International Review of Research Using HPC in the UK

December 2005

This document represents the conclusions of an international Review Panel of experts in computational science and engineering. The views expressed are entirely those of the members of that Panel.

Engineering and Physical Sciences Research Council
Polaris House, Northstar Avenue
Swindon, SN2 1ET
Wiltshire, UK

http://www.epsrc.ac.uk/

ISBN 1-904425-54-2

EPSRC
Engineering and Physical Sciences Research Council

Deutsche Forschungsgemeinschaft
DFG

CONTENTS

1. Executive Summary

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. This view was probably first expressed in the mid-1980s. It grew out of an impressive list of accomplishments in such diverse areas as astrophysics, aeronautics, chemistry, climate modelling, combustion, cosmology, earthquake prediction, imaging, materials, neuroscience, oil exploration, and weather forecasting. Today, in the middle of the first decade of the 21st century, the pace of innovation in information technology is accelerating, and consequently the opportunities for computational science and engineering abound. Computational Science and Engineering (CSE) today serves to advance all of science and engineering, and many areas of research in the future will be only accessible to those with access to advanced computational technology and platforms.

Progress in research using high performance computing platforms has been tightly linked to progress in computer hardware on one side and progress in software and algorithms on the other, with both sides generally acknowledged to contribute equally to the advances made by researchers using these technologies. With the arrival of highly parallel compute platforms in the mid-1990s, several subtle changes occurred that changed the face of CSE in the last decade.

Because of the complexities of large-scale hardware systems and the increasing sophistication of modelling software, including multi-physics and multiscale simulation, CSE increasingly became a team science. The most successful practitioners of CSE today are multidisciplinary teams that include mathematicians and computer scientists. These teams have set up a software infrastructure, including a support infrastructure, for large codes that are well maintained and extensible beyond the set of original developers.

The importance of CSE for the future of research accomplishments and economic growth has been well established. "Computational science is now indispensable to the solution of complex problems in every sector, from traditional science and engineering domains to such key areas as national security, public health, and economic innovation," is the principal finding of the recent report of the President's Information Technology Advisory Committee (PITAC) in the U.S.[1] Hence the EPSRC's decision to organize a review of the state of research using HPC in the UK was timely. The International Review Panel found that research using HPC in many areas is of the highest standing and competitive at the international level. However, in a dynamically changing and rapidly evolving field such as CSE, one cannot afford to stand still, and therefore in this report the Review Panel also point out areas for further improvements and refinements of the HPC research strategy. It is our hope that building on this report, the UK computational science and engineering community will be able to continue to maintain its position at the forefront of research using HPC.

[1] President's Information Technology Advisory Committee, *Computational Science: Ensuring America's Competitiveness* (Arlington, Virginia: National Coordination Office for Information Technology Research and Development, 2005), p. 2.

2. Introduction

2.1 THE IMPORTANCE OF RESEARCH USING HPC FOR THE UK

For the UK to compete in the global economy, it must take advantage of its excellent track record in scientific discovery and technological innovation. Both science and technology are becoming increasingly dependent on high performance computing (HPC)—indeed, computation has emerged as a distinct branch of research, with an importance matching those of theory and experiment. Computation has now become essential for the advancement of all research across science and engineering. HPC "aims to satisfy the most demanding scientific goals, to push the boundaries of researchers' ambitions and to stretch the development of hardware and software technologies in dimensions that often prove beneficial outside the [HPC] research arena."[2]

HPC enables science and engineering to tackle a class of problems and opportunities that cannot be approached in any other way. These opportunities include genome sequencing, biomedicine, and drug design; molecular engineering of new materials and devices; engineering efficient and environmentally sound technologies and energy sources; forecasting global climate change and finding strategies to deal with it; and understanding the fundamental interactions of elementary particles and the origin and fate of the universe.

Research using HPC is widely expected to be a major contributor to the UK's "knowledge economy." The concept of the knowledge economy is based on the assumption that knowledge and innovation will be crucial to future economic success in the industrialised world. Information and knowledge are replacing capital and energy as the primary wealth-creating assets, just as the latter two replaced land and labour 200 years ago. The only comparative advantage a company (or a nation) will enjoy will be its process of innovation—combining market and technology know-how with the creative talents of knowledge workers to solve a constant stream of competitive problems—and its ability to derive value from information.

The UK Research Councils recognize the importance of HPC and have made significant investments in HPC facilities and the research they support. To assess the returns on those investments, an international panel of ten computational scientists spent a week in September 2005 visiting universities, talking with researchers and students, discussing their observations, and formulating findings and recommendations. This report is the culmination of their efforts.

2.2 CONTEXT, SCOPE, AND METHODOLOGY OF THE REVIEW

In 1999, the Engineering and Physical Sciences Research Council (EPSRC)—in cooperation with relevant learned societies—began a cycle of reviews by international panels of the state of UK science in fields corresponding to its major programmes. In these reviews, a panel of internationally leading researchers benchmark the strength of the UK research activity against world competitors and highlight any gaps or missed opportunities. The panel visit a number of UK research groups and have access to a wide pool of experts and supporting data to help them reach their conclusions. Each research discipline is scheduled to be reviewed every five years.

The 2005 International Review of Research Using High Performance Computing in the UK was coordinated by the Deutsche Forschungsgemeinschaft (German Research Foundation or DFG) and was the first to review this area of research. The review was overseen by a Steering Group (see Appendix 7.1) who appointed the Chair of the Panel, selected the other Panel members (Appendix 7.2), and set the terms of reference as follows:

[2] High End Computing Terascale Resources (HECToR) Scientific Case (unrestricted version, dated 26 April 2004), p. 4.

The International Review Panel is requested to:

- report on the calibre, standing and research potential in research using High Performance Computing in the UK
- discuss the potential impact of university-based research using High Performance Computing on the UK's knowledge economy
- provide comparisons with international research using High Performance Computing
- make recommendations on future actions and/or priorities

The research groups to be reviewed (Appendix 7.4) were selected by the Steering Group from a list proposed by EPSRC in consultation with the Biotechnology and Biological Sciences Research Council (BBSRC), the Natural Environment Research Council (NERC), and the Particle Physics and Astronomy Research Council (PPARC). Criteria included usage of the CSAR and HPCx computing facilities, Research Assessment Exercise (RAE) scores, and scientific diversity.

Panel members were sent background documentation about UK science policy and funding structures, basic information about research using HPC in the UK, and information about the research groups to be reviewed (provided by the groups themselves in answers to a questionnaire). The Panel convened in London on 4 September 2005 to review roles and responsibilities before splitting up into sub-groups to visit universities and computing centres. Discussions between sub-groups and researchers were guided by a standardised review sheet (Appendix 7.5), on which panellists recorded their observations and comments. The leader of each sub-group compiled the group's comments onto a single review sheet, and all of these were distributed to the entire Panel. The Panel reconvened in London on the evening of 7 September and all day on 9 September to discuss their findings. During the following weeks, Panel members drafted sections of this report, which were compiled and edited by the Scientific Secretary.[3] Drafts were circulated for review and comments, and the report was then finalised under the supervision of the Panel Chair and accepted by the Steering Group.

This review was particularly challenging for all participants, not only because it was the first on this topic, but even more because of the diversity of the research under consideration. Computational Science and Engineering (CSE) is interdisciplinary by nature and is a relative newcomer on the academic scene. There are few degree programs in CSE and few university departments dedicated to this field; as a result, there is a scarcity of objective data on the state of the discipline in the UK, in comparison with data that can readily be obtained for more traditional disciplines. The Review Panel were keenly aware of the anecdotal nature of much of the information presented to them, and were careful in their deliberations not to draw overly broad conclusions from too few examples. Nevertheless, several common themes did emerge from the reviews of the individual research groups, and those themes form the basis for this report.

[3] In a few cases, additional information was provided to the Review Panel by research groups to resolve questions that remained open after the site visits.

3. Findings on the UK HPC Research Environment

3.1 DELIVERY OF SCIENCE THROUGH THE USE OF HPC

General Observations

During the course of this review, the panel observed a wide range in the quality of science delivered through the use of HPC in the UK but overall was very impressed. It is now widely recognized that the rapid advances in HPC have enabled realistic computer-based modelling and simulation to accelerate progress in solving the formidable mathematical models that govern the behaviour of complex natural and engineered systems, including the challenge to compute phenomena over wide ranges of temporal and spatial scales. For example, the challenge in materials research is to deduce macroscopic behaviour directly from the quantum-mechanical behaviour of the constituent atoms or molecules. For atmospheric flows, global patterns need to be resolved, as well as the intricate details of the heat, mass, and momentum exchange at the surface of the sea. In plasma physics, new insights are needed to address how large-scale magnetic confinement of high-temperature plasmas can be significantly improved via control of micro-scale turbulent losses. In particle physics, high-precision computer simulations of QCD are expected to provide detailed tests of the Standard Model and better understanding of quark confinement and the phase structure of strongly interacting matter. For astrophysics and cosmology, the task is to bring together the detailed interaction of gas with electromagnetic radiation and the gravitational potential of dark matter. In engineering applications, industrial competitiveness depends critically on the continuous development of advanced computational techniques for design optimization of complex systems. UK scientists are effectively utilizing the advances in HPC in many of these areas with results that compare favourably with international standards.

The collective strength of effective consortia formed in a number of research areas has greatly aided the delivery of new scientific results through computational models with high fidelity physics developed with significantly less abstraction than in the past. Effectively harvesting the great potential for scientific discovery made possible by HPC has also involved making progress on the formidable challenge of obtaining meaningful information from the tremendous amounts of data generated at dramatically higher rates than ever before. As emphasized with specific recommendations in Section 3.6 below, the consortium model is an organizational asset of HPC in the UK which should not only be maintained but further expanded.

Specific examples of the delivery of science through the use of HPC in the UK are highlighted in the following section. These examples of impressive advances enabled by HPC come largely from those groups (including UKQCD, Mineral Physics at UCL, and UK Turbulence Consortium) which are very proficient in both using and enhancing computational technologies for the accelerated achievement of new scientific results. Many other research groups use computational technology effectively but only in the restricted sense of delivering the scientific results using computational tools and codes developed elsewhere.

As a whole, it was observed that the level of integration of computational science with computer science in the UK has not yet reached the highest international standards. Perhaps due in some measure to present funding practices, many research groups in the UK appeared to be deficient in adapting to modern programming practices, anticipating the need for data management and visualization, and developing interoperable software and algorithm environments. If these deficiencies were remedied, we should witness the accelerated emergence of dramatically improved capabilities for making critical discoveries in many scientific domains which in the past have been considered intractable due to their extreme complexity and/or lack of available data analysis capability.

As we look toward the future, the examples of significant scientific progress observed in the current review indicate that the needs of exciting scientific applications will further stimulate the development of new computer science capabilities and innovative mathematical methods and algorithms. These advances will help harness the underlying exponential growth in technology and enable new possibilities for investigation.

Specific Accomplishments

In the course of this review, several examples of the fruitful delivery of science through the use of HPC in the UK stood out. These included:

Institute for Computational Cosmology (ICC), University of Durham: Carlos Frenk is the Director of ICC and co-PI of the Virgo Consortium, together with Simon White, Director of the Max Planck Institute for Astrophysics in Garching, Germany. The Virgo group has carried out the largest, most detailed cosmological simulations: Hubble Volume and Millennium. With 10 billion particles, the Millennium N-body simulation of the growth of structure in the standard cold dark matter model is about one order of magnitude larger than the previous largest. Science magazine called the combination of 2dF galaxy survey data interpreted via extensive use of cosmological simulations with Wilkinson Microwave Anisotropy Probe (WMAP) data the Breakthrough of the Year for 2003, "Illuminating the Dark Universe."

UKQCD: Effective utilization of HPC by the UKQCD consortium has achieved major advances in the understanding of observations from worldwide particle physics experiments. In particular, a very important and highly visible result obtained in partnership with the HPQCD, MILC, and Fermilab Lattice Collaborations is the comparison of physical results obtained in the quenched approximation with results from full dynamical simulations including the three lightest quarks. The discrepancies between results from experiments and the quenched approximation are completely eliminated in the much more realistic dynamical simulations. Although there is still an unknown systematic error in these results that ought to be clarified, this result has been widely heralded as a breakthrough in the field of lattice quantum chromodynamics.

UK Turbulence Consortium (UKTC): Significant progress in the understanding of laminar-turbulent transition (Neil Sandham's group at Southampton) and new insights into key topological issues in the physics of premixed combustion (Stewart Cant's group at Cambridge) have been attained by the UKTC's direct numerical simulations using HPC resources. In particular, direct numerical simulations of premixed turbulent flames utilizing the full available power of the HPCx system revealed that the topology is largely cylindrical in shape. This is an important finding which can be cited in justifying the application of simpler geometric models.

Multiphoton, Electron Collision, and Bose Einstein Condensates (MECBEC) HPC Consortium: In the area of atomic, molecular, and optical physics, Ken Taylor's group has gained important insights into the limits of photon absorption in time-dependent Schrödinger equation calculations of multi-photon ionization. They have also been the first to calculate the full two-electron response of helium to intense free electron laser light at vacuum ultraviolet (VUV) wavelengths.

HPC Materials Chemistry Consortium: In the area of HPC simulations of materials and nanoscience, Richard Catlow's group has achieved major progress in studies of heterogeneous catalytic processes. In the course of their development of the most advanced molecular dynamics code (DL_POLY3), simulations of several million atoms have been carried out—a world record for simulations of materials with charged ions.

Mineral Physics at University College London: David Price's group at UCL have carried out HPC *ab initio* simulations with novel results (published in Nature) which provide tight constraints on the composition and thermal structure of the Earth's core.

UK Astrophysical Fluids Facility (UKAFF): UKAFF has performed a highly visible simulation of star formation that appeared on the cover page of Nature.

These examples of impressive advances enabled by HPC come largely from those groups which are very proficient in both using and enhancing computational technologies for the accelerated delivery of new scientific results.

3.2 STANDING AND RESEARCH POTENTIAL IN COMPARISON TO THE INTERNATIONAL FIELD

The Review Panel judged that in many of the consortia and research groups visited, the scientific results compared well with the highest international standards. These groups included the HPC Materials Chemistry Consortium, the Institute for Computational Cosmology, the MECBEC HPC Consortium, the Mineral Physics Programme at UCL, the UK Astrophysical Fluids Facility, the UK Car-Parinello Consortium, the UKQCD Collaboration, and the UK Turbulence Consortium. In some cases—ICC, Mineral Physics, UKQCD, UKTC—the UK groups are playing a lead role in setting the international standards. Continued leadership and realization of the latent potential will depend on continuity in getting access to adequate computational resources, as discussed in Section 4 below.

Panel members are also aware of other internationally leading institutions in the UK that were not part of this review but should not be overlooked, including the European Bioinformatics Institute, the Hadley Centre for Climate Prediction and Research, and the European Centre for Medium-Range Weather Forecasts.

In view of the recognition that "computation has now become essential for the advancement of all research across science and engineering,"[4] it is clear that HPC in the UK has not achieved its full potential in every area of research. The Panel found a variety of possible reasons for this shortcoming:

- Many computational scientists are not working closely with computer scientists and HPC manufacturers to maximize the effectiveness of computer systems for specific scientific applications. As noted above, many research groups in the UK appeared to be deficient in adapting to modern programming practices, anticipating the need for data management and visualization, and developing interoperable software and algorithm environments.
- The issue of developing algorithms, tools, and methods that are required in common by diverse projects in computational sciences has not been adequately addressed in the UK. There appears to be no strong national computational science community to lobby for such a programme.
- The training of computational scientists in the UK is typically too narrow (problem- and code-specific) and too short for interdisciplinary training and internships, as discussed in Section 3.3 below.
- The traditionally strong assets in some fields of UK science are theory and observation rather than computation.

The EPSRC's recognition of the importance of HPC and the research community's already strong demand for HECToR indicate that the UK research environment is now ripe for fostering scientific progress through HPC. During the course of our site visits, the Panel recognized both the significant achievements and the high potential of computational science and engineering in the UK. Of course, the progress of computational science depends strongly on the available computer facilities. A new HPC system would undoubtedly attract good students and provide a boost for both existing and new computational research projects in the UK, thus contributing substantially to science and engineering worldwide.

3.3 DEVELOPING AND RETAINING TALENT

With increasing emphasis on computer and information sciences in international scientific research, the Panel assessed whether the education and training of computational scientists and engineers in the UK reflected the need for scientists who have high interdisciplinary training. The metrics employed included training and competency in numerical algorithms, expertise in software engineering, ability to adapt to modern programming practices, and efficient use of high performance computational resources.

[4] High End Computing Terascale Resources (HECToR) Scientific Case (unrestricted version, dated 26 April 2004), p. 4.

The Panel probed the plans evolved by the various consortia for training students and retaining them in the computational science area. The UK funding agencies have instituted High End Computing Studentships that provide four-year funding, including one year equivalent of training in the computational sciences over the four years for M.Sc.-Ph.D. students enrolled in this program. In Warwick and Edinburgh, students work on computational science modules and thus obtain additional training. Even though small in scale (about ten students per year), this was found to be a good mechanism for interdisciplinary training by several consortia. The Integrated Biology group had evolved computational life science modules where they trained students for six weeks. This program was widely subscribed to and was regarded as a success by life science trainees. Several of the groups, however, had no formal training mechanism for their students and relied on "during the studentship training." The Materials Chemistry and UKCP consortia expressed the desire to have more extensive training for their students, but were limited by the three-year degree constraints. There was a significant paucity of students in some consortia, where the absence of students despite the presence of resources was a barrier for research productivity. To some extent this reflects the culture of education and training in UK institutions. The Panel also were told that the difficulty of obtaining UK funding for non-domestic students imposes constraints on attracting the best talent.

Students trained in the various programs either continued to stay in the program with fellowships or moved on, in several cases to other computational jobs (mostly in the same discipline). In some areas, such as life sciences and materials simulations, the students found jobs in the respective industry. There was some attrition of students to non-science job opportunities, and this was more common in physics than engineering or life science disciplines.

The Panel observed that in the majority of the sites visited, the personnel were trained in the specific code they were using and in some cases contributed significantly to the development of the code. However, the Panel observed significant shortcomings based on the metrics stated above. The Panel observed that the training of most of the students in the consortia was centred around their specific disciplinary codes. Algorithmic sophistication was found lacking in several cases. The majority of the application codes were extensions of legacy codes, i.e., codes written in Fortran that had grown over decades, and therefore a modern software engineering perspective was lacking, even though the trainees using the codes were proficient in their respective programs. Also, in the area of visualization, modern client-server programming paradigms were not deployed. Overall, the culture of best-practices software engineering has not percolated effectively into the UK scientific community. Still, the panel observed that several of the consortia effectively used the high performance computational resources. Several codes were MPI-capable and were scaling well on multi-processor machines. In the area of data and data-driven computing, a number of groups recognized its growing importance; some are already developing Data Grid tools, but others acknowledged their unpreparedness for data organization, handling, and query and analysis.

The Panel tried to understand the reasons behind the above findings. There appear to be two major criteria: first, the culture in computational science training in the UK has historically been only problem-specific. For instance, students are trained in specialized codes such as CASTEP during their studentship. Second, the UK educational and funding system mandates that each student complete his or her Ph.D. program in three years. This leaves little or no time for formal interdisciplinary computational training. Further, it inhibits the opportunities for internship training in industry during graduate studentship.

The Panel makes the following specific recommendations:

1. The funding agencies should explore training fellowships that mandate formal curriculum in computational science. This is not just to offer training to a new generation of scientists and engineers, but also to give a stronger identity to the computational science community in the UK.

2. Institutions like Daresbury Laboratory and efforts like Collaborative Computational Projects (CCPs) should be given the resources required to educate the science consortia and university research groups in best-practices software engineering.

3. Institutions should encourage industrial internships for students and explore joint funding opportunities.

4. CCPs and other consortia should generate tools and methods that are required in common by diverse projects in computational sciences.

5. Funding should be provided on a longer time scale for computational science Ph.D. students.

3.4 IMPACT ON THE KNOWLEDGE ECONOMY

Two elements can be considered as measures of the HPC contribution to the knowledge economy: (1) technology transfer activities and (2) research collaborations between academic and industrial partners.

Technology Transfer Activities

The transfer of HPC expertise and know-how from research universities to industry provides a competitive edge in the knowledge economy. Mechanisms for this technology transfer can be any of the following:

- supporting industry in extending and deploying their high-end technology through the use of HPC
- supporting industrial efforts to port in-house, proprietary codes to dedicated HPC hardware inside the companies
- stimulating access of industry to high-end HPC facilities and intellectual resources through adequate channels, taking into account the issue of security and confidentiality critical for industry-proprietary software
- training experts in HPC software engineering, at the graduate or postgraduate level, as a basis for transferring know-how to industry
- providing dedicated training in HPC, including hardware and software engineering, databases, post-processing of data, and visualization
- providing dedicated training in problem solving through the use of HPC.

From the sites visited and reviewed, it appears that the university research groups are the main source of training in modelling, simulation, and HPC in the UK, and that the majority of students after graduating continue working on aspects of HPC either in universities or in industry. Currently, much of the demand for trained professionals in computing and modelling in the booming financial industry is satisfied by new graduates. The knowledge base on HPC related technologies is widely present within the UK research community; the good placement record illustrates this finding. However, there appears to be no global strategy towards a consistent implementation of the action lines mentioned above.

It should be noted, however, that the Computational Science and Engineering Department (CSED) at Daresbury Laboratory has developed the expertise to support code porting to HPC architectures, even though the EPSRC programme does not seem to fund dedicated efforts toward technology transfers to industry. Although the CCPs supported by the CSED have links with industry, including industrial members on working groups and the steering panel, it is unclear whether the actions mentioned above are being effectively implemented through these measures.

Research Collaborations between Academic and Industrial Partners

Some of the research consortia, such as UKAA, UKTC, UKCP, and UKQCD, have strong industrial collaborations and provide specialized expertise for the growth of UK knowledge economy. These industrial collaborations make industry aware of the benefits of HPC.

- The UK Applied Aerodynamics Consortium (UKAA) is focusing its activities towards high-end industrial applications in aerodynamics, for external flows (helicopter applications) as well as internal flows (gas turbine engine components), pushing HPC simulations of industrial applications beyond the state of the art.
- Similarly, the UK Turbulence Consortium (UKTC) is oriented towards fundamental applications in combustion and turbulence

but having a direct relevance for industrial applications.

- The UK Car-Parinello Consortium (UKCP) has developed a software system that is used worldwide and marketed to industry through a commercial joint venture.

- The UKQCD consortium has a collaboration with IBM in the development of a specialized supercomputer for lattice QCD simulations. This collaboration and the emerging computer technology laid the ground for IBM's very successful BlueGeneL systems.

Other consortia, oriented toward fundamental physics, biology, or chemistry, might not have direct links with industry. However, the experience gained within these areas might be made available to industry if more extensive interdisciplinary cross exchanges between the different consortia would be implemented. In addition, the universities should implement or strengthen such interdisciplinary cross training between different fields in HPC, so that students could learn how solutions developed for a particular problem might also be applied to different domains.

In general, the application oriented consortia have strong links with industry and provide a channel for industry to become knowledgeable about and reap the benefits of HPC.

3.5 STRENGTHS

As stated in Section 3.1 above, the panel in general were very impressed with the quality of science delivered through the use of HPC. The panel considered the formation of research consortia (discussed in more detail in Section 3.6 below) to be one of the most important factors contributing to this quality. The creation of communities with a similar agenda facilitates the flow of scientific and technical know-how among consortia members. The consortia also seem to be the main source of training in modelling, simulation, and data analysis using HPC. And the collaborations of several consortia with industry are pathways for research using HPC to have a positive impact on the UK's knowledge economy.

In general, the research groups reviewed are using computational technology effectively to deliver scientific results, and several groups are actively developing and enhancing computational technologies for the delivery of science. In particular, UK researchers' leadership in the development of scientific application codes is contributing to scientific progress worldwide. Examples of such codes (and the research groups involved in their development) include:

- GAMESS-UK, a general purpose *ab initio* molecular electronic structure program (Computational Science and Engineering Department [CSED] at Daresbury Laboratory)

- CASTEP, which uses density functional theory to compute the forces on the atoms and to simulate the time evolution ("dynamics") of molecular systems (UK Car-Parrinello Consortium)

- DL_POLY, a general purpose molecular dynamics simulation package (CSED)

- MOLPRO, a system of *ab initio* programs for molecular electronic structure calculations with extensive treatment of the electron correlation problem (University of Birmingham)

- Chroma, a software system for lattice QCD calculations which is portable and efficient on a wide range of architectures (UKQCD Collaboration)

- HiGEM, a new high-resolution integrated climate modelling code which includes atmospheric chemical influences (NERC Centres for Atmospheric Science).

The Computational Science and Engineering Department at Daresbury Laboratory plays an important role in supporting the research consortia by assisting in porting and optimising users' codes, developing new applications and algorithms, and evaluating new programming methodologies. Most of the consortia in close contact with the CSED expressed their satisfaction and sometimes enthusiasm about the support obtained from this group in areas such as e-science. In Section 3.6 below, the Review Panel recommend more opportunities for the kind of interdisciplinary communication and interactions that CSED exemplifies.

The UK has been very well served by the leading-edge HPCx and CSAR computational systems, by the special-purpose QCDOC machine, and by many university-based computational systems. The HECToR program promises to build on this excellent tradition. HPC resources are discussed in more detail in Section 4 below.

EPSRC's recently initiated High End Computing Studentships combine training at one of the UK's leading computation centres with a research project at the student's host institution, leading to a Master's degree or Doctorate in Computational Science and Engineering. The Review Panel see the studentships as a good strategy for training the next generation of computational scientists and engineers and for overcoming the boundaries and barriers between academic departments.

3.6 OPPORTUNITIES

Research Consortia

The Panel feel that the consortium structure is an excellent mechanism for granting resources to the research community.

From an *operational point of view*, this mechanism helps to smoothen the volatility in the demand for compute cycles commonly observed for research users. Consortia are able to nimbly re-apportion resources in response to sudden changes in demand caused, for example, by the progression of the research or the addition or loss of personnel, thus making resource management easier for all parties involved (servers and clients). For the computing centres (servers), the demand is more balanced, steady, and predictable, resulting in fewer compute cycles going to waste. Also, the communication between servers and clients is more coherent, as each consortium is represented by one single voice. The more coherent communication between users and compute infrastructure management makes the response to user needs as well as the improvement of services much easier.

From a *scientific point of view*, the formation of communities with a similar but not necessarily identical agenda facilitates the flow of scientific and technical know-how among consortia members.

The Panel observed that the allocation and redistribution of resources within a consortium generally appears to be a smooth process; no complaints were recorded.

Whereas some consortia live a life as pure resource owners and redistributors (e.g., ChemReact), others have much more elaborate agendas, including regular meetings to discuss scientific and technical issues. In at least one case (Materials Chemistry), we encountered a consortium with a designated internal communicator funded through the grant. Overall, the Panel observed substantial differences among consortia with respect to their presentations at the review: some consortia considered the review was a major event, whereas for others it was a sheer administrative process. Accordingly, the amount and quality of information that was conveyed varied considerably.

Funding for meetings and other consortia activities appeared to the Panel as valuable instruments for the exploitation of synergies within the consortium. However, the Panel observed some reluctance to request an allocation for consortium management support, as there was a fear of cannibalizing the allocation for the scientific agenda.

In certain instances, there is a strong overlap between the consortia and the Collaborative Computational Projects (CCPs). Some consortia even appear to be offspring of the CCPs. The CCPs represent an organizational structure which has grown organically from the very beginning of HPC in the UK, and which is still in a process of growth. However, due to the focus on scientific disciplines and the lack of well established general HPC programs, there appears to be no strong national computational science community. (Section 3.3 above discusses some of the drawbacks attached to this situation.)

On the basis of these observations, the Panel makes the following recommendations:

1. The consortium model is an organizational asset of HPC in the UK and should be maintained, if not expanded. In particular for large, distributed multi-group consortia, "coordination funds" should be automatically allocated, i.e., the proposers should be asked

to present an agenda for the management and coordination of their consortium.

2. Mechanisms for coordination of efforts between consortia need to be defined. These can be in the form of student exchanges, meetings, workshops, etc. An enhancement of the transfer of technical know-how (algorithms, software engineering, data processing) across disciplinary boundaries of the UK community is needed.

3. However, before introducing new formal structures, the role of the consortia as well as the relationship between the consortia and the CCPs needs to be revisited in order to avoid duplicate efforts or even competition.

4. The Research Councils need to keep an eye on the development of a national computational science community to make UK modelling and simulation more competitive.

Collaboration with Computer Science and Applied Mathematics

Collaborations with computer scientists have become a new element in computational science and engineering that often leads to successful new scientific accomplishments that would not have been possible otherwise. The SciDAC (Scientific Discovery through Advanced Computing) program in the US has several examples of such successful collaborations.

In our review of UK research using HPC, the Panel found several examples of outstanding and productive collaborations between applications scientists and computer scientists or mathematicians. The Reality Grid project at University College London demonstrated some first-rate examples of computational steering that required Grid middleware, understanding of networking technology, and advanced visualization. Also commendable was the inclusion of a visualization researcher from the computer science department into the Institute for Computational Cosmology at the University of Durham. However, in general, these types of collaborations were the exception rather than the norm.

The Review Panel were surprised not to find more collaborations, given the high level of competence and standing of both computer science research and applied mathematics in the UK. For example, there was no mention of collaborations with numerical mathematicians, even though the UK is home to some of the world's leading numerical analysts. The reasons for the absence of more collaborations are not obvious to the Review Panel. However, the Panel believe that the UK may be missing potentially important scientific opportunities, and therefore recommend that EPSRC devise strategies to increase the collaborations between the computer science/applied mathematics community and the researchers using HPC technology. There should be plenty of opportunities in areas such as numerical algorithms, visualization, scientific data management, and distributed computing.

4. HPC Resources

As explained in Section 2.1 above, high performance computing is essential to the future of science in the UK and around the world. To best realize scientific payoff for the funds invested, a diversity of resources and of controls are required. The Panel found a vibrant situation in the UK, spanning the spectrum from leading-edge facilities to university-based departmental and even research group facilities. This feature must be maintained as the UK goes forward.

4.1 NATIONAL LEADING-EDGE COMPUTATIONAL SYSTEMS

The UK has been very well served by the HPCx and CSAR systems, as their leading-edge systems, and by QCDOC as a special-purpose high-end machine. The HECToR program promises to build on this excellent tradition.

At the national level, leading-edge facilities are essential to support both capability computing (targeting substantial fractions of a tightly coupled resource at a single problem) and capacity computing (running many smaller, largely independent jobs simultaneously). A balance between these two is essential because both are fruitful scientifically and are often needed on the same scientific project—for example, generation of simulations may require capability computing, while the analysis of these simulations may best be done on capacity machines.

For both capability and capacity systems, there needs to be a diversity of architectures. One size does not fit all. One needs at least tightly coupled distributed-memory machines and large shared-memory systems. Arguments can sometimes also be made for special-purpose machines (e.g., the QCDOC machine in the UK). While all of these machines operate primarily in standard batch mode, they can be scheduled to allow interactive access and even co-scheduling with other facilities (e.g., for visualization). One must also adopt scheduling policies to allow significant code development. This requires rapid turnaround on smaller jobs, and may require setting aside special partitions of the machine during parts of the day, targeted exclusively for interactive code development.

The Panel found an unnatural divide between the ability of PPARC-funded researchers and EPSRC, NERC, and BBSRC researchers to access the current high-end facilities (HPCx and CSAR). The situation does not seem to be improving under HECToR. While we understand the historical reasons for this, our sense is that this hurts British science, and we urge that this divide be reexamined. We are also concerned that it is currently too cumbersome for users to casually try out the leading-edge systems. We have been told that there may be up to a six-month delay between submitting a request for such access and its granting. Since the overall machine time in such requests can be kept quite small, the impact on other users can also be quite small, and the Panel urge that an expedited mechanism be found to allow such small-scale experimental use. (Large facilities in the US can accommodate such requests within a day or two; and on the Earth Simulator in Japan, even a full-node job can be run within two days. The UK might examine how the US and Japanese systems achieve this goal while minimally impacting other users).

4.2 UNIVERSITY-BASED COMPUTATIONAL SYSTEMS

Below the leading-edge, there has been excellent progress in distributing less powerful computing facilities throughout universities. Much excellent science is carried out on these machines by, for example, the UKTC and UKCP groups at Cambridge, the ICC at Durham, the Mineral Physics group at UCL, UKAFF at Leicester, and many others. These systems also serve as development platforms for work that migrates to the national high-end facilities.

Local facilities have many advantages over national ones:

- They can be more nimble. For example, changes in research direction or approaches can be immediately attempted without formal justification.

- New casual users can be immediately accommodated.

- Attracting and training students is easier.

  - Local systems can be easily integrated into formal university-based courses involving other departments, strengthening the interdisciplinarity which the Panel feel is so desirable.

  - Students can learn systems administration.

  - Local facilities can be showcased in programs aimed at pre-college students, as is done so well at the ICC in Durham.

- Links to local storage and visualization resources are tighter and easier to implement.

- Pride of ownership increases political support throughout the nation for the larger HPC program.

Local facilities are the major route by which new users are attracted to HPC, which is so vital, as has already been pointed out.

The Panel found significant apprehension that with the termination of Science Research Investment Fund (SRIF) and Joint Infrastructure Fund (JIF) support, and under Full Economic Costing, it will be increasingly difficult to procure these local facilities in the future. It is critical that they be sustained and upgraded.

The hierarchy of computing resources is often captured in the notion of pyramid with a broad base (many more smaller systems) and a narrow top (a few leading-edge systems), as shown in Figure 1.[5] Scientific progress depends on ready access to the full spectrum of computing resources.

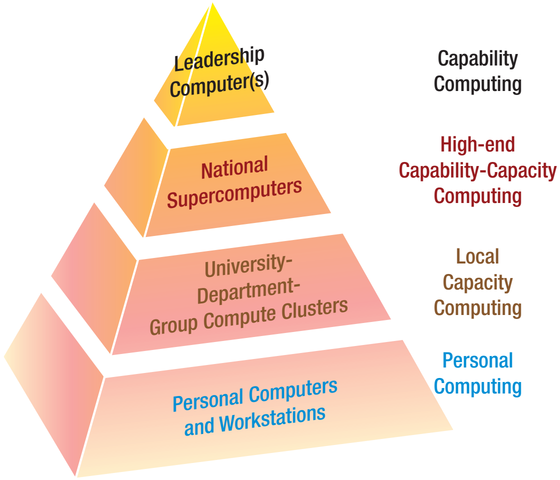

Figure 1. The spectrum of computing resources.

4.3 DATA RESOURCES

A majority of the computational science disciplines the Panel reviewed rely on numerical solutions of physics equations. Data is principally the output of these simulations. With the extraordinary increase in computing power, the amount of data output and the ability to store fine-grained data from trajectories or time series analysis of systems is beginning to produce enormous volume of data. This, by itself, would warrant a change in the way the computational science community handles data.

However, another revolution is changing the role of data in computation and has in some cases led to a shift in the data paradigm towards data-driven computing. This revolution was triggered primary by having data be the output of automated measurement, primarily in life sciences, astronomy, environmental science, climate studies, and computational particle physics. In this revolution, terabytes of data are already being produced, and soon it will be petabytes. These data, often spanning multiple dimensions and containing valuable information about functional systems, are not amenable to easy abstraction or modelling with equations of motion. Sometimes they may provide parameters and boundary conditions for physics-driven modelling.
It is mandatory for computational science to

[5] The need for a hardware hierarchy to support computational research was first articulated in a 1993 report to the US National Science Board entitled "From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing," frequently referred to as the "Branscomb report." The famous "Branscomb pyramid" was updated in January 2005 by the Subcommittee on Theory and Computation of the US Basic Energy Sciences Advisory Committee in "Opportunities for Discovery: Theory and Computation in Basic Energy Sciences."

recognize the importance of data management and analysis. The international community have already made inroads in this emerging discipline. During site visits to UK institutions, the Panel found that most of the research groups are aware of the need to focus on data, but do not have specific plans for, or a comprehensive understanding of, how to handle data flexibly and efficiently. The UKQCD consortium are the most advanced and articulated their data needs most clearly. They have initiated the International Lattice Data Grid (ILDG) and have developed Grid tools for semantic-based data access using metadata catalogues and XML schemata. UKQCD already have a middleware layer running, have successfully tested a prototype of their setup, and are now ready to go into production mode. UKTC have put the data generated from direct numerical simulations of turbulence and transition on a Web-based database, currently accessible worldwide, with ten registrations for use received between May and September 2005. The UKTC databases, under continuous development, make available statistical summary data (Moser format), complete flow field data, and associated Fortran code for I/O. The Integrative Biology Grid is arguably the most heterogeneous data collection, and its organizers are also aware of the data management problem, but are only in the early stages of planning. The CSED group at Daresbury, who provide a large computing resource, expressed the importance of planning for data needs, but believed that the e-Science programme is addressing this issue. The ICC in Durham plan to make their cosmology data publicly accessible but do not yet have a fixed concept on how to do it.

There are four general issues connected to data:

1. Data, whether from experimental high-throughput measurements or from large-scale simulations, need to be organised, stored, and disseminated. This requires establishing structured mechanisms of organising, databasing, and providing the tools for dissemination of the data. This task is highly discipline-specific, and context-specific infrastructure needs to be built on some common foundations.

2. Diverse sources of data, such as collections of distinct databases, need to be queryable in an interoperable manner, which requires metadata catalogues. This is a major area of research in computer science, and brokering of data and ontologies between databases requires planning and design. The outcome of such queries can sometimes be entirely new knowledge.

3. Often the size of the data is beyond simple visual analysis by humans. The data needs to be reduced to smaller dimensions and details, or the data needs to be mined using supervised and unsupervised machine learning tools. Often these tools warrant extensive statistical algorithms and significant computational time.

4. The most important task involves the integration of diverse data and legacy knowledge to produce new knowledge, either by combining data on an international level or by purely querying the data. Each discipline will have to develop specialized mechanisms to achieve this integration.

Based upon this analysis and our review of the research groups, the Panel recommend the following:

1. Each consortium begin to generate design plans for handling and analyzing data in their discipline. Large consortia should organise meetings to document data needs of different communities with a view to identifying new strategies that need to be adopted.

2. Specific funding and a proposal mechanism should be developed to create a workforce that develops common tools and infrastructures for data handling, organization, manipulation, and creation of user interfaces. The workforce should evaluate how much of the technology developed by the UKQCD consortium can be taken over for other consortia. The workforce should create a library of software for scientists dealing with large amounts of data.

3. Funding agencies announce targeted requests for applications to encourage interdisciplinary projects that involve data, computation, and visualization.

4.4 VISUALIZATION RESOURCES

While the reviewed research groups well understand that visualization is indispensable in computational science, the Panel observed that visualization in the UK lags behind international standards. A few groups showed us visualization labs, but they appeared to use conservative visualization devices and tools. An exception is the ICC at Durham, who showed us an impressive 3D movie of cosmology. The Panel are concerned that without an improvement in visualization sophistication (both hardware and software), hidden scientific treasures will increasingly lurk undiscovered in the massive data to be produced by the enhancement of HPC capability and capacity.

As an example of the state of the art, the Theory and Computer Simulation Centre and the Earth Simulator Centre in Japan use a two-stage visualization system: the first stage is a Cave virtual reality system, whereby overall global evolutionary behaviour is surveyed to identify interesting events; then, in the second stage, detailed analysis is made to explore these events more fully using advanced desktop visualization systems.

In the forthcoming leading-edge facilities in the UK such as HECToR, much more massive and valuable data will be produced. Therefore, the Panel recommends that the UK HPC community prepare immediately to establish a balance among leading-edge computing facilities, visualization technologies, and well-educated computational scientists.

4.5 USER SUPPORT

Examining user support was not explicitly in the mandate for this review. However, the Panel believe that effective user support is critical to productivity in HPC. Most computational science groups in a university have one or a few "go-to" people to whom the students or postdocs turn for answers to computing questions. Such support needs to be nurtured. The Panel have already argued for more support within individual university groups for people who supplement their applications expertise with software engineering or algorithm expertise.

But this is not enough. The research community also needs national, institutionalized user support, where the practitioners have a long-term career path, which is not the case in a typical university environment. This support must be provided by well-rounded professionals who can speak to the users as scientific peers, but who also have excellent understanding of algorithms, code optimization, and software engineering, as well as a mindset geared to helping others rather than a focus on independent scientific research. Such people can be a source of cross-fertilization between different research groups, recognizing that the expertise of one group can help another. The Panel were told that HECToR foresees supporting such people for a targeted portion of the total budget, which we applaud. CSED at Daresbury appears to play this role already, and the Panel often heard testimony of users' satisfaction with the support they received from Daresbury.

It is not essential that such a user support team be co-located with the leading-edge machines, but it helps. Many user issues require system responses, such as allocating large disk quotas for a restricted period, special scheduling for individual jobs, etc. It improves communication and reduces institutional suspicions if the requests for such system accommodations come from someone within the institution who represents the external user.

5. Conclusions and Recommendations

During the course of this review, the Panel observed a range in the quality of science delivered through the use of HPC in the UK but overall were very impressed. The Review Panel judged that in many of the consortia and research groups visited, the scientific results compared well with the highest international standards. With respect to computational resources, the Panel found a vibrant situation in the UK, spanning the spectrum from leading-edge facilities to university-based departmental and even research group facilities. The EPSRC's recognition of the importance of HPC and the research community's already strong demand for HECToR indicate that the UK research environment is well positioned for fostering further scientific progress through HPC. Thus the state of research using HPC in the UK is excellent, with researchers being engaged in scientific projects at the forefront of the challenges in their field, making internationally recognized, significant contributions, and having access to state-of-the-art platforms.

Research using high performance computing is a rapidly growing and dynamically changing field. The Panel therefore make several recommendations that will contribute to the continued improvement of the research environment for HPC in the UK:

1. Create a more balanced HPC infrastructure between computational technologies and intellectual resources.

The UK has a well thought-out investment strategy with HPCx and CSAR platforms in the past and HECToR in the near future. However, as high performance computing platforms become increasingly more difficult to use efficiently, the gap between users and non-users becomes larger. HPC platforms generally can be utilized more effectively if users have access to intellectual resources, for example experts in code optimization, parallel algorithms, etc., who can work with the researchers as consultants or collaborators. A relatively small group of such experts, for example associated with HECToR, can achieve high productivity gains if they are leveraged across multiple users. The productivity gains are obtained by more rapid insertion of new technology into applications codes and by realization of technology transfer between research groups.

2. Strengthen the computational infrastructure by:

a. systematically deploying leading-edge capability systems, large-scale capacity computing, and resources deployed widely at universities

The Panel believe that all elements of the computational infrastructure (the "pyramid" of Figure 1) are important for creating a healthy and productive HPC environment. In our review we saw examples of good and appropriate use of the different elements of the pyramid. This balanced approach should be continued in the future, with investments being made at all levels of the pyramid.

b. supporting and developing a state-of-the-art applications software infrastructure encompassing algorithms, data management and analysis, visualization, and best-practices software engineering.

In spite of some notable examples to the contrary, the Panel found that UK researchers using HPC are often not aware of or lack the resources to apply the latest results in numerical algorithms, data management and analysis technologies, and visualization. They are also often not aware of software engineering practices. The Panel recommend devising strategies to link computer science research in these areas more closely to the applications scientists, and facilitating innovative means for computer scientists and mathematicians to collaborate with researchers using HPC.

3. Develop human resources in HPC.

The future of CSE depends critically on the availability of highly specialized experts. Because of its interdisciplinary nature, research using HPC requires specialists who have received additional education and training beyond traditional academic disciplinary programs. While EPSRC has already taken promising initial steps such as the HEC studentships, the Panel encourages further initiatives in the UK that will lead to increasing support for the trained professionals and experts in CSE, as well as the development of formal programs for the education and training of computational scientists.

4. Bridge disciplines and build a computational science community by increasing interactions and fostering collaborations between disciplinary groups nationally and internationally.

The Panel notes that while there are several examples of excellent collaborations in individual disciplines, an all-embracing computational science community (in the sense of an "academic community") does not yet exist in the UK. The Panel recommends taking proactive steps towards the creation of such a computational science community by such means as encouraging and developing local workshops and national meetings, fostering support for a professional society in the field, and participating in leading international events such as the annual International Supercomputer Conference and the SCxy conference series.

6. Acknowledgements

The Panel are grateful to the Steering Group, in particular Harald Knobloch, for their excellent preparatory work and for their assistance throughout the review. The Panel are also very appreciative of all the logistical support provided by Linda Sayers. Hugh Pilcher-Clayton is acknowledged for briefing the Panel and answering questions about EPSRC. John Hules is acknowledged for preparing the background data and editing this report. Finally, the Panel would like to express their sincere appreciation to the universities visited, where both staff and students invested considerable efforts in ensuring that the visits were informative and enjoyable.

7. Appendices

7.1 STEERING GROUP

Professor Ron Perrott, Chair,
Queen's University Belfast
Professor John O'Reilly, EPSRC
Professor Ernst-Ludwig Winnacker, DFG
Dr. Horst Simon,
Lawrence Berkeley National Laboratory
Dr. Randal Richards, EPSRC
Dr. Christoph Schneider, DFG
Ms. Linda Sayers, EPSRC
Dr. Harald Knobloch, DFG

7.2 PANEL MEMBERSHIP

Dr. Horst Simon, Chair,
Lawrence Berkeley National Laboratory
Professor Hans-Peter Bunge,
Ludwig Maximilians University
Professor Roberto Car, Princeton University
Professor Charles Hirsch,
Vrije Universiteit Brussel
Dr. Karl Jansen,
J.v.Neumann-Institute for Computing,
DESY, Zeuthen
Professor Hans Peter Lüthi,
Swiss Federal Institute of Technology, Zurich
Professor Ralph Z. Roskies,
Pittsburgh Supercomputing Center
Professor Tetsuya Sato,
The Earth Simulator Center
Professor Shankar Subramaniam,
University of California, San Diego
Professor William M. Tang,
Princeton University

7.3 TERMS OF REFERENCE

The International Review Panel is requested to:

- report on the calibre, standing and research potential in research using High Performance Computing in the UK
- discuss the potential impact of university-based research using High Performance Computing on the UK's knowledge economy
- provide comparisons with international research using High Performance Computing
- make recommendations on future actions and/or priorities

7.4 RESEARCH GROUPS REVIEWED

Centre for Computational Science
(CCS)/RealityGrid
ChemReact Computing Consortium
CLRC Computational Science and Engineering
Department (CSED) at Daresbury Laboratory
HPC Materials Chemistry Consortium
Institute for Computational Cosmology (ICC),
University of Durham
Integrative Biology at the University of Oxford
Mineral Physics at University College
London (UCL)
Multiphoton, Electron Collision, and Bose
Einstein Condensates (MECBEC) HPC
Consortium
NERC Centres for Atmospheric Science (NCAS)
UK Applied Aerodynamics Consortium (UKAA)
UK Astrophysical Fluids Facility (UKAFF)
UK Car-Parinello Consortium (UKCP)
UK Turbulence Consortium (UKTC)
UKAEA Fusion Research at Culham
Science Centre
UKQCD Collaboration

7.5 REVIEW SHEET

Research Institution:

Reviewer:

I. International standing and research potential

Please comment on the following aspects:

1. Evaluation of Research in comparison to the international level
2. Impact of research activities to international progress in the scientific field
3. Visibility on an international level
4. Publication activities and participation in conferences
5. Research progress during the last 3–5 years
6. Impact on the UK's knowledge economy

II. Resources and instrumentation

Please comment on the following aspects:

1. Computer resources (e.g., access to computing resources, availability of architectures, allocation of resources, etc.)
2. Human resources
3. Level of funding/success in fundraising

III. Open discussion on strategic issues for promoting science using high performance computing in the UK

IV. Recommendations on future actions and/or priorities

V. Other comments

VI. Conclusion