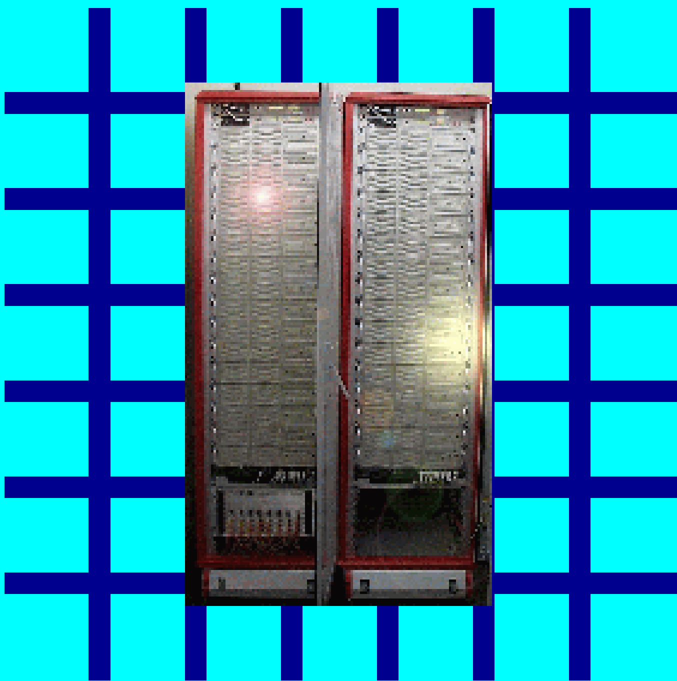
Institute *of* Physics

Newsletter

of

The Computational Physics Group



Grid Computing

March 2002

MEMBERS OF THE COMMITTEE

Peter Borcherds (chairman)	p.h.borcherds@birmingham.ac.uk
Allan Boardman (treasurer, secretary)	a.d.boardman@salford.ac.uk
Richard Ansorge	rea1@phy.cam.ac.uk
Roger Barrett	R.Barrett@surrey.ac.uk
Mike Fearn	mfearn@qinetiq.com
Andrew Horsfield (newsletter)	horsfield@fecit.co.uk
Geraint Lewis	dg.lewis@physics.org
Jim McNiff	jmn@chrdesign.com
Ian Morrison (web page)	i.morrison@salford.ac.uk
Matt Probert	mijp1@york.ac.uk
David Stephenson	david.stephenson@sun.com

Our web page can be found here: http://www.iop.org/IOP/Groups/CP/

Comments about the newsletter should be sent to Andrew Horsfield

Contents

Cover picture is of the Beowulf cluster at the Department of Physics, University of York. It is reproduced with the permission of Dr Probert

Chairman's Remarks - Keeping It Simple

Peter Borcherds (chairman)

The first personal computers (PET, Apple) appeared in the late 1970s. Their total memory was a very few kilobytes, which meant that when you wrote a program you had to think hard about what you were doing. The doubling or halving of everything inside a computer every 18 months or so (Moore's Law) means that we now have so much memory available that sloppy programming is only too easy.

When there are quantitative changes of thousands or millions of times, they may be accompanied by qualitative changes: one may be able to do different things from what was possible previously. Two examples:

1. Photographs can now be processed digitally;
2. Operational amplifiers which used to cost about a month's salary now cost about the same as resistors and are used in electronic circuits as (active) components.

Regrettably the extra speed and memory available is sometimes used only to make things unnecessarily complicated. Examples are web pages with unnecessary frills, which you cannot read unless you have the most recent version of some web browser, and emails in unnecessarily complicated formats.

There is an excellent campaign: The Campaign for a Non-Browser Specific WWW (http://www.anybrowser.org/) which advocates text only web pages where appropriate. As one who uses a computer from home on an ordinary telephone line, I would strongly support that. Too many web pages come with irrelevant graphics.

I would also support following the example of some colleagues, whose email signatures include `Please do not expect me to read attachments which are not in plain text.` Computational physicists should be in a good position to support these campaigns to keep things simple, and avoid unnecessary complications.

An example of the kind of lightweight, but powerful, application I am advocating is my wordprocessor (on which I am typing this). It can save its output in the following formats: its native format (EasiDoc), plain text, Word, HTML, PostScript, TeX and RTF. It has a powerful mathematical equation editor and can read Word files. The wordprocessor is TechWriter Professional (Icon Technology) which runs on RISCOS (e.g. Acorn Archimedes) platforms. For those using Windows platforms, there is available an emulator, VirtualAcorn, which emulates an Acorn A5000 (see http://www.virtualacorn.co.uk).

Grid Computing - What can Scientists Expect?

Dietmar Erwin (Zentralinstitut für Angewandte Mathematik)

In 1999 Ian Foster and Carl Kesselman published a book titled "The Grid: Blueprint for a New Computing Infrastructure" [1]. Based on achievements in the areas of Distributed Computing, Metacomputing, and Internet-based Computing, to name but a few, they created a compelling and powerful vision of a new computing paradigm. The metaphor of the power grid, which allows customers plug an appliance into any electric power outlet, without having to know where and how the current is produced, made Grid Computing an intuitive analogy, which was quickly accepted. As a result, many things happened rapidly:

- Researchers and developers pooled resources to make the vision become reality
- Potential users started to explore to opportunities to try the emerging Grid solutions
- Funding agencies discovered Grid technology as a key area
- Hype accompanied the serious R&D efforts.

Let us examine each aspect in greater detail. The vision of the Grid promises to link distributed resources like computers, data, applications, network connections, input and output devices, and also scientists in an seamless, secure, and easy to use way. The input/output devices include accelerators, tomographs, telescopes, satellites, caves, or data archives. Applications have been developed in the past, that use one or more distributed resources. To integrate two custom-built applications proved, however, next to impossible. Grid-enabled solutions will use standard interfaces and common protocols to allow interoperability and extensibility.

The Grid research community joins forces to create the necessary standards, protocols, services, and interfaces. It draws on the extensive experience of many research groups world-wide. Prototype implementations are being tested to accelerate the creation of solutions that meet the demands of the users. The Global Grid Forum (http://www.globalgridforum.org) has become the recognized platform for Grid research, development and use.

Early adopters of Grid technology exploit the already existing solutions and influence the directions of the future developments through the Global Grid Forum. Two European examples are: The Cactus Worm (http://www.cactuscode.org) that executes a long running simulation on participating systems. It uses idle cycles and gracefully moves on when the resources are needed locally. The implementation is based on the Globus (http://www.globus.org) toolkit. Users of German HPC centres use a graphical interface to create and run complex jobs on computers from different vendors at different sites without having to learn the intricacies of the various systems and historically grown site conventions. The UNICORE software (http://www.unicore.de) that has been developed to create this German HPC-Grid is also the basis for the European EUROGRID project (http://www.eurogrid.org).

The vision of the Grid has been formulated, but additional research, development and use in practice is needed to reach the ambitious goal. This has been recognized by funding agencies around the globe. Consequently, this provides an excellent opportunity for Grid researchers and for scientists who wish to exploit the Grid to obtain funding and support during the coming years. The European Commission alone will support Grid projects in the 6th framework with 300 million euros. As a side effect, projects without a strong Grid component might have a lesser chance to be considered.

The hype that accompanies each innovative idea can probably not be avoided. Inflated and unrealistic promises will discredit the serious work of many. The opportunistic re-labelling of existing products by vendors as Grid products is equally counterproductive. However, I am confident that Grid overselling will not take place since the potential market, which includes scientists working on expensive experiments and requiring large scale computers and petabytes of data, is much smaller than the one for e-commerce.

Therefore, the question about the killer application for the Grid is irrelevant. We should rather ask: 'Who will benefit from the Grid?' The answer is quite clear: Any researcher, engineer, and scientist who's work can be characterised by one or more of the following aspect:

- The development of applications and testing with small problems can be done on a local system.
- To solve realistic problems requires a large HPC system on the Grid.
- Different architectures are needed to process the components of a coupled application with optimal performance.
- Access to petabytes of data like those generated by CERN's LHC experiments.
- Global data mining.
- Collaborative research involving simulation and visualisation.

And as always, users will imagine novel ways to use the Grid.

[1] Foster, I. and Kesselman, C. (eds.). The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann (1999)

Dietmar Erwin is manager of the Operating Systems Department at ZAM (Central Institute of Applied Mathematice) at Forschungszentrum Jülich, GmbH. He is also project coordinator of the German UNICORE Plus project and of the European project GRIP (Grid Interoperability Project).

The GriPhyN Project

Paul Avery (University of Florida)

An important example of the application of computer Grids to physics is the Grid Physics Network (GriPhyN) project.

The GriPhyN project brings together a team of computer scientists and experimental physicists to provide the information technology (IT) advances required to enable Petabyte-scale data intensive science in the 21st century. Motivating the project is the unprecedented requirement for geographically dispersed extraction of complex scientific information from very large collections of measured data. To meet this requirement, which arises initially from the four physics experiments involved in this project, the GriPhyN team is pursuing IT advances centred on the creation of Petascale Virtual Data Grids (PVDG).

The project team is composed of seven computer science research groups and members of four physics experiments. This integrated research effort provides the coordination and tight feedback from prototypes and tests that will enable both communities to meet their goals. The four physics experiments are: the CMS and ATLAS experiments at the Large Hadron Collider which are searching for the origins of mass and probing matter at the smallest length scales; LIGO (Laser Interferometer Gravitational-wave Observatory) which detects the gravitational waves of pulsars, supernovae and in-spiraling binary stars; and SDSS (Sloan Digital Sky Survey) which is carrying out an automated sky survey enabling systematic studies of stars, galaxies, nebulae, and large-scale structure.

The data analysis for these experiments presents enormous IT challenges. Communities of thousands of scientists, distributed globally and served by networks of varying bandwidths, need to extract small signals from enormous backgrounds via computationally demanding analyses of datasets that will grow from the 100 Terabyte to the 100 Petabyte scale over the next decade. The computing and storage resources required is distributed across national, regional, and university computing centers, as well as individual desktops. The scale of this task far exceeds our current ability to manage and process data in a distributed environment, requiring fundamental advances in many areas of computer science.

To meet these challenges, GriPhyN is pursuing a program of computer science research focused on realizing the concept of *Virtual Data*. Virtual Data encompasses the definition and delivery to a large community of a virtual space of data products derived from experimental data. In this virtual data space, requests can be satisfied via direct access and/or computation, with local and global resource management, policy, and security constraints determining the strategy used. Overcoming this challenge and realizing the Virtual Data concept requires advances in three major areas:

1. *Virtual data technologies.* Advances are required in information models and in new methods of cataloguing, characterizing, validating, and archiving software components to implement virtual data manipulations

2. *Policy-driven request planning and scheduling of networked data and computational resources.* We require mechanisms for representing and enforcing both local and global policy constraints and new policy-aware resource discovery techniques.

3. *Management of transactions and task-execution across national-scale and worldwide virtual organizations.* New mechanisms are needed to meet user requirements for performance, reliability, and cost.

GriPhyN is primarily focused on achieving the fundamental IT advances required to create PVDGs, but is also working on creating PVDG software systems for community use, and applying PVDG technologies to enable distributed, collaborative analysis of data. In order to apply these advances to the experimental data analysis problems, GriPhyN is packaging them in a Virtual Data Toolkit, and using this toolkit to prototype the PVDGs and support the CMS, ATLAS, LIGO, and SDSS analysis tasks.

The challenges addressed by this program are not unique to physics, but are also encountered in biology (e.g., the human genome project), medicine (e.g., the human brain project), environment (e.g., the Earth Observing System), and many other areas. To reach these disciplines and build Grid infrastructures that can be used in a variety of settings, we are partnering with other Grid efforts such as the Particle Physics Data Grid, TeraGrid and the European DataGrid.

More infomation about this project can be found at http://www.griphyn.org

Upcoming Computational Physics Group Events

The 2^{nd} Annual Computational Physics Thesis Prize

The Committee of the Institute of Physics Computational Group has endowed two annual prizes. £500 will be awarded to the author of the PhD thesis that contributes most strongly to the advancement of computational physics. The Committee will select the recipients and its remit will be very broad, in order to capture a broad spectrum of modelling activity.

- The deadline for applications is December 31^{st}, 2002.
- The submission format is a 4 page (A4) abstract.
- The submission address is:
  Professor A D Boardman
  Hon. Sec, IoP Computational Physics Group
  Joule Physics Laboratory
  School of Sciences
  University of Salford,
  Salford, M5 4WT

Applicants must have carried out their thesis work at a University in the United Kingdom or the Republic of Ireland.

Modelling with Computers in Physics Teaching

This is a one day conference jointly organised by the Education and Computational Physics Groups of the Institute of Physics, and the School of Education at the University of Birmingham. It will be held on Wednesday 6 March, 2002, at the School of Education, University of Birmingham. The cost to attend is £10 for the buffet lunch. Numbers will be limited to 50. The meeting will start at 11:00am and finish at 4:00pm. Coffee and tea will be provided at either end of the meeting.

To book, please write to Beverly Burke, School of Education, University of Birmingham, B15 2TT enclosing a cheque for £10, made payable to the University of Birmingham.

There will be four talks and a slot at lunchtime for show and tell for which a number of laptops running Windows 98 will be available. The talks will be on:

- Modelling post 16: current practices and future prospects
- Modelling pre-16: Tools, strategies and prospects
- The place of Computer Algebra
- The life of a Computational Physicist

A plenary discussion session will give participants the chance to shape possible futures.

The Computer Grid and Computer Visualization Meeting

This two part meeting on the computational Grid, and on computer visualization will be held on Friday, 6 September, 2002, at the Institute of Physics, 76 Portland Place, London W1B 1NT.

Web page: http://www.???

Other Upcoming Meetings

Linear Scaling Electronic Structure Methods

The workshop on Linear Scaling Electronic Structure Methods will be held from 1 to 4 April, 2002 at the Institute For Pure and Applied Mathematics, University of California, Los Angeles. The purpose of this workshop is to bring together chemists, physicists, and mathematicians interested in developing new linear scaling algorithms that will have global applicability.

Web page: http://www.ipam.ucla.edu/programs/es2002

Regional Workshop on Computational Condensed Matter Physics

The Regional Workshop on Computational Condensed Matter Physics will be held from 15 to 25 April, 2002 at the Isfahan University of Technology, Isfahan, Iran. This workshop covers density functional theory (DFT), its new developments and applications.

Web page: http://www.iut.ac.ir/ccmp/

EMRS 2002 Spring Meeting

Symposium A - Atomic Scale Materials Design of the EMRS 2002 Spring Meeting will be held from 18 to 21 June, 2002 at Strasbourg, France.

Web page: http://www-emrs.c-strasbourg.fr

MATH/CHEM/COMP 2002

The 17th Dubrovnik International Course and Conference on the Interfaces among Mathematics, Chemistry and Computer Sciences will be held from 24 to 29 June, 2002 at the Inter-University Center, Dubrovnik, Croatia.

Web page: http://mcc.irb.hr

Exploring Modern Computational Chemistry

The meeting on Exploring Modern Computational Chemistry will be held from 31 July to 2 August, 2002 at the University of Nottingham.This international conference will bring together researchers in all aspects of modern computational chemistry. The programme will consist of invited lectures and a poster session, which will be an integral part of the programme. Scientific Topics will include:

- Quantum Chemistry
- Density Functional Theory
- Structure and Spectroscopy
- Intermolecular Potentials
- Statistical Thermodynamics
- Biological Applications

Web page: http://www.nottingham.ac.uk/chemistry/emc2

Conference on Computational Physics 2002

The 2002 Conference on Computational Physics will be held from 25 to 28 August, 2002 at the Hyatt Regency Islandia, San Diego.

Web page: http://www.aps.org/DCOMP

Modelling of Complex Systems

The 7^{th} Granada Seminar will be held from 2 to 7 September, 2002 at Granada, Spain.

Web page: http://ergodic.ugr.es/cp/

2002 International Conference on Simulation of Semiconductor Processes and Devices

The 2002 International Conference on Simulation of Semiconductor Processes and Devices will be held from 4 to 6 September, 2002 at the International Conference Center Kobe, Kobe, Japan. The Conference provides an opportunity for the presentation of latest advances in modeling and simulation of semiconductor devices, processes and circuits. Special emphasis is placed on practical applications.

Web page: http://www6.eie.eng.osaka-u.ac.jp/sispad/index.html

Computing Techniques in Physics

The 13^{th} summer school on Computing Techniques in Physics will be held from 16 to 21 September, 2002 at Třešt' in the Czech Republic. The principal theme of the school will be parallelization of algorithms in physics.

Web page: http://www.fzu.cz/activities/schools/epsschool

XII Computational Materials Science Workshop

The XII Computational Materials Science Workshop will be held from 23 to 29 September, 2002 at Villasimius, Sardinia, Italy.

Web page: http://www.dsf.unica.it/CMS2002