physical principles), mathematical formulation, numerical solu-



*Students thought that they mostly learn from the speakers that work in their field. They valued the opportunity to interact with them not only during their talk, but also during the whole event.*

tion, evaluation, and projection. Recent models include atmosphere, land surface, ocean and sea-ice, sulfate and non-sulfate aerosol, carbon cycle, dynamic vegetation, and atmospheric chemistry. He then presented discretization and mesh strategies in detail. Challenges here is the wide spectrum of scale, spatial from 10-3 to 107 m, temporal from one second to several years, and implementation, parallelization, and scaling on large multi-processor many-core HPC systems.

## Challenges in Computational Fluid Dynamics

Hans-Joachim Bungartz from the Technical University in Munich presented a Software Environment for Efficient Flow Simulations, starting with the wide range of computational challenges, increasing requirements, “multi-this and multi-that”, especially multi-disciplinarity, complex system architecture, software design and development, algorithms, complex (multi-physics) applications, and education. The second part of his talk was about the PDE Framework Peano, tackling the ‘memory wall’, on-chip parallelism, scalability, and the new paradigm of ‘space filling curves’. The last part was about ap-

plications: the drift Ratchet scenario in a silicon membran; and 2D and 3D flows around a sphere with drag and lift forces at low Reynolds numbers.

## Programming

David Henty from EPCC Edinburgh Parallel Computing Centre began his lecture with a surprise: he presented the 30 summary slides (in 5 minutes) of a workshop looking Towards Petaflops, and only disclosed at the end that this workshop took place at EPCC in 1999. His conclusion clearly was that the main challenges then are still the main challenges now, looking at the Peta- and Exaflops era. The second part of his presentation was on PGAS: new languages for parallel computing. The PGAS (Partitioned Global Address Space) programming model combines what we have learned so far from other models and languages, like OpenMP, MPI, UPC, CoArray



Fortran, and High Performance Fortran. He then went on to explain Unified Parallel C (UPC), a parallel extension to ISO C 99. He closed his session with several PGAS examples on how to program the derivative  $dy/dx$ , in C, Fortran, MPI, OpenMP, HPF, CAF, UPC, with drawing several lessons learned and conclusions.

The next lecture in this programming session was from Josep M. Perez from Barcelona Supercomputing Center, giving an overview on StarSs, a node-level programming model with a global view on multicore and parallel programming, suitable for a variety of parallel systems such as SMPs, GPUs, Grids, Cells, and hybrid systems. An advantage is that basically the same source code runs on any machine. He presented several programming examples, like linear algebra blocks, matrix multiplication, Cholesky, and Gauss-Seidel.

Giovanni Erbacci from CINECA in Italy closed the programming session with an Overview on Hybrid Programming. After an introduction into the evolution of modern HPC architectures and programming paradigms, he concentrated on hybrid programming based on MPI/OpenMP code using MPI for inter-node communication, and OpenMP for intra-node parallelization, for multi-node SMP (Symmetric Multiprocessor) connected by an interconnection network. With hybrid programming it is possible to decrease the number of messages by a factor of  $(\# \text{ threads}^2)$ , with increasing the length of messages by a factor of  $(\# \text{ threads})$ . He finished his talk by presenting a few examples of hybrid programming.

## Performance Analysis and Profiling

Philip Blood from Pittsburgh Supercomputer Center introduced the students to Performance Engineering of Parallel Applications. There are already many tools in the community which support performance analysis and profiling, and he mentioned POINT (Petascale Productivity from Open Integrated Tools), PAPI, PerfSuite, IPM, TAU, and Kojak/Scalasca, as examples. Then he described the overall code development and optimization process, measurement techniques, profiles, traces, and discussed in very detail the individual steps with the aid of UNRES, a code utilizing a mesoscopic protein force field to



*Students found parallel sessions on the different scientific areas stimulating, because they allowed them to look beyond their own horizon.*

study and predict protein folding pathways by means of molecular dynamics simulations. In the second part of his talk he asked the question “When will hybrid programming help my code to scale”, with the example of LAMMPS, an MPI code for Molecular Dynamics simulations. His answer was that hybrid programming can help to better balance between computation and communication at high core counts.

Bernd Mohr from Research Center Juelich talked about Trace-Based and Automatic Performance Measurement and Analysis, and he looked especially at the tools PAPI, Marmot, Vampir, Kojak/Scalasca, and Periscope. He elaborated in detail on performance measurement methods such as profiling, event tracing, trace visualizers such as Jumpshot, Paraver, Vampir, TAU, and the Intel Trace Collector and Analyzer. He then applied some of these tools to the production code UNRES, a code utilizing a mesoscopic protein force field to study and predict protein folding pathways by means of molecular dynamics simulations.

## Algorithmic Approaches and Libraries

Tony Drummond from Lawrence Berkeley National Laboratory gave an Overview on High-Performance Numerical Libraries, and focused especially on The DOE ACTS Collection of advanced computational software, which aims at making reliable and efficient software tools more widely used, and more effective in

## What students found valuable

- Parallel sessions on the different scientific areas were stimulating and allowed students to look beyond their own horizon.
- Most of the overviews on the different scientific areas were well presented.
- You mostly learn from the speakers that work in your field, also because you can directly interact with them not only during their talk, but also during the whole event.
- Especially the sessions on programming, languages, libraries, data, visualization, etc. were of broader interest, because almost everybody has to deal with them.
- You get good information about the latest trends, which you don't (yet) find in textbooks.
- Everybody agreed that the size of the class – 60 students – was ideal: small enough to still interact with everybody, large enough to be able to build groups of similar interest.

## What students missed

- It would be interesting to also hear from the speakers about efforts which did not work and were not successful, just to avoid making the same mistakes.
- Separation into parallel sessions was discussed controversially. About half of the students found it somewhat sad that they couldn't attend both sessions, because they very much liked the opportunity to learn more about other areas.

solving the nation's engineering and scientific problems. He explained the direct and iterative linear solvers, the factorization algorithms, and the libraries with their linear and non-linear equation routines such as ScaLAPACK, AztecOO, PETSc, OPT++ and TAO.

## Data Intensive Computing and Visualization

The final session of the summer school started with a presentation of John Johnson from the Pacific Northwest National Laboratory about Data-Intensive Science, i.e. problems where data is the dominating factor, with the challenges of large volumes, high throughput, relating and linking, heterogeneity, and complexity. Research areas in this field are discovering, collecting and managing data of known quality; information analysis for scientific discovery and decision; data integrity and resilience in a secure cyber infrastructure; and data-intensive methods, models, and architectures.

The next talk in this session was on Scientific Data Visualization, given by Sean Ahern from the Center for Remote Data Analysis and Visualization at the University of Tennessee. He started by explaining Visualization and its basic components and processes, and their importance for the massive data, like meshes, variables, scalar fields, vector fields, resolution, increasing dataset size, data flow networks, data decomposition and parallelization, streamline computation. He went on to explain a few end-user visualization and analysis tools for extremely large data, like VisIt, ParaView, EnSight, and Rmpi. He supported all this by really impressive graphics examples.

Uwe Woessner from HLRS in Stuttgart talked about Interactive Parallel Simulation and Visualization, with a special focus on industrial applications. He explained the different steps of a product development process, from planning to production. He described collaborative tools like COVISE and COVER, and interactive visualization for real application production codes, with many real-life examples of models, wind tunnels, visualization caves, and augmented reality, and enriched his presentation by many impressive visualizations of real-life phenomena.

## Final Session: General Discussion and Further Collaboration

The Summer School ended with a final session on students' general and discipline-specific questions and comments about the event, answers and comments by the moderators and speakers, suggestions for a similar event in the future, and how to better connect students with students and students with speakers. Finally, ideas were collected on how to stimulate cross-continental exchange and collaboration of students from the US and from Europe. The session was moderated by Wolfgang Gentsch, Giovanni Erbacci, Barry



*All students agreed that the size of the class – 60 students – was ideal: small enough to still interact with everybody, large enough to be able to build groups of similar interest.*

Schneider, and Mike Norman. Key ideas and comments from this session are summarized on page 5.

## DEISA PRACE Symposium 2011



Pictures: City of Helsinki/ Tourist & Convention Bureau's Material Bank/ Comma Image Oy and Paul Williams



### Helsinki, Finland

### 6–7 April, 2011

at the National Museum of Finland

follow [www.deisa.eu](http://www.deisa.eu) and [www.prace-project.eu](http://www.prace-project.eu) for more details

