

HPCx Quarterly Report

April – June 2005

1 Introduction

This report covers the period from 1 April 2005 at 0800 to 1 July 2005 at 0800.

The next section summarises the main points of the service for this quarter. Section 3 gives details of the usage of the service, including failures, serviceability, CPU usage, helpdesk statistics and service quality tokens. A summary table of the key performance metrics is given in the final section. The Appendices define the incident severity levels and list the current HPCx projects.

2 Executive Summary

- The Annual Plan, which was approved by STAC (Scientific and Technical Advisory Committee), should be finalised after August's Oversight Committee meeting.
- This quarter has continued to build on the successful start to 2005. Utilisation was higher than it had been for more than a year and both April and June delivered record numbers of Allocation Units.
- The system's reliability has continued to be remarkable --- there have been no hardware failures during 2005.
- EPCC successfully hosted the joint ScicomP/SP-XXL meeting from 30 May to 3 June. There were well over 100 attendees and presentations were given by representatives of the Software Engineering and Terascaling teams as well as from the HPCx user community.
- We held a Life Sciences workshop at RAL on 18 May that was attended by representatives of most of the Life Sciences consortia on HPCx.
- As the Integrative Biology consortium asked us to run courses at RAL, we followed the workshop with two co-located HPCx courses on 19-20 May. We have arranged to run 12 days of training during the next quarter, when the MSc is quiet, including 4 days at Leeds.

- The launch of the Advanced Computing Facility in Edinburgh was an opportunity to promote both HPCx and the use of HPC for scientific research to a more general audience. We are also adding further scientific highlights to the HPCx web site.
- The Terascaling team have continued the programme of consortia visits to increase our understanding of the scientific drivers and requirements. Nine such visits have now taken place.
- Two of the top 10 most heavily used codes on HPCx, DL_POLY and CASINO, achieved Gold star ratings during this quarter.
- Initial investigations of IBM's new major Fortran compiler release, xlf version 9, have been positive. We are now making this available for a period of user testing before making it the default production version.
- The Software Engineering team have been investigating RDMA, which provides significant improvements to communication performance for large messages. A technical report has been written which explains how to activate the RDMA feature and benchmarks its impact.
- Preparations are underway for the SPICE project's experiment at SC2005. This is a more ambitious follow-on to Teragyroid and will require improvements both to our port-forwarding software and to MPICH-G2.

3 Usage Statistics

3.1 Availability

3.1.1 Failures

The monthly numbers of incidents and failures (SEV 1 incidents) are shown in the table below:

| | <i>April</i> | <i>May</i> | <i>June</i> |
|-----------|--------------|------------|-------------|
| Incidents | 17 | 11 | 17 |
| Failures | 0 | 3 | 1 |

The following tables give more details on the attribution of the failures:

April

There were no failures in April

May

| <i>Failure</i> | <i>Site</i> | <i>IBM</i> | <i>External</i> | <i>Reason</i> |
|----------------|-------------|------------|-----------------|--------------------|
| 05.069 | 100% | 0% | 0% | UPS failure |
| 05.074 | 100% | 0% | 0% | Procedural failure |
| 05.075 | 100% | 0% | 0% | Firewall reload |

June

| <i>Failure</i> | <i>Site</i> | <i>IBM</i> | <i>External</i> | <i>Reason</i> |
|----------------|-------------|------------|-----------------|--------------------------|
| 05.091 | 0% | 0% | 100% | External network failure |

3.1.2 Performance Statistics

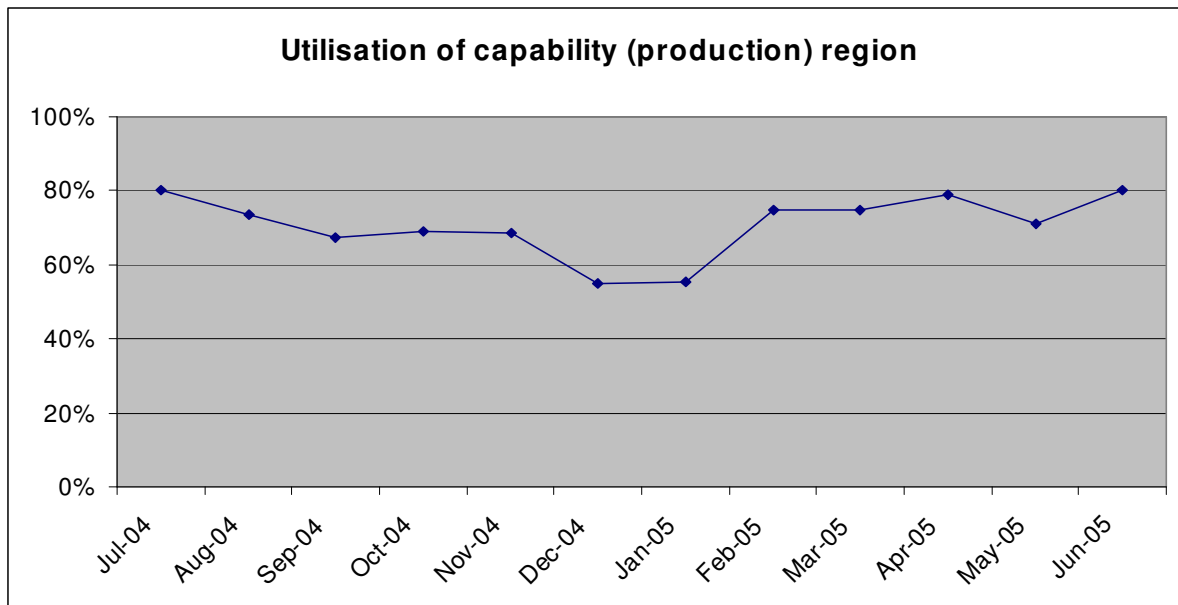
This section uses the definitions agreed in Schedule 7, ie,

- $MTBF = (24 \times 30.5) / (\text{number of failures in month})$
- $\text{Serviceability (\%)} = 100 \times (\text{WCT} - \text{SDT} - \text{UDT}) / (\text{WCT} - \text{SDT})$

| <i>Attribution</i> | <i>Metric</i> | <i>April</i> | <i>May</i> | <i>June</i> | <i>Quarterly</i> |
|--------------------|----------------|--------------|------------|-------------|------------------|
| IBM | Failures | 0 | 0 | 0 | 0 |
| | MTBF | ∞ | ∞ | ∞ | ∞ |
| | Serviceability | 100.0% | 100.0% | 100.0% | 100.0% |
| Site | Failures | 0 | 3 | 0 | 3 |
| | MTBF | ∞ | 244 | ∞ | 732.0 |
| | Serviceability | 100.0% | 98.3% | 100.0% | 99.4% |
| External | Failures | 0 | 0 | 1 | 1 |
| | MTBF | ∞ | ∞ | 732 | 2196.0 |
| | Serviceability | 100.0% | 100.0% | 95.2% | 98.4% |
| Total | Failures | 0 | 3 | 1 | 4 |
| | MTBF | ∞ | 244 | 732 | 549.0 |
| | Serviceability | 100.0% | 98.3% | 95.2% | 97.8% |

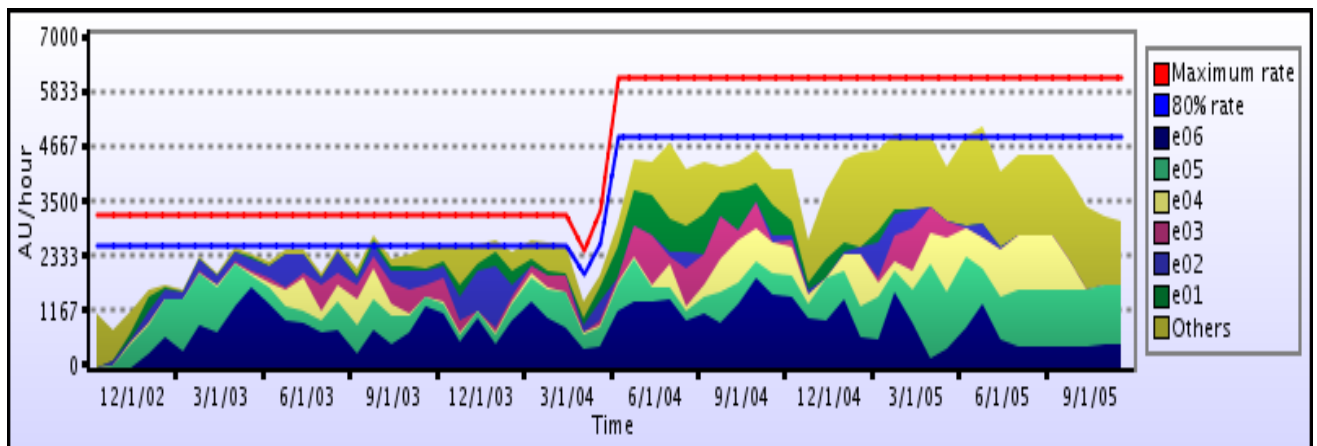
3.2 Capability Utilisation

With overall utilisation for the quarter at nearly 79%, the system is currently extremely busy. Capability utilisation for the quarter was 26.4% of the total.



3.3 Capacity Planning

Predicted Utilisation



The graph above shows the utilisation since the start of the project and the projected utilisation until the end of 2005. The scale on the y-axis is AUs per hour, where the peak that HPCx Phase 1 could currently deliver is around 3240 AUs per hour, and Phase 2 6188 AUs per hour (the upper red line in the graph). The lower line (in blue) corresponds to the more practicable 80% level.

The graph assumes that each project will use its remaining allocation pro rata with its usage profile from the SAF, which is often simply that on the original application form.

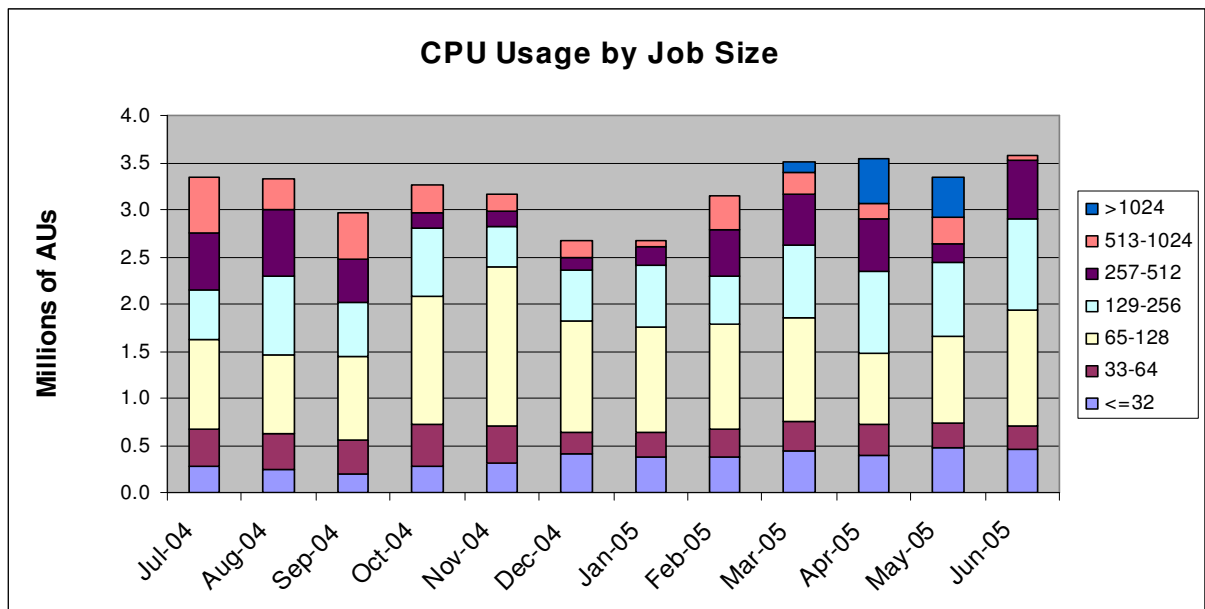
As can be seen from the graph, the projected utilisation of the existing groups falls just within the 80% limit. A number of allocations which start later in the year are not included in the graph.

Numbers of Research Consortia

There are currently 35 research consortia using the HPCx system. Three other projects have now been closed.

In addition, there are two externally funded projects.

3.4 CPU Usage by Job Size



3.5 AU Usage by Consortium

The PIs and titles for the various consortia are listed in Appendix B.

| <i>Consortium</i> | <i>April</i> | <i>May</i> | <i>June</i> | <i>Quarterly</i> | <i>%age</i> |
|--------------------|--------------|------------|-------------|------------------|-------------|
| e02 | 198366 | 3681 | 89694 | 291741 | 2.8% |
| e03 | 514485 | 355479 | 154 | 870118 | 8.3% |
| e04 | 257871 | 728655 | 454260 | 1440786 | 13.8% |
| e05 | 722366 | 1058046 | 866897 | 2647309 | 25.3% |
| e06 | 743087 | 234841 | 696616 | 1674544 | 16.0% |
| e07 | 523 | 1930 | 125 | 2578 | 0.0% |
| e08 | 43104 | 21714 | 9495 | 74313 | 0.7% |
| e11 | 64332 | 106369 | 6201 | 176902 | 1.7% |
| e14 | 1943 | 8392 | 10025 | 20360 | 0.2% |
| e17 | | 746 | 1179 | 1925 | 0.0% |
| e18 | 10001 | 1142 | 4657 | 15800 | 0.2% |
| e19 | 1471 | 50008 | 21376 | 72855 | 0.7% |
| e20 | 78076 | 36458 | 110898 | 225432 | 2.2% |
| e21 | 35 | 8934 | 1761 | 10730 | 0.1% |
| e24 | 1925 | 65 | 0 | 1990 | 0.0% |
| e25 | 0 | 64 | 7 | 71 | 0.0% |
| e29 | | 118 | 369 | 487 | 0.0% |
| z09 | 10134 | 1305 | 70 | 11509 | 0.1% |
| <i>EPSRC Total</i> | 2647718 | 2617947 | 2273784 | 7539449 | 72.0% |

| | | | | | |
|-------------------|--------|--------|---------|---------|-------|
| n01 | 175936 | 10877 | 322025 | 508838 | 4.9% |
| n02 | 115253 | 194855 | 291549 | 601657 | 5.7% |
| n03 | 127021 | 192158 | 331780 | 650959 | 6.2% |
| n04 | 75687 | 36226 | 165542 | 277455 | 2.7% |
| <i>NERC Total</i> | 493897 | 434115 | 1110896 | 2038908 | 19.5% |

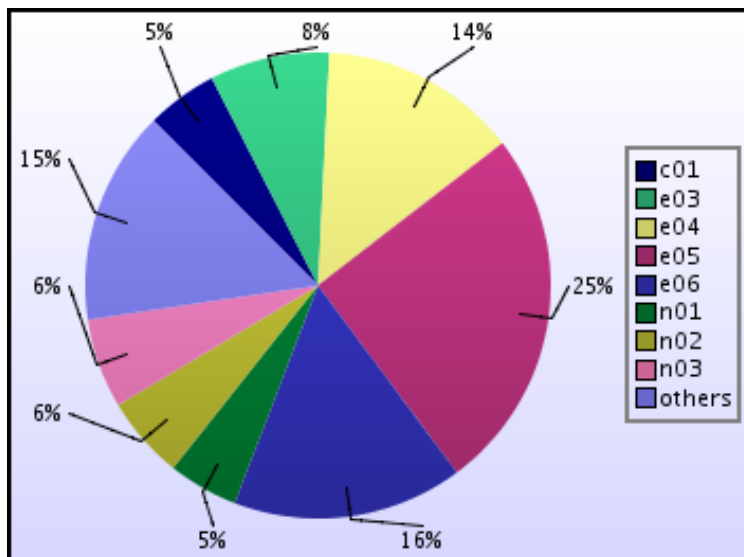
| | | | | | |
|--------------------|------|------|-------|-------|------|
| p01 | 1703 | 5655 | 35234 | 42592 | 0.4% |
| <i>PPARC Total</i> | 1703 | 5655 | 35234 | 42592 | 0.4% |

| | | | | | |
|--------------------|--------|--------|-------|--------|------|
| c01 | 237665 | 188653 | 92764 | 519082 | 5.0% |
| <i>CCLRC Total</i> | 237665 | 188653 | 92764 | 519082 | 5.0% |

| | | | | | |
|--------------------|-------|-------|------|-------|------|
| b02 | 1325 | 6738 | | 8063 | 0.1% |
| b03 | 2383 | | 5530 | 7913 | 0.1% |
| b05 | 27439 | 29444 | 3124 | 60007 | 0.6% |
| <i>BBSRC Total</i> | 31146 | 36182 | 8653 | 75981 | 0.7% |

| | | | | | |
|-----------------------|--------|-------|-------|--------|------|
| x01 | 38 | 4901 | 17900 | 22839 | 0.2% |
| x02 | 128869 | 41649 | | 170518 | 1.6% |
| <i>External Total</i> | 128907 | 46550 | 17900 | 193357 | 1.8% |

| | | | | | |
|-------------------|------|-------|-------|-------|------|
| z001 | 5950 | 16791 | 26260 | 49001 | 0.5% |
| z002 | 16 | 9 | 13 | 38 | 0.0% |
| z004 | 0 | 24 | 698 | 722 | 0.0% |
| z05 | | 4505 | | 4505 | 0.0% |
| z06 | 1521 | 1648 | 0 | 3169 | 0.0% |
| <i>HPCx Total</i> | 7487 | 22976 | 26972 | 57435 | 0.5% |



3.5.1 Discounts

There are now a number of user codes that have qualified for capability discounts. The following table shows the discounts that were awarded during the last quarter.

| <i>Consortium</i> | <i>AUs Used</i> | <i>AUs Charged</i> | <i>Discount</i> |
|-------------------|-----------------|--------------------|-----------------|
| b05 | 70086 | 60006 | 10079 |
| e05 | 2647519 | 2647308 | 211 |
| e06 | 1685855 | 1674544 | 11311 |

3.6 Helpdesk

3.6.1 Classifications

| <i>Category</i> | <i>Number</i> | <i>% of all</i> |
|-----------------|---------------|-----------------|
| Administrative | 75 | 35.4 |
| Technical | 122 | 57.5 |
| In-depth | 13 | 6.1 |
| PMR | 2 | 0.9 |
| TOTAL | 212 | 100.0 |

| <i>Service Area</i> | <i>Number</i> | <i>% of all</i> |
|---------------------|---------------|-----------------|
| Phase 2 platform | 184 | 86.8 |
| Website | 12 | 5.7 |
| Other/general | 16 | 7.5 |
| TOTAL | 212 | 100.0 |

3.6.2 Performance

| <i>All non-indepth queries</i> | <i>Number</i> | <i>%</i> | <i>Target</i> |
|--------------------------------|---------------|----------|---------------|
| Finished within 24 Hours | 154 | 78.2 | 75% |
| Finished within 72 Hours | 194 | 98.5 | 97% |
| Finished after 72 Hours | 3 | 1.5 | |

| <i>Administrative queries</i> | <i>Number</i> | <i>%</i> | <i>Target</i> |
|-------------------------------|---------------|----------|---------------|
| Finished within 48 Hours | 74 | 98.7 | 97% |
| Finished after 48 Hours | 1 | 1.3 | |

3.6.3 Experts Handling Queries

| <i>Expert</i> | <i>Admin</i> | <i>Technical</i> | <i>In-Depth</i> | <i>PMR</i> |
|---------------|--------------|------------------|-----------------|------------|
| epcc.ed.ac.uk | 65 | 49 | 5 | 0 |
| dl.ac.uk | 3 | 14 | 2 | 0 |
| Sysadm | 7 | 59 | 6 | 1 |
| Other people | 0 | 0 | 0 | 1 |

3.7 Service Quality Tokens

| <i>Date</i> | <i>Person</i> | <i>Value</i> | <i>Comment</i> | <i>Status</i> |
|-----------------------------|-----------------------------------------|--------------|--------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| May 26, 2005 10:00:15 PM | Dr. Phil P Lindan | **** | Good response from helpdesk and no big problems, thank you | |
| May 17, 2005 12:18:35 PM | Dr. Michael J Nolan | ... | my jobs seem to be held up for no apparent reason; waiting 4 days for one job to run | This resulted from a programming error, which was resolved through the helpdesk. |

4 Support

4.1 Applications Support (*Dr David Henty*)

4.1.1 Documentation

This has been an extremely stable quarter in terms of the hardware and software configuration of the machine, so no significant changes to the documentation have been required. Information on a specific upgrade to the parallel operating environment, introduced by SP12, was considered too detailed for the general service guide but was disseminated via the technical report HPCxTR0505.

4.1.2 Technical Reports

A total of four reports were planned for Q2 in the following areas:

- a) Parallel IO Techniques for Capability Computing
- b) Parallel Performance Analysis Tools
- c) Comparison of Parallel Debuggers on HPCx
- d) Achieving Capability Computing for CASTEP

where the first three are as stated in the annual plan, and **d)** was moved from Q1 as explained in the previous quarterly report.

We have produced the following four reports this quarter:

- **HPCxTR0504**: *I/O Performance on the HPCx Phase 2 System*, Michael Holden, Elena Breitmoser, Joachim Hein.
- **HPCxTR0505**: *Improved MPI with RDMA*, Alan Gray, Joachim Hein, Stephen Booth.
- **HPCxTR0506**: *DDT and Totalview on HPCx*, Mike Ashworth, Andrew Sunderland
- **HPCxTR0507**: *Towards Capability Computing with CASTEP*, Martin Plummer, Keith Refson.

Reports **04**, **06** and **07** correspond directly to the titles **a)**, **c)** and **d)** respectively. The technical reports are usually not code-specific as they are targeted at the general HPCx community. However, CASTEP is such a widely-used code that it seemed appropriate to devote one report, HPCxTR0507, to its use in a capability setting.

We chose to postpone the production of report **b)** in response to the appearance of an important new MPI performance feature in Service Pack 12, when it was

decided that it was essential to inform users immediately of the potential benefits of this upgrade. Report **05** was therefore written in its place, following detailed investigations done by the Software Engineering team, and revealed bandwidth improvements of between 25% and 100% for large messages.

There are a total of twelve reports due for this year. The profile given in the Annual Plan for 2005 had seven reports to be produced by the end of Q2, so we are currently on target.

4.1.3 Training

In Q2 of 2005 we ran the following three courses:

- **Daresbury Laboratory, 20 April:** *DL_POLY on HPCx.*
- **Rutherford Appleton Laboratory, 19 May:** *Optimisation Techniques for the POWER4 Processor.*
- **Rutherford Appleton Laboratory, 20 May:** *Improved Performance Scaling on HPCx.*

The location and timing of the courses held at RAL were chosen in response to specific requests from the Integrated Biology consortium, although they were publicised and made available to all users.

Statistics are summarised below alongside annual targets (where appropriate):

| <i>Metric</i> | <i>Total</i> | <i>Target</i> |
|---------------------------------|--------------|---------------|
| Course days | 9 | 30 |
| Number of courses | 5 | 12 |
| Different locations | 3 | 4 |
| Student-days for HPCx users | 136 | |
| Student-days for HPCx staff | 4 | |
| Student-days available for HPCx | 214 | 600 |

The number of training days in the first half of the year is below the pro-rata target of 15 days so far. However, the HPCx training schedule does not have a flat profile over the year. In particular, we do not run many courses at the start of the year due to the training commitments of the MSc in HPC, and this year the added overhead of running ScicomP in May/June made it harder for the AS team to provide a large number of courses in Q2.

The following courses are all confirmed and advertised:

- **EPCC / Access Grid, 20 July:** *Task Farming on HPCx*
- **University of Leeds, 26 July:** *Using the HPCx Service*

- **University of Leeds, 27 – 29 July:** *Message-Passing Programming using MPI.*
- **EPCC, 22 August:** *Fundamental Concepts of HPC*
- **EPCC, 23 – 24 August:** *Object-Oriented Programming for HPC*
- **EPCC, 29 August – 1 September:** *Practical Software Development*

With this schedule of courses we will be back on target by the end of Q3.

4.1.4 Workshops and Conferences

The HPCx Life Sciences Workshop was held on 18 May at Rutherford Appleton Laboratory, and was arranged to coincide with a consortium visit to the Integrated Biology group and with training courses held at RAL on the following two days. It was attended by 15 users, representing most of the Life Sciences consortia on HPCx. The format comprised brief presentations by HPCx staff, followed by talks from the attendees. However, the main aim was to promote discussion which proved extremely successful. A number of actions were taken from this meeting which have subsequently been followed up. One of the major outcomes was that several users expressed a strong interest in task-farming approaches on HPCx, and as a result the training plan has been modified with a short course now planned in Q3.

Scicomp11 and SP-XXL

Perhaps the major Applications Support activity this quarter was organising ScicomP11, the 11th meeting of the IBM System Scientific Computing User Group, and the summer meeting of SP-XXL, the organisation for system administrators of large IBM machines. Both these meetings were held in Edinburgh during the week beginning Monday 30 May. The budget for the two meetings was £15K which was covered by registration fees, sponsorship from IBM and contributions from EPCC and the University of Edinburgh. The meetings were extremely successful with very strong programmes of international speakers and an attendance of some 120 delegates.

HPCx was strongly promoted at the event including poster displays in the foyer and inclusion of the most recent edition of Capability Computing in the registration pack. The HPCx systems team was heavily involved in the SP-XXL meeting, and Stephen Booth and Ian Bush gave talks at ScicomP based on work done by the Software Engineering and Terascaling teams respectively.

Full details of the meeting including ScicomP presentations can be found at <http://www.epcc.ed.ac.uk/scicomp/>.

4.1.5 Newsletter

The fifth issue of Capability Computing, *Exploring Nature and Technology*, was produced in May as planned, and distributed at ScicomP11 in addition to the usual physical mailshot of over 3600 copies.

4.1.6 Packages

We continue to install software under the package account system, with some 50 software packages now available.

4.2 Outreach Activities (*Dr Richard Blake*)

4.2.1 Outreach to Lifesciences

Molecular Dynamics Codes

We have produced a technical note on LAMMPS that has been sent to the relevant consortium members. The work has been reported in more detail in the TeraScaling report.

Integrative Biology Project

Michael Holden has ported the Cardiac Arrhythmia Research Package (CARP) to HPCx, for the Integrative Biology project. The code has been developed to carry out large scale cardiac simulations. The code scales reasonably to 40 processors, however currently fails in PETSC for larger processors counts. This is being investigated. Current effort is focused on understand the communication overhead, to consider possible performance improvements.

Quantum Directed Virtual Evolution (b05)

In the previous quarterly report we noted that the taskfarming protocol had developed to the point where the codes and workflow were stable and robust enough to be turned over to the Durrant group for production runs. Since the last report, a new set of Density Functional Theory grids have been developed for GAMESS-UK. The new grids improve the consistency of the quadrature accuracy across the periodic table. In some cases the improved accuracy also reduces the number of steps required in a given geometry optimisation. As the molecules in the QDVE project are known to suffer from problems with quadrature accuracy, the new grids have been incorporated into the taskfarming binary and have been made available to Marcus Durrant.

Enzyme Catalysis (b02)

Extensive tests continue on the Nudged Elastic Band (NEB) and Replica Path (RP) methods in CHARMM. We have performed some parallel tests and the code currently shows reasonable speedup up to 600 processors. The main focus has been on the stability of the optimisers as it is clear now that the feasibility of these methods depends more on the optimisation technology than on parallel efficiency.

We have set up a number of test systems, ranging from simple reactions to the enzyme catalysed chorismate/prephenate rearrangement. The conclusions so far favour the NEB method over RP. This appears to be related to the angular spring forces that are employed to ensure reaction paths have small curvature in configuration space. These terms do not occur in the NEB, which instead uses a force projection method. We have shown that recent developments by the groups at NIH and MIT to enable the Adopted Basis Newton Raphson (ABNR) method to be used with NEB do indeed improve convergence but only in the later stages of the pathway optimisation. The alternative (steepest descent, SD) minimiser needs to be restarted with an extended step length in order to avoid stagnation of the optimisation. Paul Sherwood is planning to visit NIH during the summer and working on the SD minimiser will be one of the targets of this trip.

A copy of the source code incorporating the locality adjustments has been provided to the Mulholland group (b02) who have started some experiments locally prior to production runs on HPCx. In addition, we have provided data from our runs on the Chorismate/Prephenate system

Modelling of the Human Retina

A two-dimensional parallel implementation has been developed and is in the final stages of testing. Larger scale datasets are being developed for larger numbers of processors.

4.2.2 Other Outreach Activities

Operational Research

Work has continued with academics from the Operational Research group at Edinburgh University, to develop a proposal for HPCx time. The work will involve a world-breaking calculation on HPCx, with applications in financial modelling.

Industrial Outreach

Astra Zeneca and Daresbury Laboratory are on the cusp of signing a three year collaboration to design improved drug formulations. The project includes a £50K budget for access to HPCx.

Commercial arrangements are being discussed with Fluent and Abaqus.

A calling programme for up to 20 customers is being put in place with visits likely to start at the end of August.

4.2.3 Improve Public Awareness

Edinburgh's Advanced Computing Facility Opening

Edinburgh's Advanced Computing Facility was opened on the 1st of July by the Chancellor of the University, The Duke of Edinburgh. Guests were from across the world and many had non HPC backgrounds.

We developed 5 scientific posters, describing the nature of the science being carried out on HPC platforms for a non scientific audience. These covered areas such as drug discovery, nuclear fusion, new material development, mathematical modelling and systems biology. Of these, three are directly related to current (and future) HPCx consortia.

HPCx was also promoted directly through a specific postcard, the newsletter and a display of a POWER4 Multi Chip Module.

Scientific Highlights

Irina Nazarova has begun to collate scientific highlights from groups such as Cullham, Integrative Biology and OCCAM. This will add to the existing scientific highlights on the HPCx web site.

4.3 Terascaling Applications (*Dr Martyn Guest*)

4.3.1 Capability Science

Consortium Visits

We have continued the process of visiting PIs and leading users to discuss their requirements and scientific drivers. The following consortium visits have taken place:

- Patrick Briddon visited HPCx at Daresbury Laboratory for 2 days, 14th-15th April.
- Martyn Guest, Paul Sherwood and Andrew Sunderland visited Ken Taylor and Mike Finnis, Queens University, Belfast, on 27th April.
- Lorna Smith and Mike Ashworth visited the OCCAM group, NOC Southampton, on the 11th of May.
- Gavin Pringle and Kenton D'Mellow visited the e05 consortium at the RI the 13th of May.
- Lorna Smith, David Henty and Paul Sherwood visited the Integrative Biology consortium at Oxford on the 18th of May.

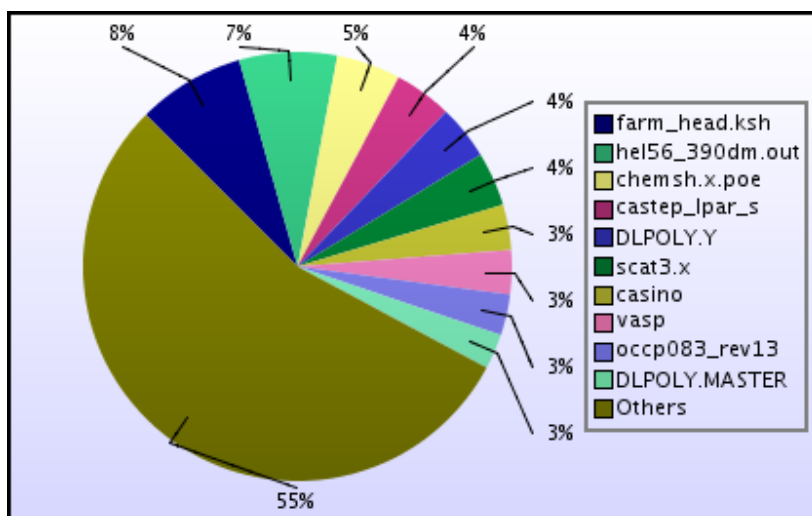
Capability Incentives

The CASINO and DL_POLY codes gained gold star ratings in this quarter.

A full list of codes which qualify for capability incentives can be found at:

<http://www.hpcx.ac.uk/services/policies/starcodes.html>

The figure below shows the percentage code utilisation for this quarter. It is clear from these figures that SIESTA (farm_head.ksh), CASTEP, VASP, HELIUM, DL_POLY, GAMESS-UK and the OCCAM code continue to be major users of the system. CASINO and ChemShell have also been used extensively in this quarter.



Percentage Code Utilisation for April - July 2005.

4.3.2 Quantum Chemistry Codes

ChemReact Consortium

- Andrew Sunderland has designed an ensemble processing harness for Stuart Althorpe's and Aditya Panda's scattering code to enable several MPI groups to be run within a job with different input and output datasets. In practice, this allows multiple parallel jobs to be run within the same LoadLeveler script, thereby increasing greatly throughput and productivity for users of this code on HPCx.

DL_POLY

- A new version of DL_POLY (3.04) has been installed and tested. This new release is some 20% faster than the previous benchmarked version (3.01), and has now achieved a gold star status on HPCx.

GAMESS-UK

- Ian Bush has written a Fortran module that allows straightforward access to shared memory segments and this has been introduced into GAMESS-UK. This will allow much larger problems to be addressed, and further should be useful in many applications where large replicated objects are required. This has required investigating the usage of features from the new Fortran standard, Fortran 2003, and any implications for their use in scientific application development. In particular the standardised method for interoperability between C and Fortran has been used to provide facilities for the Fortran developer to use the IPC facilities within standard Unix distributions for storing large data structures within shared memory segments. Further synchronisation facilities are provided, which allow coherent updates on data held in the segments by providing a critical region facility implemented using semaphores. The module may handle arbitrary numbers

of segments, shared between any number of processors in a node, and allows the Fortran developer to access them in a natural way. The performance implications of this are still being investigated, but initial results indicate that for large-scale calculations the overhead due to the critical regions is acceptably small (at the 10% or less level) due to the wider range of system sizes possible.

MOLPRO

- Andrew Sunderland has continued to investigate the performance attributes of the MOLPRO code on both HPCx and the Altix 3700 at CSAR. There appears to be little difference in the time to solution for the MRDCI benchmark on both machines – a visit to Peter Knowles and Nick Wilson at Cardiff is planned for early July to discuss next steps.

4.3.3 Computational Materials Codes

AIMPRO

- Ian Bush has started to investigate load balancing issues within AIMPRO. Profiling indicates the code spends excessive time in MPI_PROBE, but the reason behind this is not clear. This is being investigated in collaboration with the code's author, Patrick Briddon of Newcastle University.

CASTEP

- Martin Plummer has provided support to Dan Wilson (working with Richard Catlow, RI) over problems and optimisation with restarting very large Castep jobs. This work led to a general improvement in the memory management of restart jobs and also to Mr Wilson being able to complete his PhD thesis. Martin Plummer also worked with Paul Madden and research students over fast and efficient implementation of the new Wannier function module. Originally written by P Hasnip, this module required careful testing, verification of results and resolution of hidden MPI bugs as well as compatibility checks with the current version of Castep. The 'final' version is 4-5 times faster than the older 'legacy' Wannier function module and utilises memory much more efficiently, thus making realistic calculations on Wannier functions possible to the great satisfaction of Prof. Madden and his students.
- The following technical report has been published:

HPCxTR0507: *Towards Capability Computing with CASTEP*, Martin Plummer, Keith Refson.

SIESTA

- Joachim Hein has looked at the portability of this code to other platforms, such as the CSAR system and Edinburgh's Blue Gene. He has encountered problems within ScaLAPACK, with a lack of convergence on some platforms. He is currently investigating further.

VASP

- Gavin Pringle and Kenton D'Mellow visited the e05 consortium on the 13th May, following a request for a visit from HPCx to discuss this code. Kenton gave a short presentation on the current scaling and performance of VASP. He is now investigating the collective communications associated with this code, focusing on the MPI_ALLtoALL and MPI_ALLtoALLv operations.

4.3.4 Atomic & Molecular Physics

PRMAT

- Andrew Sunderland has adapted PRMAT to allow users to define a new 'uniform' energy mesh (in addition to the 'automatic' mesh definition). This allows users to sum collision strengths from different calculations with different partial waves. The new code has been made available to users.

4.3.5 Engineering Codes

UKAAC

- Andrew Sunderland has been running benchmark cases of Chris Allen's ROTORMGMBP code and Ken Badcock's PMB Code on HPCx (and other platforms). The results of this work have been presented at both an invited presentation and a technical presentation at Parallel CFD '05 at the University of Maryland.
- *Parallel Simulation of Lifting Rotor Blades*, C.B. Allen (Invited Speaker), Proceedings of Parallel CFD '05.
- *Parallel Performance of a UKAAC Helicopter on HPCx and Other Large-Scale Facilities*, A.G. Sunderland, D.R. Emerson and C.B. Allen, Proceedings of Parallel CFD '05.

4.3.6 Physics Codes

CENTORI

- Joachim Hein has been working closely with the group from Culham to use an alternative parallelisation strategy to improve the scaling of the code. Peter Knight visited EPCC on 10th-13th May for consultancy and discussion with Joachim. During this visit they implemented a mechanism to calculate the derivatives using finite difference instead of Fourier transforms. An initial investigation has been carried out to assess the performance and accuracy of this approach.

4.3.7 Environmental Codes

OCCAM

- Lorna Smith and Mike Ashworth visited the OCCAM group on the 11th of May. As part of a drive to understand the performance of key application codes on new architectures, Fiona Reid has compiled the OCCAM code on both HPCx and the Blue Gene system in Edinburgh. On HPCx the code has been run successfully on 62, 115, 128 and 256 processors. Compiling the code on Blue Gene required some minor alterations to the source code. Initial attempts to run the code failed because the executable required 640Mbytes of memory (Blue Gene only has 512Mbytes). Communication with the OCCAM developers has allowed the array sizes to be reduced in order to create a smaller executable which can run on 115, 128 and 256 processors. Attempts to run the code have resulted in an MPI error. This is currently being investigated in conjunction with the developers.

4.3.8 Life Science Codes

Lorna Smith and Paul Sherwood presented performance figures for a range of molecular dynamics codes at a visit to the system's biology consortium on the 18th of May.

LAMMPS

- Fiona Reid has written a LAMMPS benchmarking and profiling study. This has been sent to Peter Coveney and Chris Greenwell at UCL. Assuming there are no confidentiality issues the intention is that this report will be made available as an HPCx technical report. The report includes the results of both application supplied and real user supplied benchmarks ranging from 32000 to over a million atoms. These benchmarks have been run on both HPCx and CSAR. The code has also been profiled on the HPCx system using MPITrace and Vampir.

- The LAMMPS developers have implemented the MPI bug fixes suggested to prevent the code deadlocking on HPCx, see the March 31st 2005 entry at:

<http://www.cs.sandia.gov/~sjplimp/lammps/bug.17Jan05.html>

4.3.9 Compilers

The HPCx Terascaling Team has been testing IBM's latest, major Fortran compiler release, xlf version 9. No problems have been found and three codes (CENTORI, PCHAN and SIESTA) show a significant performance improvement. We plan to make the new compiler available for a period of user testing prior to its release into the production service.

4.3.10 Libraries

Eigensolvers

- Following a request by Stan Scott, Andrew Sunderland is investigating multi-threaded eigensolver performance within a node.

4.3.11 New Application Development

OpenMP

- Performance issues were identified with a user's OpenMP code. As a result the performance of a PARALLEL region encapsulating a DO versus a PARALLEL DO was investigated. The benchmarking results from May 2005 (see HPCxTR0411) suggested that the overheads associated with these directives should be of the same order of magnitude. However, re-running the EPCC Micro-benchmarks highlighted a serious performance degradation. Mark Bull contacted the compiler developers regarding this and an additional environment setting resolves the problem. This information was communicated to users in HPCx User Mailing 69 and has also been added to the User Guide.

4.3.12 Presentations

Applications Performance on the HPS, M. Ashworth, I.J. Bush, M.F. Guest, J. Hein, M. Plummer, A.G. Sunderland, ScicomP 11, 2nd June 2005, Edinburgh.

4.4 Software Engineering (*Dr Stephen Booth*)

4.4.1 General Terascaling and Optimisation Techniques

In the last quarter the IBM parallel operating environment on HPCx has been updated to Service Pack 12. This is a significant upgrade to the communications software as it now provides a usable version of RDMA. This is a performance optimisation that can significantly increase the communication performance of applications that send large messages. The software engineering team has been investigating the impact of this software upgrade and had produced a short technical report:

- **HPCxTR0505:** *Improved MPI with RDMA*, Alan Gray, Joachim Hein and Stephen Booth

This report explains how to activate the RDMA feature and benchmarks the impact it has on inter-node communication.

We have re-visited our investigation last year into planned AlltoAllv communications. This is a technique where we exploit the clustered SMP architecture of HPCx to optimise the important AlltoAll and AlltoAllv collective communications. We use a shared memory segment on each node to combine large numbers of short messages into a smaller number of longer messages. Though the total amount of data that needs to be transferred remains the same the overall reduction in the number of messages improves performance by reducing the impact of message latencies. We have produced a new production-quality implementation of this technique and used it to investigate performance on the current Phase-2 hardware and latest software. This technique is significantly more useful on the Phase-2 hardware than on Phase-1 as the larger LPAR sizes allow a greater number of messages to be combined. The technique can be used for message sizes up to about 2Kb (above this the additional cost of memory copies starts to negate the improvement in message latency). The significance of the technique was also seen to be much greater on larger processor counts. This work was presented at ScicomP 11 in Edinburgh May 31 to 3 June 2005. The presentation is available on the HPCx website at http://www.hpcx.ac.uk/research/hpc/presentations/Booth_ScicomP11.ppt

4.4.2 Advanced Data Handling and Grid Computing

We have been in discussion with the SPICE project about their software requirements. This is an ambitious project that intends to run a globally distributed computation during SC2005. Their requirements are significantly more ambitious than for the Teragyroid project. For Teragyroid the application running

on HPCx only required a single steering connection. However the compute nodes of HPCx are on a private network, only the login nodes are directly connected to the larger internet so we developed port forwarding software to allow the steering connection to be accessible remotely. The SPICE project requires the MPICH-G2 package as they wish to distribute a full MPI application across HPCx and systems on the US TeraGrid. This will require improvements to our port-forwarding software and changes to the MPICH-G2 package to allow the application to dynamically request additional ports to be forwarded while the program is running. As a first stage of this process we have been building an updated version of Globus that we will then use to build MPICH-G2.

4.4.3 Future Trends

We have started to investigate the portability of the OCCAM code to the BlueGene system. This work is progressing well. So far we have identified some minor issues with the `cpp` implementation on BlueGene and we are currently investigating a problem occurring in the OCCAM communication layer.

4.4.4 Generic Software Support

We are in the final stages of completing a new profiling tool for use on HPCx. This tool attaches to application codes using the DPCL class libraries and records CPU performance counter information. Counter information is recorded for every routine in the program. This tool is intended to fill a gap between simple tools like `hpmcount` and more complex tools like `paraver`. Unlike `hpmcount`, which only provides summary statistics for the entire program run, it provides a routine-by-routine breakdown of performance information. On the other hand it is much simpler to use than `paraver` and as it only produces summary statistics it does not require large amounts of disk space to hold trace-files. The main profiling tool is now complete and we are working on producing user documentation and a graphical viewer to visualise the output.

4.5 Operations and Systems (*Mr Mike Brown*)

The service has now reached the stability to be expected of a mature system between major upgrades.

4.5.1 Incidents

The number of failures (SEV1 incidents) continues to be low and presents no cause for serious concern.

4.5.2 Outreach

A meeting of SP-XXL, the international organisation of IBM systems administrators, took place in Edinburgh at the beginning of June. Two members of the Operations and Systems team attended the meeting, which was held under conditions of confidentiality and proved very worthwhile.

Members of the team attended a meeting at AWE in April with administrators of other IBM HPC sites in the UK, where we were able to get some insights into future IBM technology in our area.

4.6 Staffing

| <i>AV</i> | <i>January</i> | <i>February</i> | <i>March</i> |
|-----------|----------------|-----------------|--------------|
| DL | 6.2 | 5.7 | 6.2 |
| EPCC | 9.2 | 9.8 | 11.4 |
| Total | 15.4 | 15.5 | 17.6 |

| | | | |
|----------------|-----|-----|-----|
| <i>Systems</i> | 6.1 | 5.8 | 6.4 |
|----------------|-----|-----|-----|

5 Summary of Performance Metrics

| <i>Metric</i> | <i>TSL</i> | <i>FSL</i> | <i>April</i> | <i>May</i> | <i>June</i> |
|---------------------------------------------|------------|------------|--------------|------------|-------------|
| Technology serviceability | 80% | 99.2% | 100.0% | 100.0% | 100.0% |
| Technology MTBF (hours) | 200 | 300 | ∞ | ∞ | ∞ |
| Number of AV FTEs | 7.5 | 10 | 15.4 | 15.5 | 17.6 |
| Number of training days per month | 22.5/12 | 30/12 | 6/4 | 9/5 | 9/6 |
| Non in-depth queries resolved within 3 days | 85% | 97% | 98.3% | 100.0% | 96.3% |
| Number of A&M FTEs | 3.75 | 5.75 | 6.1 | 5.8 | 6.4 |
| A&M serviceability | 80% | 99.6% | 100.0% | 98.3% | 100.0% |

| <i>Colour</i> | <i>Meaning</i> |
|---------------|---------------------|
| | Exceeds FSL |
| | Between TSL and FSL |
| | Below TSL |

Note 1: The number of training days is reported as a running total since the start of the year.

Note 2: The above table includes the revised FSL targets for *training days* and *A&M serviceability*, which have been agreed with EPSRC.

Appendix A: Incident Severity Levels

SEV 1 — anything that comprises a FAILURE as defined in the contract with EPSRC.

SEV 2 — NON-FATAL incidents that typically cause immediate termination of a user application, but not the entire user service.

The service may be so degraded (or liable to collapse completely) that a controlled, but unplanned (and often very short-notice) shutdown is required or unplanned downtime subsequent to the next planned reload is necessary.

This category includes unrecovered disc errors where damage to filesystems may occur if the service was allowed to continue in operation; incidents when although the service can continue in operation in a degraded state until the next reload, downtime at less than 24 hours notice is required to fix or investigate the problem; and incidents whereby the throughput of user work is affected (typically by the unrecovered disabling of a portion of the system) even though no subsequent unplanned downtime results.

SEV 3 — NON-FATAL incidents that typically cause immediate termination of a user application, but the service is able to continue in operation until the next planned reload or re-configuration.

SEV 4 — NON-FATAL recoverable incidents that typically include the loss of a storage device, or a peripheral component, but the service is able to continue in operation largely unaffected, and typically the component may be replaced without any future loss of service.

Appendix B: Projects

B.1 Current Projects

EPSRC Projects

| <i>Code</i> | <i>Class</i> | <i>Title</i> | <i>PI</i> |
|-------------|--------------|----------------------------------------------------------------|---------------------------|
| e01 | 1 | UK Turbulence Consortium | Prof Neil Sandham |
| e02 | 1 | Ab-initio simulation of covalently bonded materials | Dr Patrick Briddon |
| e03 | 1 | Multi-photon, electron collisions and BEC HPC consortium | Prof Ken Taylor |
| e04 | 1 | Chemreact Computing Consortium | Prof Jonathon Tennyson |
| e05 | 1 | Materials Chemistry using Terascaling Computing | Prof Richard Catlow |
| e06 | 1 | UK Car-Parrinello Consortium | Prof Paul Madden |
| e07 | 2 | Turbulent Plasma Transport in Tokamaks | Dr Colin M Roach |
| e08 | 2 | Organic Solid State | Prof Sarah Price |
| e10 | 1 | Reality Grid | Prof Peter Coveney |
| e11 | 1 | Bond making and breaking at surfaces | Prof Sir David A King |
| e12 | 1 | Parallel programs for the simulation of complex fluids | Dr Mark R Wilson |
| e14 | 1 | Blade and Cavity Noise | Prof Neil Sandham |
| e15 | 2 | CSAR/HPCx Collaboration | Dr Mike Pettipher |
| e16 | 1 | Cardiac virtual tissues | Prof Arun V Holden |
| e17 | 1 | Integrative Biology | Dr David Gavaghan |
| e18 | 1 | DARP: Highly swept leading edge separations | Prof Michael A Leschziner |
| e19 | 1 | Edinburgh Soft Matter and Statistical Physics Group | Prof Michael E Cates |
| e20 | 1 | UK Applied Aerodynamics Consortium | Dr Ken Badcock |
| e21 | 1 | Intrinsic Parameter Fluctuations in Decanometer MOSFETs | Prof Asen M Asenov |
| e22 | 1 | Preconditioners for finite element problems | Prof David J Silvester |
| e23 | 1 | Exploitation of Switched Lightpaths for e-Science Applications | Prof Peter Clarke |

| | | | |
|-----|---|-----------------------------------------------------------------------------|----------------------|
| e24 | 1 | DEISA - Distributed European Infrastructure for Supercomputing Applications | Dr David Henty |
| e25 | 1 | Turbulent vortex motion in stratified flows | Dr Gary Coleman |
| e26 | 1 | Simulation of Radioprobing | Dr Charlie Laughton |
| e27 | 1 | SPICE | Prof Peter V Coveney |
| e28 | 1 | Towards the Dynome | Dr Jonathan W Essex |
| e29 | 1 | Free-surface-piercing circular cylinders | Dr Eldad Avital |
| z09 | | HECToR Benchmarking | Dr Edward Smyth |

PPARC Projects

| <i>Code</i> | <i>Class</i> | <i>Title</i> | <i>PI</i> |
|-------------|--------------|---------------------------------|-------------------|
| p01 | 1 | Atomic Physics and Astrophysics | Prof Alan Hibbert |

NERC Projects

| <i>Code</i> | <i>Class</i> | <i>Title</i> | <i>PI</i> |
|-------------|--------------|---------------------------------------------------|--------------------|
| n01 | 1 | Large-Scale Long-Term Ocean Circulation | Dr David Webb |
| n02 | 1 | NCAS | Prof Alan J Thorpe |
| n03 | 1 | Computational Mineral Physics Consortium | Dr John Brodholt |
| n04 | 1 | Shelf Seas Consortium | Dr Roger Proctor |
| n05 | 2 | Non-linear Wave-particle Instabilities in Plasmas | Dr Mervyn Freeman |

BBSRC Projects

| <i>Code</i> | <i>Class</i> | <i>Title</i> | <i>PI</i> |
|-------------|--------------|------------------------------------------------------------------|------------------------|
| b02 | 1 | Modelling enzyme catalysis | Dr Adrian J Mulholland |
| b03 | 1 | Towards a virtual outer membrane | Prof Mark S Sansom |
| b04 | 1 | Life sciences software development | Dr Jo L Dicks |
| b05 | 1 | Virtual forced evolution of catalytic transition metal complexes | Dr Marcus Durrant |
| b06 | 2 | Biomolecular computational chemistry | Prof Jonathan D Hirst |

CCLRC Projects

| <i>Code</i> | <i>Class</i> | <i>Title</i> | <i>PI</i> |
|-------------|--------------|------------------------------------------------------|--------------------|
| c01 | 1 | Daresbury Laboratory Facilities Agreement Consortium | Dr Richard J Blake |

Externally-funded Projects

| <i>Code</i> | <i>Title</i> | <i>PI</i> |
|-------------|--------------|------------------|
| x01 | HPC-Europa | Dr J-C Desplat |
| x02 | OHM Ltd | Mr Mark Westwood |

HPCx Projects

| <i>Code</i> | <i>Title</i> | <i>PI</i> |
|-------------|------------------------|------------------|
| z001 | HPCx Support | Dr Alan Simpson |
| z002 | Systems and Operations | Mr Mike Brown |
| z003 | Test Project | Dr Denis Nicole |
| z004 | HPCx Training | Dr David Henty |
| z05 | Outreach Projects | Dr Richard Blake |
| z06 | Application Porting | Dr David Henty |
| z07 | Package Installation | Dr Mike Ashworth |

B.2 Former Projects

| <i>Code</i> | <i>Class</i> | <i>Title</i> | <i>PI</i> |
|-------------|--------------|--------------------------------------------------------------|--------------------|
| b01 | 2 | Quantum Chemistry Studies of the Rusticyanin Protein Crystal | Prof Samar Hasnain |
| e09 | 2 | Molecular Properties and their Geometry | Prof Peter Taylor |
| e13 | 1 | TeraGyroid project | Dr Richard J Blake |