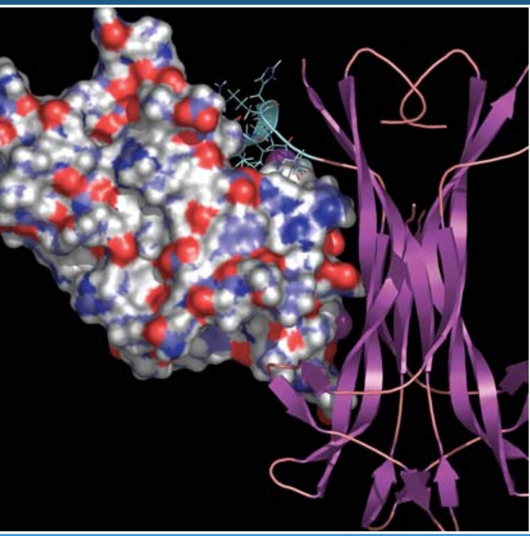
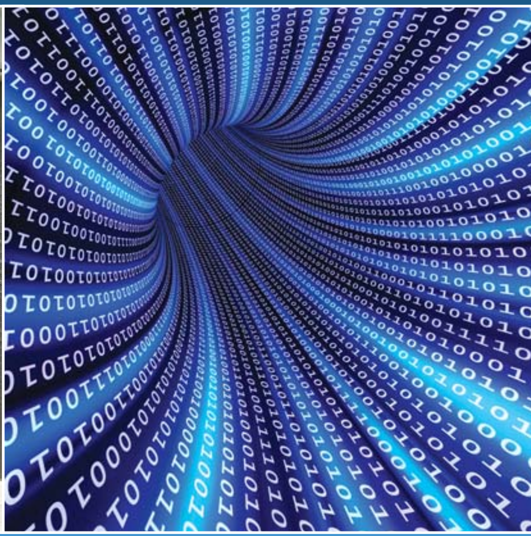
HPC-SIG REPORT 2010

UK High Performance Computing
Special Interest Group





Design, print and production
cw-design.co.uk

For queries about this report
Caroline Gardiner
Academic Research Facilitator
University of Bristol
Email Caroline.Gardiner@bristol.ac.uk
Tel +44 (0) 117 331 4375

If you need part or all of this publication in an alternative format, please contact Caroline Gardiner.

Cover images
NGF and its TrkA receptor, courtesy of Deborah Shoemark.
BlueCrystal, University of Bristol, courtesy of Timo Kunkel.

WELCOME

Welcome to this inaugural issue of the report of the UK High Performance Computing Special Interest Group[1] (HPC-SIG). The HPC-SIG was formed in 2005 in response to the significant investment for university-level High Performance Computing (HPC) supplied primarily by the HEFCE SRIF3 funding round.

Members are drawn primarily from Computing Services in the Higher Education sector with representation from related organisations such as the National Grid Service and funding bodies like EPSRC, STFC and NERC. Several non-academic institutions, for example GCHQ, are also affiliated to the HPC-SIG. The main Terms of Reference of the HPC-SIG are:

- To act as a lobby stressing the value of mid-range campus HPC provision.
- To ensure that HPC provision and research methodologies are closely aligned, promoting the academic agenda in addition to system management and support.
- To collect, disseminate and promote best practice in HPC provision, management and support.
- To coordinate and publicise training opportunities in the areas of HPC system support and usage within the UK.
- To act as a link between national HPC provision and local university level provision.
- To act as an outreach vehicle promoting the use of HPC across all academic sectors.
- To facilitate communication between academic and industrial/commercial HPC providers/users.
- To secure the role of HPC as a vital research service across all academic disciplines.
- To demonstrate the value of HPC facilities in Higher Education and to ensure that these facilities can be delivered with best possible value for money.

This report summarises the results from what we believe is the first comprehensive survey undertaken to describe the activities, measure the range and evaluate the impact of HPC facilities in UK universities. The results clearly illustrate why campus HPC, so often in the past ignored or misunderstood, is crucial to many areas of research undertaken in UK universities – providing the underpinning infrastructure to leading edge research in not only the traditional science and engineering disciplines, but in the increasingly important fields of health, energy, life sciences and also in the emerging computational research areas in social sciences and humanities.

This report will also ably demonstrate the wealth of talent and expertise that staff within the university HPC sector possess in designing, procuring, installing and running efficient, productive and cost-effective advanced high performance computing and storage facilities.

We expect that the findings contained in this report will be of value to HPC service providers in HEIs, senior management in UK HEIs such as CTOs, CIOs and Pro-VCs for Research, the major research funding bodies and those who have a general interest in the activities of a major contributor to the UK HPC ecosystem.

1 *www.hpc-sig.org*

1 INTRODUCTION

Over the last ten years, computational science has emerged as a major driver for innovation across the research spectrum. In fields as diverse as pharmaceuticals and nanotechnology, High Performance Computing (HPC) has dramatically accelerated the delivery of scientific insights; simulation and modelling supplanting traditional experimental-based methods through their cost efficiency.

At the same time, the 'data deluge' is demanding that new models for data and information management are developed, supported by novel architectures for storing and handling digital material.

21st century research is characterised by increasing levels of collaboration, across disciplines and between continents. HPC and large-scale data storage are critical enablers of this virtual laboratory and library, supported by high-speed global networks. This 'internationalisation' of HPC has seen the rapid rise of new entrants into the HPC field, with India, Brazil and China – among others – investing heavily in national programmes. Indeed, the number one spot in the Supercomputing 'Top 500' has just been taken by China, ending several years of US occupation of this prestigious position.

In these exciting times, the role of university-based HPC is more critical than ever in providing the foundation for a healthy HPC 'ecosystem' for the UK, where computational scientists and HPC-service providers work together in a highly collaborative community. Through their locality to today's research base, and the students who will become our next generation of computational scientists, universities are uniquely positioned to deliver excellent return on investment in HPC as a platform for future economic growth.

Professor David Price
UCL Vice-Provost (Research)

Contents

2 WHY HPC?

'In this new age, it's clear that whoever out-computes will out-compete.' [2]

User profile: Professor Jonathan Tennyson FRS

Department of Physics and Astronomy, UCL
www.ucl.ac.uk/phys/amopp/people/jonathan_tennyson

I have used HPC facilities at UCL for a variety of projects. The two largest projects are to compute an ultrahigh accuracy dipole surface for the water molecule and to compute a high temperature line list for ammonia.

Water vapour is the main absorber of sunlight in the earth's atmosphere and the main greenhouse gas. However it is very difficult to measure the probability of water vapour absorbing light of a particular wavelength to better accuracy than a few percent. We have developed first principles quantum mechanical methods which can calculate this probability for the majority of water absorption features to 1% or better. Doing this required performing some 12,000 highly sophisticated electronic structure calculations to give the water dipole moment as a function of geometry. These calculations would take some 20 years serial processing time, but were accomplished in three months using the capabilities of the Legion cluster.

Ammonia is a trace species in the Earth's atmosphere, but is a significant component of gas giants such as Jupiter. Theory suggests that absorption by hot ammonia will provide the spectral signature of yet to be detected Y-dwarf stars; the coolest brown dwarfs predicted and similar in character to hot gas giants and planets now observed orbiting other stars. The spectrum of hot ammonia has been simulated theoretically by calculating over 1 billion transition frequencies and probabilities. These are huge calculations which have required use of both the MPI Legion cluster and the OpenMP Unity cluster at UCL and would have been impossible without them.

2 *Matthew Faraci, 'American Business's Secret Competitive Weapon: HPC', www.forbes.com*
3 *www.compete.org/hpc*

Over the past decade there has been a revolution in High Performance Computing spearheaded by a movement away from using expensive traditional proprietary supercomputers to systems based on relatively inexpensive commodity off-the-shelf systems. A direct result of this transition is that the UK academic community has increasingly engaged in research using commodity HPC systems.

Furthermore, the advent of SRIF3 in 2004 allowed major investment into the university HPC sector resulting in a dramatic increase in the deployment and use of commodity clusters within HEIs. Today this commodity university HPC sector is by far the dominant provider of HPC resources to academics in the UK.

HPC is now acknowledged to play a key role in academia, commerce and industry. In the US the *Council on Competitiveness*[3] is a group of corporate CEOs, university presidents and labour leaders working to ensure US prosperity and enhanced competitiveness in the global economy through the creation of high-value economic activity. This Council has undertaken surveys on HPC usage and uptake in US industry and has produced several detailed and influential papers confirming that HPC is becoming indispensable for ensuring future US productivity, innovation and competitiveness. They advocate that *HPC is a proven transforming and game-changing technology*.

It is perhaps worth comparing and contrasting HPC and cloud computing. There are indeed many areas where the two overlap: both typically involve using a remote resource for example. However for this report the term refers to HPC as the use of a co-located resource optimised for parallel and/or mixed workload computing (e.g. simulation of large scale problems) rather than a distributed resource which may be used in a 'Software as a Service' regime (such as web hosting). HPC is the more mature of the two approaches, but its heritage is also based around a more specialised user community.

The UK's competitiveness in simulation, modelling and analysis of massive data will be critically dependent on continuing to have access to local and national innovative and agile HPC systems. Work in several industrial and academic disciplines is no longer viable without access to HPC. It has become an indispensable, cost effective and proven tool for many of the UK's research staff and increasingly is held to be on an equal footing with laboratory and other experimental work. As a consequence we need to be able to supply appropriate *local* HPC capability quickly to researchers who have novel or speculative ideas. National or international HPC facilities in the main are for established research programmes at the top end of user requirements and do not typically provide researchers with the required development platforms. As such these top-level services need to be 'fed' researchers and projects from the bottom up.

If the UK has the ambition that its research universities be counted amongst the top institutions worldwide and that they will be able to collaborate with international partners in the US, Europe, the 'BRICs' and other newly emerging economies, it must be able to bring world-class resources and infrastructure to the table – it must engage in this transformational technology or, for some research areas, risk being left behind to stagnate by our competitors. Well-founded world-class local university HPC facilities complemented with appropriate national facilities are the key to realising this ambition.

This unique review of the UK university HPC ecosystem has been conducted by the HPC Special Interest Group.

The HPC-SIG was formed in 2005 in response to the significant funding for university-level HPC funded primarily by SRIF3. Members are drawn, in the main, from Computing Services in the Higher Education sector with representation from organisations such as the National Grid Service and funding bodies. The HPC-SIG also has strong links with the HPC Forum, a similar organisation that caters for non-academic institutions such as AWE, GCHQ and the Met Office. Appendix B lists the 35 HEIs, plus affiliates, that form the membership. The review provides a snapshot of HPC activities throughout the HE institutions that responded to the survey request. As this is the first year of what we hope will be a regular review, some institutions had difficulty in collating all the required information within the requested timescale. We hope in future that more institutions will be able to provide input. The HPC-SIG conducted this survey in recognition of the increasing importance of having a better understanding of HPC usage and requirements and being able to show the impact of campus HPC facilities on UK research and their importance to both the UK HPC and e-Infrastructure ecosystems.

The benefit of local university HPC to the UK knowledge economy cannot be overstated - it provides the foundation of the HPC ecosystem within the UK and encourages usage within new disciplines and industry which contribute to the wider UK knowledge economy.

One of the interesting developments in 2010 has been the emergence of High Performance Computing (HPC) Wales, an ambitious 10 year development programme to build an HPC Institute and capability across Wales. The programme has been funded from a variety of sources to deliver significant benefits across the Principality and beyond through an enabling technology and the building of a skills base to support research and development projects with a range of both academic and private sector partners.

There are two distinct features of 'HPC Wales' that are worth emphasising. While funded at the level of a national HPC Facility, the project has focused not on a traditional, monolithic single supercomputing facility, but on a distributed Hub-and-Spoke model. The universities of Cardiff and Swansea will provide Hubs linking to Tier 1 Spokes in Aberystwyth, Bangor and Glamorgan Higher Education Institutes, who all form part of the Saint David's Day Alliance Group of research focused universities. There will be further Spokes in the University of Wales Alliance Group of Universities and in the Welsh Technium facilities, providing a pan Wales hub and spoke network. This network will provide a new enhanced over-arching regional structure.

The second distinctive feature is the focus on high-impact, user-valued, research outputs that have significant economic impacts. HPC capability will not in itself deliver sustainable value to the economy and therefore the establishment and development of an HPC Institute is a key objective, designed to deliver advanced research. Focused in the digital, low-carbon, health, bio-science, engineering and advanced manufacturing sectors, the purpose of the Institute will be to develop educational provision and training at all levels, building capacity and skills base to operate and take forward the optimal usage and development of HPC and supercomputing technology and solutions in the public and private sectors. HPC Wales will offer a range of opportunities, including skills-development activities, training packages and consultancy services and short-term internships to support two-way knowledge transfer and create stronger links between the initiative and the industrial community.

Staff profile:
Research Facilitator
Caroline Gardiner

Academic Research Facilitator
University of Bristol
www.acrc.bris.ac.uk

Caroline has been Academic Research Facilitator in the Advanced Computing Research Centre (ACRC) since 2006 when the Bristol HPC facility was established. The Bristol HPC system, BlueCrystal, is currently composed of two clusters comprising over 3700 cores and running IBM's General Parallel File System (GPFS). A resilient petascale research data storage facility, BluePeta, has just been installed. Caroline previously worked at the University of Bath, organising IT courses and marketing for the Lifelong Learning programme.

Caroline's role is to promote both HPC and research data storage across the university and beyond. This involves organising symposia and seminars to present research being undertaken on BlueCrystal and giving presentations to academic staff and to postgraduates to explain how BlueCrystal and BluePeta could benefit their research. She is the first point of contact for users and potential users and thus has a good understanding of the issues which are of concern to them. Through her membership of the University HPC Executive, which oversees the management of the HPC and research data storage facilities, she can therefore contribute the user perspective to assist the decision making process.

Caroline also organises a programme of half day workshops for staff and postgraduates on HPC-related topics and maintains the ACRC website. An important part of her role is the gathering of metrics which demonstrate the impact BlueCrystal is having on research at Bristol, such as the number of publications where the research underpinning the publication was undertaken on BlueCrystal and key research achievements such as the award of prizes, successful industrial collaborations and use of national and international HPC facilities.

3 CAMPUS HPC: THE PRESENT AND THE FUTURE

'In the face of serious global competition and a sobering economic climate, U.S. leadership in HPC - in hardware, software, and expertise - stands out as a true national strategic asset.' [4]

The UK has several truly world-class supercomputers in academic institutions and the significant investment by many UK universities through both SRIF3 and other funding streams is already starting to pay dividends in the form of new and novel research outcomes and the increasing diversification and growth of the computational science community.

The profile of the UK as a centre of HPC activity has been raised worldwide. The university academic HPC community is vibrant and productive and many collaborations have been formed, for example the HPC Consortia that have arisen from the Collaborative Computational Projects[5]. However, such is the progress of technology that any HPC facility needs an environment of continual investment in order to provide cost-effective, performant and competitive resources for their HPC research base. After four years a computer is significantly slower than a new system which means the user of an older system is taking years to achieve what a competitor with a new system can manage in months or even weeks.

There is an independent list which ranks the top 500 computers in the world[6] according to their performance on a popular scientific benchmark and several UK university systems feature in it. The rate of progress in super-computing and computational science can be gauged by the trend lines in Figure 2[7]. The figure demonstrates that historically there is an approximate 10-fold increase in performance every 4 years. The first *Petaflop* (1Petaflop = 1000 trillion calculations per second) system was built in 2008 at Los Alamos National Laboratory. The first system to achieve 100Petaflops is expected to be built in 2016 and the first *Exaflop* system (a quintillion or 1,000,000 trillion calculations per second) is expected circa 2019-2020. From the UK perspective, for a current Top500 university to remain competitive and remain within striking distance of the Top500 worldwide, it will need a Petaflop scale system by around 2016-2018. To maintain competitiveness internationally, a typical major research-led UK university should expect in future to be able to provide aggregate HPC performance of the order superimposed in red on Figure 2.

A similar pro-rata increase in performance will be required for all universities with ambitions to be involved with HPC based research. Without such systems it is unlikely that UK universities will be able to remain internationally competitive in strategic areas such as climate change, energy, health and engineering as:

- The computers that exist or are in planning at national HPC facilities such as HECToR[8] and its possible successor ARCHER, or systems at the proposed Hartree Centre[9] are or will be in the main focused as production systems. Generally it is accepted that it is more cost effective to develop software on local resources prior to scaling up onto larger national systems. In addition, with the current network bandwidth transporting vast volumes of data to and from remote compute facilities can be an issue.

- The performance of petascale and, in future, exascale systems will only be achieved by parallel algorithms that can scale efficiently to 100,000s or even millions of processors. It is unlikely that such software would be fully developed on desktop systems and tests to ensure scalability and verify accuracy would require some form of large parallel machine. Furthermore without competitive facilities, the UK will be unable to develop a community with the highly technical skill-set required to undertake this work and will lose influence over the direction of international research which uses HPC.

- Virtualisation, cloud computing and Green IT will likely deprecate usage of the traditional desktop computer, a major tool of many researchers, as general IT moves towards mobile and thin client models.

The UK must maintain its HPC competitiveness in industry as well as academia. Emerging economies such as China are investing heavily in HPC: Figure 1 compares the deliverable computation performance of UK and Chinese systems from the Top500 list over the last two years. Furthermore in October 2010, China took the top spot in the Top500 list, the first time in six years that the list has not been led by a US system. Such investment needs to be underpinned by skilled users and universities must play the central role in developing those skills.

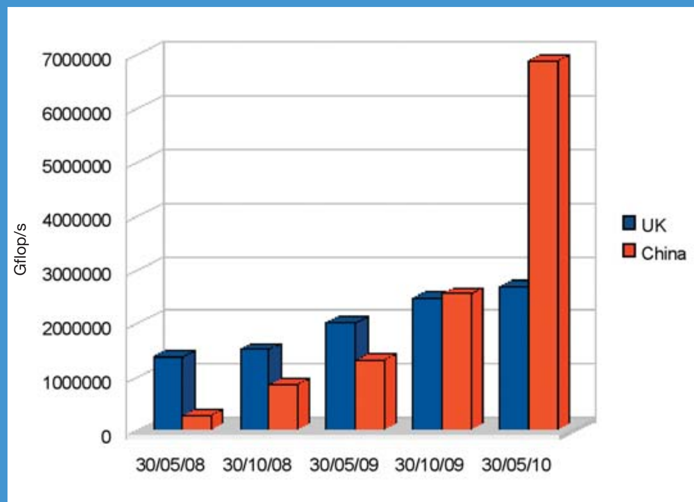


Figure 1: The emergence of China as a leading HPC user is shown by a comparison of summed peak computing power on the Top500 list over the last two years (generated for June 2010 results)[10]

4 LARGE SCALE STORAGE

'HPC is a proven innovation accelerator, shrinking time to insight and time to discovery.' [11]

Many HPC facilities not only provide processing cycles for research, but are also charged with providing large scale storage for research data. For example, the University of Bristol has just installed a 1PByte resilient, scalable, enterprise-grade facility to store its institutional research data assets.

This facility will be funded by the University of Bristol at an appropriate and sustainable scale to keep pace with researchers' data storage demands. Similar projects are in hand at UCL and the universities of Southampton, Oxford and Edinburgh.

In recent years there have been a significant number of computer technology developments. Several generations of processor design have come and gone, raising the compute power per CPU at least *six fold* – from single core in 2004 to six cores or more by 2010 - and processors that will greatly increase this performance are in the pipeline. A similar story can be told for network and storage technologies. As a consequence, there has been a vast increase in the amount of data that can be, and is being, created and manipulated by researchers. To put this into context, it is estimated that *all* the printed material in the world amounts to 200PBytes[12] (or 200,000,000GBytes). One experiment, the Large Hadron Collider at CERN - with which many UK university Particle Physics groups are involved, will produce an estimated 15PBytes of data each year – equivalent to 7.5% of the world's printed material. Furthermore the investment in campus HPC Facilities has made university researchers more ambitious in their research aspirations.

The awareness of the problems of large scale data creation, manipulation and curation is now more widespread, with funding bodies such as the Research Councils and JISC taking interest. One response to the issue is the creation of the national Digital Curation Centre[13]. Another example is provided by one of the conclusions of the 2006 report 'A Strategic Framework for High End Computing'[14], produced by the High End Computing Strategy Committee that advised the Research Councils. This stated that in order to realise the proposed strategy 'the e-Infrastructure required must be a properly balanced one; that is, it must also include provision of capacity resources, data management, data storage and data mining facilities.'

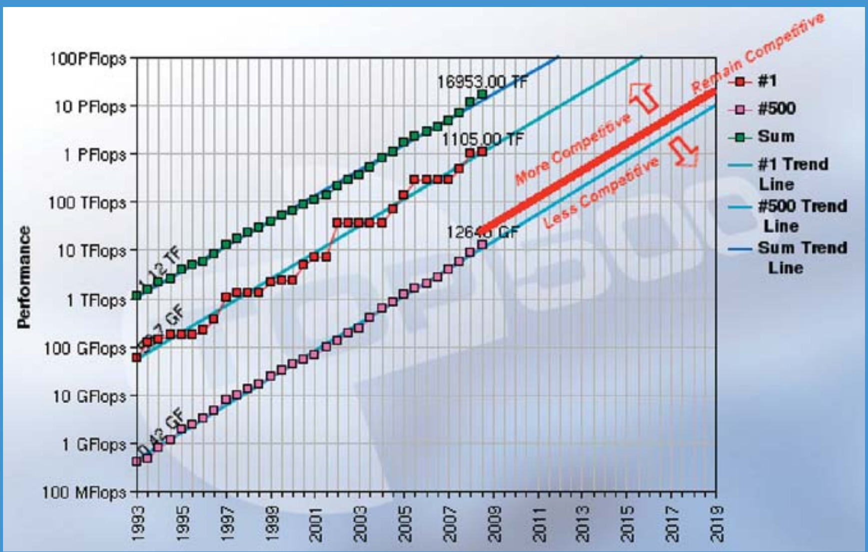


Figure 2: Top500 projected performance

Staff profile: System Administrator Dr Stuart Rankin

Senior System Administrator
University of Cambridge
www.hpc.cam.ac.uk

Stuart has worked for the High Performance Computing Service (HPCS) in Cambridge since 2007. The main HPCS system is currently a 3500 core Dell/Intel cluster with SDR and QDR Infiniband and Lustre storage. For the previous ten years he looked after various large SGI Origin and Altix NUMA systems for Professor Stephen Hawking's COSMOS Consortium, also based at Cambridge.

Stuart is currently Senior System Administrator with responsibility for the design, implementation, operational support and ongoing evolution of the software platform that underpins the Cambridge HPC Service. The role also has lead responsibility for user support and overall service delivery, and is pivotal in the management of external engineers and in the delivery of consultancy services which are important for industrial funding to the service.

Day to day system administration involves monitoring system operation, detecting, diagnosing and resolving system problems, as well as responding to requests for help from users and the creation and management of user accounts. Weekly tasks include managing tape backups, processing statistics and performing software and hardware maintenance as required. Stuart is the point of contact for all internal and external customers, and ensures that all users see a coherent and documented computational environment. He also tests new software and equipment, writes reports and white papers and gives technical presentations both internally and externally.

4 *'High Performance Computing To Enable Next-Generation Manufacturing', White paper, U.S. Council on Competitiveness, 2009, www.compete.org/hpc*
5 *www.ccp.ac.uk*
6 *www.top500.org*
7 *www.top500.org/lists/2008/11/performance_development*
8 *www.hector.ac.uk*
9 *www.stfc.ac.uk/About%20STFC/18572.aspx*
10 *Figure 1 does not include the upgraded Tianhe-1A system which has recently claimed the top spot in the Top500 list.*
11 *'Advance. Benchmarking Industrial Use of HPC for Innovation', U.S. Council on Competitiveness, 2008, www.compete.org*
12 *pcbunn.cithep.caltech.edu/presentations/giod_status_sep97/tsld013.htm*
13 *www.dcc.ac.uk*
14 *www.epsrc.ac.uk/SiteCollectionDocuments/other/2006HECStrategicFramework.pdf*

5 COLLABORATIONS AND INTERNATIONALISATION

'Computational Science, the scientific investigation of physical processes through modelling and simulation on computers, has become generally accepted as the third pillar of science, complementing and extending theory and experimentation.' [15]

**Staff profile:
Developer
Dr Michael Bane**

Senior Research Applications
and Collaboration Officer
University of Manchester
www.rcs.manchester.ac.uk/home

Michael has been working in the Research Applications and Collaborations Team (the RAC Team) within IT Services for Research (formerly RCS) since December 2008. At Manchester, high end computing support is provided by IT Services for Research within the Directorate of IT Services, with academic governance via Manchester Informatics (M*i*).

Michael's role is to help researchers at the University with their high-end computing requirements. Typically, this involves an initial meeting with a PI and postdoc (sometimes a postgrad) to get an overview of the research. This is followed by a few days' intense examination of the code or system in question after which a brief Options Report is produced laying out the expert view of the available options and the relative efforts (and outcomes) of various approaches. Generally, IT Services for Research can do a brief amount of work as part of the core service. For more in-depth work, the team encourages and supports the PI to apply for a grant to cover the costs of having a high-end computing expert assigned to the project, perhaps embedded within the School for three to six months.

Alongside the project work, Michael gets regular help-desk enquiries and provides support including facilitating access to, and applications support on, Grids (national and international), national HPC and local clusters including the Manchester Computational Shared Facility (currently consisting of a 2000 core cluster) Michael co-ordinates the IT Services for Research's research computing training, working side by side with the leader of the team, Robin Pinning, which involves overseeing the running of all courses. Michael also writes and delivers a number of courses. In November 2010, Michael organised the inaugural University GPU Club with 100 researchers attending. In any quieter moments, Michael expands his skills by dabbling in hybrid/GPU programming, evaluating performance evaluation tools and learning asynchronous PGAS languages such as CHAPEL.

Several areas of science that have traditionally been laboratory or experimentally based are now moving to an HPC base. If we couple this movement with the exploration of *Grand Challenges* in research, such as those outlined by the Hartree Centre business case – the virtual physiome, large scale simulations of biological structure, investigating new materials, climate change and CO_2 sequestration – it becomes clear that these Grand Challenge research fields will require access to very large scale computational and data storage resources.

Researchers at UK universities can therefore be expected to play a leading role internationally in emerging cross-disciplinary Grand Challenges, *provided that the UK has an appropriate sustainable HPC ecosystem*.

To forge collaborations with national and international institutions on future Grand Challenges, computational projects will require campus HPC facilities whose performance is either comparable with or within an order of magnitude of that of the prospective partner institutions.

To take just one example, the Hartree Centre model states that around five or six research projects per year will be chosen to be run on the Hartree petascale facility.

However, the software for these projects would be expected to be developed on local facilities which, in order to prove scalability, will have to be of reasonable size. This also holds true for HPC based research collaborations with other similar large scale projects such as the European Framework[16] and PRACE[17] initiatives.

CLUSTERVISION

Darwin, University of Cambridge

6 SUSTAINABILITY

Staffing

Whilst there seems currently to be a barely sufficient pool of expertise in HPC cluster design, build and management within UK universities, there is a noticeable increase in commercial and industrial requirements for staff with this HPC experience. As salaries in the commercial sector are generally higher than in the HE sector, this may lead in future to difficulties in retaining expert staff.

Also, while the UK HE sector has a worldwide reputation for ground-breaking and innovative research, for a variety of reasons, such as there being no specific funding stream to support the development aspect of R&D, it is not seen as being able to build on the research successes. To counter this there is a very high probability that in future UK universities will need access to a pool of professional HPC scientific/technical programmers not only to develop research results into professional marketable products, but also to help academics gain most out of new petascale technologies. These programmers could, for example, provide expertise in developing algorithms that can scale to hundreds of thousands of cores, fault tolerant computing, data curation and management, and performance analysis for very large scale parallel codes. However, recruitment of such personnel will be difficult as such specialists are in short supply in the UK and there is no clear career path to encourage more intake.

HPC Funding and Cost Recovery

Total annual capital expenditure on HPC as at 2010 by the 15 institutions who could supply figures is approximately £7.5M which equates to approximately 0.1% of the total combined annual income and an investment of roughly £1700 per active user. Imperial College and UCL are the strongest investors for campus HPC facilities with an annual capital expenditure of £1.5M each.

All survey respondents operate under Full Economic Costing (fEC) and the three models of cost recovery used are direct charging of users as per a Major Research Facility (MRF), charging under Indirect Rate and a hybrid model (HMRF) which aims to use elements of both MRF and Indirect.

The advantage of the MRF is that there is clear identification of income recovery and it is easy to ensure income can match expenditure. However there are a number of major issues with MRF models. There are issues with providing access for pump-priming or 'blue-sky' activities. There is significant risk to the sustainability of the facility if utilisation is not as forecast. Also the administrative burdens on the centre are much larger. In addition, there is variation between funding sources as to how HPC is treated and users find the additional work at the grant proposal stage unhelpful. In general the benefits of MRF are outweighed by the disadvantages.

The Indirect Rate model on the other hand has the advantages that recovery of HPC costs are spread across the institution and no invoicing of grant holders is required, so there is considerably less administrative overhead. The disadvantages are that there is no direct link between costs and income recovery and that calculation of the indirect rate is based on prior year costs.

UCL have around 200 registered projects with most projects equating to one user. They use a hybrid funding model - Indirect Rate on all FTE's plus Direct Charging to enable large resource users to buy a 'guaranteed' amount of resource. UCL currently operate a hierarchical fairshare resource allocation policy for all projects that are not paying for guaranteed resource. UCL's motivations for choosing this funding model are:

- to enable support of new communities who lack current funding,
- to support skills development in established HPC user communities,
- ease of administration,
- the view that as a world class research university researchers should reasonably expect HPC services to be freely available (i.e. institutional profile and attracting research talent),
- to more easily promote cross/inter-disciplinary research.

Profile of a training programme

The University of Bristol began offering half day workshops in HPC-related topics to staff and students in 2007. 950 have so far attended. The workshops are free and cover a range of topics including how to use BlueCrystal (Bristol's HPC), Linux, Matlab and programming languages including C, C++, Perl, Python and Fortran.

The tutors are all university staff or postgraduates, drawn from a range of academic departments, who see the value of the workshops in enhancing the knowledge of BlueCrystal users and thus increasing the efficient running of the machine. Feedback from the workshops is very positive with over 90% of attendees finding the workshops very useful and over 95% of attendees finding the tutors very knowledgeable. The feedback forms also provide suggestions for future topics of interest ensuring that the programme remains relevant.

15 *'International Review of Research Using HPC in the UK', EPSRC December 2005*
16 *cordis.europa.eu/fp7/home_en.html*
17 *www.prace-project.eu*

7 SURVEY ANALYSIS

'I believe that modeling, simulation, and large scale analysis with HPC is vital to maintaining an edge in American innovation. High performance computing is simply transforming business processes world wide.' [18]

Contributing Institutions

The seventeen HEIs listed in Table 2 (Appendix A) responded to the survey request. The total annual income for these institutions amounts to over £7.5 billion and they represent 50% of the HEIs in the HPC-SIG.

The HPC facilities at these institutions have computational assets of approximately 40,000 cores and serve over 4400 users with end-user and system support being provided by a total of 61 staff. The large majority (14) of facilities are managed by central IT services on behalf of the research community.

HPC Assets

There are some very large systems installed in UK universities. In fact six universities have facilities that can offer in excess of 2000 cores and the two largest facilities at Imperial College and the University of Southampton respectively provide 7000 and 8000 cores. The Southampton system has the highest reported double precision performance of UK university systems with 72TFlop/s peak and 66TFlop/s sustained on the Linpack Top500 benchmark.

Most of the systems in the UK are x86 based, run mainly Linux of some flavour and use a variety of hardware solutions from multiple vendors. Vendor representation is fairly evenly spread among the major suppliers (Bull, Dell, HP, IBM, SGI and Sun/Oracle). We note that the university HPC tier is a major factor in maintaining the diversity in suppliers to the UK HPC market.

Campus HPC facilities are in general keen adopters of novel or disruptive technologies and several sites have embraced accelerator technologies such as Clearspeed and more recently GPGPU systems. The University of Cambridge has a 32 node NVIDIA GPGPU cluster that has peak single precision performance of 120TFlop/s. Many institutions are following suit and installing smaller GPGPU clusters and there is a burgeoning interest amongst many sections of the academic community in both developing new codes for and porting existing codes to these architectures.

Experienced HPC users should also be regarded as an asset both to their home institution and the wider UK HPC effort. For this reason many centres invest a great deal of effort in teaching and training (see section 'Teaching/ training' on page 9). Fourteen HEIs are currently engaged with EPSRC to investigate the possibility of an enhanced national training programme to further improve the skill set of UK users.

Impact

One can employ several metrics to measure the success or otherwise of local HPC facilities. We shall use the following:

Income directly associated with projects employing HPC facilities.
Ten institutions could supply figures for research grant income associated with projects using their HPC facilities. For 2009, the total annual grant income was an impressive £160M.

Maintaining an individual university's position as a leading campus HPC institution.
This can be measured against peers through the independent Top500 list. While such a comparison can be informative, care must be taken not to be influenced by the 'photo opportunity' nature of this list. As such, the HPC-SIG is considering the development of an independent set of KPI's to reflect the efficiency, utility and productivity of local HPC facilities.

Number of and range of registered users of HPC clusters and storage facilities.
There are at least 4400 registered users on the 17 respondents' systems. Most were internal users, with only 5% of registered users being classed as external collaborators. The average number of active users per institution is 145. Unsurprisingly the vast majority of users are drawn from the scientific and engineering communities, though several universities are reporting an increase in interest from researchers in the social sciences and arts. Rather surprisingly given the interest in using advanced computation in areas such as health care, there seems little activity to date from the medical sciences.

Utilisation metrics for HPC clusters and storage facilities.
Most respondents report that over 90% of the workload on their machines was in the main of a parallel nature (including ensemble workloads). Long term percentage utilisation of the HPC assets varied depending on the age of the assets, with a median of 75% and with some sites reporting maximum utilisation over 90%.

Usage of national and international HPC facilities.
This can give a measure of 'scale-up' rates from local facilities to production work on larger systems. As would be expected, several sites make use of the HECToR national system. Interestingly, some also make use of European systems under the DEISA[19] project and two institutions also make use of the Jaguar system at Oak Ridge National Laboratory.

Number and quality of papers published by campus HPC facility users.
Seven institutions could provide figures for publications during the last year. The minimum number of published journal papers for an institution was 10 and the maximum 125 (University of Bristol). The total number across the seven institutions was *391 journal papers plus 66 papers in conference proceedings*. Further papers are also in course of publication. The journal papers included four published in *Nature* and two published in *Science*. A range of representative publications are given in Table 1, on the next page.

18 Richard H. Herman, Chancellor of the University of Illinois at Urbana-Champaign. www.compete.org/news/entry/525/council-on-competitiveness-idc-release-study-on-hpc-and-innovation

19 www.deisa.eu

Teaching/training

All the responding institutions offered training courses of some form with 11 institutions running in-house developed courses. The average number of course attendees during 2009 was 219, the highest 1200 and the lowest 30. Subjects covered include parallel programming techniques and methodologies, application packages, programming languages and software engineering methods. Most in-house training is offered free of charge to the end-user.

In addition a number of institutions made use of the excellent range of courses provided to UK academics and users of HECToR by the Numerical Algorithms Group (NAG)[20]. For undergraduates and taught post-graduates, five institutions either have or are preparing HPC taught courses or modules. At two institutions, HPC staff also co-supervise PhD students.

Green IT

While many HPC facilities rely on standard data centre air-cooling, several have invested significant capital in more energy efficient computer cooling solutions. Imperial College use TROX® CO_2 rear door heat exchangers, the universities of Cardiff and Bristol use the APC InfraStruXure™ in-row chiller solutions and the University of Southampton uses an IBM Idataplex™ rear-door heat exchanger solution.

The University of Sussex has recently relocated its HPC into a new energy efficient data centre, with a design PUE[21] (Power Usage Effectiveness) of 1.23. Cooling is provided by water cooled heat exchangers (USystems ColdLogic™) directly located on the rear of each cabinet.

Cardiff has a project which is monitoring energy usage by their HPC Facility and has reported an initial PUE in the region of 1.3. The median PUE of all facilities is 1.66. Seven respondents are involved in specific carbon footprint reduction schemes. Oxford Supercomputing Centre has also invested a significant effort in energy reduction both by improvement in the efficiency of their data centre and in active power management of their systems. The latter project is funded by JISC and will be released as an open source software package in the near future.

20 www.nag.co.uk
21 www.42u.com/measurement/pue-dcie.htm

Institution	Publication
Bristol	The TASC Consortium (Evans, D.M. Data Analysis Group and Manuscript Preparation group), Genomewide association study of ankylosing spondylitis identifies multiple non-MHC susceptibility loci, *Nature Genetics* 42(2), pp123-127, 2010
Cambridge	Kermode, J.R., Albaret, T., Sherman, D., Bernstein, N., Gumbsch, P., Payne, M.C., Csanyi, G., De Vita, A., Low-speed fracture instabilities in a brittle crystal, *Nature*, Volume 55, Issue 7217, pp1224-1227, 2009
Cambridge	Sebastian, S.E., Harrison, N., Palm, E., Murphy, T.P., Mielke, C.H., Liang, R.X., Bonn, D.A., Hardy, W.N., Lonzarich, G.G., A multi-component Fermi surface in the vortex state of an underdoped high-Tc superconductor, *Nature*, Volume: 54, Issue: 7201, pp200-203, 2008
East Anglia	Watson, A. J., U. Schuster, D. C. E. Bakker, N. R. Bates, A. Corbiere, M. Gonzalez-Davila, T. Friedrich, J. Hauck, C. Heinze, T. Johannessen, A. Kortzinger, N. Metzl, J. Olafsson, A. Olsen, A. Oschlies, X. A. Padin, B. Pfeil, J. M. Santana-Casiano, T. Steinhoff, M. Telszewski, A. F. Rios, D. W. R. Wallace and R. Wanninkhof, Tracking the Variable North Atlantic Sink for Atmospheric CO_2, *Science* 326(5958), pp1391-1393, 2009
East Anglia	C. Goldblatt, A. J. Matthews, M. W. Claire, T. M. Lenton, A. J. Watson and K. J. Zahnle, Nitrogen-enhanced greenhouse warming on early Earth, *Nature Geoscience*, 2(12), pp891-896, 2009
Exeter	Dobbs, C.L., Theis, C., Pringle, J.E. and Bate, M.R., Simulations of the grand design galaxy M51: a case study for analysing tidally induced spiral structures, *Notices of the Royal Astronomical Society*, Volume 403, Issue 2, pp625-645, 2010
Liverpool	Krishnan R Harikumar et al, Cooperative Molecular Dynamics in Surface-Reaction, *Nature Chemistry* 1, pp716-721, 2009
Liverpool	J. Rabone, Y.-F. Yue, S. Y. Chong, K. C. Stylianou, J. Bacsa, D. Bradshaw, G. R. Darling, N. G. Berry, Y. Z. Khimyak, A. Y. Ganin, P. Wiper, J. B. Claridge, M. J. Rosseinsky, An Adaptable Peptide-Based Porous Material, *Science* 329, pp1053, 2010
Oxford	Kiminori Maeda, Kevin B. Henbest, Filippo Cintolesi, Ilya Kuprov, Christopher T. Rodgers, Paul A. Liddell, Devens Gust, Christiane R. Timmel & P. J. Hore, Chemical compass model of avian magnetoreception, *Nature*, Volume 453, pp387-391, 2008

Table 1: Representative publications

8 CONCLUSIONS

SWOT Analysis

Strengths

- Campus HPC is close to the end-user so research staff can build a long term working relationship with the HPC service provider. Indeed, many academics are on the management committees governing their local HPC facility.

- Campus HPC provides a local pool of HPC expertise within a university which helps enable a quick turn-around of technical issues or problems.

- Campus HPC provides the agility to respond to new research ideas, teaching drivers and the adoption of new technologies.

- Data locality. Campus HPC minimises possible issues with network bandwidth limitations or bottlenecks in transferring large datasets to and from remote HPC resources.

- Most campus HPC systems have proven to be very productive and cost-effective in comparison with some non-campus facilities.

- It is easy to ensure that the activities of campus HPC facilities align with institutional research and teaching strategy.

- Collaboration – the SRIF3 shared procurement process ably demonstrated the negotiating and purchasing power of universities collaborating with each other.

- The UK campus HPC service providers have an acknowledged breadth and depth of expertise in designing, building and running advanced computing and storage facilities. This strength and maturity of service translates into the strong and stable infrastructure supporting our HPC based researchers.

Weaknesses

- A major weakness is the funding model for campus HPC facilities. There appears to be a lack of a level playing field compared to some nationally funded facilities.

- Similarly, the effects of the variation of Full Economic Costing models as applied to HPC facilities can affect the efficacy of campus HPC provision.

- There is some difficulty in measuring the success, in terms of research outputs and/or value added, to enable campus HPC impact analysis due to lack of standardised methods.

- There are some difficulties in buying capacity outside of the institution (e.g. from an HPC cloud) as this usually requires capital expenditure to be converted into operating expenditure.

- It is proving difficult to retain the expert HPC support staff required to efficiently run HPC facilities due to the lack of an appropriate career development path and better salaries offered outside of academia.

- Most of the current endeavour in HPC focuses mainly on research. There are limited funding opportunities for the development of HPC technology, software and applications which could provide income streams.

- Variable quality of underpinning estate infrastructure can also affect the efficacy of campus HPC provision.

Opportunities

- Ensuring the continuation of a strong HPC ecosystem will continue to bring in world class researchers to the UK and ensure their retention.

- The HPC community have long embraced shared services – in fact members of the HPC-SIG have collaborated on shared service proposals in the past and are continuing to explore areas of common advantage and opportunity. This may lead in the fullness of time to regional HPC consortia with members drawn from both academia and local industry. The HPC Wales initiative is a prime example.

- To substantiate the previous point, there is growing interest from commercial, financial and industrial sectors in exploiting HPC. Regional HPC consortia could aid local SMEs to embrace HPC at little cost and risk to themselves.

- Campus HPC facilities can react quickly to changes in technology and in many cases lead the uptake and exploitation of new and novel technologies.

- Many campus HPC facilities are closely involved with the training and teaching of university staff and students – educating the next generation of computational scientists on campus with the latest technology and software assets.

- There are obvious opportunities for reuse and repurposing of data through data locality.

- Campus HPC facilities could also offer cloud-based HPC and storage services either by themselves or by forming consortia with similarly minded institutions.

Threats

- Staffing is a major issue – lack of sustained investment may hamper the retention of both research and support staff. Lack of established career development opportunities for support staff is also a major concern. There is evidence that support staff employment opportunities are increasing outside of academia.

- Fragmented and short-term funding streams are an obvious threat to sustainability.

- Embracing cloud based HPC could lead to a variety of issues such as finding funding for 'blue sky' or speculative research ideas and for student teaching in HPC technologies. It may also lead to an erosion of the skills base of local HPC support staff.

- The current funding pressure is increasing the danger of increased fragmentation of the HPC ecosystem. A lack of connectivity between national and local facilities will prove detrimental to the UK economy.

- The lack of a vibrant UK HPC ecosystem will lead to the UK falling behind international competitors and inevitably lead to a loss of reputation and scientific standing.

- Furthermore, the UK would become a secondary market for vendors and losing its substantial negotiating power in large IT procurements.

Final Remarks

In these difficult times, it is clear that world-class research and innovation will be an important factor in maintaining the UK's competitiveness in the world economy. Computational science is now accepted as the third pillar of science alongside theory and experimentation and as such having a well-founded and well-balanced HPC ecosystem will be crucial in helping preserve the UK's stature as a major centre of research excellence and innovation. Indeed, the use of HPC and computational science in new emerging areas in the digital economy such as health, bio-informatics, advanced manufacturing and energy highlight its increasing importance to 21st century society.

This report has highlighted some of the activities at campus HPC facilities. The breadth and depth of some of the research highlights ably demonstrate the need for such local facilities. Campus HPC facilities have many advantages such as:

- allowing HEIs to attract and retain the best academic staff, and allow them to develop leading-edge research,
- service maturity of university HPC - many facilities are embedded within wider service oriented organisations such as local IT service departments,
- enabling exploration of unfunded or new 'blue-sky' ideas and the associated development and testing of new software,
- enabling teaching of HPC methods and technologies to both undergraduate and postgraduate students,
- enabling researchers to quickly explore and adopt new or novel technologies.

In order to maintain the position of our world-class facilities relative to global competitors and to ensure our competitiveness in attracting external funding, particularly for large projects such as EU Framework projects, the UK needs to ensure sustainable funding for HPC at *both* campus and national levels in a co-ordinated fashion. This is vital for long term planning and to provide a stable state-of-the-art campus and national HPC base to nurture and expand applied computational research within the UK for the foreseeable future. Indeed, it is clear from the HPC-SIG survey that there is very strong support amongst the community of campus HPC service providers for a UK leadership-class system.

The survey also suggests that to ensure maximum exploitation of capital assets it is becoming increasingly obvious that there is a requirement for the development of a professional scientific/technical programmer career structure. If the investment in capital assets is not matched by a similar investment in appropriate training programmes and the associated development of an appropriate career pathway, there is a danger that highly talented staff will be recruited away by our global competitors who can and do offer these.

Several of the issues highlighted herein such as sustainability, the long term curation of vast amounts of data, and developing and maintaining the skills base, have also been raised in a recent report, 'Delivering the UK's e-Infrastructure for research and innovation'[22]. We broadly agree with the recommendations and conclusions of the RCUK report and we hope that this report can similarly inform decision makers about the UK campus HPC ecosystem.

Finally, there is a real spirit of collaboration within the HPC-SIG. In the past several members submitted a proposal for a joint HPC venture to the HEFCE shared services initiative. This collaborative interest is spurring discussions between members on reducing expenditure by, for example, sharing data centres, collaborating on disaster recovery and business continuity plans, and developing regional HE HPC clouds. The community is watching the current HPC Wales collaboration with some expectation. If successful, HPC Wales could provide an exemplary model for building sustainable, collaborative, shared regional HPC services which collaborate with and offer services to local academic and industrial stakeholders.

22 *www.rcuk.ac.uk/documents/research/esci/e-Infrastructurereviewreport.pdf*

9 CASE STUDIES

University of Bristol
Professor Nello Cristianini and Dr Marco Turchi
patterns.enm.bris.ac.uk/smart-dissemination-workshop

SMART is an European project that studied how to connect methods of modern Statistical Learning with Statistical Machine Translation and Cross-Language Information Retrieval.

Statistical methods are promising, in that they achieve performances equivalent or superior to those of rule-based systems, at a fraction of the development effort. There are, however, some identified shortcomings in these methods, preventing their broad diffusion. As an example, even though lexical choice is usually more accurate with Statistical Machine Translation (SMT) systems than with their rule-based counterparts, the text they produce tends to be less grammatical. As a second example, SMT systems are trained in batch mode and do not adapt by taking user feedback into account. Finally, in Cross-Language Information Retrieval tasks, query words are most often translated independent of one another, thus giving up possibly relevant contextual clues.

SMART addressed these and other shortcomings by the methods of modern Statistical Learning. The scientific focus was on developing new and more effective statistical approaches while ensuring that existing know-how was duly taken into account.

Inside the SMART project, we present an extensive experimental study of a Statistical Machine Translation system, Moses, from the point of view of its learning capabilities, and we discuss learning-theoretic aspects of these systems, including model selection, representation error, estimation error and hypothesis space. Learning Curves are obtained, by using High Performance Computing, and extrapolations of the projected performance of the system under different conditions are provided. More than 1,000 experiments were run using different training set sizes (from 12 k to 22,000 k training points), data domains (legal and news) and language pairs (French, Chinese, Spanish to English). As far as we know, we obtained the most accurate Learning Curves in Statistical Machine Translation.

University of Oxford
www.earth.ox.ac.uk/research/groups/geodynamics/home

A case study about the support provided by the Oxford Supercomputing Centre (OSC) for a final year undergraduate project.

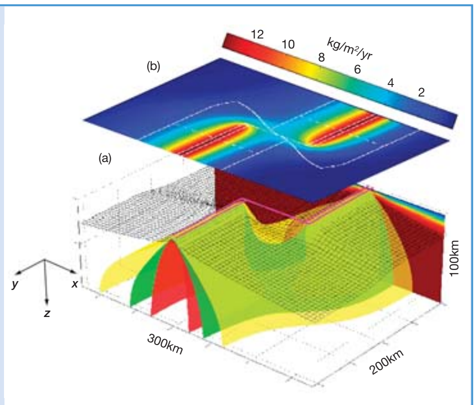


IMAGE COURTESY OF SAM WEATHERLEY

Sam Weatherley, supervised by Dr Richard Katz, in the Department of Earth Sciences undertook his 4th year project in 2008-2009. Sam used OSC hardware to run three dimensional numerical models of the creeping flow of the Earth's mantle. His objective was to investigate the dynamics of a mid-ocean ridge. Mid-ocean ridges are plate-tectonic boundaries beneath the ocean, where sea-floor plates diverge; new crust is created in the gap between them.

The project was motivated by observations of the global mid-ocean ridge system by a colleague in the US: there is a consistent asymmetry between opposite sides of the mid-ocean ridge near offsets of the ridge (these offsets are known as transform faults). This asymmetry is observed to be correlated with the direction of migration of the ridge over the mantle.

Sam used OSC hardware to compute the temperature and creeping flow of mantle rock beneath the ridge. In his models, mantle flow is driven by plate divergence and ridge migration. He used the calculated patterns of mantle temperature and flow to predict mantle melting, melt transport, and thickness of the sea-floor crust. His predictions produced an asymmetry that is consistent with the observations, and hence showed that the dynamics of the mantle are expressed in terms of a subtle but observable feature of the sea-floor.

University of Liverpool
Professor Alan Nahum
ctuprod.liv.ac.uk/CRUKCentre/nahum.html

A case study on using NW-GRID for a project at the Clatterbridge Centre for Oncology.

We are now extremely close to being able to take a radiotherapy treatment plan (i.e. the characteristics of each radiation beam and the CT representation of patient anatomy) and simulate it on the highly sophisticated Vancouver 'BEAMnrc' MC system (specifically adapted to advanced radiotherapy delivery techniques, known generally as intensity modulation) via the NW-Grid. This will provide highly accurate dose distributions for selected treatments carried out at our radiotherapy centre (one of the largest and best equipped in Europe) which will then be compared in detail with the doses obtained from analytical methods; we term this whole process 'Monte-Carlo based Quality Assurance of Advanced Radiotherapy Techniques' (MCQAART). As far as we are aware we will be the first radiotherapy centre in the UK to have implemented MCQAART, and this will have been made possible by the use of the NW-Grid.

University of Oxford
Dr Philip W. Fowler and Professor Mark S.P. Sansom
sbcb.bioch.ox.ac.uk

A case study about using supercomputing to probe the mechanical properties of proteins for a project undertaken by the Structural Bioinformatics and Computational Biochemistry Unit, Department of Biochemistry.

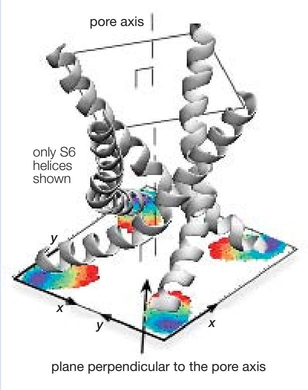


IMAGE COURTESY OF PHILIP FOWLER AND MARK SANSOM

Potassium ion channels are proteins that sit in cell membranes and conduct potassium when the voltage across the membrane is greater than a certain threshold. They behave therefore like biological transistors and are important in the generation of action potentials in e.g. the brain, along nerves and the heart. When the threshold voltage is reached, part of the protein moves down in the membrane and pushes on four alpha-helices which swivel shut, much like the iris diaphragm of a camera. The aim of the research is to calculate a map of how the free energy varies when the helices undergo this iris-like motion. We used NAMD2.7, a parallel classical molecular dynamics code, to run 60-90 different simulations, each of which had the ends of the helices constrained in a different position. All this data is then combined to yield a single map showing where the free energy minima are and any kinetic barriers. Our preliminary result is drawn above and as we can see there is a single free energy minimum which tell us that when the channel is closed it is frustrated which, amongst other things, explains the different opening and closing kinetics measured by experimentalists.

We ran the NAMD2.7 simulations on SAL, the OSC's Intel Xeon Nehalem cluster. Each simulation was only 10 ns long but running on 8 cores still took nearly two weeks to complete, generating over 2 GB of data in the process. Seven years ago, a single 10 ns simulation of a protein was considered cutting-edge, now it is possible to run 60 of them to probe the mechanical properties of proteins, like the potassium ion channels. It is essential that we begin to do more quantitative studies of this type to allow much tighter integration with experiment and so that we can try to provide parameters as inputs to higher-level models.

NAMD2.7 is a second-generation classical molecular dynamics code specifically designed for simulating the dynamics of proteins using the CHARMM forcefield. Unlike the first-generation codes, it was designed to be parallelised from the start and has demonstrated good scaling up to 8192 cores on an IBM BlueGene/P system. It is freely available for academic use and has extensive documentation and tutorials as well as a popular mailing list.

APPENDIX A: SURVEY OF HPC-SIG

The survey was carried out between March and August 2010. The 17 respondents given in Table 2 answered a total of 44 questions about their HPC facilities and how they are managed and funded.

Institution
Cardiff University
Imperial College London
Loughborough University
Queen's University Belfast
Queen Mary, University of London (QMUL)
University College London (UCL)
University of Birmingham
University of Bristol
University of Cambridge
University of East Anglia
University of Exeter
University of Liverpool
University of Manchester
University of Oxford
University of Sheffield
University of Southampton
University of Sussex

Table 2: Survey Respondents

A.1 Management of HPC facilities

The respondents were asked to describe the management of their HPC facility and whether it is managed as a central IT department, an academic facility or another model. Most HPC facilities are currently managed as a central IT department.

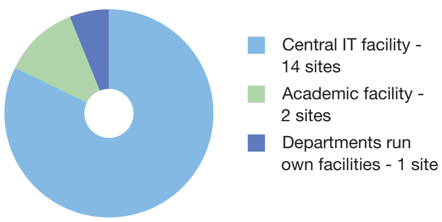


Figure 3: Management of HPC facilities

An example of each model is:
- Loughborough University is managed as a central IT department.
- The University of Exeter is managed as an academic facility.
- Queen Mary, University of London is run by an academic department.

A.2 Income and Expenditure

A.2.1 Funding and cost recovery models

Respondents were asked to describe their funding model as a Major Research Facility (MRF) funded via grant income, an indirect model where funding is provided centrally by the institution through top slicing of income or a hybrid model which combines elements of both the above models. The responses show a mixed pattern, with an indication that hybrid models are becoming more popular.

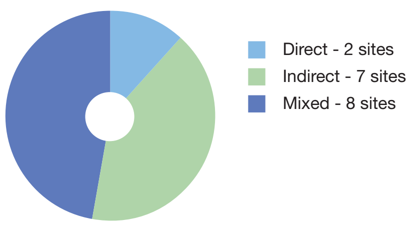


Figure 4: Funding models

An example of each model is:
- University of Oxford – directly funded as a MRF, underwritten by the university with the view to becoming self-sustaining.
- University of Southampton – indirectly funded through top slicing, with a percentage of research income allocated to HPC as academic infrastructure.
- University of Sheffield – a hybrid model where the facility is supported by an annual capital investment of £75K per annum which funds 59% of the facility; 32% of the facility is funded by investment from individual research groups and 9% is funded by groups purchasing resource for dedicated usage.

A.2.2 fEC rates per cpu hour

Respondents were asked to indicate their current level of Full Economic Costing (fEC) rate per cpu hour where appropriate. 11 of the 17 respondents provided a rate.

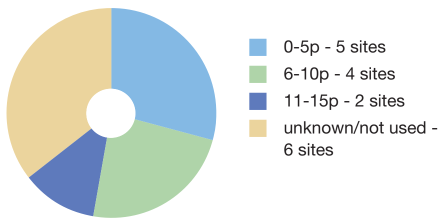


Figure 5: fEC rates per cpu hour

A.2.3 Income

Respondents were asked to indicate their institution's total grant income where there is an HPC element plus details of any commercial or rental income (as at 1 January, 2010).

Ten sites provided grant figures with a maximum income of £59m and a minimum income of £200k.

Three institutions derive commercial income from their HPC facilities; details are confidential.

Two institutions rent out machine room space, of which one does so internally for disaster recovery purposes.

A.2.4 Capital expenditure per annum

Respondents were asked for their capital expenditure as a representative annual spend, based on an average three year timescale.

All but two respondents provided a figure, although concern was expressed by several institutions about future levels of investment. The figures show a substantial variation, with the median capital expenditure figure of £400k as a representative annual spend over three years, the minimum spend £50k and the maximum £1.5m.

A.2.5 Non salary operating costs per annum

When asked about non salary operating costs, a range of responses was given; most institutions provided total non-salary operating costs, but no breakdown of depreciation, electricity, software licences, staff training and other costs.

Electricity costs were difficult for some institutions to obtain, with some electricity usage paid centrally. Four individual responses were given showing an average cost of £140k.

Some software license costs are built in to procurement costs or are funded from grants. Three individual responses were given, showing an average spend of £31k.

Some training is also built into procurement costs or is funded centrally.

A.2.6 Depreciation

Responses indicate that computer equipment is depreciated over an average term of 3.6 years and machine rooms over an average term of 10 years.

Facility	Users from institution using the facility
Daresbury, Blue Gene and others	Liverpool
DEISA (including Jülich and Max Planck)	Manchester, Bristol
HECToR	Bristol, Cardiff, Liverpool, Manchester, Oxford, Sheffield, UCL
Jaguar (Oak Ridge)	Oxford, UCL
National Grid Service	Liverpool, Manchester, UCL
NWGrid	Liverpool, Manchester
PRACE	Liverpool
Shaheen (Saudi Arabia)	Oxford
Teragrid	UCL

Table 3: Use of regional, national and international resources

A.3 Users

A.3.1 Total, internal, external and active users

Respondents were asked to give a total number of users, and then a breakdown between internal and external users plus an assessment of the number of users who can be classed as active, i.e. the Principle Investigator (PI) of a project and users within a project where jobs have been submitted to the HPC facility within the last 12 months.

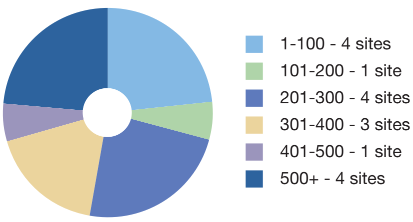


Figure 6: Total number of users

Internal users were 95% of total users and external users (who are collaborating with a researcher at the institution) only 5%.

The average number of active users was 145; 3 institutions were unable to provide a breakdown.

A.3.2 Use of regional, national and international resources

Respondents were asked for details of users or projects using national or international resources.

Seven institutions gave responses – at least one user or project is using the facilities as listed in Table 3, above.

A.4 Staffing

A.4.1 HPC total staff numbers

Respondents were asked how many HPC staff they have and to list the roles and percentage Full Time Equivalent (FTE).

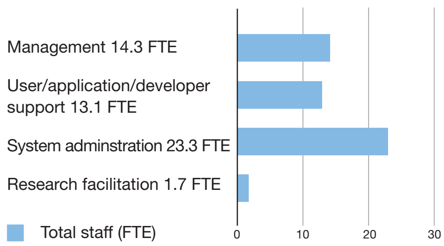


Figure 7: Breakdown of HPC staff roles and FTEs

The average number of staff was 3.63 and the median 3. The highest number of staff was 8.1 and the lowest 1.

There is therefore a substantial variation between institutions, but overall the figures indicate that HPC staff numbers are low in relation to the number of users and the size of the machines.

Profile of a project to reduce carbon footprint Sheffield

This is a summary of the project, *Optimising Energy Use In the Hounsfield Road Computer Centre*, undertaken by S Richardson, MSc student and his supervisor, Dr. S Beck in the Department of Mechanical Engineering (May 2008).

An energy audit was undertaken for the Hounsfield Road machine room; this is the location of the central HPC facility. Data was gathered and analysed from a diverse range of sources. The building was found to consume a significant amount of electricity both to run and cool the computer hardware, whilst at the same time requiring large amounts of energy to heat nearby offices.

With a view to finding ways of capturing the heat from the servers, a computational fluid dynamics package was used to simulate the flows of air and heat within the room where they are housed. This provided an overview of where heat energy from the servers is going, although a flaw in the model meant that data could not really be used to calculate meaningful estimates of the amounts of energy available.

From the data recorded on site it was estimated that the total cooling being provided to the hardware is in the region of 92kW. It was concluded that a heat capture scheme, possibly using a heat pump, could be implemented to provide significant savings from both a financial and an energetic view point.

Early in 2011 the university expects to sponsor another student to take this work forward.

The breakdown of roles indicates that the majority of HPC staff are undertaking management and system administration of the facilities, with a much smaller number specifically undertaking application and developer support, although in some cases the System Administrators also undertake an element of application and development support.

The importance of staff to the success of HPC is highlighted in A.9. When asked what would make their HPC facility more effective, ten out of seventeen respondents (59%) said more staff.

A.4.2 Academic staff

Respondents were asked to provide details of any academic HPC staff.

The University of Bristol has an HPC lecturer, based in the Department of Computer Science and the University of Oxford is hiring a lecturer for the new academic year in HPC and Visualisation.

With some institutions already providing credit bearing courses and others planning to do so, it is likely that there will be an increase in the number of academic HPC staff in the next few years.

A.5 Teaching and Training

A.5.1 Training courses

Respondents were asked to detail what training is offered, e.g. informal half day workshops, use of external training materials, commercial courses, courses run by NAG.

All institutions provide some training, with 11 institutions running in-house courses. The average number of attendees in 2009 was 219, the highest 1200 and the lowest 30.

A number of institutions run courses given by the Numerical Algorithms Group (NAG), www.nag.co.uk. These training courses are provided free of charge to HECToR users and UK academics whose work is covered by the remit of one of the participating research councils (EPSRC, NERC and BBSRC). Most HPC training is provided free of charge to users.

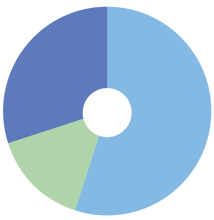


Figure 8: Training courses provided

A.5.2 Topics taught

Respondents were asked to list the topics taught as part of their training courses. Table 4 lists the results.

Using HPC, MPI, managing files	Applications	Programming languages
Condor	Abaqus	C
Introduction to HPC	Ansys	C++
LaTex	Image based modelling	CUDA
Linux	Matlab	Fortran
Make	Mathworks distributed computing toolbox	Perl
MPI (both in house and run by NAG)	Visualisation – Avizo, AVS/Express,	Python
OpenMP		R
Subversion		

Table 4: Subjects taught as part of HPC training programmes

A.5.3 Credit bearing courses

Respondents were also asked about any credit bearing courses offered and whether there are plans to offer them in the future.

Institution	Existing credit bearing courses	Plans for credit bearing courses
Imperial	Gives ancillary credit for the parallel programming courses to higher degrees	
Manchester	Runs MSc modules: Grid and e-Science, Computer Animation, Visualization for HPC (all School of Computer Science), Biomechanics (School of Life Sciences)	
Queens, Belfast	Offers an optional final year parallel processing module	A joint MSc is planned with Trinity College, Dublin
Sheffield	5 credit units are offered as part of the Doctoral Development Programme	
Bristol	A level M HPC module was offered from October 2010	A MSc is planned from 2011
Oxford		Credit bearing courses are under discussion

Table 5: Credit bearing courses

A.5.4 PhD supervision

At the University of Bristol, HPC staff co-supervise 2 PhD students and at the University of Manchester, HPC staff co-supervise 12-15 PhD students.

A.6 Machine statistics

A.6.1 Summary of HPC assets

Respondents were asked to provide details of their HPC assets.

Asset	Usage
Operating system	Mainly Linux
Cores	Highest 8000 Lowest 64 Total cores for all 17 respondents 39448
Interconnect	Mainly Infiniband
Storage/File system	NFS, Panasas, GPFS and Lustre all used
Peak performance	Highest 72 TFlop/s; lowest 1.2 TFlop/s. 120 TFlop/s (single precision) peak on a 32 node GPU cluster; 1 TFlop/s (double precision)
Sustained performance	Median 8.3 TFlop/s; highest 66 TFlop/s; lowest 2 TFlop/s (data not supplied for all machines)
Accelerator cards	5 institutions
Visualisation	4 institutions
Large scale storage	2 institutions
Windows cluster	5 institutions

Table 6: HPC assets

A.6.2 Sustained performance

This table summarises sustained performance for the 10 sites that provided figures; it does not include the University of Cambridge GPU cluster which achieved a sustained performance of 120 TFlop/s.

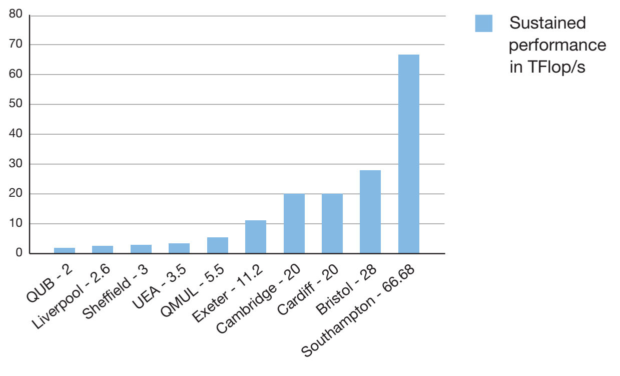


Figure 9: Sustained performance

A.6.3 Largest HPC facilities

The following facilities all have more than 2000 cores:

Institution	Number of cores
Southampton	8000
Imperial	7032
Cambridge	4180
Bristol	3744
University College London	2656
Oxford	2496
Cardiff	2048

Table 7: Largest HPC facilities

A.6.4 Vendors

Respondents were asked to provide the vendor (and integrator where appropriate) for each of their systems.

Vendors	Total number of machines (including those with an integrator)
Bull	3
ClusterVision	1
Dalco	1
Dell	6 (4 with ClusterVision, 1 with Streamline)
HP	2
IBM	4 (3 with ClusterVision, 1 with OCF)
SGI	7
Streamline	1
Sun	4 (1 with Esteem, 3 with Streamline)
Supermicro	1
Viglen	2 (1 with Streamline, 1 with NVIDIA)
Total number of systems	32

Table 8: Vendors

A.6.5 Usage

Some respondents gave a detailed breakdown of the largest research areas by usage on a percentage basis.

Institution	Usage by research area
Sheffield	Mechanical Engineering 25% Physics 24% Computer Science 18% Biological Sciences 15% Applied Maths 5% Electrical and Electronic Engineering 2% Control Engineering 1% Chemistry 1%
Liverpool	Chemistry 40% Surface Science 40% Engineering 6% High Energy Physics 1 % Radiology 1% Human Anatomy 1% System/Test/Development 10% Other 1%
Queen Mary University of London (QMUL)	Astronomy 80% Engineering 20%
Oxford	Mathematical, Physical & Life Sciences 80% Medical Sciences 10% Social Sciences and Humanities 10%
University College London (UCL)	Surface Science & Catalysis 27% Nanoscience & Defects 22% Earth Materials 20% Molecular Quantum Dynamics & Electronic Structure 8% High Energy Physics 6% Astrophysics & Remote Sensing 4% Bioinformatics & Comp. Biology 1% Unspecified 5%
Bristol	Chemistry 40% Geographical Sciences 18%, Physics 11% Mathematics 10% Computer Science 7% Electrical & Electronic Engineering 4% Biochemistry 3% Engineering Maths 2% Aerospace Engineering 1% Others 4%

Table 9: Usage of HPC facilities by research area

A.6.6 Usage statistics

Respondents were asked about the mix of work load running on their machines with a breakdown sought between serial and parallel jobs. It was suggested that parallel in this context can be running ensembles of serial jobs.

Most respondents report that their machines run over 90% parallel workloads (including ensemble jobs.)

A wide range of maximum job times were reported, with a median of 4.5 days and an average of 9 days. The maximum length allowed was 41 days.

Long term percentage utilisation of the HPC assets varied depending on the age of the assets, with a median of 75%. Maximum utilisation at one institution was 91%.

A.7 HPC facilities

A.7.1 Machine rooms

Respondents were asked how many machine rooms contained HPC equipment.

Responses indicated a mix of institutional shared machine rooms and dedicated HPC machine rooms. A number of sites are looking at expanding machine room resources, recognising the need to refresh machine room infrastructure. Most systems are housed in institutional machine rooms, but three sites have dedicated HPC machine rooms. Most sites have and use at least two machine rooms to provide for at least an element of business continuity.

The University of Sussex has recently relocated its HPC into a new energy efficient data centre, with a design PUE of 1.23. Cooling is provided by water cooled heat exchangers (USystems ColdLogic™) directly located on the rear of each cabinet (to cool the equipment within each cabinet rather than the entire room). The thermally isolated modular room is designed to operate between 24-27°C to maximise the free cooling throughout the year (external cooling plant includes a large Dry Air cooler and resilient industrial standard chillers for the warmer days). Power efficiency (99% even under part load) is provided by resilient line interactive UPS units (Eaton) housed in a separate plant room.

A.7.2 Cooling

Within their HPC machine rooms 11 institutions use standard Computer Room Air Conditioning (CRAC), five use advanced water cooling (three employ water cooled hot-aisle containment systems and two employ water cooled rear door heat exchangers) and one uses a mixture of CRAC and the TROX® CO_2 cooling system.

A.7.3 Machine room statistics

Respondents were asked to provide statistics for their machine rooms:

- Power density per rack in kilowatts indicated a median of 12.25 kW, with a maximum of 20 kW and a minimum of 5 kW.
- All machine assets are covered by Uninterruptible Power Supply (UPS) at 6 sites; the head/login nodes and storage only are covered by UPS at 10 sites.
- Power Usage Effectiveness (PUE) provides the ratio of the total amount of power used by the data centre facility to the power delivered to the IT equipment (i.e. not for cooling and other overheads). Ten responses were given, showing a wide range with most sites between 1.3 and 1.8 and a median of 1.66. There was one outlier of 3.07, representing an older machine room.

A.7.4 Carbon footprint reduction

Respondents were asked if they are involved in any projects to reduce carbon footprint.

Seven sites are involved in projects:

- University of Sheffield – to automate shutdown of worker nodes with a low level of utilisation; an on-campus project reviewing recovery of waste heat from data centre cooling systems; optimisation of machine room layout to optimise air conditioning requirements.
- University of Liverpool – proposals submitted to Salix to replace old hardware with new multi-core systems. ROI over 5 years looks feasible; have just won the Green ICT award at the Green Gown Awards – includes the Condor pool.
- University of Birmingham – actively using Adaptive Computing's MOAB Green Computing options.
- University of Oxford – JISC funded to write software which powers down under-utilised HPC resources.
- University of East Anglia – have worked with a Sustainable ICT Service Provision (SISP) project on improving compute suite efficiency.
- University of Manchester – investigating replacing old research computing equipment with more efficient newer equipment with investment from a 'Revolving Green Fund', half funded by the university and half from the Carbon Trust. Business cases in preparation.
- Cardiff University – measuring PUE; increasing air temperature; turning off unused nodes automatically. On the mainstream IT side, we are virtualising 90% of the main server-based infrastructure and implementing increasingly aggressive power saving measures on all PCs. We are also implementing an asset management package to gather more data on actual PC power consumption.

A.8 Research outputs

A.8.1 Publications in 2009

Respondents were asked how many publications have been published in 2009, or are in course of publication, listed by type (e.g. journal article, book chapter, conference paper), where at least an element of the content reflected research undertaken on their HPC facilities.

Seven institutions provided figures. The minimum number of published journal papers was 10 and the maximum 125, with a total number across the seven institutions of 391 journal papers plus 66 conference proceedings. Further papers are in course of publication. The journal papers included four published in *Nature* and two published in *Science*.

A.8.3 Key research achievements

Respondents were asked for details of any key research achievements, which help to demonstrate that the facility is good value, such as prizes won, awards given or contributions to HPC/software development.

Six institutions provided examples including:

- University of Sheffield – model wave propagation in the Solar Corona providing an insight into a mechanism for Coronal heating.
- University of Bristol – a team from the Department of Engineering Mathematics won Best Model Prize at the 2008 and 2009 international competition on Genetically Engineered Machines at MIT, Boston (iGEM).
- University of Liverpool – a consortium of academics named APEMEN (Agent-based Predictive Environment for Modelling Expansion in the Neogene) based around the North West Grid (NW-GRID) has developed evidence-based models for research into primate and human evolution. They simulated locomotion and carried out gait analysis to understand the energy costs and locomotor capabilities of extinct and extant species using metric evidence from fossil remains and observational field work. Circa 170,000 core hours were used, with the results only being possible through the use of the HPC facility. The research was featured on the BBC website – news.bbc.co.uk/1/hi/sci/tech/6956867.stm.

A.9 HPC wish list

All respondents were asked what would make their HPC facility more effective.

Ten would like more staff for system administration and particularly for application support and code development. The other responses reflected a desire for more equipment through a mix of more nodes, storage, visualisation and check point mechanisms.

APPENDIX B: MEMBERS OF THE HPC-SIG

The following are currently full members of the HPC Special Interest Group:

- Aston University
- Cardiff University
- Cranfield University
- Durham University
- Imperial College London
- Institute of Cancer Research (ICR)
- Kings College London
- London School of Hygiene and Tropical Medicine
- Loughborough University
- Queen Mary, University of London (QMUL)
- Queen's University Belfast
- University College London (UCL)
- University of Bath
- University of Birmingham
- University of Bristol
- University of Cambridge
- University of Central Lancashire (UCLAN)
- University of East Anglia
- University of Edinburgh
- University of Exeter
- University of Glasgow
- University of Lancaster
- University of Leeds
- University of Leicester
- University of Liverpool
- University of Manchester
- University of Nottingham
- University of Oxford
- University of Reading
- University of Sheffield
- University of Southampton
- University of Strathclyde, Glasgow
- University of Sussex
- University of Warwick
- University of York

The following are affiliate members of the HPC Special Interest Group:

- Cancer Research UK
- Engineering and Physical Sciences Research Council (EPSRC)
- GCHQ
- Natural Environment Research Council (NERC) Research Centres
- Science and Technology Facilities Council (STFC) Research Centres

APPENDIX C: GLOSSARY

BBSRC
Biotechnology and Biological Sciences Research Council
www.bbsrc.ac.uk

Cluster
A compute cluster is normally a collection of commodity of the shelf computers grouped together to form a single compute resource.
See, for example, www.beowulf.org

Core
The computational processor of a modern multi-core CPU.

CPU hour/Core hour
One hour's worth of computing on a computational processor. Most CPUs nowadays are multi-core so the Core hour is gradually replacing CPU hour in common parlance.

EPSRC
Engineering and Physical Sciences Research Council
www.epsrc.ac.uk

fEC
Full Economic Costing
www.jcpsg.ac.uk/index.htm

FTE
Full time equivalent

MRF
Major Research Facility (see fEC)

NAG
Numerical Algorithms Group
www.nag.co.uk

NERC
Natural Environment Research Council
www.nerc.ac.uk

NW-Grid
North West Grid is a collaboration between Daresbury Laboratory and the Universities of Lancaster, Liverpool and Manchester.
www.nw-grid.ac.uk

Petabyte
10^{15} bytes

PFlop/s or Petaflop
10^{15} Floating Point Operations per second

PUE
Power Usage Effectiveness – a metric used to measure the efficiency of a data centre by dividing the total power delivered by the total IT equipment power usage.

PRACE
Partnership for Advanced Computing in Europe
www.prace-project.eu

STFC
Science and Technology Facilities Council
www.stfc.ac.uk

Terabyte
10^{12} Bytes

TFlop/s or Teraflop
10^{12} Floating Point Operations per second

Teragrid
TeraGrid is an open scientific discovery infrastructure combining leadership class resources at eleven partner sites to create an integrated, persistent computational resource.
www.teragrid.org

UPS
Uninterruptible Power Supply

White Rose Grid
A collaboration between the Universities of Leeds, Sheffield and York who are engaged in eScience, Grid and cloud computing.
www.wrgrid.org.uk

High Performance Computing
Special Interest Group

www.hpc-sig.org