High performance research

Boosting results with advanced computing techniques

Our genetics, chemistry and biomedicine software collaborations

Also in this issue

Accelerated 3d modelling of buildings
From the directors

Welcome to the Summer issue of EPCC News. We hope you enjoy it.

After almost a year in development, we are pleased to announce that we have launched our new website! A few of our goals were to make our website faster, simpler to navigate and easier to maintain. Most of all though, we wanted to help our collaborators and clients get to know us better. We hope we have succeeded. On our new website you will also find our blog (see below), which conveys the wide range of our activities. We would like to thank Basestation for their excellent contribution to our website project.

Through our work in a number of exascale projects, we had often talked about starting a new conference to bring together those involved in solving the particular software challenges associated with programming exascale systems. And so in April we ran EASC2013: First Exascale Applications and Software Conference. We were delighted to welcome over 130 international delegates here to Edinburgh and thanks to the success of the event we are now looking forward to EASC2014 which will be held at KTH in Stockholm.

Finally, we want to mention the APOS-EU project which has successfully concluded after two years. Funded as part of the EU's Framework 7 “EU-Russia” call to foster collaborative research between European and Russian scientists, the project was led by EPCC. APOS-EU was extremely enjoyable for us and we have forged some good friendships with fellow HPC researchers in Russia. We hope to continue our collaborative research in the near future.

We would like to know what you think about this newsletter and our website: feel free to send us any comments you may have.

Всего наилучшего!

Mark Parsons & Alison Kennedy
EPCC Executive Directors
a.kennedy@epcc.ed.ac.uk
m.parsons@epcc.ed.ac.uk

EPCC’s new blog

Want to find out what we’re doing? Read our new staff blog: www.epcc.ed.ac.uk/blog

The EPCC blog was launched alongside our redesigned website. Posts are written by staff and cover the wide range of EPCC’s activities, from technical work to training and outreach.

Recent subjects include:
- Preserving digital art
- A new MPI library written in C#
- A report on an OpenACC event in Sweden.

Contact us
www.epcc.ed.ac.uk info@epcc.ed.ac.uk +44 (0)131 650 5030

EPCC is a supercomputing centre based at The University of Edinburgh, which is a charitable body registered in Scotland with registration number SC005336.
Scaling to thousands of GPUs on Titan

We have been among the first researchers to take advantage of the massive amounts of computing power available on Titan, the world’s fastest supercomputer, based at Oak Ridge National Laboratory.

The full machine will boast 18,000 GPUs, and just under half of these have been made available recently. We have shown that our highly scalable “Ludwig” soft matter physics application can efficiently take advantage of at least 8192 GPUs in parallel.

This application, which is able to simulate a wide range of complex fluids, will soon be put into production use on Titan to improve our understanding of the behaviour of systems containing the combination of liquid crystals and colloidal particles, with potential benefits for a wide range of technological areas. The code uses a combination of Lattice Boltzmann and finite difference techniques.

These latest performance results were recently presented at the EASC2013 conference in Edinburgh and the work was described in EPCC News issue 72.

We’ve won gold!

EPCC has been awarded a Gold in the University’s Sustainability Awards

EPCC has been involved with the awards for a few years now. Last year we won a Silver and this year we achieved the top award, Gold.

This has taken a lot of work by everyone at EPCC. We’ve had to think about so much: from small things like turning off lights and promptly reporting repairs, to larger issues like thinking about how we best use our staff and our computer resources.

Sustainability is something we are committed to in EPCC because to us it means much more than simply putting things in the correct bin: it concerns most areas of our business. Being sustainable allows us to adapt and change along with the area in which we operate.

Find out more
www.ed.ac.uk/about/sustainability/on-campus/awards
Profiles

Two of our colleagues talk about their work here at EPCC

I conduct a research programme on soft condensed matter physics and complex fluids, collaborating with a number of researchers at EPCC and the University of Edinburgh, the Centre for Numerical Algorithms and Intelligent Software as well as external partners.

Our field of research requires sophisticated modelling techniques, often involving multiscale approaches and high-performance computing. This is because soft materials tend to self-organise into more complex structures which are difficult to predict, but crucially influence their macroscopic behaviour.

A good part of our research is based on EPCC’s own Lattice Boltzmann simulation package “Ludwig”. Ludwig has been around for more than a decade and was born out of the necessity for a suitable code environment where researchers can implement their own specific models without having to start software development from scratch. Over the years, Ludwig’s functionality has grown considerably. This can make software management a difficult task, but also offers unprecedented opportunities to the researcher.

Together with EPCC’s Kevin Stratford and Alan Gray, we are developing models for charged soft condensed matter and new programming models for heterogeneous computing architectures. With our recent simulation of liquid-crystalline composite materials, we were among the first to take advantage of the massive amounts of computing power available on the “Titan” supercomputer based at Oak Ridge National Laboratory (see previous page). A public release of Ludwig is scheduled for next year.

Oliver Henrich
Advanced Fellow EPCC
ohenrich@epcc.ed.ac.uk

Thanks to EPCC and the local research infrastructure, we can draw on both scientific and computational expertise and address substantially new aspects of soft condensed matter.

After completing my PhD in Computer Science and specialising in programming language design for HPC, I worked in industry for a number of years as an IT consultant. I was keen to get back into the university environment and the challenges that it provides.

I now work in EPCC’s Applications Group where I am investigating innovative approaches to scaling current algorithms which are used as the building blocks of many scientific problem solvers. Not only are supercomputers becoming ever more powerful, but we are more ambitious in what we do with them. To meet future challenges, it is therefore critically important that we improve the ways in which we solve problems.

I am also involved in an outreach activity where we demonstrate the power of supercomputing by allowing the public to race virtual dinosaurs at science events. The dinosaur demo is built upon a simulation code created by the Faculty of Life Sciences at Manchester University and with their help we have modified it to allow people to design their own creatures and run the simulations on HECToR, the UK national supercomputing service.

Dinosaurs are a great illustration of HPC because, whilst paleontologists have developed plenty of theory, the fact that the animals themselves have been extinct for 65 million years means there is no way to examine this in practice. Instead, to test the theory, simulations of these creatures can be used. And to keep the public interested, there is the competitive element of who can design the fastest dinosaur and what this would look like.

Nick Brown
Applications Developer
n.brown@epcc.ed.ac.uk

The work at EPCC is so varied and there is plenty of opportunity to take part in unique projects and work with some really interesting people.
SPRINT: taking biomedical analysis from the desktop to supercomputers and the cloud

SPRINT website
www.r-sprint.org/

The statistical environment, R, is key to many of the workloads in biomedical sequence analysis and systems biology.

Making R scalable as datasets grow is of great interest to researchers. To achieve this, SPRINT was established in 2008 to provide a set of easy to use, 'drop-in' parallelised functions allowing R users simple access to high performance computing (HPC).

Since the Apple Mac is the preferred platform for many researchers in this field, in December 2012 SPRINT was migrated to Mac OS X to enable researchers to exploit its multi-core capability. SPRINT now offers such researchers an easy route from their desktop up to supercomputers and the cloud.

With that in mind, the SPRINT team recently assisted staff at the Cancer Research UK Cambridge Research Institute to install SPRINT on their researchers’ Apple Macs. Also, in April 2013, three members of the SPRINT team spent the day at the Institute discussing the use of HPC in R data analyses with various researchers. The team also gave a seminar on SPRINT, its future development and its use in post-genomics workflows at the Division of Pathway Medicine. This seminar was attended by researchers from Cancer Research UK, the University of Cambridge and the European Bioinformatics Institute.

To further help researchers, in conjunction with PRACE (the Partnership for Advanced Computing in Europe), the SPRINT team organised the second SPRINT training course in May 2013. This used HECToR, the UK’s national supercomputing service, as the platform for the training course attendees. The course included an introduction to HPC concepts and an exercise to familiarise attendees with running a job on HECToR. The rest of the course focused on running SPRINT examples on HECToR as well as topics on installation, features, setup and use of SPRINT on a desktop machine.

Registration for the course had to be closed 10 weeks early due to the high volume of applicants, who came from Heriot-Watt, Edinburgh, Dundee, St Andrews, Roslin, Cambridge, Sussex, UCL, Institute of Cancer Research, Glasgow, Oxford, Bristol, Nottingham and Leeds. A further course is therefore planned for the near future.

SPRINT is a collaborative project between EPCC and the University of Edinburgh’s Division of Pathway Medicine.

Recent journal articles about SPRINT by members of the project team:


M. Piotrowski et al., 2013, “Exploiting Parallel R in the Cloud with SPRINT.” Methods of Information in Medicine 52 (1), 80-90.


The Altmetric score for an article is one measure of the quality and quantity of online attention that this article has received. The original December 2008 SPRINT article in BMC Bioinformatics now has an Altmetric score of 9.23. This means it is in the top 10% of all 1.2m articles tracked by Altmetric and is ranked 80th of all the 1,073 articles in the journal itself.

Terry Sloan
tms@epcc.ed.ac.uk

The statistical environment, R, is key to many of the workloads in biomedical sequence analysis and systems biology.

Making R scalable as datasets grow is of great interest to researchers. To achieve this, SPRINT was established in 2008 to provide a set of easy to use, 'drop-in' parallelised functions allowing R users simple access to high performance computing (HPC).

Since the Apple Mac is the preferred platform for many researchers in this field, in December 2012 SPRINT was migrated to Mac OS X to enable researchers to exploit its multi-core capability. SPRINT now offers such researchers an easy route from their desktop up to supercomputers and the cloud.

With that in mind, the SPRINT team recently assisted staff at the Cancer Research UK Cambridge Research Institute to install SPRINT on their researchers’ Apple Macs. Also, in April 2013, three members of the SPRINT team spent the day at the Institute discussing the use of HPC in R data analyses with various researchers. The team also gave a seminar on SPRINT, its future development and its use in post-genomics workflows at the Division of Pathway Medicine. This seminar was attended by researchers from Cancer Research UK, the University of Cambridge and the European Bioinformatics Institute.

To further help researchers, in conjunction with PRACE (the Partnership for Advanced Computing in Europe), the SPRINT team organised the second SPRINT training course in May 2013. This used HECToR, the UK’s national supercomputing service, as the platform for the training course attendees. The course included an introduction to HPC concepts and an exercise to familiarise attendees with running a job on HECToR. The rest of the course focused on running SPRINT examples on HECToR as well as topics on installation, features, setup and use of SPRINT on a desktop machine.

Registration for the course had to be closed 10 weeks early due to the high volume of applicants, who came from Heriot-Watt, Edinburgh, Dundee, St Andrews, Roslin, Cambridge, Sussex, UCL, Institute of Cancer Research, Glasgow, Oxford, Bristol, Nottingham and Leeds. A further course is therefore planned for the near future.

SPRINT is a collaborative project between EPCC and the University of Edinburgh’s Division of Pathway Medicine.

Recent journal articles about SPRINT by members of the project team:


M. Piotrowski et al., 2013, “Exploiting Parallel R in the Cloud with SPRINT.” Methods of Information in Medicine 52 (1), 80-90.


The Altmetric score for an article is one measure of the quality and quantity of online attention that this article has received. The original December 2008 SPRINT article in BMC Bioinformatics now has an Altmetric score of 9.23. This means it is in the top 10% of all 1.2m articles tracked by Altmetric and is ranked 80th of all the 1,073 articles in the journal itself.

Terry Sloan
tms@epcc.ed.ac.uk

The statistical environment, R, is key to many of the workloads in biomedical sequence analysis and systems biology.

Making R scalable as datasets grow is of great interest to researchers. To achieve this, SPRINT was established in 2008 to provide a set of easy to use, 'drop-in' parallelised functions allowing R users simple access to high performance computing (HPC).

Since the Apple Mac is the preferred platform for many researchers in this field, in December 2012 SPRINT was migrated to Mac OS X to enable researchers to exploit its multi-core capability. SPRINT now offers such researchers an easy route from their desktop up to supercomputers and the cloud.

With that in mind, the SPRINT team recently assisted staff at the Cancer Research UK Cambridge Research Institute to install SPRINT on their researchers’ Apple Macs. Also, in April 2013, three members of the SPRINT team spent the day at the Institute discussing the use of HPC in R data analyses with various researchers. The team also gave a seminar on SPRINT, its future development and its use in post-genomics workflows at the Division of Pathway Medicine. This seminar was attended by researchers from Cancer Research UK, the University of Cambridge and the European Bioinformatics Institute.

To further help researchers, in conjunction with PRACE (the Partnership for Advanced Computing in Europe), the SPRINT team organised the second SPRINT training course in May 2013. This used HECToR, the UK’s national supercomputing service, as the platform for the training course attendees. The course included an introduction to HPC concepts and an exercise to familiarise attendees with running a job on HECToR. The rest of the course focused on running SPRINT examples on HECToR as well as topics on installation, features, setup and use of SPRINT on a desktop machine.

Registration for the course had to be closed 10 weeks early due to the high volume of applicants, who came from Heriot-Watt, Edinburgh, Dundee, St Andrews, Roslin, Cambridge, Sussex, UCL, Institute of Cancer Research, Glasgow, Oxford, Bristol, Nottingham and Leeds. A further course is therefore planned for the near future.

SPRINT is a collaborative project between EPCC and the University of Edinburgh’s Division of Pathway Medicine.

Recent journal articles about SPRINT by members of the project team:


M. Piotrowski et al., 2013, “Exploiting Parallel R in the Cloud with SPRINT.” Methods of Information in Medicine 52 (1), 80-90.


The Altmetric score for an article is one measure of the quality and quantity of online attention that this article has received. The original December 2008 SPRINT article in BMC Bioinformatics now has an Altmetric score of 9.23. This means it is in the top 10% of all 1.2m articles tracked by Altmetric and is ranked 80th of all the 1,073 articles in the journal itself.
CP2K-UK: Enabling a step change in UK materials science

EPCC is leading a new project that aims to improve the usability of the CP2K code and create a thriving developer community.

CP2K [1] is a freely available open-source program for atomistic simulation, best known for its implementation of the ‘Quickstep’ [2] Density Functional Theory method. However, CP2K provides a much wider range of capabilities, from classical potentials and semi-empirical models to hybrid DFT functionals, and Møller-Plesset 2nd order perturbation theory.

Built on top of these models are an extensive set of simulation tools including molecular dynamics, Monte Carlo, nudged elastic band, path integrals and free energy methods. Significant effort has been spent on obtaining good parallel performance on modern HPC architectures, and calculations on over 60,000 CPU cores have been demonstrated [3]. As a result, there are a growing number of research groups in the UK who use CP2K spanning the fields of Computational Chemistry, Materials Science, Solid State Physics, and Biochemistry. CP2K is currently the second most heavily used code on the national service HECToR, and usage is growing - to date, HECToR compute time worth over £3.15 million has been consumed by CP2K calculations.

Although CP2K is a well-engineered, high performance toolkit with a wide range of applications, it has two drawbacks that hinder wider uptake and more efficient research using the code. Firstly, CP2K is an intimidating prospect for new users (and developers) due a lack of high level documentation, the large number of options available via the complex input file format, and the size of the code (over 900,000 lines of Fortran 95). What expertise exists is isolated within individual research groups. Secondly, while CP2K is designed to be extensible and allow users to add functionality required for their own research, to do so is a complex task and requires detailed collaboration with key members of the existing development team. While there are many capabilities that UK users require to be added to the code in order to tackle key research problems, they have been unable to do so thus far.

To overcome these issues EPSRC has funded a five-year project, led by myself as Principal Investigator, with co-Investigators Prof. Lev Kantorovich (KCL), and Dr. Ben Slater and Dr. Matt Watkins (UCL) to improve the usability of CP2K, build a sustainable community to share best practice in using CP2K, and embed development capability within UK research groups. Starting in September 2013, the ‘CP2K-UK’ project will offer training, knowledge transfer and networking activities for users and developers of CP2K.

laid Bethune
ibethune@epcc.ed.ac.uk

If you are interested in learning more about the capabilities of CP2K, how you can apply it to your research, or need new functionality added to the code, we’d love to hear from you!


via a series of annual meetings. We will also directly improve the usability of the code by developing user-level documentation, case studies, and integrate the code with widely used pre- and post-processing tools. Most importantly we will develop new algorithms within CP2K improving the performance of code and widening the range of problems it can be used to tackle. As well as benefitting users directly, this will demonstrate that with the support of the CP2K-UK network, it is possible for new developers to quickly gain the skills necessary to modify and extend the code to suit their own needs.

We have already begun forming a wide network of interested users of CP2K, including researchers at Imperial College London, Queen’s University Belfast, The University of Cambridge, and the University of Edinburgh in addition to research groups closely associated with the project’s investigators at University College London and King’s College London. We have also established collaboration with of the existing CP2K developers centred around the groups of Prof. Jürg Hutter and Prof. Joost VandeVondele in Zürich, who are fully supportive of our effort to establish a vibrant community of CP2K users and developers in the UK.

**CP2K user feedback**

"The research in my group is almost completely dependent on CP2K. The area of our research is computational electrochemistry. Atomistic simulation of electrochemical interfaces is at the frontier of condensed phase computational science, requiring the latest in electronic structure calculation methods in combination with extensive dynamics sampling. CP2K is one of the very few packages that allows us to carry out such calculations on currently available facilities for large scale computation (such as HECToR)."  **Prof. Michiel Sprik, University of Cambridge**

"My group already makes extensive use of the CP2K code. Its ability to scale well to larger systems and the extremely efficient hybrid functional implementation have allowed us to study problems in disordered systems like amorphous silica that would have been out of reach (or at least required excessive computational resources) using other codes.

"Despite these benefits, we have been held back from fully exploiting the code by some key issues that this [project] nicely addresses. I and my colleagues are looking forward to attending the meetings of the network. I believe this would be an excellent opportunity for my students and other group members to share experience and good practice as well as establish scientific links with other UK groups on using CP2K."  **Prof. Alexander Shluger, University College London**

---

Charge density difference induced by Co-Salen molecule interacting with NiO(100) surface. Reprinted with permission from A. Schwarz et al, J. Phys. Chem. C 2013 117 (2), 1105-1112. Copyright 2013 ACS.
Thinking about HPC operating systems

CRESTA’s most recent piece of work focused on the operating systems used by the largest supercomputers, past and present, and considered the future of operating systems and system software for HPC machines.

Modern operating systems have evolved incrementally from simple beginnings in the 1970s when the primary goal was time-sharing a single processor so that it could run two programs at the same time. The operating system and, more generally, the system software in a modern computer is now responsible for the smooth and efficient use of multiple CPUs, possibly with features like simultaneous multi-threading and out-of-order execution, by dozens of programs all vying for limited resources. Consequently, scheduling algorithms are now incredibly complex. Programs can collectively ask for and use more memory than is physically present in the computer because the memory management component supports virtual memory: using a file on a hard-disk to store data that doesn’t fit into physical memory. Device drivers continually manage every hardware device in the computer, responding to events such as data arriving from a network connection or working out where the mouse-pointer should be positioned on the screen when the mouse is moved by the user.

Most of these activities are not needed for HPC applications running on HPC machines. Dynamically scheduling processes requires running processes to be paused, moved and re-started; this overhead is generally avoided by pinning them to specific CPU cores. Virtual memory is very slow because hard-disks are very slow – and a scalable parallel application can obtain more physical memory by requesting more processors. Responding to regular hardware events can impact the performance of applications far more than predicted by just adding up the time taken to deal with the interruptions. If the interrupts resonate with application activities then the negative effects on performance can be amplified in ways that are still not fully understood.

Current research and development in this area can be categorised into three approaches:

1. Start with a standard operating system – a full-weight kernel (FWK) – and strip it down by removing unnecessary system activities and services.
2. Start from scratch and build a new minimal operating system – a light-weight kernel (LWK) – with limited functionality, usually for specific hardware.
3. Start with two existing operating systems – a FWK and a LWK – and merge them by modifying both so that they can inter-operate.

Within the Top 500 supercomputers, the most common of these approaches is to use a full-weight kernel; of these

As part of the co-design activities within CRESTA, we are examining the full software stack – from the six representative applications, through programming models and compilers, to the operating system – and evaluating the readiness for exascale computing.
Disruptive technologies may be needed to solve scalability and performance issues, in operating systems and hardware, without sacrificing usability.

Current research includes:
- CNL (Compute Node Linux), Cray
- CNK (Compute Node Kernel) and FusedOS, IBM
- Catamount and Kitten, Sandia Laboratories
- Singularity, Microsoft

machines, 4 use Microsoft Windows and 469 of them use Linux. The Compute Node Linux (CNL), part of the Cray Linux Environment (CLE), is a particular example of the first approach, where standard SUSE Linux is customised to significantly reduce unnecessary overheads.

The operating system for the BlueGene series of machines, Compute Node Kernel (CNK), is a good example of the light-weight kernel approach. Its design and creation was influenced by two decades of operating system research at Sandia National Laboratories, which is currently working on a light-weight kernel called Kitten and has previously produced Catamount and Portals.

IBM now has a prototype for a new operating system, called FusedOS, which merges two existing operating systems so that compute-intensive application processes run on a light-weight kernel and any system calls that cannot be handled locally are forwarded and handled by a full-weight kernel.

The current trend in operating system research and development of re-implementing existing APIs is likely to continue. This development is incremental and driven by advances in hardware as well as the necessity to improve the operating system to make full use of current technologies. Unfortunately, improvements that enhance scalability of the operating system often reduce usability. This method of operating system development will provide scalability for the immediate future but it is likely to be limited by the original design decisions of modern HPC technology. Developments in hardware, operating systems, programming models and programming languages are all interdependent, which leads to cyclical improvements rather than novel approaches. The abstractions that have held true for hardware for several decades are no longer adequate to describe modern hardware. For example, procedural languages such as C and FORTRAN, assume single-threaded, sequential processing and memory isolation enforced by hardware protection. Operating systems now depend on this hardware protection mechanism to isolate the memory spaces for different processes, which requires an expensive context-switch when transferring control from one process to another. This cannot be avoided unless a disruptive technology breaks the dependency by introducing a novel way to protect process memory spaces. Similarly, disruptive technologies may be needed to solve other scalability and performance issues, in operating systems and hardware, without sacrificing usability.

Find out more:
www.cresta-project.eu
Fed4FIRE: First Call for Future Internet Experiments

The Fed4FIRE project which started in October 2012 with the aim of federating existing infrastructure test-beds for Future Internet Research and Experimentation (FIRE) has launched its first open call for experiments.

Under the open call, which opened on 1st May and closes at the end of July, researchers can apply for up to €80,000 to fund experiments of up to 1 year’s duration.

The call is targeted at researchers with innovative ideas for experiments that cross Future Internet domains. In addition, further facilities may join the federation so that its scope can be enriched.

Experimentally-driven research is a key factor for growing the European Internet industry. To enable this type of activity, a number of projects for building a European facility for Future Internet Research and Experimentation (FIRE) have been launched (see BonFIRE article on page opposite). Each project targets a specific community within the Future Internet ecosystem, such as wired and wireless IP networks, cloud computing, software-defined networks and innovative Internet services.

Fed4FIRE aims to provide a framework that will allow features of these multiple test facilities to be accessible in single experiments. Fed4FIRE will provide tools and mechanisms to discover, control and monitor resources in a trustworthy environment. The results will benefit researchers by making the experiment lifecycle much easier to manage in a heterogeneous environment, while infrastructure operators will see increased use of their facilities by ambitious experiments.

Fed4FIRE has put forward sample scenarios to inspire experimenters. Case-studies include the smart stadium, in which services based on real-time data are provided in the context of a stadium hosting a major sporting or cultural event. Such services will rely on interacting heterogeneous infrastructures, service platforms and different access networks. In addition to the new commercial opportunities of such a scenario, there are clear public safety benefits, and opportunities to interact with other infrastructures such as transport systems.

Other Fed4FIRE use-cases include providing resources for large-scale online gaming with dynamic requirements, public safety and information management during an emergency and a scenario whereby data about the urban environment is collected and shared by a large population. We anticipate that researchers will come up with other imaginative and challenging ideas for experiments.

Successful applicants to the first Fed4FIRE open call will be able to access the federated infrastructure early in 2014. A second open call is planned for 2014. The Fed4FIRE project continues until 2015, and is expected to be self-sustaining into 2016 and beyond.
BonFIRE is a multi-cloud facility created to foster experimentation and testing of Cloud and distributed applications.

BonFIRE is a multi-cloud facility created to foster experimentation and testing of Cloud and distributed applications. You can read in previous issues of EPCC News and on the EPCC website about how BonFIRE includes 6 sites across Europe, allows advanced Cloud and network experimentation and has already hosted large-scale, impacting research.

Now you can use BonFIRE yourself, and access to the resources will cost you nothing. You can apply for peak usage of up to 50 cores and 250GB of storage for a period up to four months.

**Benefits of BonFIRE**

Some money could buy you these on a public cloud, but BonFIRE offers things you can’t get elsewhere.

For example, BonFIRE tells you on which physical host your VMs are running and gives you monitoring data from these physical hosts. It tells you the exact specification of the compute clusters and lets you place your VMs exactly where you want them. You can also get exclusive access to our physical hosts and even use our Controlled Contention and Malicious Patterns generators to really test your software. If you are interested in network effects on your application, you can use the Virtual Wall facility to emulate the network conditions that concern you and see how your software reacts to network outage and starvation. And BonFIRE’s complementary graphical, command-line, programmatic and declarative interfaces make it easy for you to get started and progressively escalate the complexity of your experiments.

So, if you are frustrated with black-box computation, limited by the size of your own experimentation cluster, excited by the opportunity to take your application to its limits, or even curious to play on a multi-cloud experimentation facility, try the BonFIRE Open Access.

Join BonFIRE!

The 3-step process is really easy and quick to follow and we would love to receive your application. [www.bonfire-project.eu/involved](http://www.bonfire-project.eu/involved)
Each human shares 99.9% of their genome with every other human, and that is why we all look similar to each other and unlike, for example, a cabbage (although it is interesting to note that 57% of our genome is shared with a cabbage). But since there are several billion building blocks, a 0.1% difference still corresponds to many variations (and that is why you don’t look exactly like anyone else unless you have an identical twin, who shares your genome exactly). The term for such a variation in a single G, A, T or C building block is a “single-nucleotide polymorphism” (SNP, pronounced snip).

These variations are not only responsible for differences in visual appearance, but also play a major role in other characteristics such as health, personality and intelligence. Our genetic makeup does not completely determine which traits we display, but is instead involved in a complex interplay with our environment and experience. A few traits are highly dependent on specific SNPs, but most are influenced by many different SNPs. For these so-called complex traits, the small effects of each of the SNPs combine to form a large overall genetic contribution.

EPCC is working with Dr Albert Tenesa’s research group at The University of Edinburgh Roslin Institute to obtain a better understanding of the genetic contribution to a variety of complex traits. The initial focus is on colorectal cancer, one of the most common forms of cancer in the UK, but the techniques and software are applicable to a wide range of traits in humans and animals.

The research involves studying genetic variation across many people: typically the variation of hundreds of thousands of SNPs across several thousand individuals. Some of the individuals have tested positive for the disease (or other trait of interest), and the others negative. We create a “Genetic Relationship Matrix” (GRM): a numerical measure of how closely the variants match across each pair of individuals, i.e. how “related” they are in the genetic sense. This can be correlated with the trait occurrence information in a statistical analysis to determine the overall genetic contribution. With enough accuracy, we can also determine the contribution of each SNP: information that could potentially be used in a clinical setting to predict the disease risk for a certain individual, with preventative measures prescribed as necessary.

This analysis is computationally demanding. As a starting point, we used the open source package

Our DNA influences a wide range of factors important in our daily lives. There are great potential benefits to understanding its intricacies, not least in healthcare, but this involves unravelling complex associations. By developing software that effectively exploits modern computational resources, we are helping to make the problem tractable.
“Genome-wide Complex Trait Analysis” (GCTA) [1]. This versatile package offered the required functionality, but proved restrictively computationally prohibitive for the large numbers of SNPs and individuals required. A single run for our data set took 17 hours, and many runs were required.

We used the GCTA software for a starting point to create a new "mutated" package Advanced Complex Trait Analysis (ACTA), which reduced the runtime, for the above case, from 17 hours to 11 minutes (on a single node of a compute cluster). ACTA obtains this performance improvement through a range of methods including the restructuring of algorithms, the removal of redundant operations and the use of multithreaded linear algebra libraries plus OpenMP directives to fully exploit modern multi-core chips [2]. ACTA is now freely available [3], and researchers at The University of Edinburgh and beyond are using it. Meanwhile, we have been working hard to exploit HPC resources even more effectively.

High-end Graphics Processing Units (GPUs) offer impressive computational performance, and can be programmed to tackle the most computationally demanding parts of the algorithms. This is no trivial task, due to the architectural complexities of GPU-accelerated systems. We have recently created a version of ACTA able to exploit such systems. We are currently fine-tuning the application and expect it give a several-fold increase in performance over the optimised CPU version.

Standard analysis using ACTA involves analysing all the SNPs in the dataset together. However, a more sophisticated “Regional Heritability” approach can be used to discover correlations more effectively. This involves partitioning the genome into a large number of overlapping “regions” (each corresponding to a specific DNA sub-sequence), performing a separate analysis for each, and finally combining the results. The high computational cost of this approach can be tackled head-on using parallel computing resources. We have modified ACTA to run across multiple compute nodes in a parallel cluster, using the Message Passing Interface (MPI), where each node is responsible for a subset of regions. The multiple CPU or GPU cores internal to each node work together to analyse each region. This software is also in the final stages of development and will be released soon.

IES is the world’s leading provider of software and consultancy services which make the built environment more energy-efficient, reducing overheads and CO2 emissions in projects around the globe. Its technologies and products rely on the detailed 3D modelling of buildings and their surroundings, and the company was keen to offer faster run-time capabilities for its SunCast software, which architects, designers and planners use to analyse the sun’s shadows and the effects of consequential solar gains on the thermal performance and comfort of buildings.

EPCC has been working with IES to accelerate the performance of SunCast by optimising the software and using parallel programming techniques. As a direct result of enabling SunCast to fully exploit the multi-core potential of modern computers (from ordinary laptops, to workstations and servers), this work will not only give IES and its customers greater flexibility in undertaking larger, more complex projects, but also analyse many different geometric options at early stages in the design process.

**Reduced run times**

The project with EPCC will reduce the time-to-solution for a typical customer with a 4-core laptop from several hours to less than a single hour to process exactly the same scenario. Rather than losing a day waiting for the SunCast model to run, the customer will now be able to either look at multiple options in the same time, or just do one run much more quickly.

In addition, SunCast will also be able to be run on large high-performance computers. For example, a SunCast calculation that used to take two weeks to run on a Windows desktop will now be able to operate on EPCC’s 1500-core supercomputer INDY, with the results available in a matter of minutes.

Commenting on the project being undertaken with EPCC, Ronnie Galloway, Senior Development Manager at IES said: "Supercomputing Scotland has done a super job for us! It has given us higher performance tools and enabled us to study a larger number of scenarios."

**Supercomputing Scotland**

Supercomputing Scotland is a partnership between Scottish Enterprise and EPCC. It assists Scottish companies in the use of EPCC’s computing facilities and expertise to support strategic growth plans.

The programme is designed to provide an in-depth assessment of a company’s HPC capabilities and potential for using HPC. This detailed assessment then provides the basis for an application for financial assistance from Scottish Enterprise.
undertaken with EPCC, IES Technical Director Dr Peter Thompson said, “IES is delighted to be working on key technical developments in high performance computing with EPCC. This work is an important part of our plans to ensure that we remain at the forefront of technology and continue to provide world class services in modelling and simulation in relation to energy efficiency of buildings. The structure of the Supercomputing Scotland programme together with support from Scottish Enterprise has meant the feasibility studies undertaken by EPCC could be converted quickly to live projects.”

David Smith, director of technology, engineering and creative industries for Scottish Enterprise said: “Supercomputing Scotland supports companies to be more innovative and gain competitive advantage by providing access to the world-class advanced computing technology and expertise at EPCC.

“We have established this joint partnership with EPCC because we recognise the importance of high-performance computing for Scottish companies. The recent success story of Integrated Environmental Solutions is testament to how Supercomputing Scotland helps companies to make use of high-performance computing technology and expertise to realise their strategic growth ambitions.”

Forthcoming collaboration

In addition to the Supercomputing Scotland project, IES and EPCC will also work together on a new European Commission FP7-funded programme called FORTISSIMO. This will work to develop high performance computing (HPC) capabilities for SMEs in Europe using the Cloud by partnering closely with a number of European HPC centres and modelling and simulation software providers. Final details of the project are currently under negotiation.
EPCC team enters the ISC’13 Student Cluster Challenge

The Student Cluster Challenge (SCC) at ISC (International Supercomputing Conference) is a competition where student teams from around the world build clusters, competing against each other to achieve the maximum performance whilst remaining within a strict power limit of 3kW.

Cluster performance will be assessed by the HPCC[1] (HPC Challenge) benchmark suite and a set of real world benchmarks, including Gromacs[2], MILC[3], WRF[4] as well as two secret applications which will be announced during the competition.

Nine student teams will participate in the SCC competition at ISC’13. Our student team, which consists of four students studying for the MSc in HPC at EPCC, will be the first team from the UK to participate in this competition. Viglen Ltd and Boston Ltd, as the team’s industrial partners, will provide our team with hardware, technical training and support.

You can find the EPCC SCC Team at ISC’13 booth 491. The competition will run from midday on Monday, June 17 until Wednesday, June 19.

What the student team says

The Student Cluster Challenge will involve us in solving problems and developing skills that are essential not only for our current studies, but for our future careers as well.

The competition will give us an opportunity to work on a massively parallel cluster system which consists of emerging architecture running taxing benchmarks. This will give us a unique chance to gain hands-on experience of high performance computing, and to experiment with cutting-edge hardware, whilst applying the theory we have learned in the classroom.

By competing as a team in the Cluster Challenge, we feel that the competition will drive us to work harder, experiment more, and develop more innovative ideas about cluster design, and HPC programming techniques.

We will also enjoy meeting other students from around the world, to exchange ideas and socialise, as well as engaging with the vendors and participants who will be present at ISC’13.

By strengthening our collaboration with Software Carpentry, we help train researchers to confidently and efficiently develop software that contributes to correct, reproducible and reusable research.

One year ago, four hardy members of The Software Sustainability Institute sweltered for two days in a cramped, airless lecture theatre on a scorching day in central London and experienced our first Software Carpentry boot camp. Two weeks later we helped to run the first UK-led boot camp at Newcastle University.

Now, one year on, after helping to deliver boot camps in Newcastle, Oxford, Edinburgh, Munich, Tubingen and Virginia, we are the coordinators for Software Carpentry in the UK. In our role, we help UK researchers organise boot camps for their research groups, institutions and communities. We help the local organisers of a boot camp to create a syllabus; attract instructors and helpers; manage registration; advise on publicity; and provide support in all aspects of organising a boot camp.

We are also continuing to help deliver boot camps across the UK. In early April, at the EGI Community Forum in Manchester, I ran a day of boot camp highlights. These were taster sessions to give attendees a flavour of Software Carpentry's hands-on, researcher-focused and “live-coding” approach to teaching.

A week later, I returned to The University of Manchester where I led a Software Carpentry boot camp alongside my Institute colleague Aleksandra Pawlik, and David Jones from the Climate Code Foundation. We had 24 attendees who unanimously agreed that they had enjoyed the boot camp. A fortnight later, EPCC’s Mario Antonioletti was a boot camp instructor for the Oxford doctoral training centres. You can read about his experiences on our blog: www.epcc.ed.ac.uk/blog

In the coming months, we’ll be helping to deliver boot camps for Southampton University and Bath University.

Mike Jackson
m.jackson@epcc.ed.ac.uk

Comments from Bootcamp attendees...

“A good two days! Thanks :)

“Within my field...there are already lots of workshops covering the specific software packages but nothing to do with generally using computers to automate stuff so this camp filled a big gap in the teaching available to me.”

Greg Wilson, Founder of Software Carpentry...

“We are pleased to have the Software Sustainability Institute take over coordination of UK boot camps. The Institute shares many goals and values with Software Carpentry, and we believe this partnership will benefit both organisations.”

Find out more
www.software.ac.uk
www.software-carpentry.org
www.epcc.ed.ac.uk/blog

Street art in Manchester.
Fifth PRACE seminar for industry users

The principle aim of this seminar in Bad Boll, Germany, was to enable potential industry users of HPC to learn what PRACE (Partnership for Advanced Computing in Europe) can do for them.

Highlights from the keynote speakers included Suzy Tichenor from Oak Ridge National Laboratory who talked about their engagements with industry using HPC, and Christophe Gümbel of Porsche AG who talked on the importance of HPC for vehicle development within Porsche. These were followed by parallel sessions on Open Source in Industry, and the PRACE Technology Watch.

The social event in the evening was a tour of the local museum which specialises in ichthyosaur fossils as well as other rare and exotic prehistoric creatures, followed by an excellent meal in a local restaurant. This event gave an opportunity for HPC contacts to be formed and consolidated outside of the presentation environment.

Day two started with success stories from businesses who had worked with PRACE, followed by focussed topic parallel sessions on both the automotive and aerospace industries, plus an overview of PRACE’s work with emerging applications of interest to industry. Further success stories followed, and then parallel sessions on how Large Companies and SMEs could access the PRACE programmes.

The session ended with an award for the “Most Innovative European Industrial HPC Application”, which went to the company CERFACS for their work on AVBP, a Large Eddy Simulation tool used by the aeronautical, energy and automotive industries on major super computing architectures in Europe.

The seminar gave an opportunity to elaborate on the industrial collaborations offered by PRACE. The PRACE Open R&D Access Programme allows industrial users to conduct industrial, open research, based on a peer review process, on PRACE’s resources. PRACE is committed to adapting this model to the needs of industry by implementing an SME support programme (SHAPE), an Industrial Advisory Committee and other initiatives, all presented and discussed at the seminar.

PRACE

The mission of PRACE is to enable high impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society.

Next year’s seminar will take place in Barcelona from 22–23 May, 2014.

For more information on gaining access to PRACE resources, see www.prace-ri.eu/HPC-access

There are also various calls for applying for access. See: www.prace-ri.eu/Call-Announcements

Paul Graham
p.graham@epcc.ed.ac.uk
EU-Russia project: APOS-EU

Now that APOS has come to a successful end, it is time to take stock and see if our objectives were met.

Michele Weiland
m.weiland@epcc.ed.ac.uk

For the past two years, EPCC led APOS-EU, one half of a project funded by the European Commission as part of its Framework 7 “EU-Russia HPC” call. The other half of the project, APOS-RU, was sponsored by the Ministry of Education and Science of the Russian Federation and led by the Keldysh Institute for Applied Mathematics at the Russian Academy of Sciences.

APOS was involved in a range of activities, mostly focusing on specific problems in a small set of scientific applications. One such activity looked at improving the scalability of the PETSc library by introducing hybrid communication, another compared the performance of a range of implementations (CPU, GPU and FPGA) of a seismic modelling application. More detail on our technical work and research can be found on our website [1].

In this project the second objective – fostering collaborative research – took the same importance as the technical work. Prior to APOS-EU, none of the partners involved in the project had any experience of collaborating with fellow HPC researchers in Russia. Today, having visited Moscow on a number of occasions and worked closely with APOS-RU, we feel that the types of relationships that were established between the European and Russian scientists as part of APOS have been long overdue. Both sides have vast amounts of knowledge that need to be shared: our colleagues in Russia are extremely gifted mathematicians with in-depth understanding of numerical methods, while the APOS-EU partners excel in the more hands-on aspects of HPC. Joining the two will be the key to solving some of the problems faced by HPC when progressing towards Exascale.

All in all, APOS-EU was an extremely enjoyable project to work on and a good opportunity to venture into a previously unfamiliar world. We hope to be able to build on what we learned and continue our collaborative research in the near future.

[1] www.apos-project.eu
Innovative Learning Week 2013

For one week in February, the University of Edinburgh suspends traditional lectures, tutorials and labs, and supports a programme of alternative learning activities: Innovative Learning Week.

EPCC runs the MSc in High Performance Computing (HPC), teaching students the tools and techniques they require to program modern parallel computers. During Innovative Learning Week, students from the MSc in HPC took part in three activities: developing an exhibit to explain HPC concepts to 7-12 year old children, writing a mobile app for learning numbers in Japanese, and building small computing clusters out of scrap (and not so scrap) computing hardware.

The six-strong outreach team worked with Sara Shinton from the Bang Goes the Borders Science Festival which is held in September in Melrose. They developed four exhibit ideas for EPCC’s next outreach demonstration at the festival (see EPCC blog).

Gaurav Saxena and Weronika Filinger worked together to write a simple mobile app to learn numbers in Japanese using a game-like structure. At the start of the week they knew nothing about Android, a bit about XML, and a bit of Java; at the end of the week they had an app working on an Android phone!

The other MSc in HPC students participated in a Cluster Building Challenge. In friendly competition with the ISC team (see p16), who are building a cluster out of ARM processor boards, four more teams of 4 students each built a cluster from a variety of old, discarded hardware.

The Rocks on Jimmy team had five 32-bit nodes from a scrapped cluster that had been used in the National e-Science Centre. Using the Rocks cluster building software, they got to 21 Gflops and 22 Mflops/watt on the Linpack HPL benchmark. Not quite as green as the ISC team (aka team ARMaggedon), who achieved 103 Mflops/watt with their cluster of 7 PandaBoard worker nodes and one Raspberry Pi head node.

The PELiCaN team took a different approach and used different software (Pelican). They happened to have a 24-way and an 8-way network switch and decided to go all out for speed, eventually using 26 of the 32-bit nodes. With some tuning they reached 67 Gflops overall and kept the room nice and warm by dissipating 3kW of power.

Mark Filipiak
m.filipiak@epcc.ed.ac.uk

Mark has also written about Innovative Learning Week on the EPCC blog:
www.epcc.ed.ac.uk/blog

MSc in High Performance Computing
Places are still available for the 2013/2014 MSc in HPC. Find out more: www.epcc.ed.ac.uk/msc
What our students say...

Gaurav Saxena, Mobile App Team
I would start by saying that it was an amazing experience for me. I was a member of the mobile application development team. On Monday we knew nothing about Android, xml and a bit of Java and seriously had no idea as to how we would make ‘anything’ for the Android platform. Well, we did make something and seeing it run on the mobile device gave me and the team an amazing feeling of achievement.

I understand that the work that we did was very elementary and incomplete, but it is a certainly a start. EPCC’s Alistair Grant motivated us at each step and with his wonderful sense of humour kept us going despite repeated failures and disappointments. It was like a crash course in swimming, when you’re thrown into the water and, if you can stay afloat, the inner smile appears. Hence I feel Innovative learning Week is an amazing idea where students are exposed to various activities.

Dimitrios Kapnopoulos, PELiCaN Cluster Team
Although this week was more busy than the others, I really enjoyed it. All these months here, we had been working on HECToR and Morar (computing cluster used for MSc teaching in EPCC) and yet we hadn’t seen what a cluster looks like or how it can be built. We had the chance to see how people work in groups, how they try to find solutions and the diversity in how people think of a way to solve everything. Furthermore, although I have been running Linux on my computer for quite a few years, I hadn’t had the chance to work on that level before.

The two other teams had different hardware and different challenges. The Atom Smasher team used four low-power, 64-bit Atom processors to run disk-bound data intensive jobs using Hadoop Map/Reduce and then added some higher-power 32-bit nodes to make a hybrid cluster.

And the WSoGMM team had a totally heterogeneous system: one Sony laptop, one Dell desktop, one low-power Atom desktop, and a rather old Dell PowerEdge rack-mountable server. Their main challenge was to get good load balancing between the very different processors.

At the end of the week there was a ‘show and tell’ session by all the teams, attended by EPCC staff and other students.

Innovative Learning Week was a great success with the students, who enjoyed working on something completely different for a week. We certainly plan to have similar activities next year.
EASC13: First Exascale Applications and Software Conference

It is an exciting time in software developments for exascale, and the conference provided an ideal opportunity for the community to engage in interesting discussion.

Over 130 delegates attended this EPCC-hosted conference, which brought together those involved in solving the software challenges associated with programming exascale systems.

The conference was located in a beautiful setting on a University of Edinburgh campus, with Arthur’s Seat as a backdrop. Following an introduction from Dr Mark Parsons, EPCC’s Research & Commercialisation Executive Director, the first of our six keynote speakers presented.

Prof. Bill Tang, from the Princeton Plasma Physics Laboratory, described the challenges associated with scaling fusion energy applications to extreme levels and really set the agenda for the application challenges associated with exascale.

Prof. Satoshi Matsuoka, from the Tokyo Institute of Technology, followed in the afternoon, providing an interesting description of Japan’s impressive efforts towards exascale.

The second day saw two further keynotes. The first was from Prof. Jack Dongarra, who looked at future trends in HPC and how these changes will impact future software. The second was given by Prof. Peter Coveney, who described the challenges for computer modelling and simulation at the exascale before introducing a range of potentially disruptive simulation techniques that allow effective use of exascale resources.

This day also saw a panel session with a distinguished panel from Europe, Japan, Russia and the US. These include Boris Chetverushkin, Satoshi Matsuoka, Jack Dongarra, Barbara Chapman, Sudip Dosanjh and Per Oster. Current topics range from how many of the challenges at exascale have actually been solved at the tera and peta scale levels; the future of MPI; solving the challenges for memory bound applications; political and funding issues.

The last day of the EASC2013 conference began with a keynote speak from George Mozdzynski, from the European Centre for Medium-Range Weather Forecasts (ECMWF) and from Vladimir Lorna Smith

l.smith@epcc.ed.ac.uk

It is an exciting time in software developments for exascale, and the conference provided an ideal opportunity for the community to engage in interesting discussion.
Voevodin, Deputy Director on Science in Research Computing Center of Moscow State University. Prof. Voevodin described Russia’s approach to exascale challenges in education. Together with Professor Voevodin and Chetverushkin, there was a notable Russian engagement with the conference, demonstrating the great interest in this area in Russia. For example, a few of the presentations resulted from the APOS project, a joint EC-Russia collaboration. Further details on this project can be found in Michele Weiland’s APOS article on p19.

George Mozdzynski began by describing the societal importance of ECMWF’s Integrated Forecasting System (IFS), highlighting the role played in the early prediction of Hurricane Sandy, which made landfall in New Jersey on 30th October 2012. From 23rd October ECMWF predicted the storm would strike the East Coast of the US, with most other models predicting the storm would move out to sea. Hurricane Sandy was a major meteorological event and George clearly demonstrated the importance of this early prediction.

Having described the importance of their model, he then moved on to describe their requirements for exascale-size systems and articulate the challenges associated with exploiting these systems. As a key partner in the CRESTA project, George described his work in this project to address these challenges.

There was also a series of interesting and diverse talks around the conference, with parallel sessions covering topics such as tools for exascale, preparing applications for exascale and the I/O challenges. Disruptive technologies proved to be a key theme in many talks and discussions.

In summary, this proved to be a timely and interesting conference and one to be repeated: our colleagues and CRESTA partners from PDC and SeRC at the Royal Institute of Technology (KTH), Sweden’s largest technical university, will be hosting the Second EASC conference in April 2014.
PRACE Advanced Training Centre: 2013/14 curriculum

As a PRACE Advanced Training Centre (PATC), EPCC runs around 35 course-days of HPC training per year, with attendance free for all.

The PATC year runs from September to August, and we have recently finalised our proposed programme for 2013-14. Dates are still to be confirmed, but the course list (in approximate date order) is:
- GPU Programming with CUDA and OpenACC
- PGAS Programming with UPC and Fortran Coarrays
- Software Carpentry
- Parallel Systems: Getting the Benefits
- ARCHER System Workshop
- Data staging and data movement with EUDAT
- Tools for Large-Scale Parallel Debugging and Profiling
- Parallel Materials Modelling Packages
- Statistical Analysis for Post-Genomic Data: Parallel Computing with R
- Message-Passing Programming with MPI
- Advanced OpenMP
- ARCHER System Workshop
- Programming the Xeon Phi
- Parallel CFD packages

PGAS13 conference visits Edinburgh

This is the first time the conference is being held outside the US and EPCC is very excited by the prospect of hosting PGAS2013!

PGAS programming models offer a shared address space model that simplifies programming while exposing data/thread locality to enhance performance. The PGAS conference is the premier forum to present and discuss ideas and research developments in the area of: PGAS models, languages, compilers, runtimes, applications and tools, PGAS architectures and hardware features.

The conference will present research in areas such as: applications; performance; developments in programming models and languages; and tools, compilers, and implementations.

If you are working on research in any of the related areas, please think about submitting a paper or extended abstract and contribute to the conference programme - the deadline for submitting is mid-June. Or if you have an interest in PGAS and the role it plays in reaching the Exascale, this event will be for you as well.

Edinburgh is a beautiful city, so learning about new trends in HPC can also easily be combined with a spot of sightseeing!

The local organising committee is already working hard on making PGAS2013 a success. For more information and details on the event see: www.pgas2013.org.uk