

# HPCx Quarterly Report

## July - September 2004

### 1 Introduction

This report covers the period from 1 July 2004 at 0800 to 1 October 2004 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

- We have now agreed with EPSRC to increase the maximum number of concurrent consortia on HPCx to 40. As part of this agreement, there will be three additional science support posts funded from 2005 to 2008, which should allow us to continue to provide high quality support for all users.
- After installation of the improved HPS microcode, we repeated the implementation tests benchmark runs and reported these to EPSRC. The HPS latency was reduced from 10 microseconds to around 6 microseconds and half of the user benchmark codes saw improvements in excess of 10%.
- Utilisation has remained at around 75% since the upgrade to Phase 2 with around one-third of this being on capability jobs; we expect this position to continue throughout the rest of the year. As well as major users from Materials and Chemistry, there is significant capability usage from Engineering, Atomic & Molecular Physics and from the Terascaling team, as we work to ensure that a wide range of applications codes can effectively exploit large numbers of processors
- During this quarter, good performance improvements have been reported for SIESTA, VASP, NWChem and H2MOL. We have also begun working

on the CENTORI fusion code and have made available two CFD codes for the UKAAC consortium.

- Another 4 technical reports have been published during this quarter taking our annual total to 13. Given our target of 12, we are clearly well ahead of schedule.
- The Software Engineering team has used a significant amount of effort this year on general terascaling techniques on HPCx-class systems. As well as direct benefits to specific codes, this work has formed the basis of our “Improved Performance Scaling” course; a version of this course has been accepted as a tutorial at SC2004.
- IBM held a workshop in Manchester to allow users to input into IBM’s priorities for system software; this will now be repeated annually.
- HPCx are represented on the steering committee investigating possible experiments with the US Teragrid sites following on from the very successful Teragyroid demonstrator.
- As a result of the Life Sciences outreach activity, BBSRC usage of HPCx is now increasing and a number of groups are beginning to report interesting results.

### 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>July</i> | <i>August</i> | <i>September</i> |
|-----------|-------------|---------------|------------------|
| Incidents | 2           | 1             | 3                |
| Failures  | 21          | 22            | 5                |

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

##### *July*

| <i>Failure</i> | <i>Site</i> | <i>IBM</i> | <i>External</i> | <i>Reason</i>                             |
|----------------|-------------|------------|-----------------|---|
| 04.151         | 100%        | 0%         | 0%              | Emergency power down after fire condition |
| 04.169         | 0%          | 100%       | 0%              | Maintenance session overrun               |

##### *August*

| <i>Failure</i> | <i>Site</i> | <i>IBM</i> | <i>External</i> | <i>Reason</i>                     |
|----------------|-------------|------------|-----------------|-----------------------------------|
| 04.188         | 0%          | 100%       | 0%              | Inter-node communications problem |

##### *September*

| <i>Failure</i> | <i>Site</i> | <i>IBM</i> | <i>External</i> | <i>Reason</i>   |
|----------------|-------------|------------|-----------------|---|
| 04.195         | 0%          | 0%         | 100%            | Network failure at Manchester                         |
| 04.208         | 100%        | 0%         | 0%              | Tape archive down due to problems with cooling system |
| 04.213         | 0%          | 100%       | 0%              | Login node failed, backup took over                   |

### 3.1.2 Performance Statistics

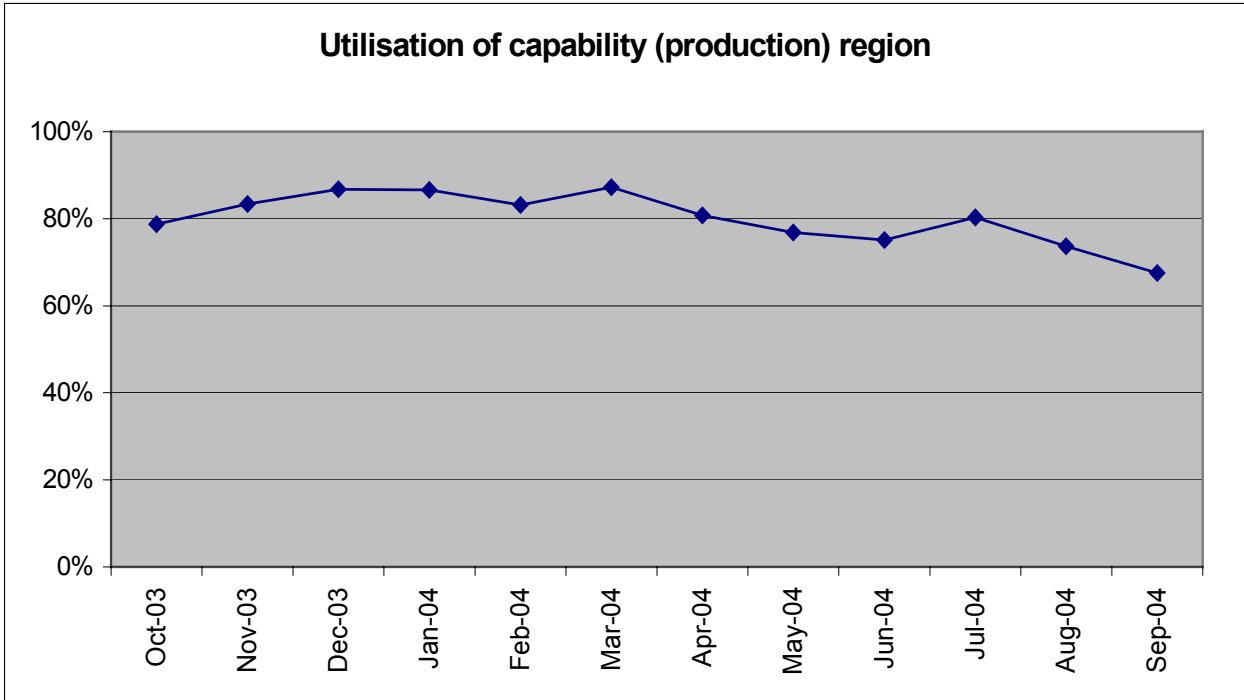
This section uses the definitions agreed in Schedule 7, ie,

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

| <i>Attribution</i> | <i>Metric</i>  | <i>July</i> | <i>August</i> | <i>September</i> | <i>Quarterly</i> |
|--------------------|----------------|-------------|---------------|------------------|------------------|
| IBM                | Failures       | 1           | 1             | 1                | 3                |
|                    | MTBF           | 732         | 732           | 732              | 732              |
|                    | Serviceability | 99.8%       | 99.7%         | 99.9%            | 99.8%            |
| Site               | Failures       | 1           | 0             | 1                | 2                |
|                    | MTBF           | 732         | ∞             | 732              | 1098             |
|                    | Serviceability | 98.8%       | 100.0%        | 99.4%            | 99.4%            |
| External           | Failures       | 0           | 0             | 1                | 1                |
|                    | MTBF           | ∞           | ∞             | 732              | 2196             |
|                    | Serviceability | 100.0%      | 100.0%        | 99.9%            | 99.9%            |
| Total              | Failures       | 2           | 1             | 3                | 6                |
|                    | MTBF           | 366         | 732           | 244              | 366              |
|                    | Serviceability | 98.6%       | 99.7%         | 99.4%            | 99.2%            |

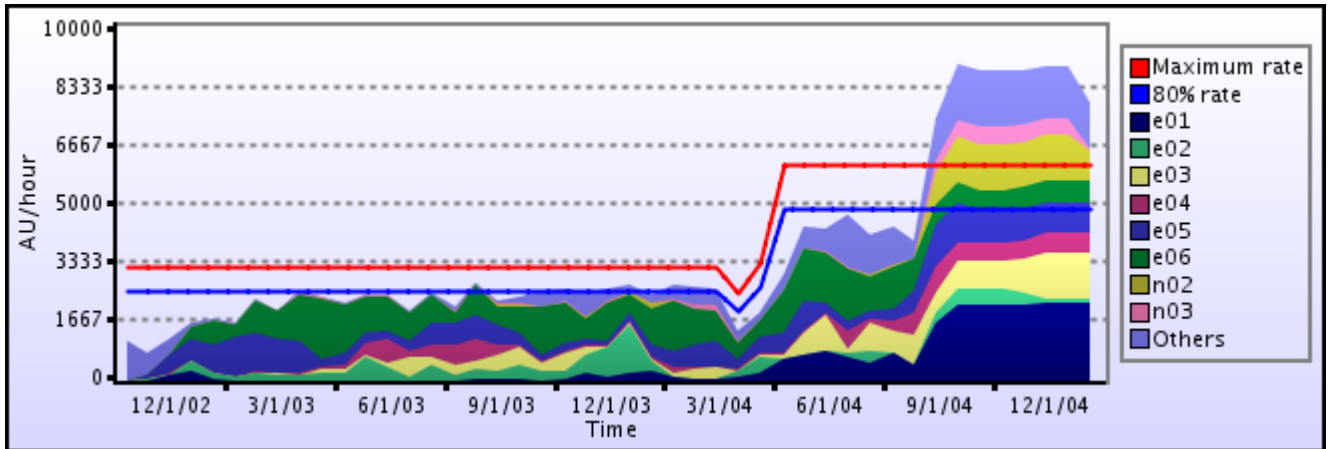
### 3.2 Capability Utilisation

Utilisation has continued at around 75%, with a slight slackening of demand in September resulting from a change in work pattern by a few projects.



### 3.3 Capacity Planning

#### *Predicted Utilisation*



The graph above shows the utilisation since the start of the project and the projected utilisation until February 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.

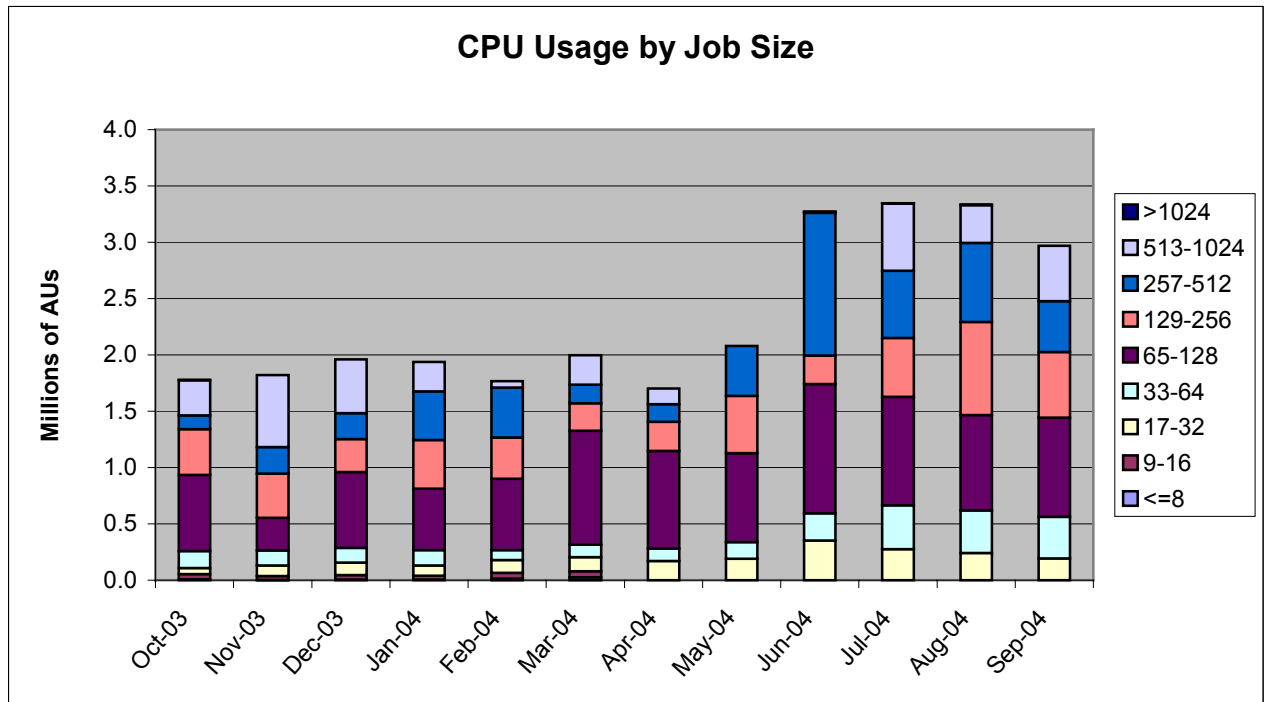
The graph suggests that the service will be substantially overloaded through the rest of this year.

### Numbers of Research Consortia

There are currently 33 research consortia using the HPCx system. Three other projects have now been closed.

In addition, there is one externally-funded project.

### 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>July</i> | <i>August</i> | <i>September</i> | <i>Quarterly</i> | <i>%age</i> |
|--------------------|-------------|---------------|------------------|------------------|-------------|
| e01                | 592083      | 448187        | 507557           | 1547827          | 16.1%       |
| e02                | 3353        | 199293        | 62               | 202708           | 2.1%        |
| e03                | 571474      | 388820        | 483636           | 1443930          | 15.0%       |
| e04                | 36023       | 257788        | 222886           | 516697           | 5.4%        |
| e05                | 399790      | 176033        | 305227           | 881051           | 9.2%        |
| e06                | 1123969     | 826416        | 772623           | 2723008          | 28.3%       |
| e07                | 20758       | 3609          | 45655            | 70022            | 0.7%        |
| e08                | 15485       | 12988         | 3938             | 32412            | 0.3%        |
| e10                | 10071       | 19647         | 114688           | 144406           | 1.5%        |
| e11                | 4119        | 0             | 0                | 4119             | 0.0%        |
| e12                | 26569       | 13034         | 6235             | 45837            | 0.5%        |
| e15                | 5749        | 543           | 4637             | 10928            | 0.1%        |
| e17                | 0           | 0             | 18199            | 18199            | 0.2%        |
| e18                | 25889       | 122           | 1115             | 27125            | 0.3%        |
| e20                | 198259      | 497011        | 108554           | 803825           | 8.4%        |
| <i>EPSRC Total</i> | 3033590     | 2843490       | 2595014          | 8472093          | 88.1%       |

|                   |        |        |        |        |      |
|-------------------|--------|--------|--------|--------|------|
| n01               | 90570  | 152822 | 57315  | 300707 | 3.1% |
| n02               | 14184  | 16072  | 34377  | 64633  | 0.7% |
| n03               | 8020   | 24416  | 10572  | 43008  | 0.4% |
| n04               | 273    | 33396  | 9785   | 43454  | 0.5% |
| <i>NERC Total</i> | 113047 | 226707 | 112049 | 451803 | 4.7% |

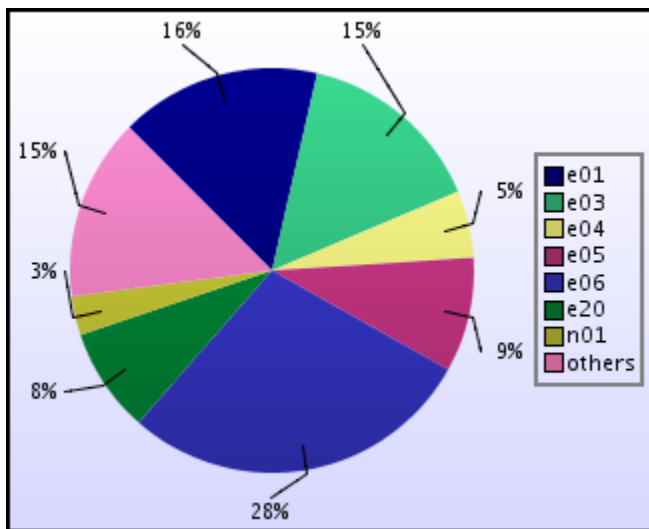
|                    |      |      |      |       |      |
|--------------------|------|------|------|-------|------|
| p01                | 1601 | 1249 | 8359 | 11208 | 0.1% |
| <i>PPARC Total</i> | 1601 | 1249 | 8359 | 11208 | 0.1% |

|                    |       |       |       |        |      |
|--------------------|-------|-------|-------|--------|------|
| c01                | 82101 | 78823 | 98548 | 259472 | 2.7% |
| <i>CCLRC Total</i> | 82101 | 78823 | 98548 | 259472 | 2.7% |

|                    |       |       |       |        |      |
|--------------------|-------|-------|-------|--------|------|
| b02                | 107   | 66567 | 17756 | 84431  | 0.9% |
| b03                | 10725 | 2231  | 0     | 12957  | 0.1% |
| b06                | 31    | 1938  | 0     | 1969   | 0.0% |
| b07                | 15    | 1693  | 7919  | 9626   | 0.1% |
| <i>BBSRC Total</i> | 10879 | 72428 | 25675 | 108982 | 1.1% |

|                   |       |       |       |        |      |
|-------------------|-------|-------|-------|--------|------|
| z001              | 65268 | 62879 | 69905 | 198052 | 2.1% |
| z002              | 4709  | 3529  | 4191  | 12428  | 0.1% |
| z004              | 4348  | 19706 | 981   | 25035  | 0.3% |
| z06               | 5187  | 756   | 1827  | 7771   | 0.1% |
| <i>HPCx Total</i> | 79512 | 86869 | 76905 | 243286 | 2.6% |

|                       |      |       |       |       |      |
|-----------------------|------|-------|-------|-------|------|
| x01                   | 5230 | 31125 | 31192 | 67546 | 0.7% |
| <i>External Total</i> | 5230 | 31125 | 31192 | 67546 | 0.7% |



### 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> |
|-------------------|-----------------|--------------------|-----------------|
| e01               | 1574656         | 1547826            | 26829           |
| e10               | 161492          | 144406             | 17086           |
| e17               | 21411           | 18199              | 3211            |
| x01               | 71181           | 67546              | 3635            |

## 3.6 Helpdesk

### 3.6.1 Classifications

| <i>Category</i> | <i>Number</i> | <i>% of all</i> |
|-----------------|---------------|-----------------|
| Administrative  | 125           | 34.5%           |
| Technical       | 220           | 60.8%           |
| In-depth        | 15            | 4.1%            |
| PMR             | 2             | 0.6%            |
| TOTAL           | 362           | 100.0%          |

| <i>Service Area</i> | <i>Number</i> | <i>% of all</i> |
|---------------------|---------------|-----------------|
| Phase 1/2 platforms | 315           | 87.0%           |
| Website             | 14            | 3.9%            |
| Other/general       | 33            | 9.1%            |
| TOTAL               | 362           | 100.0%          |

### 3.6.2 Performance

| <i>All non-indepth queries</i> | <i>Number</i> | <i>%</i> | <i>Target</i> |
|--------------------------------|---------------|----------|---------------|
| Finished within 24 Hours       | 307           | 89.0     | 75%           |
| Finished within 72 Hours       | 344           | 99.7     | 97%           |
| Finished after 72 Hours        | 1             | 0.3      |               |

| <i>Administrative queries</i> | <i>Number</i> | <i>%</i> | <i>Target</i> |
|-------------------------------|---------------|----------|---------------|
| Finished within 48 Hours      | 124           | 99.2     | 97%           |
| Finished after 48 Hours       | 1             | 0.8      |               |

### 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 | 85           | 79               | 6               | 0          |
| dl.ac.uk      | 2            | 33               | 3               | 1          |
| Sysadm        | 38           | 107              | 6               | 1          |
| Other people  | 0            | 1                | 0               | 0          |

### 3.7 Service Quality Tokens

| <i>Date</i>               | <i>Person</i>                    | <i>Value</i> | <i>Comment</i> | <i>Status</i> |
|---------------------------|----------------------------------|--------------|----------------|---------------|
| Sep 7, 2004<br>6:46:33 PM | <a href="#">Dr. Jens Harting</a> | * * * * *    |                |               |

## 4 Support

### 4.1 Applications Support (*Dr David Henty*)

After a very busy second quarter (which included the HPCx Annual Seminar, a workshop, development of a new course and the Phase 2 upgrade) this quarter has been somewhat quieter. However, we have continued to make steady progress against the yearly workplan, with concrete arrangements now in place to meet all the targets for the applications support team by the end of Q4.

#### 4.1.1 Documentation

Online documentation has required few changes this quarter after the substantial updates required for the Phase 2 transition. The FAQ now contains a lot of useful information, and we plan to reorganise it with a new index to make it easier to navigate.

#### 4.1.2 Technical Reports

A total of three reports were due for Q4 in the following areas:

- a) Scalability of OpenMP
- b) Communications performance with release 2 HPS software
- c) Applications performance on the HPS

We have actually produced four reports this quarter:

- **HPCxTR0410:** *Improved parallel performance of SIESTA for the HPCx Phase2 system*, Joachim Hein.
- **HPCxTR0411:** *OpenMP Microbenchmarks V2.0*, Fiona Reid, Mark Bull.
- **HPCxTR0412:** *Communications on SP7 and SP9*, Adrian Jackson.
- **HPCxTR0413:** *Profiling H2MOL on an IBM p690+ Cluster*, Lorna Smith, Mark Bull, Andrew Sunderland.

Reports **11** and **12** correspond directly to the titles **a)** and **b)**. We decided to move the production of report **c)** to Q4 so we can do a complete review of applications performance on the full SP9 HPS release rather than an intermediate study on SP7 (installation of SP9 has been slightly delayed). To compensate for this delay we have brought forward a report called *Parallelisation techniques: a case study*, originally planned for Q4. Reports **12** and **13** are both in this area, covering two applications that have very different parallelisation techniques. Both reports perform comprehensive studies of the parallel

performance of the codes in question (using VAMPIR and other tools), identify communications bottlenecks and implement performance improvements.

In the first three quarters of 2004 we have produced a total of thirteen reports, four more than the target. Of the total of twelve report titles identified in the 2004 Annual Plan, nine have been produced so far. There has been some change in the ordering of reports: two of the planned Q4 reports have already been published, whereas one report from Q2 and another from Q3 are still to be completed. However, we are on target to produce all the planned 2004 titles by the end of Q4, with at least four reports in addition to the original twelve.

### 4.1.3 Training

In Q3 of 2004 we ran the following two courses:

- **Edinburgh, 8 July:** *Improved Performance Scaling on HPCx.*
- **Imperial College, London, 24 September:** *Introduction to Parallel Crystal*

The first course was developed from scratch and was delivered for the first time this quarter. It focuses on tools and techniques for achieving optimal parallel performance on HPCx, covering a wide variety of topics including advanced lectures on mixed-mode programming and memory affinity.

The second course was given as a session during the conference *MSSC2004: Ab initio Modelling in Solid State Chemistry* from 20-24 September. It covered parallelisation of the Crystal package, illustrated with examples from HPCx. For details of the event see:

<http://www.cse.clrc.ac.uk/events/MSSC2004/TIMETABLE.pdf>

The statistics for the first three quarters are summarised below alongside the annual targets (where appropriate).

| <b>Metric</b>                   | <b>Total</b> | <b>Target</b> |
|---------------------------------|--------------|---------------|
| Course days                     | 24           | 30            |
| Number of courses               | 11           |               |
| Different courses               | 10           | 12            |
| Different locations             | 4            | 4             |
| Student-days for HPCx users     | 248          |               |
| Student-days for HPCx staff     | 18           |               |
| Student-days available for HPCx | 493          | 600           |

The summer is always a quiet time for training as people are on leave and there are so many other meetings and events to contend with. All the Q4 courses are planned and advertised, and with these runs we will meet all the targets for 2004.

#### **4.1.4 Workshops and Conferences**

The Second HPCx Annual Seminar was held in Q3 on 9<sup>th</sup> July. However, it was fully covered in the last quarterly report so there is nothing to add here. Next year we plan to hold the third annual seminar alongside the Daresbury Machine Evaluation Workshop in December, a format that worked very well for the inaugural HPCx event in December 2003.

EPCC is hosting the European meetings of the international ScicomP and SP-XXL conferences (for users and system administrators of IBM supercomputers) from 31 May to 3 June 2005. Although this is not formally being organised as part of the HPCx service, we will make every attempt to promote HPCx and its users.

#### **4.1.5 Newsletter**

The fourth issue of Capability Computing was published in time for shipping to Supercomputing 2004. It focuses on capability science and should prove valuable publicity material for UK computational science research as well as the HPCx service itself. We have sent some 3,500 copies through the standard mailshot to the UK, Europe and Overseas.

#### **4.1.6 Packages**

We have installed a centralised version of SIESTA, an application whose performance was significantly improved by the HPCx Terascaling Team (see Joachim Hein's report HPCxTR0410).

## **4.2 Outreach Activities (*Dr Richard Blake*)**

Over the past quarter the outreach activities have progressed in the following areas.

### **4.2.1 Life Sciences Projects**

The Life Sciences projects at Bristol University, Oxford University and the John Innes Centre are now well under way. Support requirements for the other two projects will be developed in October. Major updates on the various projects can be found in Appendix C.

### **4.2.2 Other Life Sciences Activities**

In terms of other Life Sciences support activities, Paul Sherwood (Daresbury Laboratory) attended the CHARMM developers meeting at Harvard (9-10th July, 2004), which included a detailed discussion of performance aspects of this code. Performance data from HPCx were presented and compared with results from other IBM (e.g. the SDSC Datastar) and non-IBM systems, and some suggestions for performance improvement obtained. Sherwood then visited the National Institutes of Health to work with the Brooks group (developers of the Replica Path and NEB functionality and CHARMM). This trip led to the identification of a long-standing problem in the pre-release version of this code which was then fixed in time for the August release. This new released code (c31b1) is currently being tested on HPCx and should be available to projects in the next few weeks.

Paul Sherwood and Richard Blake will meet with Charles Laughton in October to progress the proposals to establish a Consortium for Biomolecular Simulations: Structure Function relationships of Biomolecules.

### **4.2.3 Commercial Codes**

Fluent, CFX and Abaqus have been ported to HPCX Phase 2, the first two in support of the UKAAC Consortium. Fluent are organising a series of seminars with the academic community to explore future applications and developments for the software. Richard Blake has offered to provide a talk overviewing the HPCx service and demonstrating the codes performance on capability class problems. Stimulating academic use of commercial codes on HPCx will provide demonstrations that can then be promoted to industry.

#### **4.2.4 Follow-up to HPCx Industry Day**

Following on from the HPCx industry day and discussions with commercial software vendors a 'proposition' based on Access, Codes and Solutions is being developed and a list of 20 customer sites has been identified. A calling programme will be put in place over the next quarter.

#### **4.2.5 Teragyroid Experiment**

A report on lessons learnt from the Teragyroid experiment undertaken last November was submitted in April 2004. Richard Blake gave a presentation summarising the lessons learned at the RealityGrid Workshop. The authors have been invited to submit the report for publication in the Journal of Scientific Programming (October 2003). Richard Blake has been invited to sit on a Steering Committee, representing HPCx, which will overview a Call for Proposals for future experiments. The first full meeting of the team will be on 13 October 2004.

#### **4.2.6 JCSR Working Group**

Richard Blake has been invited to sit on the JCSR Working Group chaired by Prof J Annett to explore the academic community's needs for visualisation. The Working Group will arrange a 'user requirements meeting' on 19 October with a report to be submitted to JCSR in November. The outputs from the user consultation will be used to inform CCLRC's programme of visualisation activities. These currently include outline discussions with vendors on the scale and scope of cluster and industrial strength visualisation software and hardware.

## 4.3 Terascaling Applications (*Dr Martyn Guest*)

### 4.3.1 SP7 and SP9 benchmark testing

Following improvements to the HPS microcode contained in Service Packs SP7 and SP9, benchmark runs of AIMPRO, CASTEP, DL-POLY, H2MOL and PCHAN have been repeated. Results have been reported to EPSRC. This work included assessment of the new MP\_TASK\_AFFINITY option, which replaces the VSRAC tool.

### 4.3.2 Computational Materials Science

#### *VASP*

- Following IBM's fix of the problems associated with running this code on Phase 2, an extensive performance analysis has been carried out. Gavin Pringle visited a member of the consortium, Dario Alfè (UCL), to discuss potential optimisations. Since his return, work has focused on finding the optimal parallelisation strategy for HPCx (VASP is parallelised in three ways – initial studies show an 8.4% speed-up can be obtained on 128 processors by choosing the correct strategy). Other work includes investigating the memory profile of this code.

#### *Castep*

- Castep scripts, executables and documentation have been updated. The Castep executables are HPCx-optimised equivalents of the latest distribution 'Castep3.0'. Work has continued on looking at particular problems associated with modelling of organic liquids. In response to a user request (from Matt Probert, York) there has been further development of the Castep-specific task-farming facility, with the addition of HPCx-specific communications optimisations. The task-farming facility will not be made generally available, until ratified by the Castep Developers' Group.

The optimisations of Castep for HPCx were reported at the HPCx Annual Seminar in "Terascaling Applications on HPCx" by Martin Plummer.

### 4.3.3 Molecular Simulation

#### *NAMD*

- We had experienced problems with the scalability of this code for large number of processors using earlier versions of the switch microcode. For a number of different benchmark systems, ranging from 23,000 to over 300,000 atoms, we have shown that these problems have been overcome with the latest microcode release (Service Pack 7).

## SIESTA

- We have completed a detailed performance analysis of the SIESTA code. The analysis showed that the performance of the diagonaliser was the main bottleneck inside the code. Changing the processor grid in the diagonaliser from one to two dimensions led to a dramatic improvement of the code's performance. On 128 processors the code now performs over 3 times faster for our benchmarking configuration. We have published these results in a technical report, available on the HPCx website. The optimised routines have been made available to HPCx users, to be incorporated into their production codes, via the helpdesk. We have communicated our results to the SIESTA development team and the improvements will be included into the forthcoming SIESTA release.

J. Hein, "Improved parallel performance of SIESTA for the HPCx Phase2 system", HPCx Technical report: HPCxTR0410. See:

[http://www.hpcx.ac.uk/research/hpc/technical\\_reports/HPCxTR0410.pdf](http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0410.pdf)

## GROMACS

- We have reported the bug identified in GROMACS 3.2.1 to the developers and a bug fix has now been released. Following this, we have investigated the scaling of this code for one of our user groups.

## MD Code Comparison

- Work has focused on developing a GROMACS input file for the joint AMBER CHARMM benchmark – to allow the performance of this code to be compared directly to that of AMBER, NAMD and CHARMM. An initial input file has been developed, and is currently being investigated for scientific correctness. We have also installed the LAMMPS code and are investigating the scaling of this code for one of our user groups.

## 4.3.4 Molecular Electronic Structure

The release of the SP9 LAPI library has led to a major performance improvement of all Global Array (GA) based application codes (NWChem, GAMESS\_UK and MOLPRO). With a marked improvement in the efficiency of the ACCUMULATE GA function, good scaling is now apparent, e.g. in NWChem, DFT and CCSD(T) calculation times to solution on 128 processors have improved by factors of 2.3 and 1.8 respectively compared to the SP7-based implementation.

## GAMESS-UK

- The distributed memory SCF code has been merged with the task-farming code required in the Life Sciences OVDE project. Optimisations to the distributed data code itself have reduced the memory requirements, improved performance and extended the size of system amenable to treatment. They

included, (i) introduction of distributed data parallelism into generating the initial estimate of the wavefunction, (ii) removal of some data structures from the SCF code plus removal of certain global communications, and (iii) avoidance of some communications in the matrix creation routines.

An initial look at load balancing issues has revealed potential for increased performance, as the load balancing currently limits the scalability of the integrals. It is more noticeable in the new distributed memory SCF driver, since the load balancing is static, rather than dynamic as in the GA version of the code. All the developments above have been reported in "The GAMESS-UK Electronic Structure Package: Algorithms, Developments and Applications", Martyn F. Guest, Jens M.H. Thomas, Paul Sherwood, Ian J. Bush, Huub J.J. van Dam, Mol. Phys (submitted)

### 4.3.5 Engineering

#### *CENTORI*

- We have begun work on the CENTORI code, one of the main codes from the UKAEA (Culham) Consortium. This code simulates the fluid flow inside a tokamak, as used for fusion physics. Initial profiling of CENTORI shows the present version of the application spends most of its time in Fourier transforms. Joachim Hein and Lorna Smith visited the development team in Culham/UK in September. The main points of discussion were possible strategies for future optimisations (additional parallelisation, use of different libraries). After returning from Culham, we have focused on the Fourier transforms. Initial benchmark tests comparing the performance of the presently used FFT routines to the standard library FFTW look promising.

#### *CFX and Fluent*

- Scripts have been created to facilitate the use of Fluent (VMPI) and CFX (PVM) via LoadLeveler batch jobs on HPCx. CFX and Fluent users, primarily from the UKAAC Consortium, have been given advice and assistance with running jobs. Benchmark tests have been carried for CFX and Fluent datasets. New licences have been obtained for CFX and Fluent.

### 4.3.6 Physics

#### *H2MOL*

- A detailed performance analysis has been carried out on this code. The code makes substantial use of optimised numerical libraries and scales well to 496 processors. The majority of the communications are point to point. A 6% performance improvement has been obtained on 120 processors by replacing the buffered send routines with non-blocking send and receive operations. We have published these results in a technical report, available on the HPCx website.

L. Smith, J. M. Bull, A. Sunderland, "Profiling H2MOL on an IBM p690+ Cluster", HPCx Technical report: HPCxTR0413. See:

[http://www.hpcx.ac.uk/research/hpc/technical\\_reports/HPCxTR0413.pdf](http://www.hpcx.ac.uk/research/hpc/technical_reports/HPCxTR0413.pdf)

#### *LUDWIG*

- Following the success of our optimisation work from last quarter, Ludwig has been demonstrated to run efficiently on the phase 2 machine and will scale from 128 to 1024 processors for lattice sizes of 512x512x512 and beyond. No benchmarks have been taken since the installation of SP7, and we are waiting for SP9 to rerun. Performance for smaller problems (256x256x256 lattice on 128 processors) will also be investigated.

### **4.3.7 Serial Code Efficiency**

Serial efficiency data has been collected for approximately 15 user codes on the Phase 2 system. This data is currently being collated and analysed for an HPCx Technical Report.

### **4.3.8 Libraries**

#### *Eigensolvers*

- The eigensolver test suite has been expanded to include PDSYEVX for PESSL/ScaLAPACK performance comparison. Ian Bush's Matrix modules (see the GAMESS-UK work) have been incorporated into the diagonalisation suite to create a generalised ScaLAPACK API.

The comparative performance of different eigensolvers was reported at the HPCx Annual Seminar in "Parallel Symmetric Eigensolver Performance on HPCx" by Andy Sunderland.

#### *Benchmarking*

- A numerical libraries benchmark suite is being developed. See the Software Engineering report for further details.

### **4.3.9 Tools**

#### *DDT*

- A new release has been obtained from Streamline (who now have their own POWER4 system for testing and development). There are problems with DDT attaching to running LoadLeveler jobs, which are currently under investigation by Streamline.

#### *Paraver*

- A new licence has been installed for Paraver.

#### 4.3.10 Presentations

M.F. Guest, "Scientific Application Performance on the HPCx IBM p690+ Cluster", University of Cardiff, 24 August 2004.

M.F. Guest, "Large Scale Electronic Structure Calculations in the Study of Biomolecular Reactivity", presented at MGMS September 2004, "Towards Accurate Calculation of Biomolecular Recognition and Reactivity", University of Manchester, 8 September 2004.

M.F. Guest, "Performance Engineering for Scientific Applications", presented at the European HPTC Workshop, Paris, 22 September 2004.

J. Hein, "HPCx Phase2 vs. Phase1: A performance comparison", HPCx Annual Seminar, Edinburgh, 9<sup>th</sup> July 2004

J. Hein, "HPCx Phase2 vs. Phase1: A performance comparison", NIC, Forschungszentrum Jülich, Germany, 7<sup>th</sup> September 2004

M. Plummer, "Terascaling Applications on HPCx", HPCx Annual Seminar, Edinburgh, 9<sup>th</sup> July 2004.

P. Sherwood, "Supporting Computational Chemistry on the HPCx system", HPC-EUROPA Workshop: State-of-the-art computational chemistry: an overview, Edinburgh, 17<sup>th</sup> September 2004.

L. Smith, "Current Trends in HPC", HPC Inform Meeting, University of Strathclyde, 16<sup>th</sup> September 2004.

L. Smith, "Applications on HPCx", HPC-EUROPA Workshop: State-of-the-art computational chemistry: an overview, Edinburgh, 17<sup>th</sup> September 2004

A.G. Sunderland, "Parallel Symmetric Eigensolver Performance on HPCx", HPCx Annual Seminar, Edinburgh, 9<sup>th</sup> July 2004

## **4.4 Software Engineering (*Dr Stephen Booth*)**

### **4.4.1 Low Level Communications**

A good understanding of the low-level communication performance is vital to the efficient use of systems like HPCx. We have been running a number of low level communication benchmarks to evaluate the communication performance of the phase-2 system and to compare this with that of the Phase-1 system. This has been concentrating on comparing various low level communications including:

- Low level performance of MPI and LAPI
- Use of shared memory segments.
- Optimised collectives using split communicators.
- The optimised collectives provided by IBM.

We will publish the results of this investigation as a technical report once we have final performance results for the SP9 update to the microkernel.

### **4.4.2 Shared Memory Techniques**

The standard environment for developing parallel programs in a shared memory environment is OpenMP. OpenMP is a set of language extensions (compiler directives) and is available for both C and Fortran. We have recently updated the OpenMP micro benchmark suite to support the additional features in OpenMP version 2.0. This benchmark suite has been run to compare the OpenMP implementation on HPCx, the SGI Altix and a Sun E15K. The results have been compiled into the following report:

- HPCxTR0411 "OpenMP microbenchmarks version 2.0"

### **4.4.3 Grid Computing**

No major development of the Grid Computing infrastructure was scheduled during this quarter.

### **4.4.4 Data Handling**

We have been re-running the MPI-IO performance tests on the HPCx Phase-2 system.

#### **4.4.5 General Terascaling Techniques**

We have been investigating general terascaling techniques applicable to systems built out of clustered SMP nodes. Much of this work was developed as part of the course "Improved Performance Scaling on HPCx". However, many of these techniques are generally applicable to other systems with a similar architecture and we will be presenting a tutorial on this material at Supercomputing 2004 in Pittsburgh. An expanded version of this material will also form the basis of a HPCx technical report which is currently under development.

We are developing a scientific library benchmark suite to identify potential performance problems with the scientific libraries on HPCx. Currently this is targeting the following areas:

- Eigensolvers
- FFTs
- Matrix-Matrix multiplication

#### **4.4.6 Systems Programming**

We have been continuing to extend and develop the SAF system, which supports the administrative website and the database. The major developments undertaken this quarter are to improve the performance of the report generation code. The SAF stores user accounting data in an SQL database. The database table storing this information now contains over 18000 records. As this database table grows, the SQL queries used to generate the accounting reports will start to take longer. However, by far the majority of database queries are only concerned with the most recent data. We have therefore modified the SAF to allow older data to be stored in a separate table, improving the performance of the most common types of query.

## **4.5 Operations and Systems (*Mr Mike Brown*)**

### **4.5.1 Staffing**

There has been no change in staffing levels, although the coverage still remains substantially in excess of the "core hours" contractual requirement.

### **4.5.2 Test & Development System**

There is unfortunately no Test and Development system available to support the Phase 2 service. The Phase 1 Test and Development system was highly valued and made a real contribution to enhancing the quality of service under Phase 1. The unavailability of such a facility under Phase 2 is regrettable.

### **4.5.3 Maintenance Sessions**

Regular maintenance sessions were taken up until the closure of the Phase 1 service, and their need continues on the replacement Phase 2 service.

### **4.5.4 File Archive**

The file archiving system under TSM is in full service.

### **4.5.5 Outreach**

The Operations and Systems Group was represented at the IBM SP-XXL summer meeting in Austin, and at the meeting of the UK IBM HPC user group in Edinburgh early in September. The value of these meetings (where operational staff from sites with large-scale systems can meet, present, compare problems and solutions) and interact directly with IBM, cannot be overestimated.

## 4.6 Staffing

| <i>AV</i> | <i>July</i> | <i>August</i> | <i>September</i> |
|-----------|-------------|---------------|------------------|
| DL        | 5.2         | 4.6           | 5.6              |
| EPCC      | 9.5         | 6.5           | 7.6              |
| Total     | 14.7        | 11.2          | 13.2             |

|                |     |     |     |
|----------------|-----|-----|-----|
| <i>Systems</i> | 6.2 | 6.0 | 6.3 |
|----------------|-----|-----|-----|

## 5 Summary of Performance Metrics

| <i>Metric</i>                               | <i>TSL</i> | <i>FSL</i> | <i>July</i> | <i>August</i> | <i>September</i> |
|---|------------|------------|-------------|---------------|------------------|
| Technology serviceability                   | 80%        | 99.2%      | 99.8%       | 99.7%         | 99.9%            |
| Technology MTBF (hours)                     | 200        | 300        | 732         | 732           | 732              |
| Number of AV FTEs                           | 7.5        | 10         | 14.7        | 11.2          | 13.2             |
| Number of training days per month           | 22.5/12    | 30/12      | 23/7        | 23/8          | 23/9             |
| Non in-depth queries resolved within 3 days | 85%        | 97%        | 100.0%      | 100.0%        | 99.0%            |
| Number of A&M FTEs                          | 3.75       | 5.75       | 6.2         | 6.0           | 6.3              |
| A&M serviceability                          | 80%        | 99.6%      | 98.8%       | 100.0%        | 99.4%            |

| <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 provisionally 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 Decananometer 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         |

## 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  |
| b07         | 1            | Simulation of Radioprobing                                       | Dr Charlie Laughton    |

## 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 |

## 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 |

## Appendix C: Life Sciences Projects

### C.1 Modelling Enzyme Catalysis: Mechanism and Dynamics

Fred Claeysens and Adrian Mulholland  
School of Chemistry, University of Bristol

Understanding how enzymes 'work' is a fundamentally important problem in biology. Enzymes are outstandingly efficient natural catalysts. Better understanding of the principles by which they achieve these catalytic properties promises technological spin-offs such as routes to new drugs (e.g. in the design of enzyme inhibitors); analysis of the effects of genetic variation and mutation (e.g. in predicting individual metabolism of pharmaceuticals); design of new catalysts (e.g. biomimetic catalysts or engineered enzymes). Modelling has a central role to play in these developments: unstable species such as transition states and reaction intermediates are crucial to questions of reactivity, and cannot be studied directly by experiment in systems as complex as enzymes. Combined quantum mechanics/molecular mechanics (QM/MM) methods allow enzyme reactions to be modelled: a small region at the active site (where the reaction happens) is treated by a quantum mechanical electronic structure method, and interacts with the protein and solvent environment, which are included more simply (though in atomic detail) by an empirical 'molecular mechanics' force field. This approach combines the simplicity and speed of the MM treatment of the protein structure with the flexibility and power of a quantum chemical treatment (which allows modelling of bond breaking and making, and electronic polarization). Until recently, QM/MM investigations of enzymes have been limited to approximate, error-prone and limited semi-empirical models. Through the application of high-performance computing, it is possible to extend beyond these approximate semi-empirical levels of QM treatment to treat systems where more sophisticated QM methods (such as those based on density functional) are required, e.g. metalloenzymes such as cytochrome P450 (see below). These demanding calculations bring a new level of accuracy to bear on the problem, essential for quantitative conclusions and comparison with experiment. They also make possible the study of new enzyme classes of vital biological importance (e.g. in drug metabolism).

This project has helped us to develop links with companies including Vernalis/British Biotechnology, in the field of enzyme modelling. It has also contributed to obtaining follow-on support from BBSRC (a Committee Studentship and a new e-Science Project, IntBioSim; these projects will run in parallel with this work; we have also been invited to join the BBSRC BioSimGrid e-Science project which will be useful for making our results accessible to the biochemical community). It has also contributed to a current proposal to EPSRC on QM/MM method development for enzyme modelling. AJM has been invited to contribute a review to *Drug Discovery Today* based in part on the work of this project. Work from the project has been presented in a number of invited international conference lectures and seminars in the last year (details below), including presentations to industry. A particular aim has been to reach out to the biological and biochemical community, to demonstrate the practical benefits that state-of-the-art HPC modelling can provide to basic enzymological problems. This is illustrated by the successful applications to BBSRC described above, and invitations to present to the Biochemistry Society, the British Biophysical Society, and the Pro-Bio Faraday Partnership for biocatalysis.

To date, we have modelled the following enzyme systems on the HPCx:

- **Chorismate mutase:** a bacterial enzyme, central to current biological debates on the fundamental nature of enzyme catalysis. We have significantly extended beyond earlier modelling (which used e.g. the T3E at CSAR and local machines), to high levels of QM/MM theory (hybrid density functional theory), analysing multiple reaction pathways. We have also carried out QM/MM molecular dynamics simulations to analyse conformational contributions to catalysis. Thirdly, **using the HPCx, we have carried out what we believe are the**

**highest level QM/MM calculations performed to date for an enzyme** (local CCSD(T) calculations with the VTZ basis set, using a modified version of the MOLPRO program). These calculations are ongoing. Together, these results demonstrate significant transition state stabilization in catalysis, casting light on an important current controversy in enzymology.

- **Cytochrome P450** The cytochromes P450 are a ubiquitous family of haemproteins, which are capable of a variety of oxidation reactions. Cytochromes P450 are involved in a number of vital processes, such as biosynthesis and degradation of xenobiotics, carcinogenesis and drug metabolism. On HPCx, we have carried out calculations for human cytochrome P450 2C9. These include single point QM/MM calculations of the reactive 'compound I' form of the enzyme, and QM/MM structural optimisation of the enzyme with a docked model substrate, using the CHARMM22 MM forcefield with QM treatment at the UB3LYP/6-31g level +LANL2DZ for iron. Modelling of the reaction of real drugs in human P450s is being carried out.
- **TEM1  $\beta$ -lactamase from *E.coli*** Class A  $\beta$ -lactamases are the most common family of a group of bacterial enzymes which cause resistance against the most important and common antibacterial drugs, the  $\beta$ -lactam antibiotics such as penicillin. Using the HPCx, we have been able to extend significantly beyond our previous, preliminary, studies for the first step of the reaction (acylation of the enzyme), which applied lower level (semiempirical QM/MM) calculations. We have performed ab initio QM/MM (HF/3-21G-CHARMM22) optimisation of the Michaelis (substrate) complex of the TEM1  $\beta$ -lactamase, involving a large QM region (56 QM atoms and 3228 MM atoms). Recently, we have gone further, applying energy calculations at the hybrid density functional theory level to generate potential energy surfaces for acylation. The results identify a mechanism of antibiotic breakdown. These studies have demonstrated the need for high-level QM/MM modelling. Based on the modelling, we have suggested modifications of existing  $\beta$ -lactam antibiotics that could improve their stability against lactamases, and so could provide a route to overcoming bacterial antibiotic resistance.
- **Threonyl t-RNA synthetase** Aminoacyl-tRNA synthetases are centrally important enzymes in protein synthesis. Aminoacyl-tRNA synthetases catalyse the aminoacylation (charging) of tRNA. Using HPCx, we have investigated threonyl-tRNA synthetase from *E. coli*, complexed with reactants. We have carried out QM/MM modelling of the reaction pathway for threonyl-tRNA synthetase at the hybrid density functional theory level (B3LYP/6-31G-CHARMM22 with GAMESS-UK/CHARMM). This involved a series of QM/MM energy minimizations for 32 points along the reaction path (each consisting of 300 steps of steepest descent followed by 400 of adopted basis Newton-Raphson optimisation)

### *Method implementation and testing*

We have concentrated on QM/MM calculations on the HPCx via the CHARMM/GAMESS-UK interface. Some initial setup and testing was carried out on our local Myrinet cluster, to establish stable and useful simulation systems prior to modelling on the HPCx. After compilation of the CHARMM/GAMESS-UK code (work of Dr. Paul Sherwood, Daresbury Lab, who has contributed very helpfully to this project) we carried out QM/MM calculations at ab initio and density functional levels as described above, and also performed QM/MM dynamics simulations (see below) on the HPCx. Local MP2 calculations were also run on the HPCx with the MOLPRO program. Local MP2 reaction profiles were obtained via these calculations with VQZ basis sets and local CCSD(T) calculations with the VTZ basis set, producing the highest level of theory QM/MM calculations to date as outlined above.

B3LYP/6-31G(d)-CHARMM27 QM/MM calculations are being performed on the HPCx on chorismate mutase with CHARMM/GAMESS-UK. To investigate the scalability, 50 steps of a dynamics run were performed. These test runs are highly demanding calculations: full molecular dynamics simulations with a high level of QM/MM theory, demanding extensive HPC resources. The QM region consisted of 24 atoms and the MM region of 7086 atoms. For the chorismate mutase system there is an increase in computational speed up to 128 processors, but little benefit

above this level as currently implemented. However, the scaling behaviour is highly dependent on the enzyme system studied. For the  $\beta$ -lactamase enzyme described above, HF/3-21G-CHARMM22 calculations (QM:56 atoms/MM:3228) showed good scaling up to 512 processors. This indicates that the size of the QM region (i.e. the number of atoms to be treated QM) may be a determining factor. This is being explored. We found a considerable improvement in performance in moving to Phase 2 on the HPCx.

### *HPCx resources used to date*

In the year to date, a total of 84755.58 AU were used on the HPCx. Approximately 80% of the budgets are still available. This is in line with our scientific goals for the coming year. Having set up and tested various enzyme systems for QM/MM reaction modelling, and classical molecular dynamics simulations, intensive calculations on these important enzymes are in progress.

### *Workplan for Year 2*

We have demonstrated that high-level QM/MM calculations can be performed on enzyme reactions. We will extend our studies to examine reaction pathways in the biologically and medically important enzymes discussed above. The results will give detailed insight into the fundamental catalytic mechanisms in each case. The results will be tested by correlated ab initio QM/MM calculations. Studies of the of iron metalloenzyme lipoxygenase, an important factor in cell signalling, are beginning (QM models have been constructed prior to QM/MM modelling). In collaboration with Dr. Sherwood, we plan to test the newly developed replica path method in GAMESS-UK/CHARMM, which promises to increase the scalability and efficiency of the calculations. For example, 20 separate points on the path are modelled **simultaneously**. This allows calculation of a complex reaction pathway through one linked calculation. Through this method, we will employ for example 640 processors. This approach will be compared with current techniques for enzyme reaction pathway calculations. In addition to the QM/MM studies of reactions, classical molecular dynamics simulations of enzymes are underway. Dr. Claeysens has prepared two enzymes for classical molecular dynamics simulations with CHARMM. He is also exploring the use of NAMD as an alternative. We aim to contribute to the benchmarking efforts of the HPCx Terascaling Team in this area, by developing a standard protein molecular dynamics benchmark. The scalability of these calculations will be assessed.

### *Publications*

- 'Differential Transition State Stabilization in Enzyme Catalysis: Quantum Chemical Analysis of Interactions in the Chorismate Mutase Reaction and Prediction of the Optimal Catalytic Field.' B. Szeferczyk et al. *J. Am. Chem. Soc.*, accepted for publication (JA049376t).
- 'Mechanism and Structure-Reactivity Relationships for Aromatic Hydroxylation by Cytochrome P450' C.M. Bathelt et al. *Organic and Biomolecular Chemistry*, in press (2004) (B410729B)
- 'Conformational effects in enzyme catalysis: QM/MM free energy calculation of the 'NAC' contribution in chorismate mutase' K.E. Ranaghan & A.J. Mulholland *Chem. Commun.* (2004) (10), 1238-1239
- 'MM and QM/MM Modeling of Threonyl-tRNA Synthetase: Model Preparation, Testing and Simulations', J. Zurek et al. *Structural Chemistry*, 15, 405-414 (2004).

### *Other articles*

The following articles have featured work carried out in this project:

- 'Much ado about enzyme mechanisms' *Chem. Eng. News*, 82, 35-39 Feb. 2004.
- 'Reactions in Action' *University of Bristol Research Review*, March 2004

### *Reports in progress:*

- 'Modelling enzyme reaction mechanisms, specificity and catalysis', A.J. Mulholland, *Drug Discovery Today*

- 'Modelling biological systems' A.J. Mulholland in Royal Society of Chemistry Specialist Periodical Report *Chemical Modelling Applications and Theory. Volume 4* (Dr. A. Hinchliffe, Ed.)
- 'DFT-QM/MM study of chorismate mutase shows unequivocal evidence of specific transition state stabilization', F. Claeysens et al., to be submitted to *Journal of the American Chemical Society*.
- 'QM/MM studies of chorismate mutase beyond the DFT level of theory', F. Claeysens, K.E. Ranaghan, F.R. Manby, J.N. Harvey and A.J. Mulholland, to be submitted to *Journal of the American Chemical Society*.
- 'QM/MM modelling of the deacylation mechanism of the TEM1  $\beta$ -lactamase from *E. coli*' J. Hermann *et al.*, to be submitted to the *Journal of Molecular Biology*.

### **Presentations**

The following invited lectures have or will feature work from this project:

- 'Modelling Biological Catalysis', NSF/Polish Chemical Society/Czech Chemical Society Workshop, Prague, September 2005.
- 'Quantum Chemistry Applied: from  $H_3$  to Biocatalysis', a conference in honour of Per Siegbahn's 60<sup>th</sup> birthday, Stockholm, 18-21 June, 2005.
- 'Protein Dynamics and Function' British Biophysical Society/Biochemistry Society, Leicester, January 2005
- 'From Structural Genomics to Drug Discovery', Parma, Italy, September 2004.
- 'Quantum/Classical Calculations in Chemistry and Biophysics' American Chemical Society National Meeting, Philadelphia, PA, August 2004.
- 'Computational enzymology', NSF/Polish Chemical Society/Czech Chemical Society Workshop on Modelling and Design of Molecular Materials, Wroclaw, Poland, September 2004.
- 'Computational enzymology: insights into catalysis from modelling' Gordon Conference on Computational Chemistry, Plymouth, New Hampshire, July 2004.
- 'Quantum enzymology', 2004 ISQBP President's Meeting, Como, Italy, May 2004
- UK/US N+N Workshop on *High Performance Computing for Biomolecules & Materials*, Washington DC, April 2004.
- HPCx Industry Day, CLRC Daresbury Laboratory, Cheshire, April 2004.
- 'High-performance computing applied to biological problems: modelling enzyme catalysis'. UK Computational Chemistry Working Party Meeting, King's College London. January 2004.
- 'Modelling enzyme-catalysed reaction mechanisms', Biocatalysis Meeting of Faraday Pro-Bio Partnership, Edinburgh, November 2003.

### **Invited seminars at:**

- ETH, Zürich (December 2004)
- Interdisciplinary Project in Cellular, University of Warwick, (February 2004)
- University of Surrey, (November 2003)

## C.2 Towards a Virtual Outer Membrane

Prof. Mark S.P. Sansom & Dr. Jorge Pikunic  
Department of Biochemistry, University of Oxford

### *Progress Report – the First Year*

Our work is aimed at progressing molecular dynamics simulations of biological systems from the single molecule to the sub-cellular level. We see this as an essential first step towards biomolecular systems biology (i.e. integration between different levels of description of complex biomolecular systems) [1].

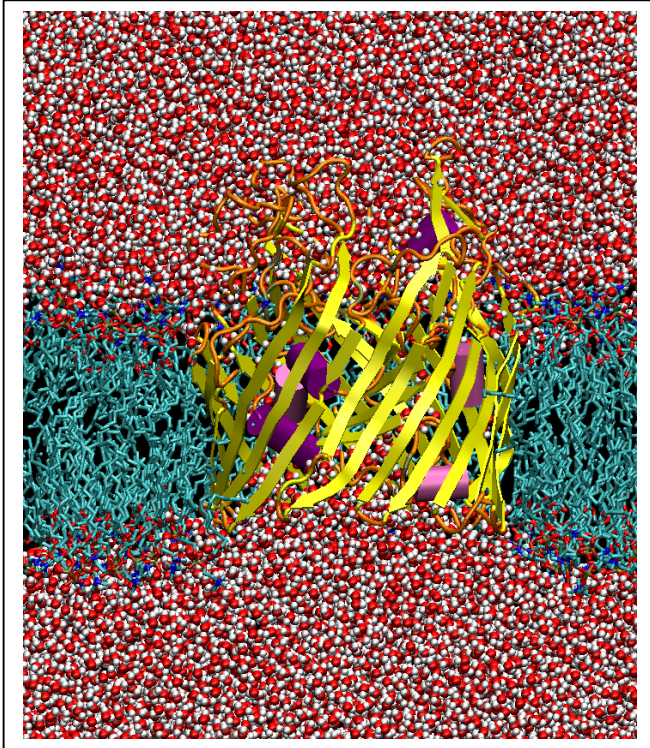
Capability computing is essential in achieving our goals for this project. Therefore, our first step was to select the molecular dynamics simulation package with best scalability for the size and type of systems that we are studying. We have chosen the outer membrane protein BtuB, the vitamin B<sub>12</sub> transporter in *E. coli*, as a benchmark system for the scalability study. The atomic interactions involved in this system resemble those that we expect to have in our model outer membranes, making it a suitable benchmark. The size of the resulting system (see Figure 1) is ca. 100,000 atoms. We are collaborating with Dr. Joachim Hein, from EPCC, in measuring the performance of the available simulation packages (Gromacs, NAMD, CHARMM, Amber, and DL-POLY) on the HPCx supercomputer. Results of this study will not only be useful to us. They will also provide a valuable guide to other supercomputer users. Preliminary results indicate that NAMD scales linearly up to ca. 170 processors on HPCx. By contrast, Gromacs and CHARMM scale linearly only up to tens of processors.

Aside from the suitability of BtuB as a benchmark system, this protein and other TonB-dependent transporters are of biomedical importance as potential targets for new antibiotics, due to their ubiquity. However, the function of these proteins, particularly the release mechanism of vitamin B<sub>12</sub> into the periplasm, is still not well understood. We have performed 10 ns of molecular dynamics simulations of our benchmark system starting from the crystal structure of BtuB. We found that the motion of the N-terminus, in the vicinity of the segment responsible for signalling the presence of vitamin B<sub>12</sub> in the periplasm, is strongly correlated to that of two extra-cellular loops. The mobility of these loops is dramatically reduced upon vitamin B<sub>12</sub> binding, “locking” the N-terminus in a specific conformation. Principal component analysis of the trajectories reveals the details of the concerted motion between the extra-cellular loops and the N-terminus (Figure 2), revealing the possible transmembrane signalling mechanism that ultimately triggers the machinery of vitamin B<sub>12</sub> release into the periplasm.

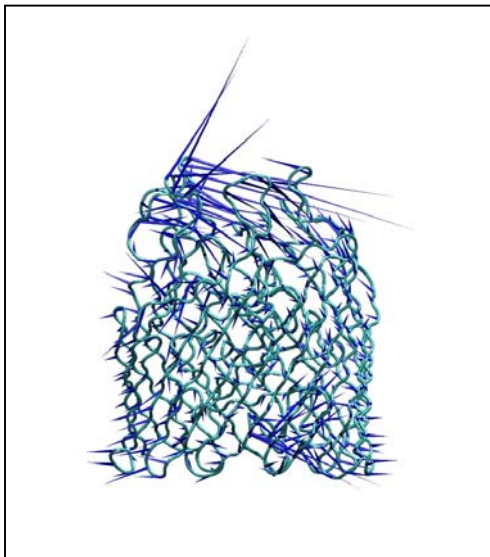
The extracellular leaflet of outer membranes is composed mainly of lipopolysaccharides (LPS). For the sake of greater biological realism, we have built a molecular model of lipid A, the only conserved motif of LPS, from *E. coli*. The topology and interaction parameters of our model are based on the CHARMM force field, and thus can be easily used within NAMD to perform large scale molecular dynamics simulations of lipid A in conjunction with other biomolecules. We are currently assembling our virtual outer membrane, with lipid A in the outer leaflet and phospholipids in the inner leaflet. We plan to characterise this model membrane (i.e., structure, dynamics, and thermodynamics) and study the conformational dynamics of single proteins embedded in it. We will then insert a matrix of proteins to complete a patch of outer membrane.

## References.

- [1] Arinaminpathy, Y., Biggin, P.C., Bond, P.J., Domene, C., Pang, A. and Sansom, M.S.P. (2003) Large scale biomolecular simulations: current status and future prospects. *Proc. UK e-Science All Hands Meeting 2003 (ISBN 1-904425-11-9)* pp. 901-907
- [2] D.P. Chimento, A.K. Mohanty, R.J. Kadner, M.C. Wiener, *Nature Structural Biology* **10**, 394 (2003)



*Figure 1.* Benchmark system: BtuB (shown in cartoon representation) embedded in a lipid bilayer, solvated with water.



*Figure 2.* Porcupine plot of the 1<sup>st</sup> eigenvector (dominant mode of motion). The extra-cellular loops move towards (or away from) the centre of the protein, named the hatch domain, while the periplasmic turns move away from (or towards) the hatch domain. This motion controls the position of the N-terminus, and provides a description of the basis of the transmembrane signalling mechanism.

## C.3 Quantum Directed Virtual Evolution

Marcus Durrant  
John Innes Institute

Recent support from Daresbury Laboratory has focussed on:

### *Job submission and execution*

Since July we have worked on the task farming harness, based on an approach in which many individual GAMESS-UK calculations are bundled together to form a large HPCx batch job (option 3 in the list provided in the July report). The task-farmed job creates a number of MPI communicators at the outset, and a single process is assigned to dynamic task distribution. The harness is now running on the HPCx system, and we are making some adjustments to improve robustness, particularly for those situations when some of the individual job inputs contain fatal errors, or where GAMESS-UK fails to complete the calculation for unexpected reasons. The code incorporates the new GAMESS-UK SCF kernel implemented within the MPI/ScaLAPACK framework by Ian Bush; this promises to provide better scalability for parallel execution than the original MPI-based code, which did not incorporate parallel linear algebra. Some additional work (including some reported in the last period) was needed to ensure the new driver had all the functionality required for the current project.

### *Output Parsing*

We have written python scripts to extract the required data and provide the structure data in PDB format.