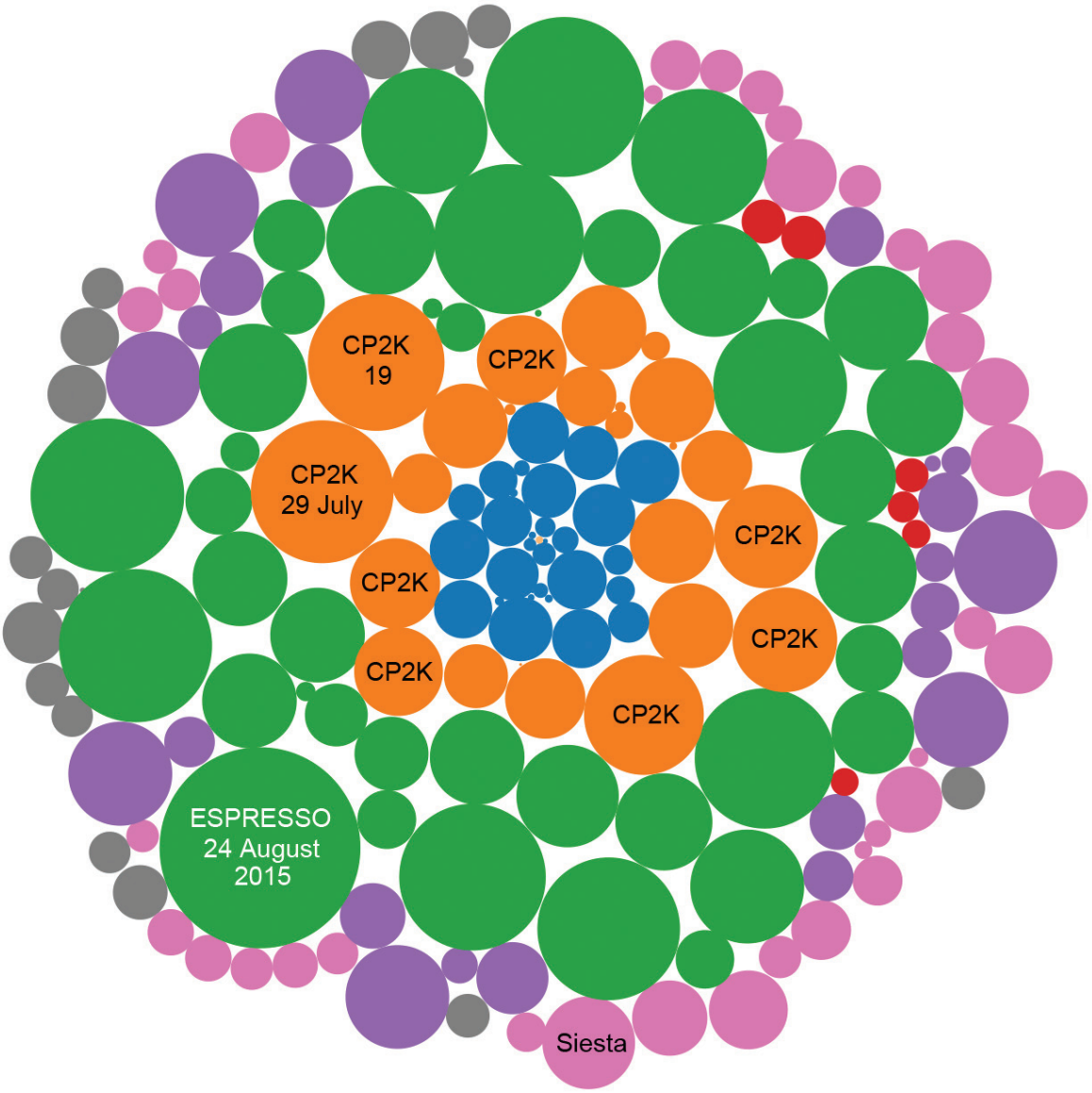
Science Highlights 2014 - 2015

Scientific Computing Department (SCD)



Science & Technology Facilities Council

Contents

Designed and layout by STFC Graphic Designer Andrew Collins
Produced by RCUK's internal service provider.

Welcome

Welcome to this year's highlights from the Science and Technology Facilities Council's Scientific Computing Department. I have now been in post for around 18 months, and in that time have worked with the staff to develop a set of strategic objectives, around which the future of the department is now focussed. It is clear to me that computing capability is increasingly becoming the limiting factor in enabling scientific delivery, not least from the facilities that STFC provide, such as the ISIS neutron muon source, the Central Laser Facility, or the Diamond Light Source. In a time of continuing austerity, it is clear that we not only have to be smarter in how we work, but we also need to work in much closer collaboration in order to deliver that urgently needed computing capability to the science users of all disciplines.

There is a very large range of projects on show here, which I hope are as striking and fascinating for you as they are for me. These highlights are a window into the department's work, and while they give a very good representation of the work we do, they do not describe everything that the Department is involved in. So if there is anything here which interests you, or if you would like to know more about the Department, and its capabilities, please contact any of the authors of these articles, or me at david.corney@stfc.ac.uk.

As before, I would like to finish by thanking the staff for all they have done and are doing to continue to make the department, their department a success. As a young department we have become used to rapid change and evolving quickly to adapt to the developing needs of our scientific user communities. I am excited by the future, all that needs to be done, and greatly looking forward to working with the staff and the science communities to continue to develop and deliver the computing solutions required to unlock scientific discovery in the coming year.

Thank you and regards,

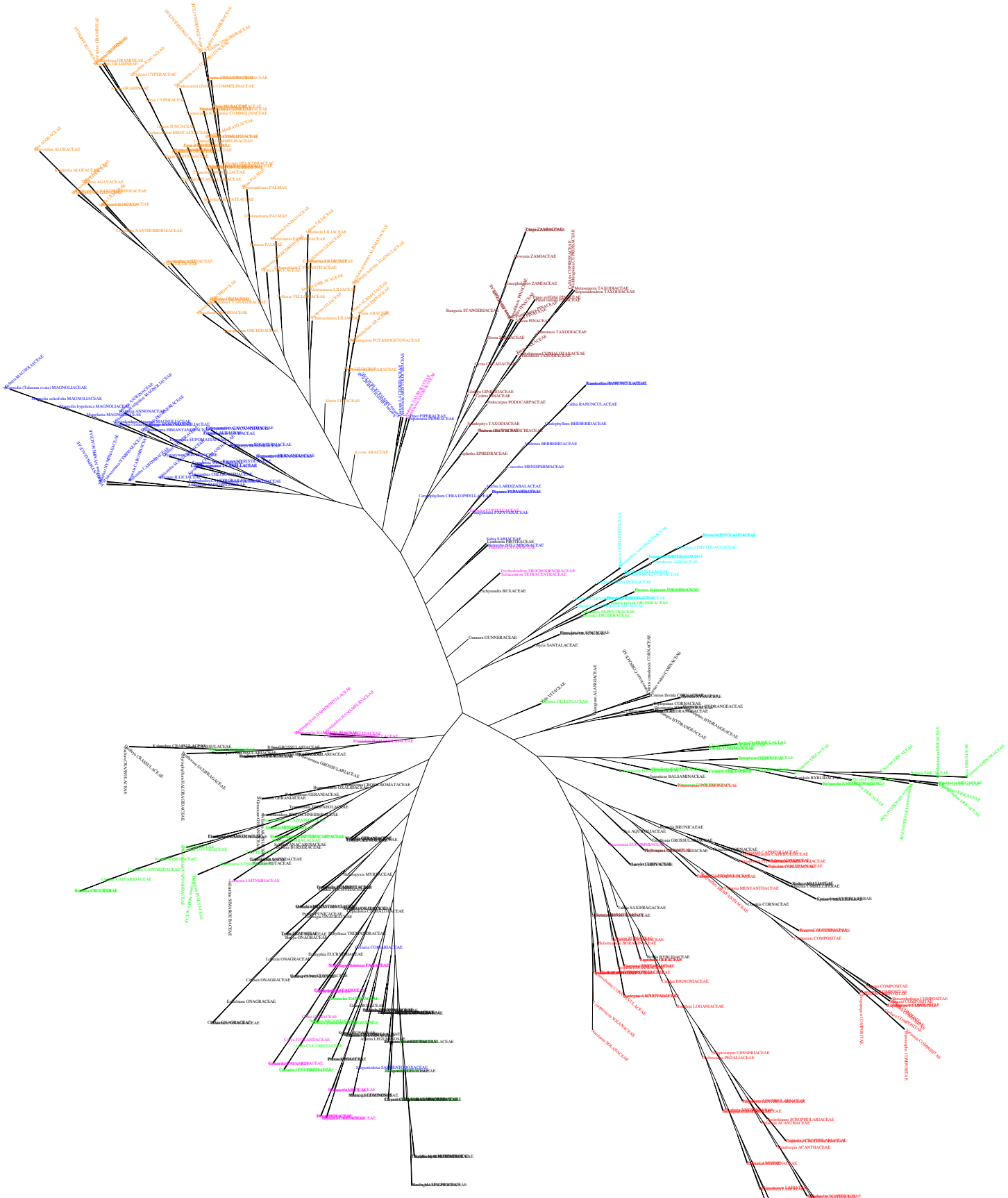
David Corney

Acting Director of the
Scientific Computing Department

As the Acting Director of the Hartree Centre, I greatly value the close working relationship between the Hartree Centre and the rest of the Scientific Computing Department. This report highlights several of the areas of research led by the Hartree Centre, and as we move into the next phase of the Hartree Centre, we will be developing new areas of research jointly with IBM. These research projects will be guided by our core mission of helping UK industry gain competitive advantage by working with the Hartree Centre. These are exciting times, and the work of the Scientific Computing Department and the Hartree Centre are a vital contribution to the mission of STFC, the growth of the UK economy, and science projects on a global scale.

Peter Allan

Acting Director of the Hartree Centre



Using HPC to Deduce the Tree of Life

Ever since we have known species share common ancestors, we have sought to discover the pattern of relationships among species. The pattern of relationships is typically represented as a bifurcating tree, whose tips represent extant species and internal nodes represent hypothetical ancestors. The study of these relationships – phylogeny – has applications in the prioritisation of conservation efforts, the search for natural products in related species, tracing the origin and spread of bacterial and virus pathogens, and understanding of gene function.

Until the late 20th Century, phylogenetic research was speculative, due to the limited quantity of data that morphology can provide. Two technological developments in the 20th Century revolutionised the field. One was the availability of computers, allowing the replacement of subjective reasoning in phylogeny reconstruction with algorithmic and statistical approaches. The other was the availability of DNA sequences, allowing a step-change in the quality and quantity of data upon which to base phylogeny reconstruction. More recently, however, the advent of next generation sequencing technology is placing a strain on computational phylogeny. In the 1990s, a PhD student might sequence a single gene from each of 20 species, and the insights would be novel and astounding. Now, the student might sequence the entire genome of these same species, with enormously greater opportunities for discovery and certainty. With continuing developments in DNA sequencing technology, within a decade such a student will be able to sequence the entire genome of several thousands of species.

The bottleneck has shifted from DNA sequence availability, to algorithms and software able to extract phylogenetic relationships from these data. To find the optimal phylogeny based on DNA sequences, ideally one would evaluate all possible tree topologies. However, the number of topologies increases factorially with the number of extant species, and for 54 species there are more possible unrooted bifurcating topologies than atoms in the universe. The only way forward is a combined high performance computing (HPC) and advanced algorithmic approach. This will search the range of solutions in an effective and rapid manner, and will maximise the potential of increasingly available large computing resources.

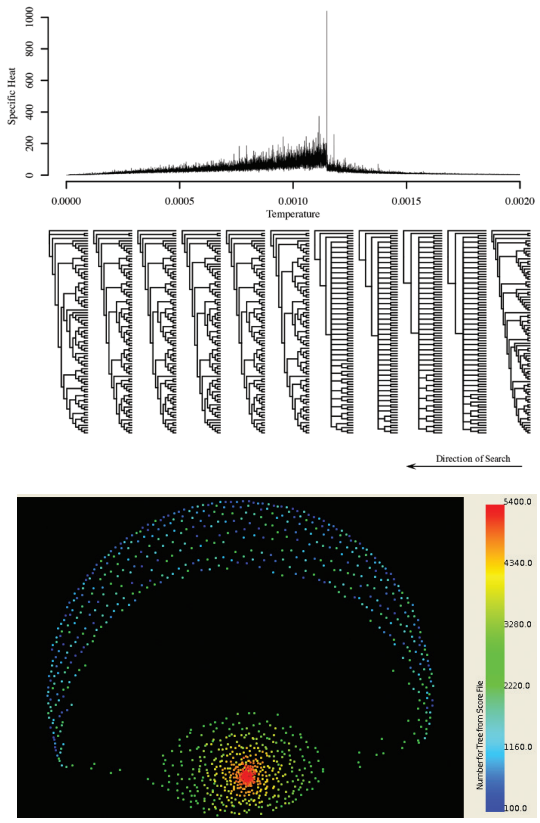


Figure 2: The simulated annealing algorithm as implemented in LVB. The search proceeds from right to left, as the temperature is decreased. The y-axis shows the specific heat which is proportional to the variance of the cost function, peaks of which reveal phase transitions in the tree topology. Inset: a consensus of the solutions examined by the algorithm at different temperatures illustrates the resolution of the final tree structure and demonstrates the significant role of the phase transition period.

Figure 1 (opposite): A milestone in phylogeny research - the first large phylogeny reconstruction based on DNA sequences. The tree indicates relationships among 500 seed plants (Chase *et al*. 1993, Ann MO Bot Gard 80:528-580). Labels are coloured to show major groups in a pre-DNA classification system. Many of these pre-DNA groupings are scattered across the tree, indicating they do not represent evolutionary relationships correctly (e.g. Hamamelidiae, pink).

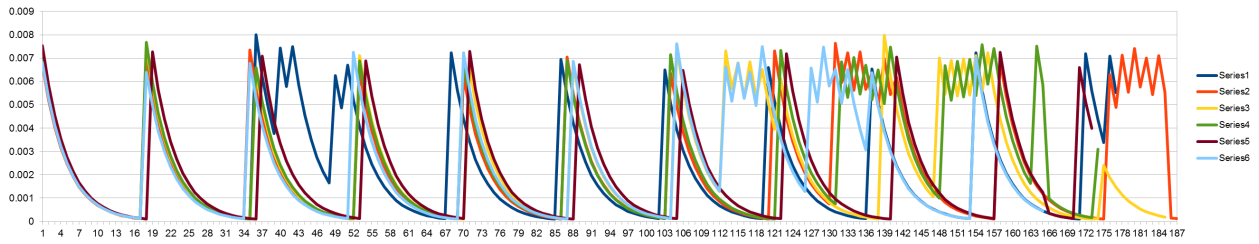


Figure 3: Parallel searches, spread over multiple nodes. The six curves represent different compute nodes, with peaks occurring at the beginning of new searches.

LVB is software for phylogeny reconstruction developed at the University of St Andrews, which uses a simulated annealing heuristic. This allows a good (realistic) result to be obtained, by looking at only a small fraction of possible tree topologies. At each step, a new tree is generated by making one of two changes to the existing tree: either a local change (nearest neighbour interchange) or a global change (subtree pruning–regrafting). The principle of maximum parsimony means that the shortest possible tree that explains the data is considered best. During the search, changes that decrease or do not alter the length of the current tree are always accepted. Changes that increase the length are accepted with probability $p = \exp[-(1/T)\Delta H]$, where T is the current temperature and ΔH is the increase in homoplasy index. Temperature falls as the search progresses. Phase transitions are often observed, as the tree undergoes major rearrangement towards the final topology (see Figure 2).

With support from the STFC Global Challenge Concepts fund, we have investigated two parallelisation strategies to enable the application of LVB code to large phylogenies. Firstly, a parallel search has been implemented in which a large number of simulated annealing runs are performed each starting with a different initial tree. On a compute cluster, these searches are distributed over the available nodes. Typically, far more searches are attempted than there are nodes available, so when a compute node has finished a search, it will launch the next search from the list. Comparison between parallel searches (implemented via MPI messages) allows unproductive searches to be identified, and these are terminated early.

Access to Blue Joule at the Hartree Centre has allowed us to push the parallelisation to large scale. We have computed phylogenetic trees for up to 28,000 species, extracted from a data matrix for 55,473 flowering plant species, running parallel searches on 1024 nodes.

LVB keeps the current set of best trees in memory, and when a new tree is created of the same length it must be checked to see if it already exists in this set. For complex cases, there can be thousands of trees held on the tree stack, and this check can become the most expensive step of the search. Following our earlier work on using the MapReduce paradigm to cluster nucleotide sequences, we have developed a MapReduce algorithm for clustering trees in order to identify duplicates. The algorithm is implemented using the MapReduce-MPI library from Sandia Lab which uses MPI to handle the shuffle / collate step (http://mapreduce.sandia.gov/).

Tests show that this can improve runtimes for LVB significantly. LVB has been run on Blue Joule with up to 8000 sequences, and 64 nodes available to the MapReduce algorithm.

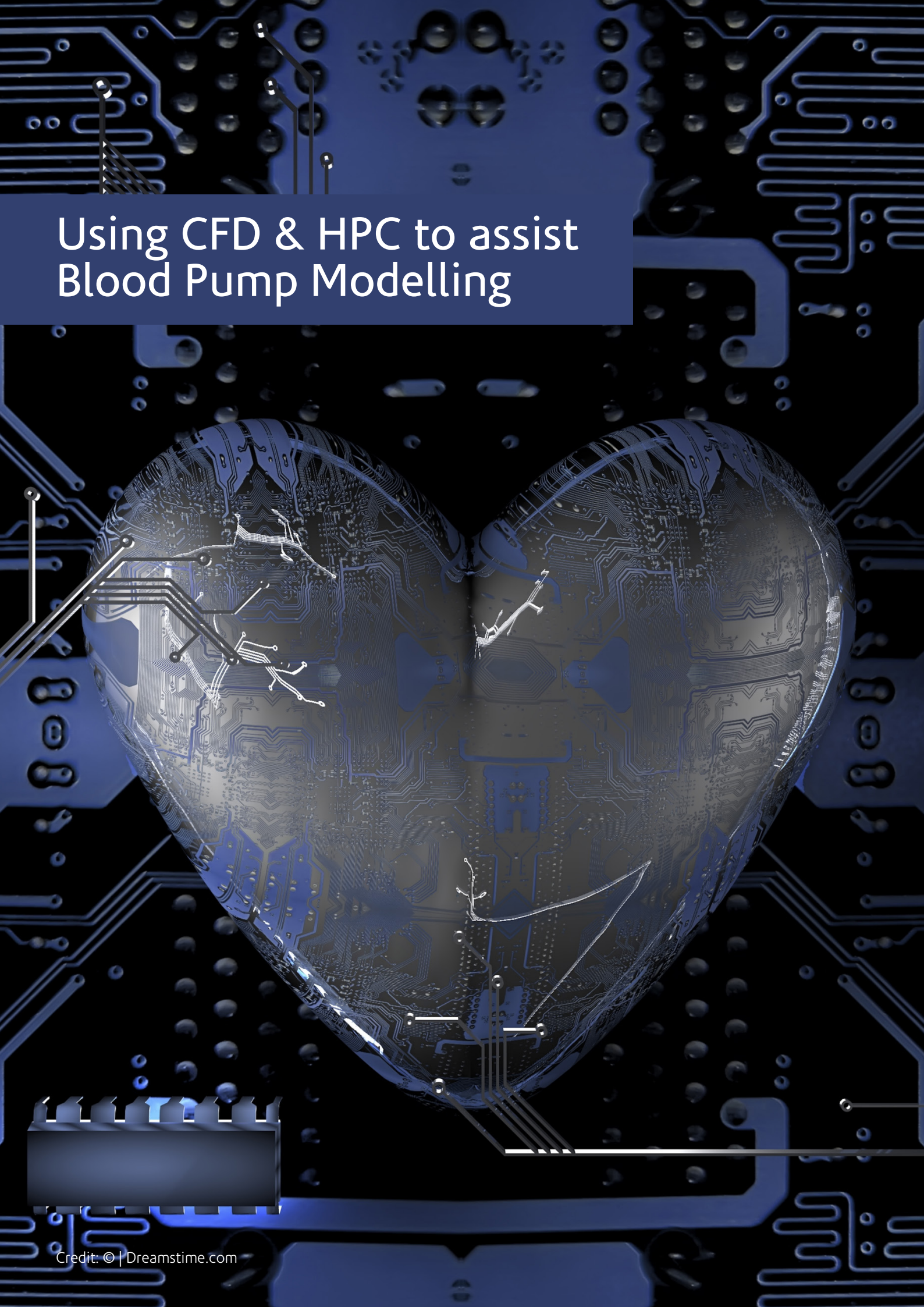
The two approaches address different bottlenecks in the LVB workflow, and we are now working on combining these into one unified code. Dynamic load-balancing is required to assign nodes and/or compute time to either parallel searches, or to resolving large numbers of similar trees. With these algorithmic advances, we aim to tackle ever larger phylogenies, and thus exploit the deluge of sequence data that is now being generated.

Authors
C. Sik Kim, M. Winn, STFC Daresbury Laboratory. M. Pinheiro, D. Barker, M. Strobl, T. Meagher, University of St Andrews

References

[1] Barker, D. (2004) "LVB: parsimony and simulated annealing in the search for phylogenetic trees." Bioinformatics, 20, 274-275
[2] V. Sachdeva, C.S. Kim, K. E. Jordan and M.D.Winn, Proceedings of the Thirteenth IEEE International Workshop on High Performance Computational Biology, "Parallelization of the Trinity pipeline for de novo transcriptome assembly" DOI: 10.1109/IPDPSW.2014.67

Using CFD & HPC to assist Blood Pump Modelling



Credit: © | Dreamstime.com

Heart failures account for one of the most frequent diseases occurring worldwide, often requiring heart transplant, but too few donors are in general available. These heart conditions are nowadays more and more often treated by means of Ventricular Assist Devices, such as blood pumps, whose purpose is to prolong the life of patients awaiting a donor. Computational fluid dynamics has been used to study the flow in a model blood pump. Its characteristics i.e. geometry, operating conditions have been set by the US Food and Drug Administration as part of their second Computational Round Robin. The work presented here was carried out in the framework of a joint project between FH Aachen, Juelich Forschungszentrum, EDF and STFC.

Introduction

Ventricular Assist Devices (VADs) are used to help patients with serious heart conditions (see Figure 1). They must be designed to reduce potential haemolysis and thrombosis. VADs should also operate at low and high flow rates and under various pressure heads to mimic heart behaviour [Behbahani et al.]. Computational fluid dynamics (CFD), in conjunction with high performance computing (HPC), offer tools to show detailed insight into the complex patterns of blood flow that determine hydraulic performance. CFD is capable of quantitative haemolysis prediction if used with an appropriate blood damage model. It can also help engineers identify recirculation areas and other regions with an increased probability for blood clotting. To assess CFD and hopefully increase confidence in using it for designing complex medical devices, the US Food and Drug Administration (FDA) has proposed a series of benchmark studies. Their second Computational Round Robin deals with the flow in a model blood pump, under various mass flow rates and rotation angles. Some results using *Code_Saturne* are presented here.

Function of the pump

The main function of a blood pump [Behbahani et al.] is to achieve adequate hydraulic performance while maintaining good haematological compatibility. The pump needs to operate at a desired flux and at a given pressure head (resistance); haematological requirements arise because non-physiological flow conditions in a pump often lead to haemolysis (haemoglobin release from red blood cells) and thrombosis (clotting of blood). Haemolysis, beyond a certain level, and thrombosis, are the two major life-threatening factors for patients depending on VADs.

FDA Computational Round Robins

The purpose of the project is to determine how CFD can be effectively used to characterise fluid flow and to predict blood damage in medical devices. For comparison and validation of the computational simulations, three selected laboratories will perform quantitative flow visualisation measurements on physical models. The FDA will collect the data from the simulations, analyse the results in a blind test (i.e. the identification of the person who sent in the data will be removed and replaced with a unique identifier), and then compare and present the collective results.

Configuration – Operating conditions

Figures 2 and 3 show a sketch of the model blood pump and of the computational domain, respectively. The diameter of the pump is 6 cm and it is about 2 cm high. Several flow rates and rotation angles are investigated leading to 6 simulations in total.

Figure 1: Example of a VAD: the HeartMate II heart pump (Courtesy from bidmc) http://www.bidmc.org/Centers-and-Departments/Departments/Cardiovascular-Institute/About-the-CVI/CVI-In-the-News/2013/September/LVAD.aspx

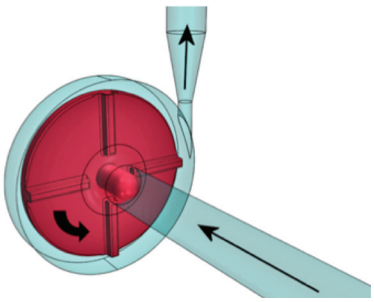


Figure 2: Model blood pump.

Our approach

Code_Saturne (http://www.code-saturne.org) is used to perform all 6 simulations. It is an open source CFD software where the Finite Volume method is used to compute the Navier-Stokes equations. Meshes for the rotor (MR) and the stator (MS) are generated independently. MR rotates at each time step and is glued in parallel to MS, before the governing equations are solved on the newly built mesh. Turbulence is modelled by a second-order closure Reynolds-Averaged Navier-Stokes (RANS) model.

Results

Quasi-steady results are obtained after 15 pump revolutions and the pressure drop across the pump is considered to assess convergence of the solution and to characterise the energy loss of the VAD. Investigations of the velocity field show that there exist tip vortices in the wake of the pump blades (see Figure 4) that rotate at high velocities. Strong velocity gradients and shear layers are observed in the outflow region potentially leading to haemolysis. Furthermore, investigations of the wall-shear stress reveal the existence of thin boundary layers at the blade tips. These results show that designs should be more blood sensitive to reduce the risk of haemolysis and thrombosis.

Conclusions – Future work

The simulations were carried out under the conditions set by the FDA, and advanced RANS modelling was used. FDA publication of the results has been delayed and we are still waiting for the comparison of our simulations with other software data and experimental results.

The same four institutions, namely FH Aachen, Juelich Forschungszentrum, EDF and STFC have got compute time on JUQUEEN (Juelich) to continue this project and to assess Large Eddy Simulation as a turbulence model, before simulating haemolysis.

Figure 3: Sketch of the model blood pump.

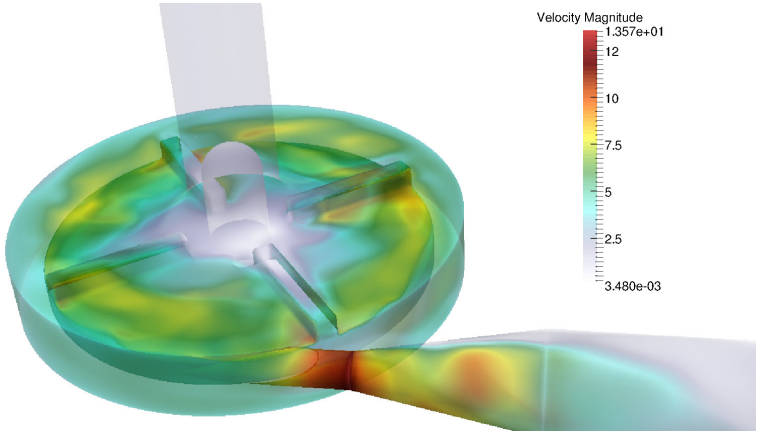


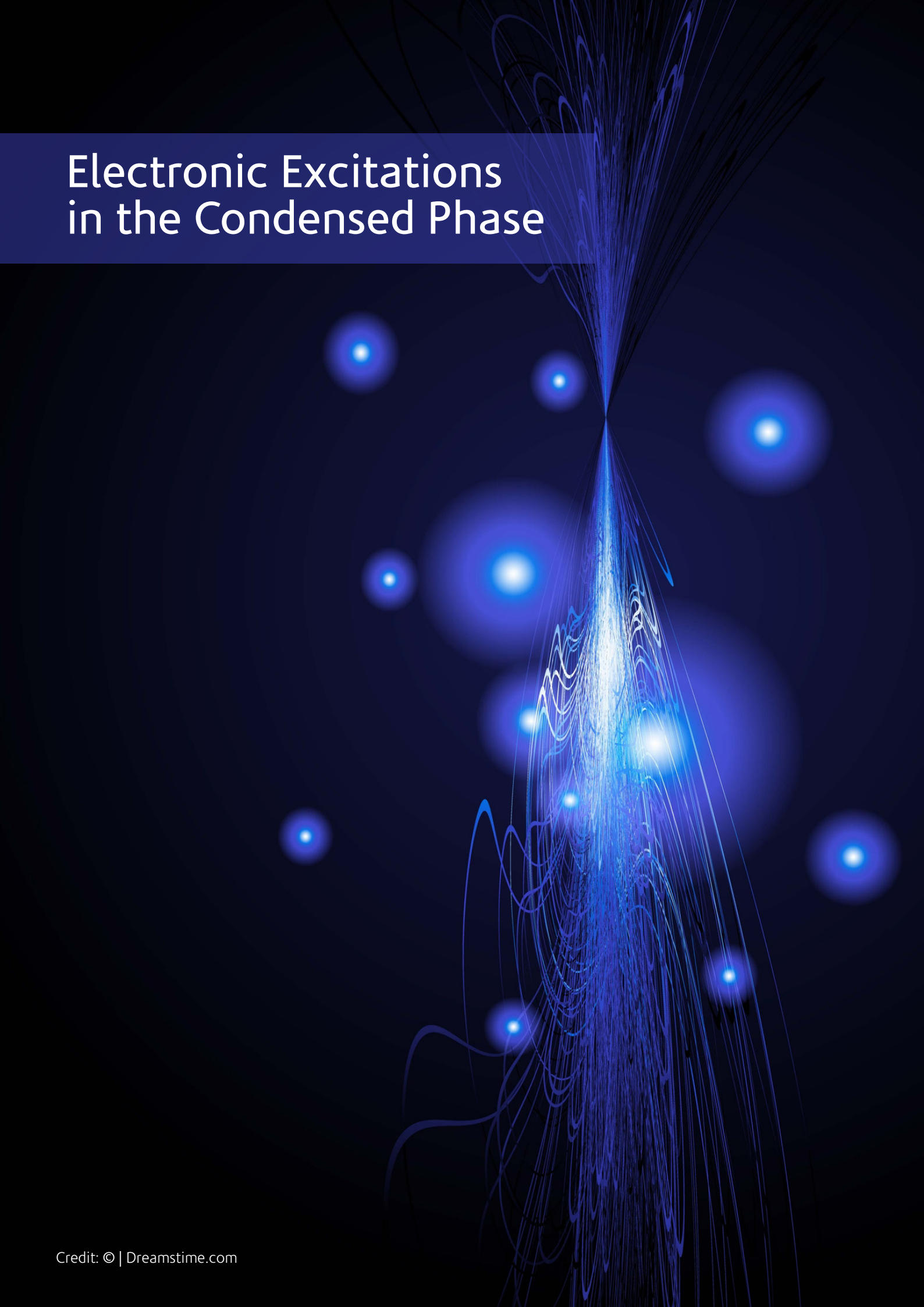
Figure 4: Velocity Field.

Authors
C. Moulinec, D. R. Emerson, STFC Daresbury Laboratory
The authors would like to thank M. Behbahani, J-H Goebbert, S. Rible, V. Marinova, I. Kerroumi (FH Aachen); M. Nicolai, P. Crosetto, A. Lintermann (Juelich Forschungzentrum) and Y. Fournier (EDF) for their contribution. The authors would like to thank the Hartree Centre and Juelich Forschungzentrum for providing time on their machines.

References

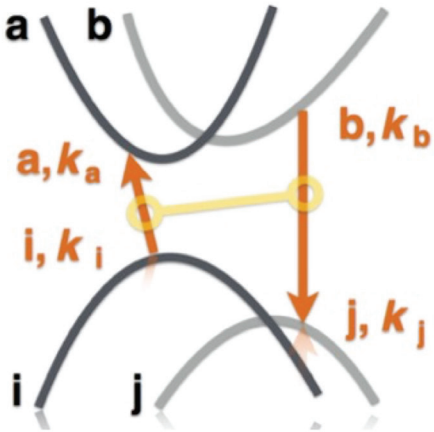
M. Behbahani, M. Behr, M. Hormes, U. Steinseifer, D. Arora, O. Coronado and M. Pasquali. A Review of Computational Fluid Dynamics Analysis of Blood Pumps. European Journal of Applied Mathematics. 20:4 pp. 363-397, 2009.

Electronic Excitations in the Condensed Phase



Credit: © | Dreamstime.com

The interaction of electromagnetic radiation (visible light, UV and IR radiation, X-rays, etc.) with matter is governed by the rules of quantum mechanics. When a photon incident on a sample is absorbed, one or more electrons are excited from their equilibrium state in normal conditions, the ground state Ψ_0, to one of the system's electronically excited states Ψ_n. The character of the excited state can be quite different from the ground state, e.g. in the way atoms are arranged or in its response to external perturbations. The ability to account quantitatively for the physics of electronic excitations is important for the interpretation of experimental spectroscopic data and for the prediction of complex phenomena occurring in catalytic processes, solar energy conversion and radiation damage in materials and biological systems, like proteins and DNA in living cells.



$$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

Figure 1 The exciton problem in TD-DFT. Electron-hole pairs (red) created by excitations from occupied (i, j) to unoccupied (a, b) one-electron bands in a crystal couple via a generalised eigenvalue equation (Casida's equation). This coupling creates bound states lower in energy than the empty bands. The number of electron-hole pairs determines the size of the matrices A and B and it can grow prohibitively large in calculations on infinite periodic systems.

There are two fundamental challenges in the physics of excited states than need to be addressed:

1. the mechanism through which an excited electron interacts with the positive hole created by the absorption of a photon (the ***exciton*** problem);
2. the atomic relaxation following the creation of the excited state Ψ_n.

These two classes of phenomena can manifest themselves differently in finite systems (molecules, small atomic clusters, etc.) and in condensed phase samples (e.g. crystals, surfaces, polymers, nanostructures, liquids and large biological molecules). The formation of excitons is particularly important in infinite crystals, e.g. the crystalline semiconductors used in various classes of photovoltaic devices, as well as at crystal surfaces and in polymers.

An exciton can be represented as an electron-hole pair, in which the electron and the hole carry a negative and a positive charge respectively. Effectively, this system is an exotic atom created by photoexcitation within a crystal lattice, similar to a hydrogen atom in which the nucleus is replaced by the positive hole. Excitons are the quasi-particles responsible for the transfer of energy in solar cells, LEDs and semiconductor circuits and they are involved in various natural processes, including photosynthesis.

The formation of an exciton is a quantum many-body process, in which virtual electron-hole pairs screen the bare excited electron and hole (Figure 1). The screened electron-hole pair is held together by Coulomb interactions. The development of methods capable of describing these phenomena fully from first principles, i.e. with no need of external empirical parameters, has been the focus of work carried within the Theoretical and Computational Physics group over several years. One of the successes of this effort has been the development of an original and powerful method for computing electronic excitations in solids based on the time-dependent density-functional theory (TD-DFT) framework. [1] This approach exploits a mathematical

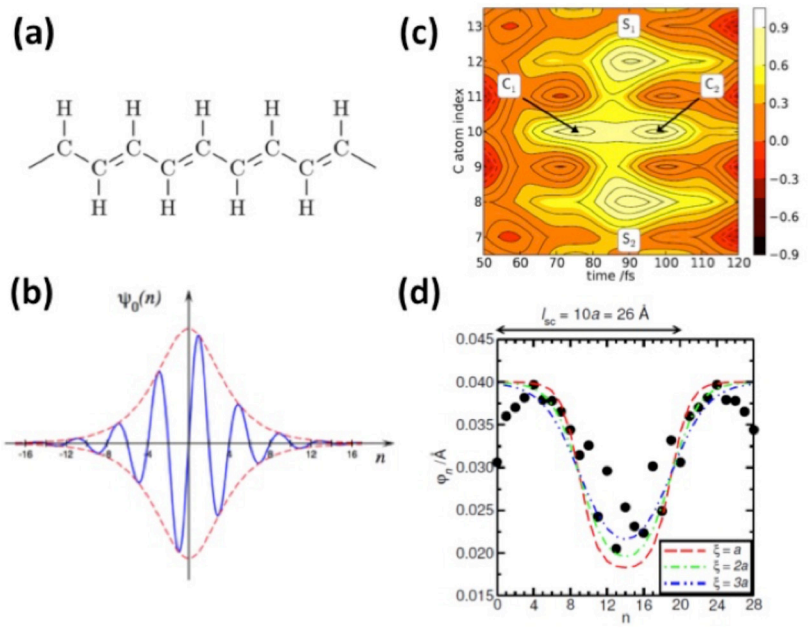


Figure 2 Photoexcited dynamics in trans-polyacetylene. (a) The structure of polyacetylene. (b) A soliton wavefunction. (c) Soliton collision at room temperature: electronic correlation determines the appearance of non-linearity in the soliton dynamics; the figure shows a 2-bounce resonance of a soliton and an antisoliton, an intriguing phenomenon observed in certain classes of non-fully integrable partial differential equations with chaotic solutions. (d) Condensation of solitons at zero temperature: a soliton and an antisoliton form a bound pair stabilized by electronic correlation.

procedure to recast the direct diagonalisation of large matrices describing the electron-hole screening into a more manageable self-consistent field (SCF) minimization of a suitable functional. Effectively, the SCF avoids the complexity of the direct coupling of elementary excitations, and allows calculations to be carried accurately and efficiently on three-dimensional infinite crystals. This method has been implemented in the CRYSTAL software, one of the most advanced electronic structure codes for extended systems currently available, whose development has been actively supported by STFC/CCLRC for almost 40 years, and it has been applied successfully to several problems including optical absorption in semiconductors, [1-2] the design of new materials for photovoltaics [3] and the nature of low-energy excitations in alkali halide crystals. [4]

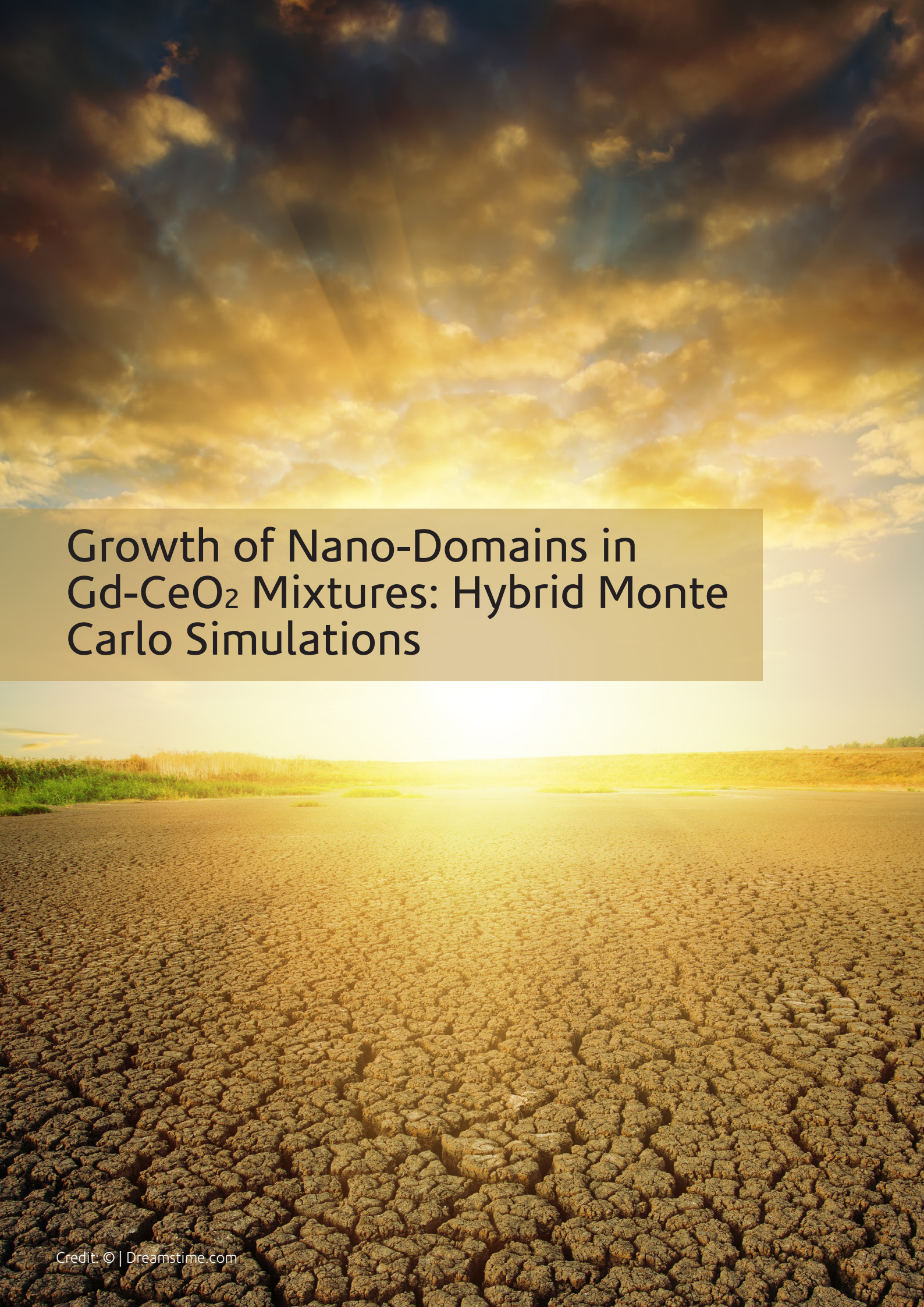
The second challenge, the nuclear relaxation following photoexcitation, is currently an open problem, with important implications in photochemistry, solar energy production (exciton transport in a crystal lattice) and biology (exciton trapping and energy transfer, photoinduced defect formation in macromolecules). Once a system is excited to Ψ_n, the forces acting on each atom change relative to the ground state, and the atoms rearrange to reach new equilibrium positions. The nuclear relaxation may induce the population of excited states other than Ψ_n, and the forces acquire multi-state (non-adiabatic) character. The inclusion of non-adiabatic effects for excited state relaxation in extended systems within the TD-DFT framework requires substantial methodological and algorithmic developments. Our preliminary work in this field has been focussing on polymeric quasi-1D systems, in which hopping between excited states occurs relatively infrequently. An interesting example is trans-polyacetylene (PA), an important prototype material for applications in photovoltaics, optoelectronics, nonlinear optics and electromagnetic shielding. We have used TD-DFT excited-state dynamics to examine the photoinduced formation of topological defects (solitons) in PA and their dynamics at room temperature (Figure 2). [5] Our work shows that electron-electron correlation in excited states affects profoundly the properties of the solitons, and brings about exotic non-linear effects in their dynamics. This work is currently being extended to biological macromolecular crystals, in collaboration the Diamond Light Source.

Authors
L. Bernasconi, Rutherford Appleton Laboratory

References

[1] L. Bernasconi, S. Tomić, M. Ferrero, M. Rérat, R. Orlando, R. Dovesi, N. M. Harrison, First-principles optical response of semiconductors and oxide materials, Phys. Rev. B 83, 195325 (2011)
[2] L. Bernasconi, R. Webster, S. Tomić, and N. M. Harrison, Optical response of extended systems from time-dependent Hartree-Fock and time-dependent density-functional theory, J. Phys.: Conf. Ser. 367, 012001 (2012)
[3] S. Tomić, L. Bernasconi, B. Searle and N. M. Harrison, Electronic and optical structure of wurtzite CuInS2, J. Phys. Chem. C 118, 14478 (2014)
[4] R. Webster, L. Bernasconi and N. M. Harrison, Time-dependent density-functional theory description of exciton binding in alkali halide crystals: The role of Hartree-Fock exchange, J. Chem. Phys. 142, 214705 (2015)
[5] L. Bernasconi, Chaotic soliton dynamics in photoexcited trans-polyacetylene, J. Phys. Chem. Lett. 6, 908 (2015)

Growth of Nano-Domains in Gd-CeO_2 Mixtures: Hybrid Monte Carlo Simulations



Credit: © | Dreamstime.com

Due to growing concern about global warming and dwindling supplies of fossil fuels, increasing interest is being directed towards alternative sources of energy. Popular and promising alternatives are fuel cells, and a substantial amount of research and development has been undertaken to improve the materials used in, and the design of, these devices. For example, in solid oxide fuel cells (SOFCs) the oxidant (e.g. air, O_2) is reduced at the cathode. The oxide ions produced from this process are then transported through the solid electrolyte material, ideally a purely ionic conductor, to the anode where the fuel (e.g. H_2, hydrocarbons) is oxidised. Electrons generated by this process then flow from the anode to the cathode, completing the circuit and generating power. Lowering the operating temperature of SOFCs to an intermediate temperature range of 600 °C - 800 °C is an ongoing area of intense research and gadolinium doped ceria (CeO_2) is a key material suggested as a suitable electrolyte for this purpose [1].

The efficiency of a solid electrolyte material is linked to how freely the oxide ions can move through it and thus the ionic conductivity. The conductivity of CeO_2 can be enhanced by doping with a cation with a charge that is different from that of Ce. In gadolinium doped ceria (GDC) a Gd^{3+} ion replaces a Ce^{4+} and to maintain charge neutrality an oxygen vacancy is created for every two Gd^{3+} dopants. Thus the formula of the solid solution can be written as $Gd_xCe_{1-x}O_{2-x/2}$ where x can take values from 0 to 1. This implies the ionic conductivity should increase steadily with the number of dopant ions. But it does not, the experimental behaviour is considerably more complex and the ionic conductivity is not a simple function of dopant concentration, but passes through a pronounced maximum at $x_{Gd} = 0.1$ [2]. In addition, the conductivity is very sensitive to the thermal history of the sample and is observed to degrade if the sample is exposed to elevated temperatures [3].

In order to investigate the thermodynamic and structural properties of materials containing large numbers of defects (either dopants and/or vacancies) it is absolutely vital to sample a large number of different configurations of atoms. In conventional Monte Carlo (MC) it is possible to exchange the position of ions chosen at random in order to sample multiple configurations. Unfortunately, in Gd doped CeO_2 sampling of different configurations will be very poor due to the different size and charge of the cations and it is essential to resort to a different approach, i.e. some sort of relaxation is required to allow the atoms move when ions are exchanged and take account for the very different immediate local environments of the Gd and the Ce.

We have developed a computational strategy in which a short molecular dynamics (MD) simulation provides the necessary relaxation. This method improves the efficiency of the simulation dramatically and makes simulations on these systems feasible.

The structural properties of $Gd_xCe_{1-x}O_{2-x/2}$ can be determined over the entire solid solution and allows our simulations to be compared with powder diffraction and EXAFS experiments. For example, figures 1 and 2 display calculated bond lengths determined from EXAFS and diffraction data.

The conductivity can be calculated [4] for atomic configurations determined from the HMC simulations and compared to configurations in which the cations are randomly distributed. We have assumed that the sintered samples in reference [3] can be represented by a random configuration of cations (and vacancies) and the aged samples by configurations obtained from the HMC simulations. Our results are in good agreement with the experimental results in reference [3] and the calculated conductivity of the HMC configurations is greatly reduced for $x \geq 0.15$ (figure 3). This is the composition at which we observe the formation of a network of Gd-rich ions (figure 4). Our results suggest that the growth of Gd-rich domains have significant impact on the conductivity of GDC, preventing the passage of oxide ions through the electrolyte. Indeed, we attribute the ageing of GDC electrolytes to the growth of these domains.

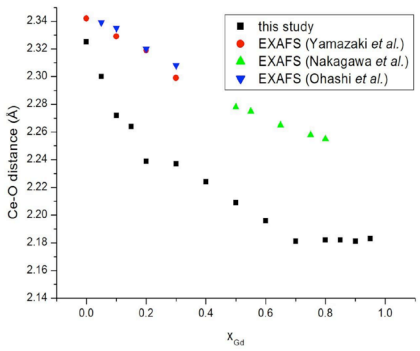


Figure 1. Ce-O distances as a function of composition for $Gd_xCe_{1-x}O_{2-x/2}$. The experimental data are from Ohashi *et al*. [5], Nakagawa *et al*. [6], and Yamazaki *et al*. [7].

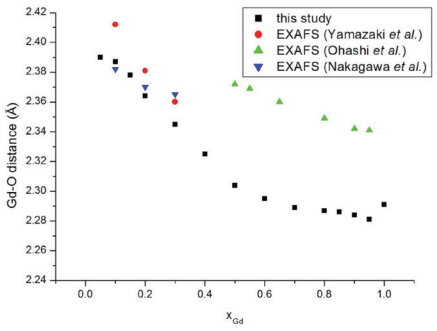


Figure 2. Gd-O distances as a function of composition for $Gd_xCe_{1-x}O_{2-x/2}$. The experimental data are from Ohashi *et al*. [5], Nakagawa *et al*. [6], and Yamazaki *et al*. [7].

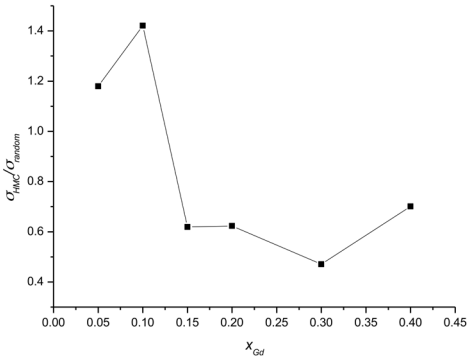


Figure 3. The ratio of ionic conductivity obtained from HMC starting configurations, σHMC, and those obtained from random configurations of cations, σrandom. The ionic conductivity is assumed to be entirely from oxygen ion diffusion.

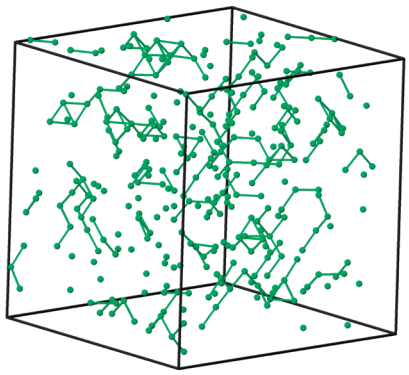


Figure 4. The distribution of Gd ions in GDC for x_{Gd} = 0.15. Only the Gd ions are shown and where a next-nearest neighbour interaction exists a "bond" has been inserted (this "bond" is only for graphical purposes and there is no other inference as to physical interaction).

Authors
J. Purton, N. Allan and D. Gunn, STFC Daresbury Laboratory

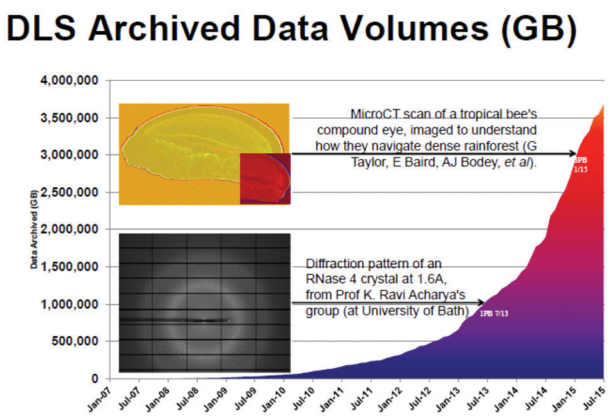
References

[1] Sun, C.; Li, H; Chen, L. Nanostructured ceria based materials: synthesis, properties, and applications. Energy &Environmental Science, 2012, 5, 8475-8505.
[2] Steele, B.C.H. Appraisal of Ce1-yGdyO2-y/2 electrolytes for IT-SOFC operation at 500 °C Solid State Ionics, 2000, 129, 95-110.
[3] Zhang, T.S.; Ma, J.; Kong, L.B.; Chan, S.H.; Kilner, J.A.Aging behaviour and ionic conductivity of ceria-based ceramics: a comparative study. Solid State Ionics, 2004, 170, 209-217.
[4] Kilo, M.; Argirusis, C.; Borchardt, G.; Jackson, R.A. Oxygen diffusion in ytria stabilised zirconia-experimental results and molecular dynamics calculations. PCCP, 2003, 5, 2219-2224.
[5] Ohashi, T.; Yamazaki, S.; Tokunaga, T.; Arita, Y.; Matsui, T.; Harami, T.; Kobayashi, K. EXAFS study of Ce1-xGdxO2-x/2. Solid State Ionics, 1998, 113-115, 559-564.
[6] Nakagawa, T.; Osuki, T.; Yamamoto, T.A.; Kitauji, Y.; Kano, M.; Katsura, M.; Emura, S. Study on local structure around Ce and Gd atoms in CeO2-Gd2O3 binary system. J. Synchrotron Radiation, 8, 2001, 740-742.
[7] Yamazaki, S.; Matsui, T.; Sato, T.; Arita, Y.; Nagasaki, T. Defect structures in doped CeO2 studied by using XAFS spectrometry. Solid State Ionics, 2002, 154-155, 113.

Data Services Group

Credit: © | Dreamstime.com

The data services group runs a range of data services – data storage providing backups of science data, working repositories, archives, and file servers – as well as databases which manage the metadata for all the data, plus other services such as user accounts and accounting.



The data services group's storage services are based on CASTOR (the CERN Advanced Storage Manager), which is a petabyte-scale storage system developed at CERN for the high-energy physics community. The largest user of CASTOR is the LHC Tier 1, which uses CASTOR to store ~11PB of data on disk and ~12.5PB data on tape in the robot. This storage is used both for long-term archival and as fast-access disk-only storage to support the batch farm. Preparation for the recent LHC run 2 led to a concerted push for upgrades, both on the operating system and software level. We encountered a number of problems specifically related to the way CMS use the storage system. The problems manifested themselves in three ways; extremely poor CMS job efficiencies, long open times and an inability to use our site for xroot fallback. While the long opening times remain an issue, the CPU efficiency for the worst case workflow has improved from 30% to 80%, comparable to other Tier 1s, and the "federated access" efficiency has improved from near 0% to about 70%.

The Tier 1 service has recently been extended to another community: it has been adopted by the DiRAC project to back up about 5PB of data, starting with Durham University's Institute for Computational Cosmology. The data transfers for the backups are using grid technology to ensure we can move large volumes of data efficiently.

The 'facilities' instance provides tape backup and archival for local RAL users, including the Diamond Light Source and the Central Laser Facility. The data volume for this instance is around 7.5PB, almost entirely on tape, and usage is expected to expand significantly as new data-heavy detectors are installed at Diamond. This service also includes the ICAT data catalogue.

The graph above shows the rapid, almost exponential, increase in data volume over time held on hundreds of tapes within one of the tape robots at RAL – although recent growth looks more linear than exponential. In any case we are well placed to support this increase in data volume with a currently theoretical maximum storage of 80PB per tape robot. Moreover, we have recently got another Horizon2020 project called ESiWACE in which we will investigate the long term future (the next 10-15 years) of tape storage for climate modelling; within STFC, also RAL Space are a part of this project and will be looking at metadata management.

Back to the more imminent future, developments include collaborations on the CEPH storage, which is slated to become the replacement for CASTOR's disk only storage. The advantage of CEPH is that we can use erasure coded storage, so won't need to rely on replication of data, nor for the recovery of data from tape if a disk server goes down. This work is in collaboration not just with the PSCS group but also with other departments in STFC and with other sites using, or planning to use, CEPH. We plan to retain and reengineer CASTOR as a dedicated tape endpoint for Tier 1 and facilities users, to support the archiving and backup use cases.

EUDAT2020

Another one of the group's Horizon2020 projects is EUDAT2020. The project builds on the success of EUDAT and aims at extending existing services to new communities and adding new services which have been identified in surveying of partners. The current services from EUDAT cover a broad section of data management needs, from the simple B2DROP, a simple Dropbox like service based on OwnCloud to a more managed regime in B2SAFE where data is replicated amongst several sites for both security and performance. EUDAT2020 will extend the services to include B2HANDLE, allowing users to get PIDs (persistent identifiers) for their own purposes without having to use other EUDAT services and B2ACCESS providing federated identity management. In addition, there are plans to provide a simple RESTful web interface to the services to allow programmatic access for communities requiring this. Closer collaboration with other e-Infrastructures is also planned, including EGI, HelixNebula and OpenAire.

So what does the future hold for data services? Whether fixing motorcycles or software, the Zen principle applies. Data services rely on highly complex software products such as Oracle databases and CASTOR; some of the facilities' data workflows are quite complex. We continue to expand the group's expertise in maintaining and improving services while also delivering data services to the projects such as the Horizon2020 projects EUDAT2020 and ESiWACE. Every user community is different and their needs will change with time. By combining the expertise in running services with our collaborations with other groups and projects, we should be able to provide high end scientific data services for many years to come.

Authors
J. Jensen, R. Appleyard, C. Prosser,
S. de Witt STFC Rutherford Appleton Laboratory

Research Data for Environmental Applications

Credit: NASA Goddard Space Flight Center Image by Reto Stöckli

As well as its work supporting the data needs of STFC's facilities, the Research Data Group is also active in the area of environmental data, working with RAL Space, the Centre for Environmental Data Analysis (CEDA) and other partners to maximise the value from this important resource.

There are a number of challenges and opportunities around environmental data. Much of it is not reproducible - for example observations of the Earth taken from satellites that reflect conditions at a particular time and place and cannot be repeated. The data has considerable potential for reuse and repurposing, and can be combined with other datasets to allow new analyses and interpretations of important issues such as climate change and flood risk assessment. And it comes with all the challenges of "big data" in general, including large volumes and high rate of growth.

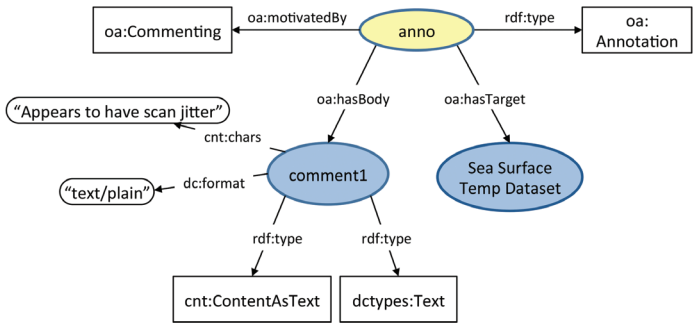
For environmental scientists, one of the key challenges is identifying and locating existing datasets suitable for their purposes. Research Data Group is working with CEDA as part of the CLIPC project to set up a "vocabulary server" allowing consistent and reliable querying of datasets. A vocabulary of environmental terms (such as "sea surface temperature") has been defined and related both to other terms and to pre-existing vocabularies, using SKOS and OWL to represent relationships between terms. This means that environmental scientists can perform searches across many datasets using different descriptive terms and be confident of finding the data they need.

The CHARMe project, funded by the European Commission, has developed a system to allow data users (the environmental scientists) to make annotations on datasets concerning their usefulness, quality and limitations, allowing other members of the user community to judge the value of the data for their own purposes. A central node, running on the JASMIN infrastructure, manages the collection of comments, and individual sites run client software allowing users to view and add comments. Special attention has to be paid to managing comments and tracing provenance, for example if comments are deleted. The Research Data Group provided technical support.

ESPAS is another European project, building the e-infrastructure necessary to support the access to observations, the modeling and prediction of the near-Earth space environment. As an infrastructure with demanding usage requirements, good software development practices are essential, and Research Data Group has been providing advice on best practice software lifecycle management.

More broadly, Research Data Group is, along with CEDA, representing STFC in the recently founded Institute for Environmental Analytics (IEA). This organisation is based at the University of Reading and is funded by HEFCE's Catalyst Fund to help bridge the innovation gap in the market for information about weather and climate change. Other partners include the Met Office, universities, technology suppliers such as Airbus Defence & Space and Microsoft Research, and end-user organisations (Sainsburys and Lighthill Risk Network). The IEA will be conducting strategic R&D and undertaking training and engagement activities with a view to transforming into a self-sustaining leader in the field.

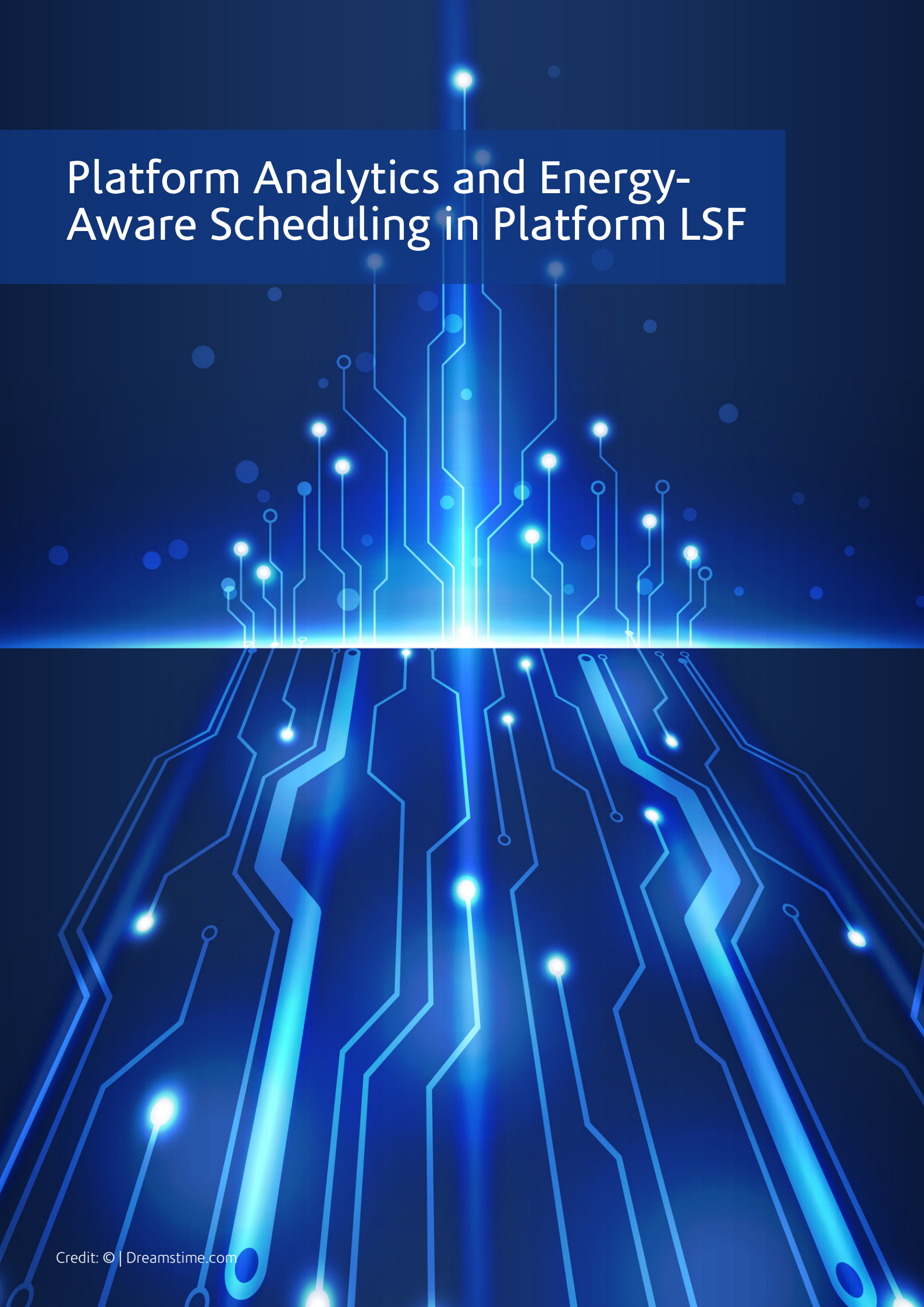
In last year's report, the SCAPE project on long-term preservation of digital data was described. Within the environmental domain, RDG has been involved in another project with this focus, SCIDIP-ES. Focussing on users in Earth science, and working with ESA, the project upgraded a set of prototype preservation services into scalable, robust e-infrastructure components to support digital preservation of all types of digital objects, founded on the OAIS (Open Archival Information System) standard, the ISO standard for long-term archives.



This is an example of using CHARMe to comment on a dataset

Authors
S. Lambert, A. Wilson,
STFC Rutherford Appleton Laboratory

Platform Analytics and Energy-Aware Scheduling in Platform LSF



Credit: © | Dreamstime.com

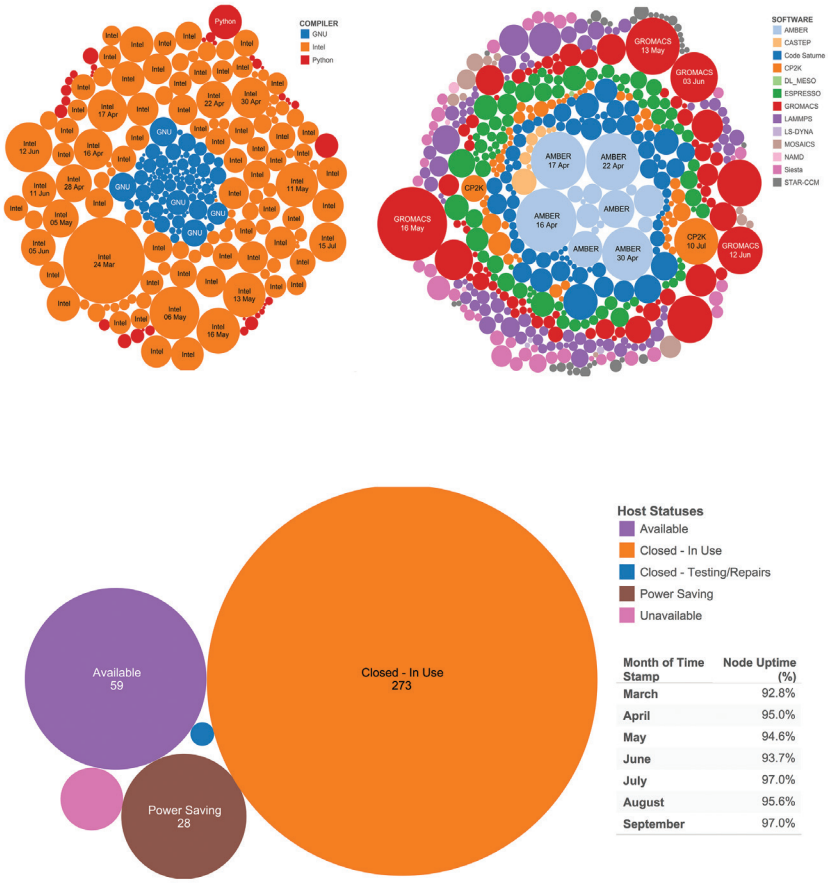
HPC provides a platform for performing computationally demanding calculations for a huge variety of disciplines. Every day hundreds (if not, thousands) of these calculations are sent to a wide range of systems at STFC. We have recently acquired the analytical software, IBM Platform Analytics (PA), which allows us to investigate a significant number of metrics of the systems, which we can use to gain valuable knowledge and insight. The data analysis and visualisation is performed by Tableau, which in turn reads data stored in HP Vertica databases.

Platform Analytics at STFC:

One of the key use cases for PA is the ability to access the health and capacity of the systems. We can then use this information to react to the changing dynamics of the clusters. This feedback has been used to change the cluster to better fit the workloads that are using it. From analysing the historical cluster capacities of our systems in PA, we identified that one system was heavily oversubscribed, and another undersubscribed. We then reconfigured the scheduler, LSF, to send jobs to a cluster dynamically based on the current utilisations.

We can move to a finer granularity to gather trends on a job-by-job basis, which is particularly useful to external customers who require metrics on the job, group and project levels. We are able to design custom dashboards that host a variety of data and statistics, which are generated using PA (Tableau) designer. The data is presented in the form of graphs and plots, which can be filtered dynamically to quickly assess multiple dimensions of data. Platform Analytics is also being utilised by Helpdesk support, who are able to query job details prior to forwarding the query onto the operations teams, which speeds up the query resolution process. The granularity can be extended to the amount of metrics gathered from the cluster. For example, a project investigator may want to know how many pieces of software are being used on the cluster for licensing purposes. Platform Analytics collects a number of metrics from the systems as standard, but to enhance the functionality of the software, custom loaders and ETL (extract, transform and load) calculations are added. For the software dashboards, we have designed an ETL that extracts the binaries of commonly used software and which version of MPI is used.

Another key part of HPC operations is visibility. We have generated dashboards that are on display internally within the organisation that show the cluster capacity, utilisation and workloads on the HPC.

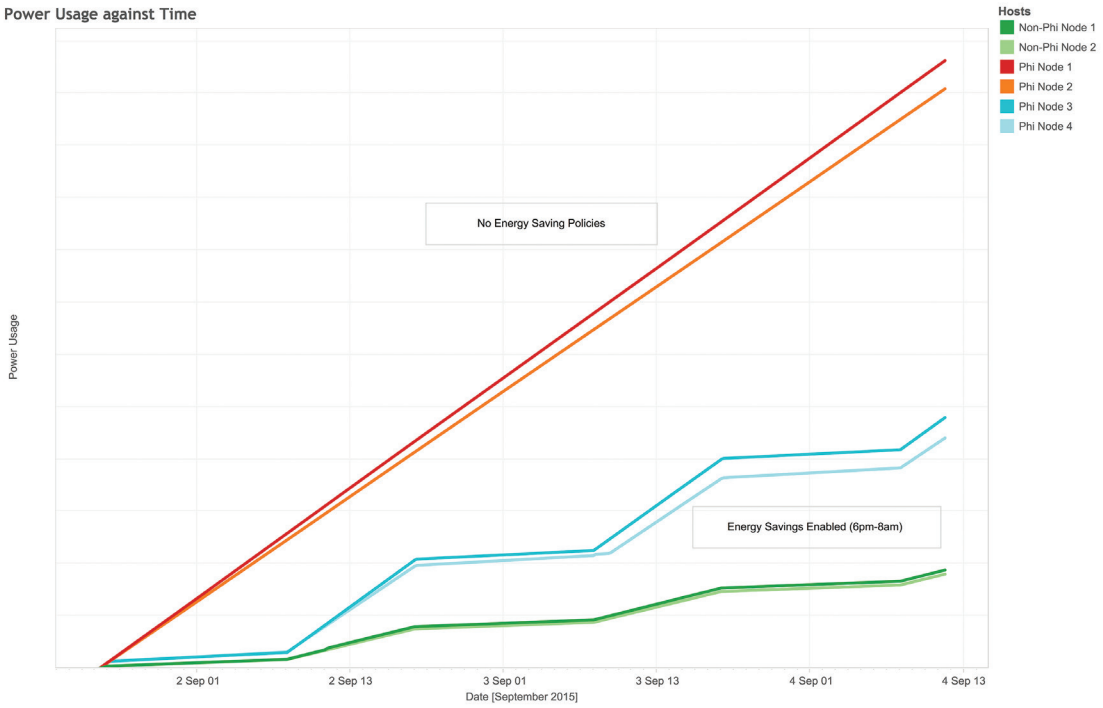


Month of Time Stamp	Node Uptime (%)
March	92.8%
April	95.0%
May	94.6%
June	93.7%
July	97.0%
August	95.6%
September	97.0%

Investigating power usage using Platform Analytics:

There has been an interesting overlap between the analytics and energy efficient computing (EEC) projects, where we have been able to visually interpret and understand the benefits of initiating energy saving policies on the systems. We have been investigating using S3 suspend states for nodes when they're idle during the night. We have two methods of gathering power metrics from the systems, one from the scheduler, LSF, and the other from querying the energy counters via the IPMI interface. In the graph below we can see the power usage of iDataplex nodes with and without power savings enabled. It's evident that there is a clear power saving when utilising S3 states. Interestingly, the nodes hosting Intel Xeon Phis can be seen as are a major contributor to the total energy of the node.

We are able to relate power consumption with a number of different metrics in PA, including software usages. We are able to see which pieces of software are using the most resources with respect to both the hardware and power usage.



Author
J. Baker, STFC Daresbury Laboratory

Supporting e-Infrastructures

Credit: © | Dreamstime.com

The Research Infrastructure Group (RIG) develops and hosts a number of production services that are highly visible in a number of national and international scale compute infrastructure projects. These services form part of the core operational services for many compute sites across the UK and EU.

Impact for EGI sites

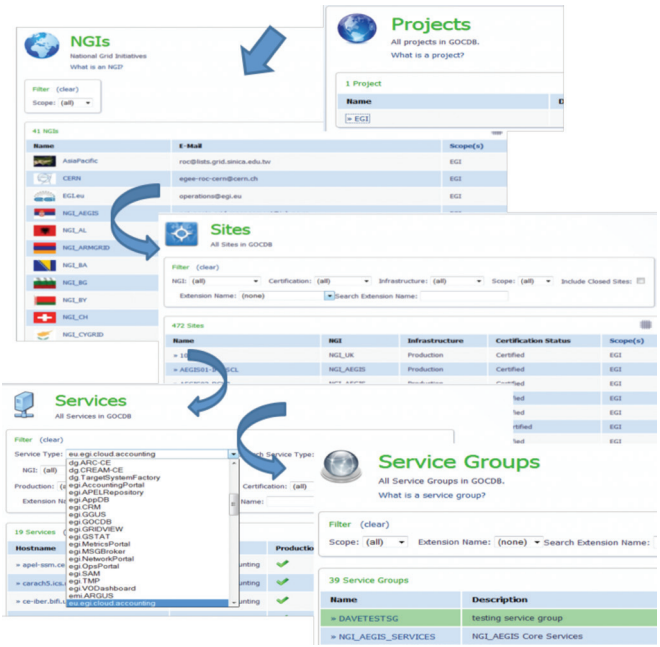
During 2015, the European Grid Infrastructure project (EGI) (http://www.egi.eu) conducted a survey to measure the level of impact and criticality of their core operational services, which included both GOCDB and APEL (described below). Both services were consistently recognised as having both high impact and a high level of criticality. For example, the resource centres that participate in EGI and provide compute resources regarded GOCDB as the most critical operations service (94.12% of responders), while APEL was regarded as the most critical service by user communities (76.92% of responders). Both services also scored highly in other related metrics.

GOCDB

GOCDB stands for 'Grid Operations Centre Database' and is used to record computing sites and services that participate in the European Grid Initiative (EGI) and/or the Worldwide LHC Computing Grid (WLCG).

The GOCDB holds technical details about the computing sites and resource centres, including service endpoints, internet IP addresses, contact details and roles of different users, and periods of downtime when services become unavailable due to maintenance.

The data in the GOCDB is accessible to authenticated users who can browse the web portal interface, and is programmatically available via an API which is queried by computer systems from across the world.



Drilling down through the GOCDB

Whilst traditionally being core to the EGI and WLCG projects supported by EGI and H2020 funding, the GOCDB project is now forging wider links with other large collaborative projects such as EUDAT (http://eudat.eu) as well as looking at ways to support Bioinformatics communities.

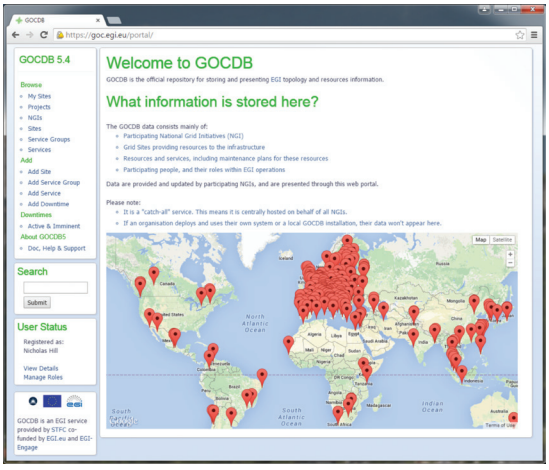
APEL

The Accounting Processor for Event Logs (APEL) is an accounting system. It allows individual scientists, site administrators, and VO (Virtual Organisation) managers to monitor compute resources used across large grid infrastructures.

As the name implies, the APEL tools collect accounting data by parsing log-files from a massive network of computers distributed across the world. These computers/services participate in the EGI and WLCG infrastructures as well as from sites belonging to other Grid organisations that collaborate with EGI.

The accounting information is collated into a central accounting database where it is processed to generate statistical summaries that are available through the EGI/WLCG Accounting Portal. The statistics are available for viewing in different levels of detail by Users, VO Managers, Site Administrators and anonymous users according to well defined access rights.

At any time, the APEL database holds accounting records and summaries for the preceding 18 months, which equates to 450GB of data and over 530 million records. In the last year, accounting information was received from 314 separate sites at the rate of around 3 million records a day.



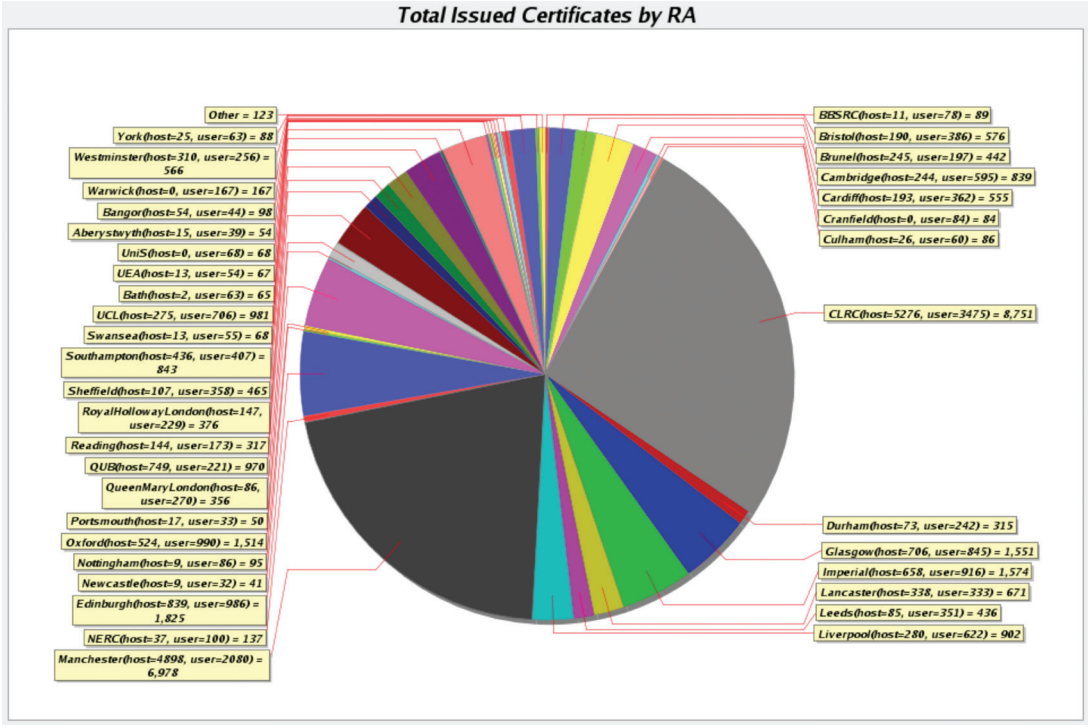
The GOCDB web portal

As with the GOCDB project, APEL has traditionally been core to the EGI and WLCG projects supported by EGI and H2020 funding but is now forging wider links with other large collaborative projects such as the H2020 funded INDIGO-DataCloud project (https://www.indigo-datacloud.eu).

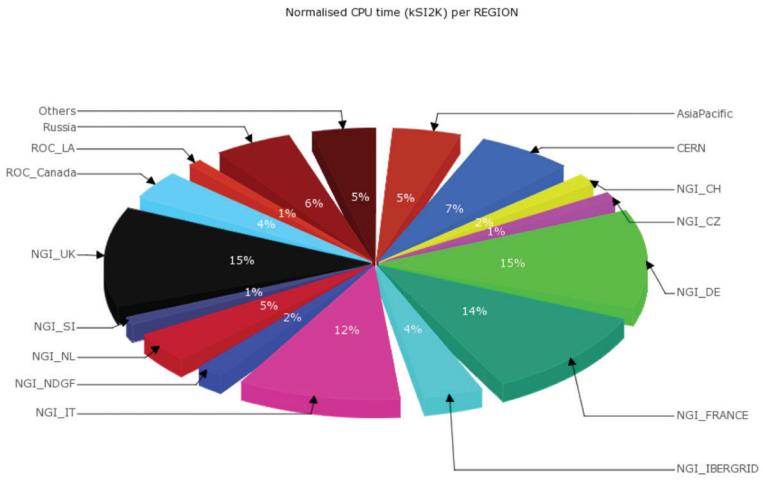
UK eScience Certification Authority

The UK eScience Certification Authority service (UKCA) (http://www.ngs.ac.uk/ukca) is essential for the UK's participation in a number of highly visible e-infrastructure projects, including the WLCG and EGI projects. Communities from these projects require secure two-way communication that also provides a high level of assurance when authenticating its users and servers. This is achieved using "x.509" digital certificates issued by a trusted authority. The UKCA provides this service for the UK academic and research communities and is trusted by the Interoperable Global Trust Federation (https://www.igtf.net). This ensures our certificates can be used to authenticate users with a variety of Grid, Data, Compute and Cloud resources around the world.

A recent survey conducted by the UKCA revealed that the service has 2553 active certificates, with approximately 800 personal certificates, 1700 host certificates and a number of additional special-use certificates.



Breakdown of certificates issued by the UK Certificate Authority to different UK academic institutes



Accounting portal plot of CPU time using APEL accounting data

Authors
D. Meredith, N. Hill
STFC Rutherford Appleton Laboratory

Computing for Physics Research beyond the "Standard Model"

The CERN Large Hadron Collider (LHC) switched on for the start of "Run 2" on the 3rd of June 2015. Since the restart the LHC has been providing collisions to its experiments at 13 TeV, almost double the collision energy of its first run. The LHC will run round the clock for the next three years producing even greater volumes of data for processing on the Global LHC Computing Grid of which the GRIDPP Tier-1 Centre at RAL is a major partner.

To meet the increased capacity requirement of Run-2, Tier-1 storage capacity has already increased to provide 18 Petabytes of disk storage, spread over 400 disk servers and 24 Petabytes of tape storage in an SL8500 tape robot. The data is received from CERN on a dedicated 20Gb/s Optical Private Network (OPN) and is then available to be processed on a compute cluster of 13,000 cores – as well as additional capacity on the Scientific Computing Department Cloud.

Further LHC Discoveries

The Higgs boson (announced in July 2012), as proposed within the Standard Model, is the simplest manifestation of the Brout-Englert-Higgs mechanism. Other types of Higgs bosons are predicted by other theories that go beyond the Standard Model. The extra energy of the LHC collisions in run 2 should significantly increase the chance of Higgs boson being created, this will allow the properties of the Higgs boson to be measured precisely.

The Standard Model of particle physics has been incredibly successful in providing experimental predictions, it does leave some phenomena unexplained and it is not yet a complete theory of fundamental interactions. Many extensions to the Standard model have been predicted such as Supersymmetry, Exotic particles and extra dimensions. Most of these theories predict the existence of new particles which the LHC experiments will be uniquely placed to search for.

Cloud Computing in SCD

In spring of this year SCD launched an Infrastructure-as-a-Service cloud computing service, integrated into the Tier 1 management and monitoring infrastructure. This currently consists of 896 cores, 3.5TB of memory, backed by a 0.7PB Ceph cluster for the underlying storage.

The cloud service will provide part of STFCs commitment to the INDIGO Datacloud and West-life H2020 projects, as well as supporting developing requirements in LHC computing. Spare capacity is already being used elastically by virtual machine on demand web portal has also been launched to allow members of SCD to quickly access VMs for testing and development work, increasing the efficiency of other projects. We are also participating with the cross STFC Virtualisation Working Group.

There are currently over 80 users of the cloud service. With the exception of a small number of power users, they are using our web portal (cloud.stfc.ac.uk). As of September 2015, over 50 unique users had VMs running and we consistently have over 250 VMs running (including virtual worker nodes running LHC computing jobs). Over 13,000 VMs have been launched since February.

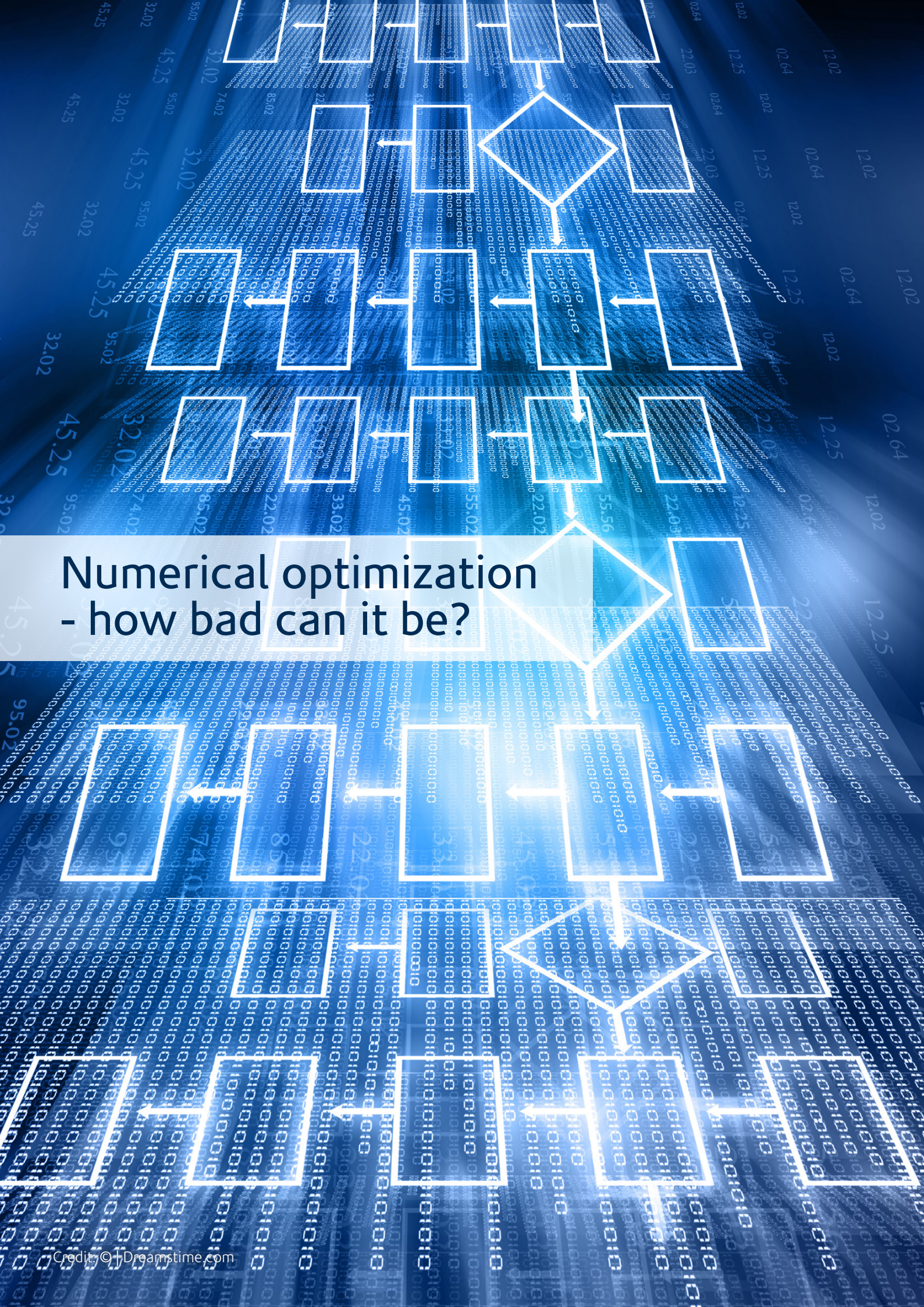
High capacity Ceph Object Store

During Run 2 the data taking rates of the LHC experiments will increase substantially. The extra energy will produce larger events which are more complex to analyze. SCD are deploying a Ceph backed object store to cope with the increased data taking challenges. Object stores provide a much simpler feature set compared to traditional file system storage. This allows them to be scale to the multi-petabyte level comfortably.

To allow the LHC experiments to access their data in the same way as before, SCD, in collaboration with others, is working on providing XrootD and GridFTP gateways. An S3/Swift compatible API is also being provided, which may provide a better long term solution to accessing the storage. SCD will provide a production quality object store for data storage by April 2016 running in parallel with the existing Castor storage for a year before the Castor disk only storage is completely phased out by 2018.

Author
I. Collier, STFC Rutherford Appleton Laboratory

Numerical optimization - how bad can it be?



One of the key concerns of the Numerical Analysis Group is the solution of nonlinear optimization problems. That is to say, the minimization of some provided function f of many parameters x, where perhaps the parameters are constrained to take on certain values. The Group provides optimization software through its GALAHAD library, and this is used worldwide by thousands of scientists, engineers, economists and planners; typical applications include data fitting by application scientists, option valuation in financial markets, efficient distribution of resources by public utilities and exploration of potential energy surfaces. Our users feel confident that our methods will work, since the algorithms we implement are covered by reassuring convergence theory and sound numerical practice.

When we apply a numerical method to solve any scientific problem, we are normally concerned with how well it works. Does the algorithm solve the problem? Does it solve it accurately? An important question that we rarely ask, though, is whether the problem at hand will prove to be difficult for our algorithm. Indeed, given an algorithm, is there a fiendish example that might act as its nemesis?

In its most primitive, unconstrained form, a minimum of f will occur when its gradient $g(x)$ is zero, and optimization methods generally claim success when the gradient is small, or more formally when $\|g(x)\| \leq \varepsilon$ for some small, user desired value $\varepsilon > 0$. All of our favourite optimization methods are iterative, that is, from a given initial guess x_0, a sequence $\{x_i\}$, $i > 0$, is generated, and the method terminates at the first iterate x_k for which $\|g(xk)\| \leq \varepsilon$. Since each iteration requires some, and often considerable, computational effort, the question we posed above may then be rephrased as determining how large k might be.

This question is essentially impossible to answer unless we impose some regularity on the functions we aim to minimize. If our function is smooth, does not "wiggle" too much (formally, its gradient is Lipschitz) and is bounded from below, one can show for some of our most popular methods that the largest possible number of iterations k will be no more than a constant multiple of ε^{-2}; the constant takes into account how much wiggle there is, and the range of possible f values. To put this into context, if ε is a very typical 10^{-6}, the bound is some multiple of a trillion iterations. Even worse, it is possible to find a function for which this frightening bound is actually achieved (see Figure 1). While the bound applies to the much-derided method of steepest descent, more worryingly it also applies to most provably convergent variants of the ever-popular Newton's method such as those embedded in linesearch and trust-region frameworks.

Can we do better? Fortunately, yes, another variant of Newton's method, in which bad behaviour is avoided

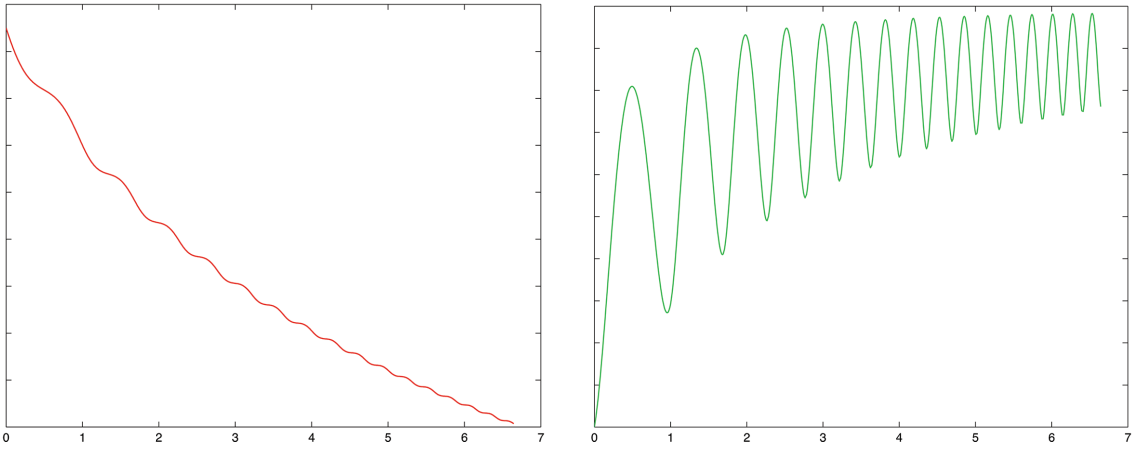


Figure 1: A bad function f for Newton's method (left) and its gradient (right). The example is constructed so that each iterate x_i falls at a place where f is locally steepest.

by regularization, can be shown to require at most a multiple of $\varepsilon^{-3/2}$ iterations, although even here examples exist for which the bound is attainable. Again, in the context we mentioned above, we might still require a billion iterations, but this is a million times better than for the more traditional approaches. The worst-case bound can be improved if we are able to compute higher-order derivatives; if we can find derivatives of order up to p, there are methods that only require of the order of $\varepsilon^{-(p+1)/p}$ iterations. Moreover, if we restrict our attention to better-behaved functions, things improve further; for convex functions the bound drops to the order of ε^{-1} iterations, while for strictly-convex functions it becomes a very modest, order $\log_2 |\log_2 \varepsilon|$ iterations.

Should we worry? Almost certainly not! Vast empirical evidence suggests that good implementations of Newton-like methods, such as those in GALAHAD, are extremely reliable and unlikely to require anything remotely approaching the huge numbers of iterations we mentioned above; tens of iterations are far more typical in anything other than really hard cases. This suggests something deep about the relationship of average and worst-case optimization problems, and we are currently investigating this issue. Whether one day some physicist's or economist's model will coincide with one of our bad examples is a different question entirely.

Author
N.Gould, STFC Rutherford Appleton Laboratory

References:
N. I. M. Gould, D. Orban and Ph. L. Toint, "GALAHAD–a library of thread-safe fortran 90 packages for large-scale nonlinear optimization". ACM Transactions on Mathematical Software 29(4) (2003) 353-372. See http://www.galahad.rl.ac.uk.

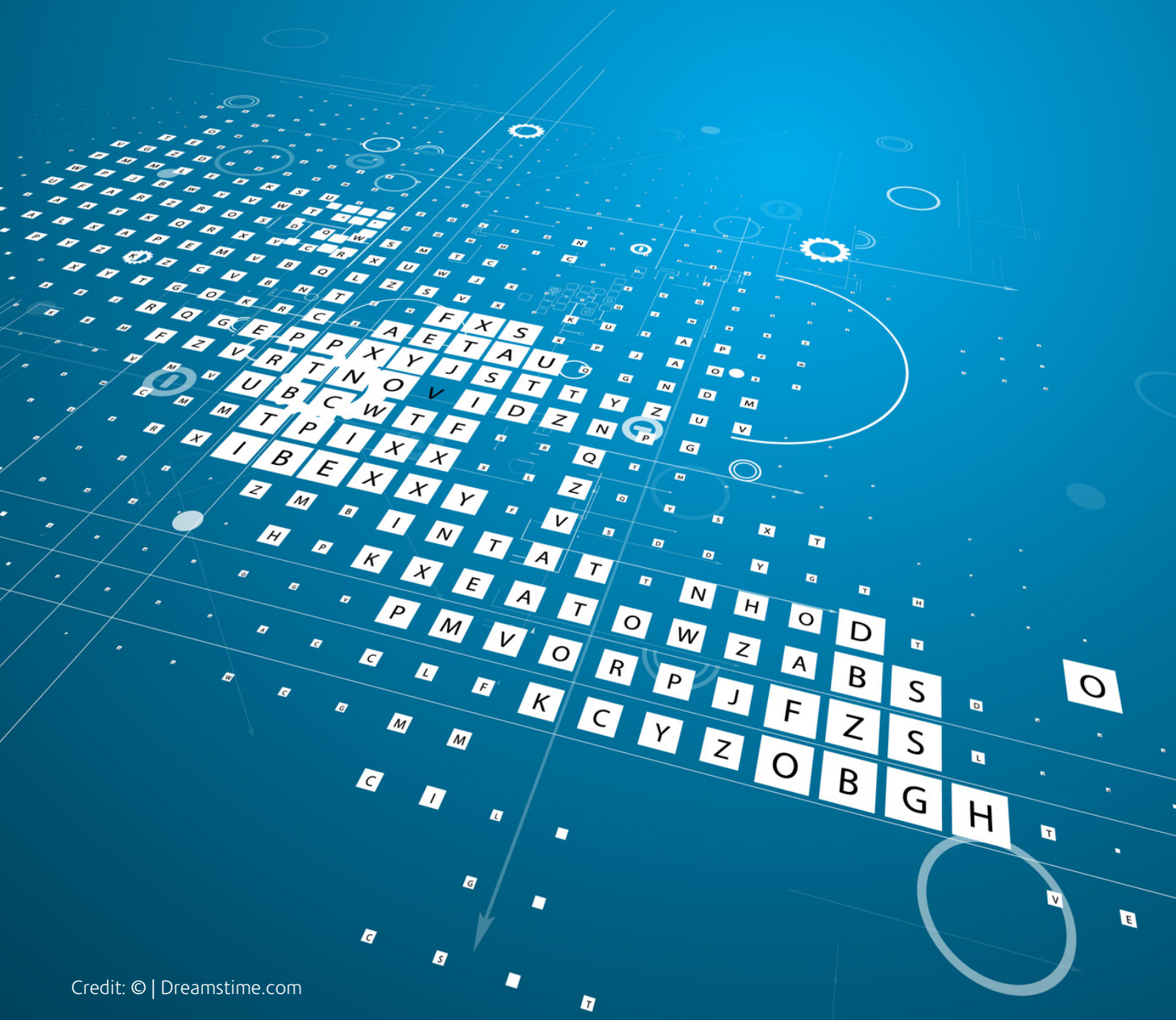
C. Cartis, N. I. M. Gould and Ph. L. Toint, "On the complexity of steepest descent, Newton's method and regularized Newton's methods for nonconvex unconstrained optimization problems".

SIAM Journal on Optimization 20(6) (2010) 2833-2852.

C. Cartis, N. I. M. Gould and Ph. L. Toint, "Adaptive cubic regularisation methods for un-constrained optimization. Part II: worst-case function and derivative-evaluation complexity" Mathematical Programming 130(2) (2011) 295-319.

C. Cartis, N. I. M. Gould and Ph. L. Toint, "On the evaluation complexity of constrained nonlin-ear least-squares and general constrained nonlinear optimnization using second-order methods".
SIAM Journal on Numerical Analysis 52(3) (2015) 836 851.2

Supporting the Software Development Process through the Software Engineering Support Centre



Credit: © | Dreamstime.com

The Software Engineering Support Centre (http://www.softeng-support.ac.uk/) is EPSRC funded and aims to support the wider academic community through provision of tools and specialised advice. The main community focus is the Collaborative Computational Projects which are also funded by EPSRC.

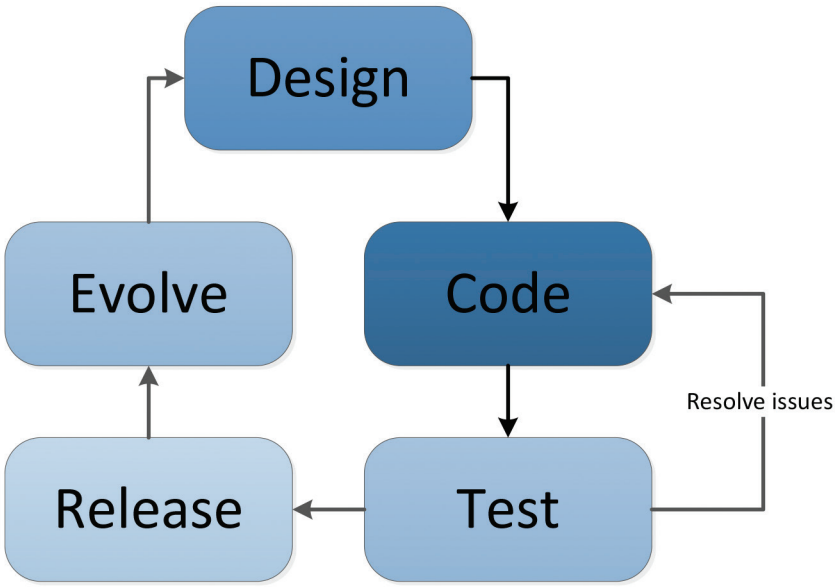
The software development process follows through a number of stages, depending on the complexity and lifespan of the code being developed as demonstrated in the diagram below. Ensuring that the code produced meets the aims of the development, is well managed, robust and reliable in operation is an important facet of development and can be supported by a variety of tools. The main tool provided by SESC to support these aims is CCPForge (https://ccpforge.cse.rl.ac.uk/gf/). CCPForge supports the good management of the software development process through access to code repositories, which ensure the code is in one place and developments by a wider team can be managed effectively, provides a mechanism to track issues under development through the tracker and can provide web pages and other features to build a community around the development. The added benefits brought by using the tools provided by SESC is the expertise provided by the team who provide the service, both through their knowledge of CCPForge and through their software engineering experience.

A User Survey undertaken in summer 2015 showed that the most popular programming languages in use by CCPForge users were Fortran, Python, C and C++. There was also feedback on the growing requirement to use Python to interface to programs originally developed in Fortran, C or C++. This feedback has helped SESC to assess priorities for support and quality assurance tools.

Developments in the final quarter of 2015 include Jenkins support for automated builds and the integration of quality assurance tools into Jenkins to enable CCPForge users to automate testing to improve the quality of the software provided. If you are a CCPForge user and interested in becoming an early adopter, please contact the CCPForge team.

In an associated project, Software Reuse, Repurposing and Reproducibility, which is funded by JISC, has been examining the issues involved in persistent identification of software to enable citation of software and how software can remain reproducible over the longer term. It is hoped that some of this work will be incorporated into services provided by SESC in 2016. For further information about this project see http://rrr.cs.st-andrews.ac.uk/RRR/ or listen to the podcast https://www.jisc.ac.uk/podcasts/research-data-springsoftware-reuse-repurposing-and-reproducibility-22-sep-2015

Finally, in June 2015 Professor Chris Greenough, who had led SESC, and its predecessor SESP, since its inception retired. We acknowledge all his hard work in setting up and running the service. SESC and the wider Software Engineering Group is now led by Catherine Jones.



Author
C. Jones,
STFC Rutherford Appleton Laboratory

Targeting Future Application Performance on Intel Xeon Phi

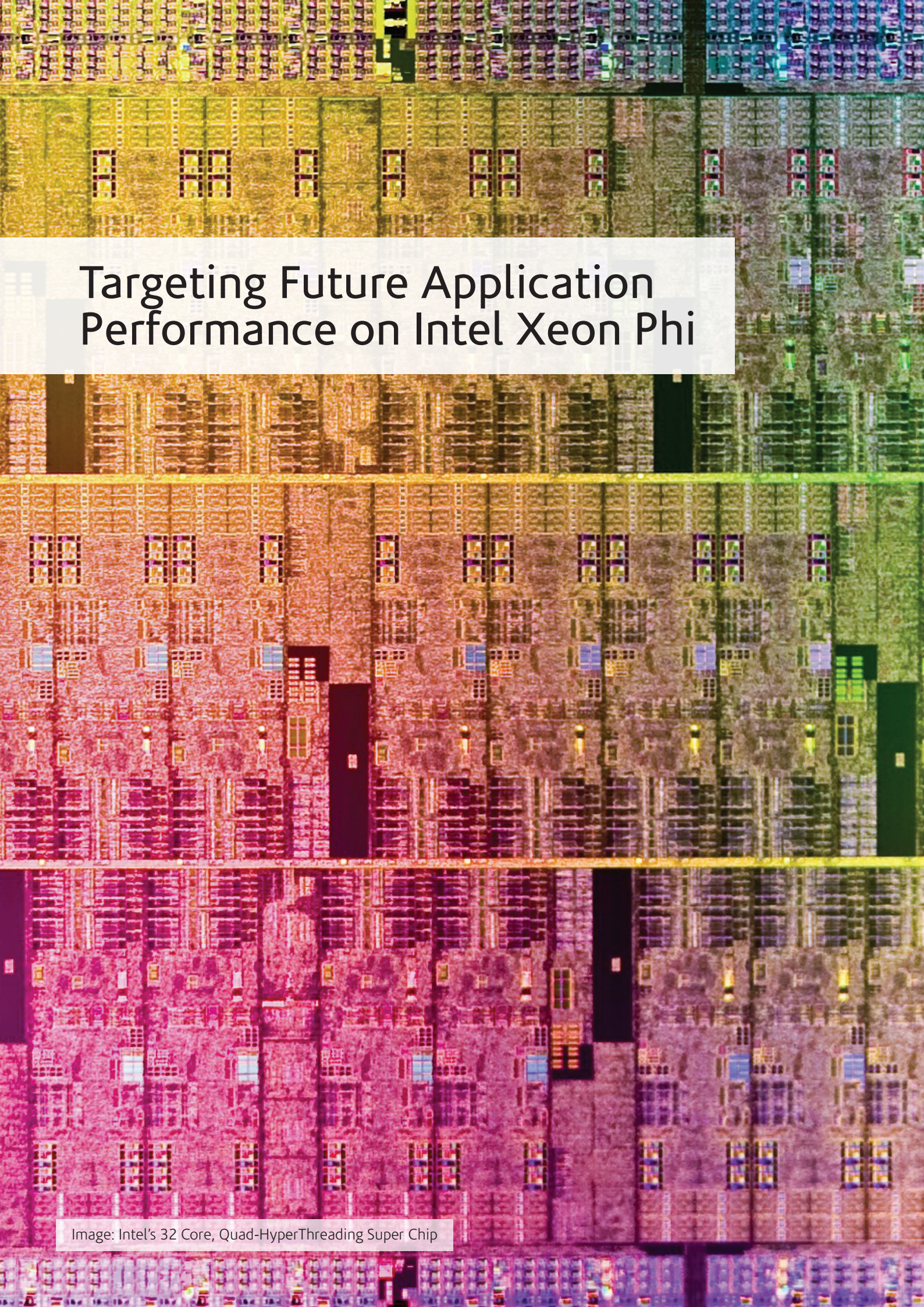


Image: Intel's 32 Core, Quad-HyperThreading Super Chip

The Intel Parallel Computing Centre (IPCC) at the Hartree Centre is concerned with porting codes to the Xeon Phi architecture. The Xeon Phi is a highly parallel architecture requiring efficient use of methodologies such as MPI, OpenMP and SIMD in order to achieve application performance. The porting project covers a wide range of different codes with differing application areas, some highlights of which are shown below.

With power limits restricting any further increase in the speed of single compute cores, scientific codes need to become increasingly parallel in order to properly exploit emerging architectures. Intel's Xeon Phi is such an architecture, providing 60 cores each equipped with 4 hardware threads for a total of 240-way parallelism. Additionally the Xeon Phi uses 512 bit wide vector processing units making proper exploitation of data level parallelism key to code performance. The current generation Xeon Phi (Knights Corner) is supplied as a co-processor in a similar manner to a GPU. The soon to be released next generation chips (Knights Landing) will be supplied as a host processor enabling them to use all available system memory and avoid PCI bus communication bottlenecks.

OSPRay is an **o**pen source, scalable, and **p**ortable **ray** tracing engine that enables visualisation workflows on standard CPUs and now on the Xeon Phi. As part of the Hartree Centre's Intel Parallel Computing Centre (IPCC) we use the OSPRay software stack to improve visualization workflows used by STFC projects. The first workflows improved are the Visualization applications which run in the Hartree Centre's older Sandy Bridge system. This machine only contains 2 graphical nodes, and when the visualization applications are scaled to multiple processes IO quickly becomes the performance bottleneck. The newer Ivy Bridge iDataplex system has considerably better IO but it does not contain graphical nodes with GPUs. For this reason visualization workflows were not carried out on the new system. OSPRay enables the Visualization workflow to be executed completely with CPUs and/or Xeon Phi enabling use of the newer iDataplex. The Heating, Ventilating and Air Conditioning (HVAC) model shown in Figure 1 demonstrates a 3x speed up using the newer hardware and OSPRay compared to utilising the older GPU equipped system. A new use case aims to use the STFC SCARF system to provide in-situ visualization of the data generated in the IMAT laser facility in less than 20 minutes in order to guide which will be the next experiment to run. Initial tests show promising results.

The Met Office's **Unified Model** (UM) is used for weather and climate modelling. The current porting and optimisation project focuses on the calculation of radiative fluxes on the Xeon Phi. The radiation calculation is already well written for application on vector machines like the Xeon Phi and benefits from OpenMP threading. The porting has involved correctly

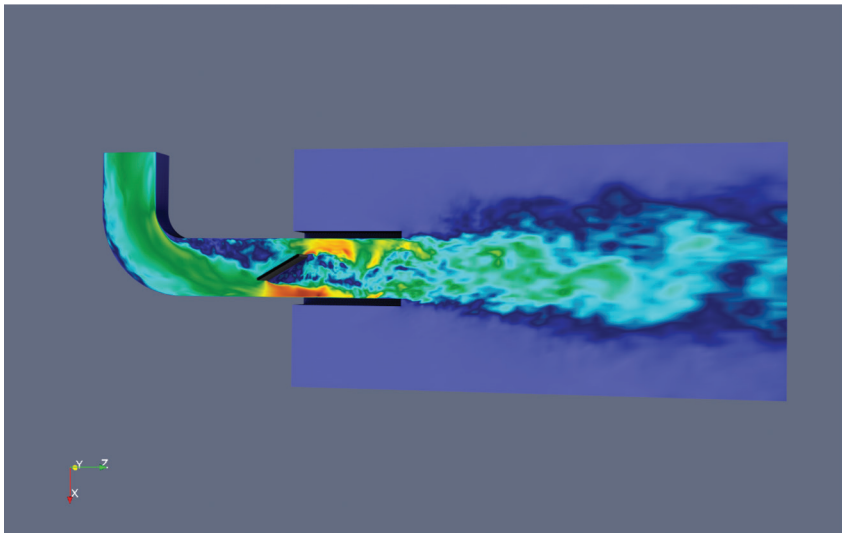


Figure 1: HVAC rendered with OSPRay

tuning the calculation tile size for the new architecture, removing scalar remainder loops, applying array alignment to improve data loads, and relaxing the strict floating point model to allow use of the Xeon Phi's vector instruction set. With these optimisations applied the time to calculate the radiative fluxes is faster on a single Xeon Phi than the same calculation on two Ivy Bridge Xeon E5-2697 v2 CPUs.

DualSPHysics [1] is a Smoothed Particle Hydrodynamics numerical model developed to study free-surface flows where Eulerian methods can be difficult to apply, such as waves. In recent years development has focused on the CUDA version of DualSPHysics and this is currently the recommended version of the software. In the IPCC we are working on porting DualSPHysics version 4 to the Xeon Phi architecture. We have introduced several optimizations to achieve a better vectorization and extended the OpenMP coverage to most of the runtime. Although these optimisations are targeted at the Xeon Phi they have also improved the performance of DualSPHyics on more standard CPUs as can be seen in Figure 2. There is still plenty more work to be done but these early results show promise, and a number of further improvements have been proposed in order to close the performance gap with the GPU version.

ParaFEM [2] is a parallel Finite Element solver which has been applied to problems ranging from analysis of dinosaur tracks to advanced materials for fusion reactors (ITER project). Optimisation efforts have been directed at native execution on the Xeon Phi with a particular focus on the preconditioned conjugate gradient solver. This portion of the code performs many vector-matrix operations on matrices with a typical size of 60x60. Through application of Intel Math Kernel Library (MKL), data alignment and streaming stores this hotspot shows improved performance on the Xeon Phi compared to two Xeon CPUs.

As well as these code porting projects the Hartree Centre's IPCC also makes its Xeon Phi hardware available to UK academics through a series of **Access Programmes** and provides training on code optimisation for the architecture. In partnership with Intel the team has also contributed a chapter to the second volume of the "High Performance Parallelism Pearls" book [3].

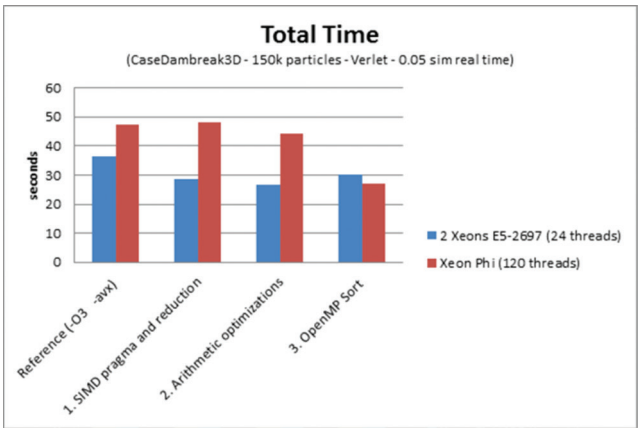


Figure 2 : Comparative performance of DualSPHysics

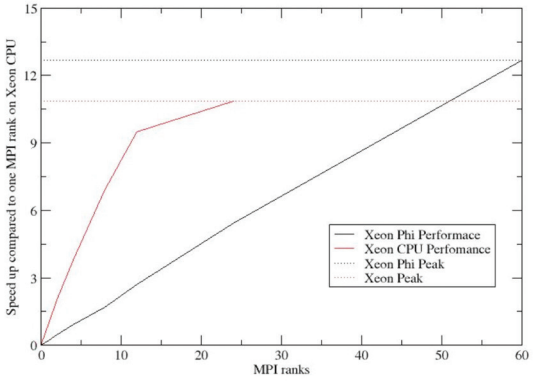


Figure 3: ParaFEM comparative performance

Author
L. Mason, S. Siso, STFC Daresbury Laboratory, L. Margetts, University of Manchester

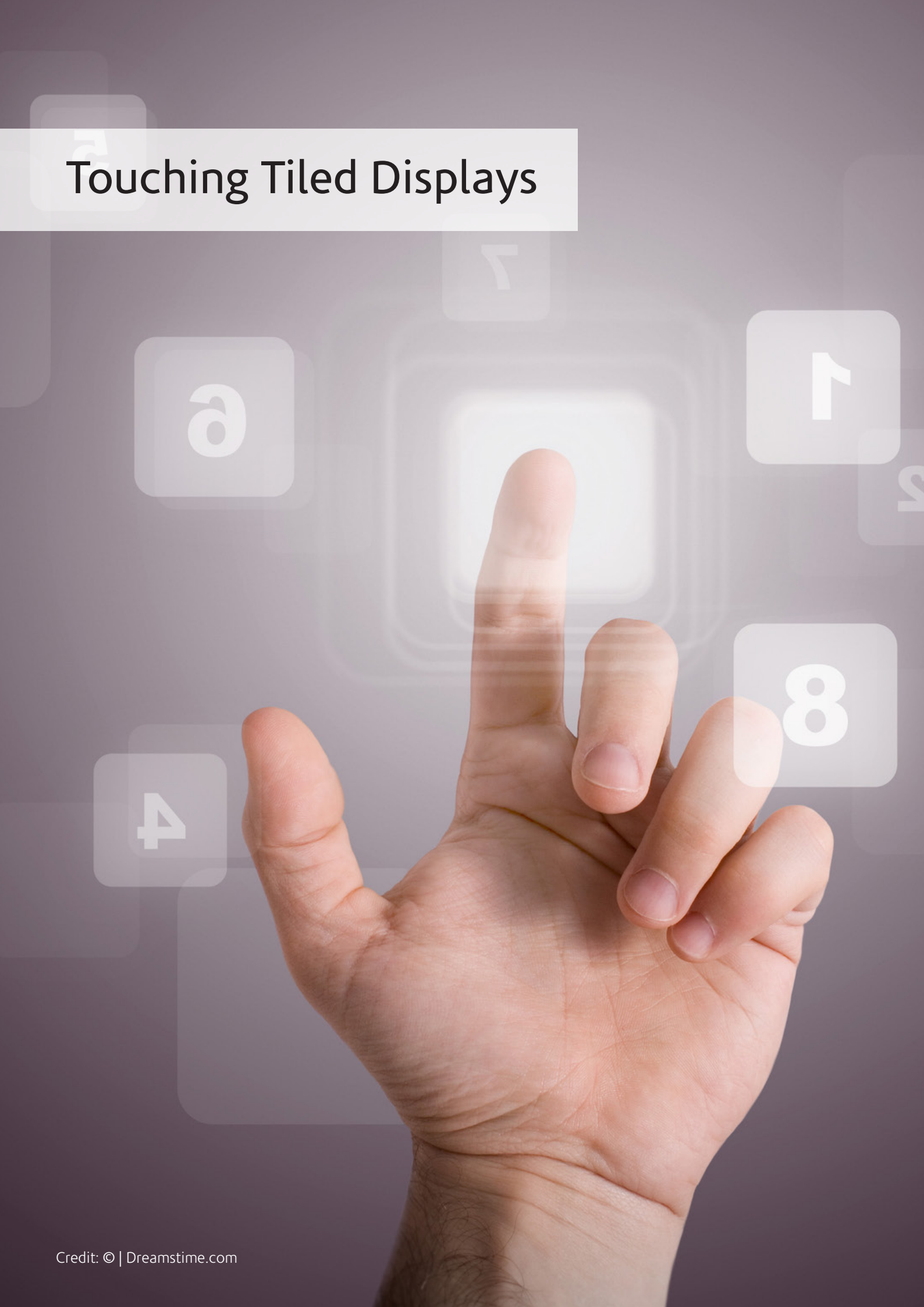
References:

[1] Crespo A.J.C, Domínguez J.M., Rogers B.D., Gómez-Gesteira M., Longshaw S.M., Canelas R., Vacondio R., Barreiro A. and García-Feal O., DualSPHysics: open-source CFD solver based on Smoothed Particle Hydrodynamics (SPH), Computer Physics Communications, 2014.

[2] Smith IM, Griffiths DV and Margetts L, "Programming the Finite Element Method", 5th Edition, Wiley, 2014

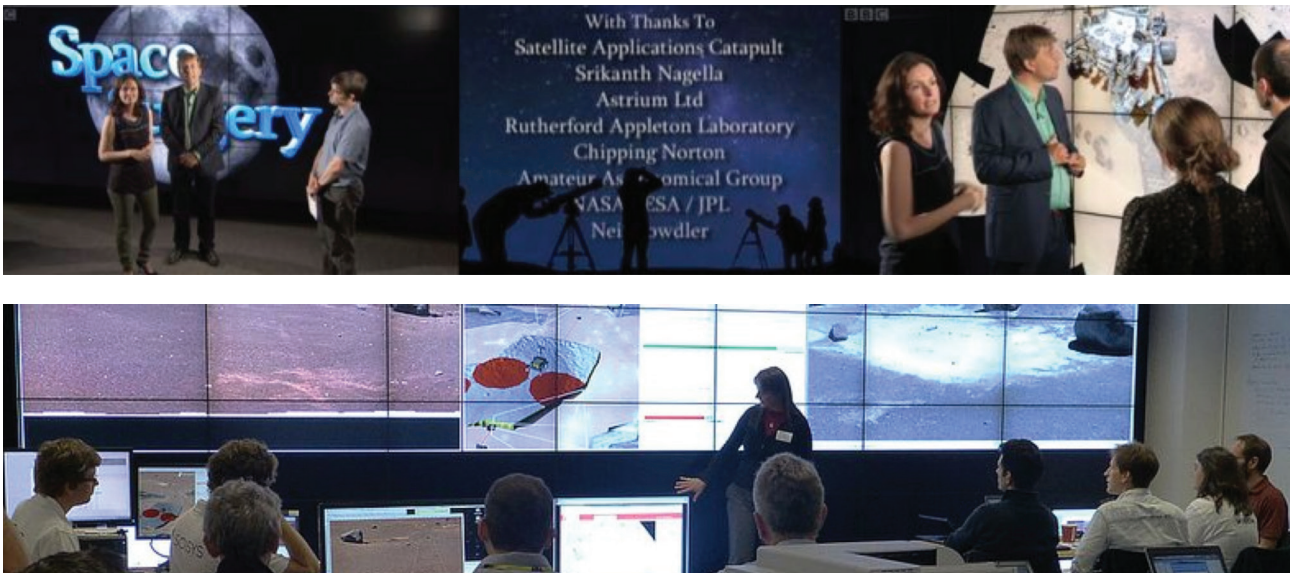
[3] James Reinders and Jim Jeffers, "High Performance Parallelism Pearls: Multicore and Many-core Programming Approaches". Boston, MA, USA: Morgan Kaufmann Publishers Inc, 2015.

Touching Tiled Displays



Credit: © | Dreamstime.com

The visualisation group within SCD is working to support the high-end visualisation centres within STFC, with the key objective to consider the human-in-the-loop as an integral part to pre- mid- and post-data visualisation needs from the major facilities, archived data stores and computational simulations. This it is believed is a key component to increasing the efficiency of the major STFC facilities allowing researchers' work flows to be controlled, changed and even stopped on the fly. In this highlight article we consider the interactive post-visualisation stage that has resulted in us being commissioned to build small and medium touchable visualisation systems.



Demonstrations of use from 2013 of the original Tiles Display system, at the TSB Space Applications Catapult; BBC Sky at Night sessions and the SAFER test connection to Chile. These allowed mixed forms of display but can have limited interaction as not atouchable.

Scientists have been collecting and generating large amounts of data that includes multi-variable and multi-dimensional date, and as part of their data analysis they want to visualise and make sense of this data. One way of visualising these large multi-dimension datasets is to use big displays such as display walls. Display walls (also called Tiled Displays) are a group of display monitors/TVs often arranged in a grid format to provide single images. There are several ways of visualising the data on a display wall and one of the common modes of use is to compare two or more datasets where the display area is virtually split into multiple regions. This gives an instant visual impact and enables detecting patterns and difference in the data.

The important aspect of this visualisation process is not just about being able to look at the static image but to interact with the data to get more information. The traditional way of interacting is with a keyboard and mouse but this can be a problem when you have such a large display area. This is one of the main reasons for moving to touch tiled displays.

The visualisation group has several years of experience in designing and building visualisation system including large tiled displays. We have designed and installed two display walls in the Satellite Applications Catapult (SAC formerly International Space Innovation Centre). One of the first systems we have installed is a 28 screen display wall with each screen 46" in size giving a total resolution of 9562 x 3072 pixels (~29.4mpixels). The second main system installed is an 18 screen display wall with each screen being 55" in size giving a larger total resolution of 11520 x 3240 pixels (~37.3mpixels); this later system is now being used at SAC as a Command and Control Centre.

Last year we were involved in the design and installation of a true touch tiled display system for the Environmental Systems Science Centre (ESSC), at the University of Reading. The requirement for them was a multi-purpose, large display system with the users being able to interact with the data up close. The multi-purpose aspect of the system requirements enabled it to be a meeting room, display PowerPoint presentations, video conference and be a scientific data

visualisation centre. This was a challenge as all of the products supplied by traditional display wall suppliers only provide a hardware video controller that has video signal inputs and video outputs but do not run any traditional operating system. The other requirement of interactivity is also a challenge as traditional display walls are not touch enabled.

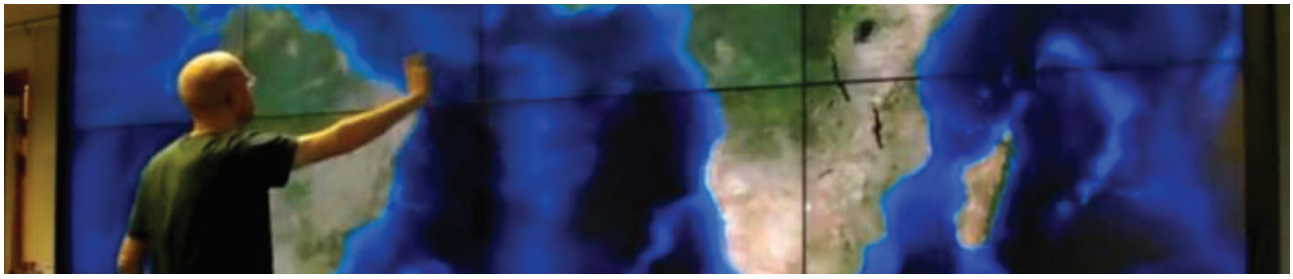
We have designed the display wall so that it fills the room meaning it could accommodate eight 55" panels that were arranged in a 4x2 format. For driving this display we selected a Datapath controller that runs the Windows operating system with video capture cards. This enables running simple windows applications and video conferencing applications such as IOCOM Visimeet. One disadvantage of this is that the graphics cards in this system cannot run graphics intensive scientific visualisation applications. To overcome this we installed a graphics workstation with 2 x NVIDIA K5000 cards that is capable of running graphics intensive scientific applications.

To enable the touch interactivity we selected the U-Touch one glass solution that is overlaid on top of the tiled display. This provides seamless touch interactivity across the tiled display. It was a challenge to get such a big single glass panel into the room and was a similar challenge to the installation of a glass screen in DL visualisation room. The whole system was placed on a movable platform so that the panels can be pulled forward to get to the back for servicing thus allowing the panels to be placed close to the wall. To allow reaching the top of the screens a raised platform was built.

With the newly funded CCPi (ccpi.ac.uk) we are coordinating open source and self-build hardware solutions for products from large tiled displays to smaller kiosk designs. There are future opportunities in using this expertise in creating visualisation systems for other Labs and Universities. The visualisation group is now actively involved in creating a subgroup that specialises in touch technologies across the UK. With the involvement of a range of CCPs we have installed and advised on touch ready software in a number of sites. As well as at the University of Reading, this includes; DL (Hartree Centre main room at Daresbury Labs), Manchester Museum and ICAM (International Centre for Advance Materials), RCaH (Research Complex at Harwell), the DLS (Diamond Light Source) and the NHM (Natural History Museum).

This project is a potential component part of the 'Increase SCD Impact' goal, which has brought a range of staff together under the banner to "Enable remote analysis and visualisation to give scientists/engineers rapid insight from large data; and make more efficiently use of STFC imaging facilities and HPC cycles"

Command-and-control centre examples on the newer Tiled Display wall setup at the SAC in 2015; showing live mapping and shipping modelling examples with links to remote interface wands but not directly touchable.



Dr Jon Blower and colleagues, in the University of Reading, who commissioned the software setup and assistance required to build their new tiled touch-screen videowall, utilising a single glass front surface, which was all installed and completed on time in 2014. https://www.youtube.com/watch?v=vOFOxMTKSTY

CCP and other visualisation software installed on the DL touchtable that has been used for EPSRC review meetings and collocated to the Hartree centre screens.

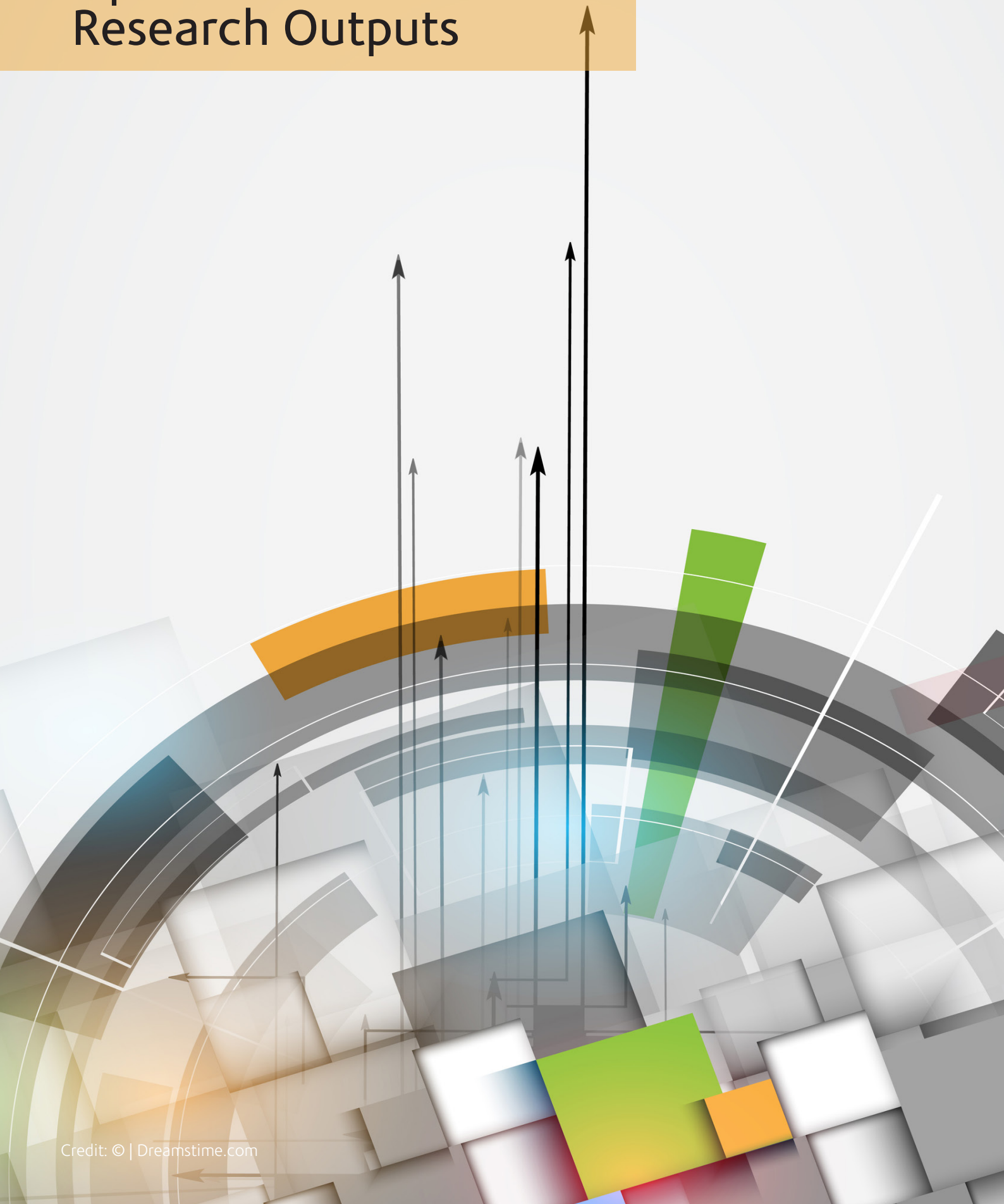
Cheltenham Science Festival 2015: Over 700 visitors have explored the smaller touchscreen system but some people needed extra assistance in operating the early device and took a little longer than expected to become confident. Also the first version of the custom built kiosk creating a self-contained automatic unit, was launched at the 2015 Tomography for Scientific Advancement symposium (www.nhm.ac.uk/tosca).

Authors
M. Turner, R. Allan, D. Bailey, STFC Daresbury Laboratory,
R. Fowler, S. Nagella, E. Ovtchinnikov, STFC Rutherford Appleton Laboratory

References:
Original build: "Building a Video Wall for Earth Observation Data" Bennett, Victoria; Haffegee, Adrian; Matthews, Brian; Nagella, Srikanth; Shaw, Andrew; Styles, Jon

URI: http://dx.doi.org/10.2312/LocalChapterEvents/TPCG/TPCG11/041-048 2011 ISBN 978-3-905673-83-8

Open Access to STFC's Research Outputs



Credit: © | Dreamstime.com

STFC's Chadwick & RAL Libraries connect people with information to create new knowledge. The Libraries have a lead role in giving external access to STFC research outputs. The Libraries' position in Scientific Computing Department is ideal for maximising the synergy between the information management skills of Library staff and the information and data management technologies developed and deployed throughout the Department. One of our success criteria is to ensure that "STFC generated research outputs are discoverable and accessible in line with funders' and STFC's open access policies".

Open Access

Free and open access to the outputs of publicly funded research offers significant social and economic benefits as well as aiding the development of new research . There are various models for achieving open access, some of which allow more immediate access to the publication and less restrictions on re-use of its content. The Chadwick & RAL Libraries support STFC authors through the processes of Open Access publishing, with the aim to make as many of their publications as possible Open Access.

'Gold' Open Access

'Gold' Open Access – immediate, free, and unrestricted open access provided by the publisher – usually involves payment of an article processing charge (APC) by the author to the publisher who will then make the paper freely and immediately available on their website under a Creative Commons CC-BY license . STFC staff who are corresponding authors (submitting a paper for publication) can apply for funds to pay for APCs through the Chadwick & RAL Libraries. We provide guidance on sources of funding, different versions of licenses, and liaise with publishers to make the process as streamlined as possible.

'Green' Open Access

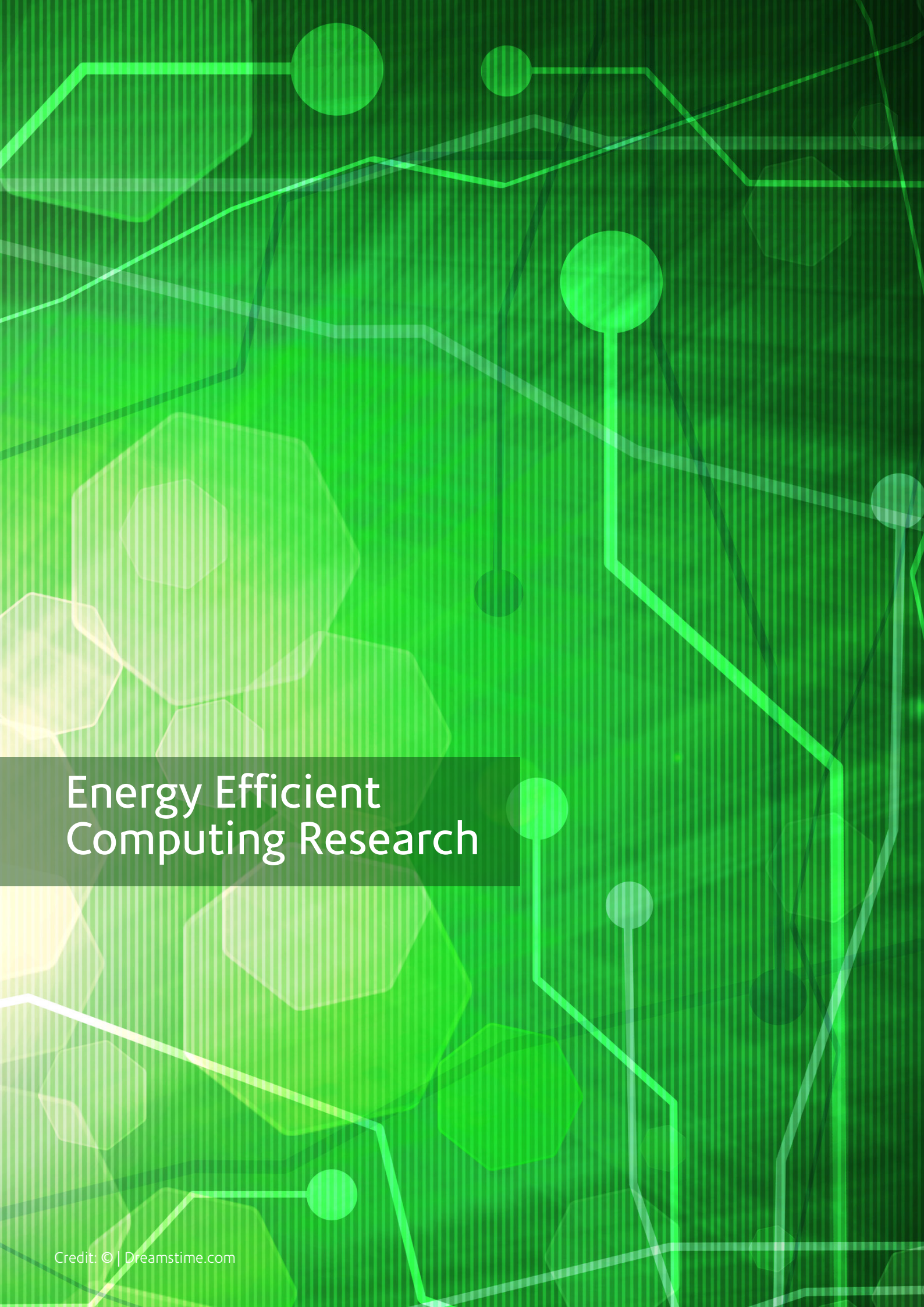
'Green' Open Access – immediate or delayed free open access provided through self-archiving – is also supported by the Chadwick & RAL Libraries. STFC Staff can deposit their Author Accepted Manuscript (AAM) - the accepted, peer-reviewed version of their paper - with the Libraries on acceptance of the publication by a journal. The Library makes the AAM available in ePubs , STFC's repository, after any embargo period defined by the publisher. Full-text content in ePubs is discoverable through a public search interface and via indexing and harvesting by search engines and other repository sites.

Future Plans

Building on our success in supporting open access publishing of research papers, Library staff will be working with colleagues in Scientific Computing Department and major stakeholders to develop support services for Open Access Data, and linking research publications with the datasets produced by, or used in, the reported research.

Authors
D.Franks, STFC Daresbury Laboratory

References:
RCUK Policy on Open Access and Supporting Guidance http://www.rcuk.ac.uk/research/openaccess/policy/
Creative Commons 'About the licenses' http://creativecommons.org/licenses/
ePubs https://epubs.stfc.ac.uk/index



Energy Efficient Computing Research

The Energy Efficient Computing (EEC) Research Programme at the Hartree Centre aims to investigate innovative new approaches to the development of both hardware and software. Building upon the expertise already present within the Scientific Computing Department and through collaboration with research and industry partners its aim is to act as a centre of excellence, enabling scientific advances and delivering real economic benefit for the UK.

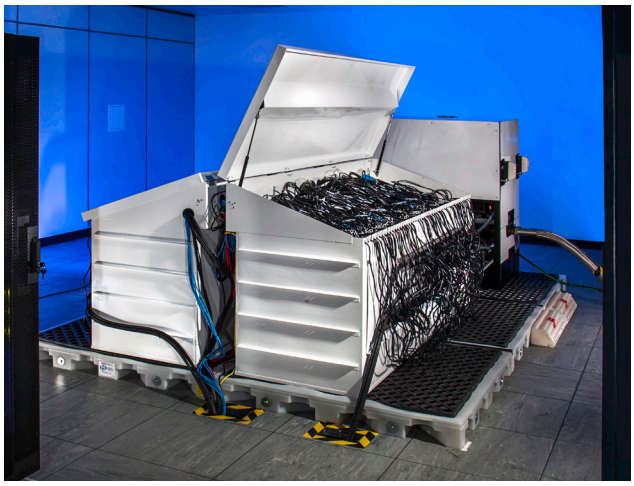
The Drivers for Energy Efficiency within HPC

As supercomputers become increasingly heterogeneous, relying on accelerators to drive performance and data throughput, whether through Graphical Processing Units (GPUs), Intel's Xeon Phi or the integration of Field Programmable Gate Arrays (FPGAs) there is an increasing need to profile the capabilities of these evolving architectures, not just in terms of overall performance but also in terms of their energy consumption with regard to a range of computing tasks and workloads. No longer can we be solely concerned with the number of Floating Point Operations per Second (Flops) a supercomputer can do – the world is changing and energy consumption, data throughput, and application resilience at scale are vitally important if we are to realise the opportunities afforded to us by Exascale levels of computing capability.

Towards the end of 2013 the Hartree Centre was granted £19M of capital investment from the UK Government, specifically for research and outputs in the area of Energy Efficient Computing research architectures and software. Following investment of this capital grant it is arguable that the Hartree centre now has one of the most architecturally diverse hardware estates, at scale, anywhere in the world.

The Energy Efficient Computing programme at the Hartree Centre is addressing one of the fundamental challenges facing High Performance Computing – how do we do more computation with less energy? While this is a significant problem for HPC it also applies with equal measure to the wider computing sector. Energy Efficient Computing may until recently have played only a minor role in the industries history but the issues it deals with are vital to the evolving architectural landscape: The increasing number of smart enabled devices and the development of the Internet of Things (IoT) is leading to incredible growth in the production of data, which in turn is driving a need for large scale storage and computer services moving into the cloud.

As of 2014 there were 1.75 billion smart phones in use globally and it is anticipated that by the end of



2015 we will reach two billion personal computers – consuming approximately 280 gigawatt hours of electricity per year. It's worth pausing to consider that these numbers don't take into consideration embedded digital devices, wearable technologies or the IoT which are rapidly emerging sectors. The growth in hardware is also mirrored by data growth which current estimates indicate will reach 40 billion terabytes by 2020.

The manufacture of digital devices in all there forms whether embedded low power systems or high performance supercomputers, along with the creation, transmission, storage and use of data have all had a significant effect on global energy consumption – in UK datacentres alone it is currently estimated to cost £5.6 billion annually

This energy cost has global implications as we attempt to reduce our reliance on fossil fuels and limit the effects of climate change - a 2008 report by McKinsey suggested that data centres produced about 80 Megatons (0.2%) of all CO2 emissions in 2007, but with the rapid growth of this technology could rise to as much as 340 Megatons of CO2 by 2020, over a four fold increase.

The management of data in all stages of the information lifecycle - from creation through to destruction, and the need to analyse, integrate and share data often in real time - has driven a surge in demand from consumers and businesses alike for data storage and analytics. The impact in terms of global energy consumption has been dramatic. Jonathan

Koomey's report of 2011 to US Congress estimates that data centre energy usage by the end of 2015 will reach 600TWHr at a cost of £45Bn per annum.

How is the Energy Efficient Computing Programme Addressing these Issues?

The Energy Efficient Computing research programme is investigating a number of areas to identify opportunities for improving the energy efficiency of HPC hardware and applications. Our approach is a co-design one that views close coupled integration between hardware and software as essential to achieving best performance at minimum energy. Allied to this we are identifying opportunities to seed best practice within the wider commercial world through engagement with our industrial partners, ensuring tangible benefits for our clients. Areas of focus within the programme include:

- Energy Efficient Code Optimisation
- Code Modernisation – Portability and Scalability
- Energy Aware Scheduling
- Novel Cooling approaches
- Data Centre Monitoring and Management
- Hardware and software profiling
- Portable Benchmarking
- Platform Assessment, Comparison and Evaluation

Developing the Hartree Centres Energy Efficient Computing Research Capabilities

The focus for the EEC Programme at the Hartree Centre is the specially constructed 'Green Machine Room'. This computer room has been equipped with twin water cooling loops as well as conventional air conditioning and free cooling capabilities so that we can better understand the impact of cooling on the overall efficiency of our systems. In order to develop baseline metrics we are in the process of fitting all the machine rooms at the Daresbury laboratory with advanced monitoring equipment. This Data Centre Infrastructure Management system will provide granular energy, temperature and humidity data across all three machine rooms, with the ability to monitor and analyse data at the machine, system and node level. This combined with data from our building management system and Platform Utilization manager will enable us to develop energy and performance profiles for all our hardware configurations running a range scientific application workloads.

Authors
N. Morgan, STFC Daresbury Laboratory

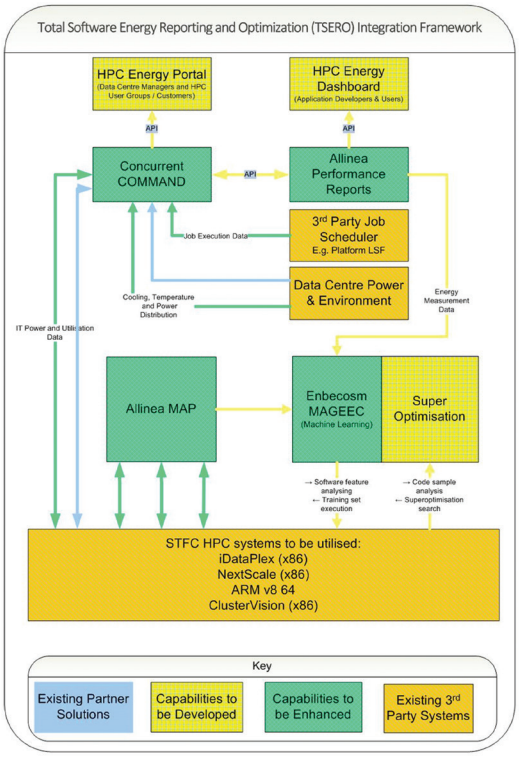
References:
ftp://public.dhe.ibm.com/software/uk/itsolutions/optimiseit/greenhub/carbon-managment/pov-mckinsey-report.pdf http://www.mediafire.com/file/zzqna34282frr2f/koomeydatacenterelectuse2011finalversion.pdf

Collaboration and Projects

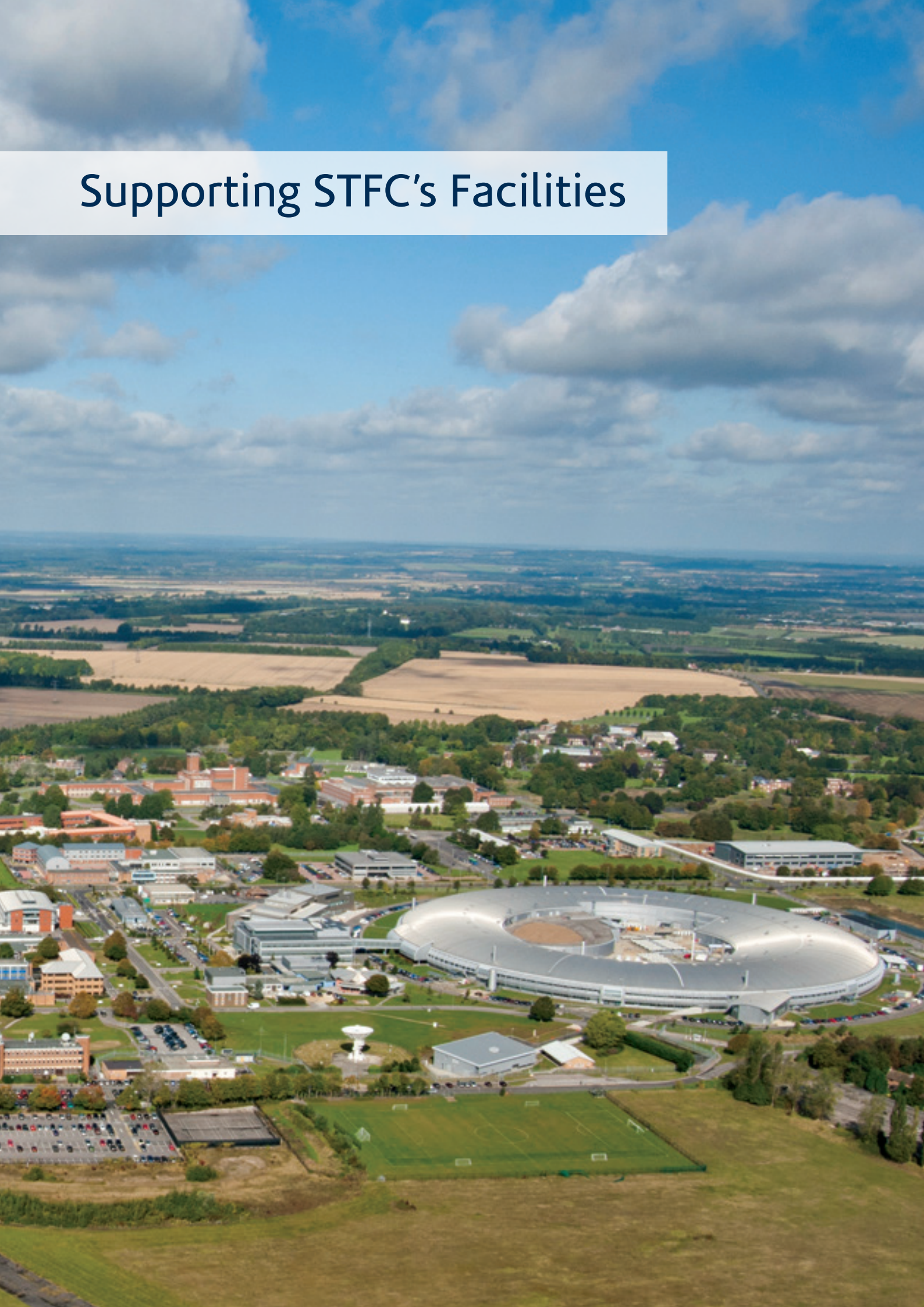
Working in collaboration with other research institutions and industry partners is essential to the success of the EEC programme and has led to the creation of a number of research projects with a diverse array of partners including: IBM, Lenovo, ARM, Pathscale, Allinea Software, Concurrent Thinking, and Embecosm to name just a few. Key projects include:

Total Software Energy Reporting and Optimisation (TSERO)

In June of this year we commenced work on the 2 year TSERO project, funded through Innovate UK as part of the 'Scaling up Energy Efficient Computing Call' this collaboration between the Hartree Centre, Allinea Software (Application Profiling and Optimisation Tools), Concurrent Thinking (Data Centre Infrastructure Management) and Embecosm (Compiler Development) is looking to develop an integrated infrastructure suite of monitoring and profiling tools that combine machine learning and compiler optimisation through a technique known as 'Superoptimisation' to drive maximum performance from applications and the hardware they run on.



Supporting STFC's Facilities



Large-Scale Analytic Facilities are today a vital resource for advanced experimental science. Such facilities provide intense beams of particles which can be used in conjunction with highly specialised instruments to probe into the heart of matter, acting as highly accurate microscopes looking down as far the molecular building blocks of materials. STFC support and runs several of these facilities on behalf of the UK's science community, allowing new breakthroughs across physics, chemistry and biology.

In particular, STFC funds the ISIS Neutron and Muon Source, the Diamond Synchrotron Light Source and the Central Laser Facility, all on the Rutherford Appleton Laboratory campus. Together these provide a unique collection of powerful instruments for studying how the world around us is formed.

All science in the 21st century uses computing as a vital tool and STFC's facilities are no exception. Consequently the Scientific Computing Department works closely with these facilities to provide data and compute infrastructure and expertise.

Facilities instruments are electronic devices which generate data arising from the impact of the particle beams on the samples being studied. If facilities are to get the best science from the use of their instruments, this data needs to be kept safely for scientists to inspect and analyse to derive scientific insight. Scientific Computing Department provides data storage facilities within its data centres to store and archive this data. By using the tape and disc capability of SCD, facilities can store their data safely and efficiently, and allow scientists to retrieve their data for subsequent use. The amount of data which facilities are generating is growing all the time; SCD stores over 3.5 PetaBytes for the Diamond Light Source alone.

Scientists also need to be able to run programs which process this data to see how it captures the physical properties of material. Further, they also wish to run models and simulations - programs which use mathematical theories to mimic real physical processes within the computer memory; by comparing computer models with actual experimental data, they can better interpret the results to work out what is going in the material. But all these complex programs are difficult to run on normal computing equipment. SCD provides facilities with access to large computing resources. In particular, the SCARF cluster provides some 4500 computing cores for facilities scientists; this can run many programs in parallel, or split programs across many cores running at once to make it run faster. Thus scientists can carry out these computations much more quickly.

SCD provides a range of other support for facilities scientists, from tools for capturing the experiment and visualising the results, through code development in the simulation and modelling software to providing support for the distribution and access of scientific papers.

Looking to the future, the computing challenges which the facilities are facing are becoming ever greater. The instruments are becoming ever more powerful, so the data they produce continues to increases; SCD has to keep up with the data collection rate and volume. Then processing this data becomes a larger problem too, so there is more need for super-computing resources. Also as the questions scientists ask become more sophisticated, the computational models and software also has to become more powerful. So as the science of the facilities becomes ever more data and computational intensive, access to the high-performance resources needs to become routine, part of how the facilities operate. SCD can provide the capability for facilities to meet these challenges and support the science.

Author
B. Matthews, STFC Rutherford Appleton Laboratory

Contacts and Further Information

Mr David Corney	Director of the Scientific Computing Department	david.corney@stfc.ac.uk
Mrs Shirley Miller	Personal Assistant	shirley.miller@stfc.ac.uk
Mrs Jean Pearce	Personal Assistant	jean.pearce@stfc.ac.uk
Mr Damian Jones	SLA Project Support Officer	damian.jones@stfc.ac.uk
Mrs Karen McIntyre	Administrative Assistant	karen.mcIntyre@stfc.ac.uk
Miss Helen Walker	Administrative Assistant	h.walker@stfc.ac.uk
Mrs Esme Williams	Administrative Assistant	esme.williams@stfc.ac.uk

Division Heads

Dr Juan Bicarregui	Data Division	juan.bicarregui@stfc.ac.uk
Dr Peter Oliver	Technology Division	peter.oliver@stfc.ac.uk
Dr Andrew Sansum	Systems Division	andrew.sansum@stfc.ac.uk
Prof Paul Sherwood	Applications Division	paul.sherwood@stfc.ac.uk

Hartree Centre

Dr Peter Allan	Director, Hartree Centre	peter.allan@stfc.ac.uk
Mrs Laura Johnston	Personal Assistant	laura.johnston@stfc.ac.uk

Group Leaders

Dr Mike Ashworth	Applications Performance Engineering	mike.ashworth@stfc.ac.uk
Mr David Cable	High Performance Systems Group	david.cable@stfc.ac.uk
Mr Ian Collier	Petascale Computing and Storage	ian.collier@stfc.ac.uk
Prof Dave Emerson	Engineering and Environment	david.emerson@stfc.ac.uk
Mrs Debbie Franks	Library Services	debbie.franks@stfc.ac.uk
Dr Tom Griffin	Research Data	tom.griffin@stfc.ac.uk
Mr Nick Hill	Research Infrastructure	nick.hill@stfc.ac.uk
Mrs Catherine Jones	Software Engineering Support Centre	catherine.jones@stfc.ac.uk
Dr Brian Matthews	Facilities Programme	brian.matthews@stfc.ac.uk
Dr Barbara Montanari	Theoretical and Computational Physics	barbara.montanari@stfc.ac.uk
Mr Neil Morgan	Energy Efficient Computing	neil.morgan@stfc.ac.uk
Ms Alison Packer	Data Services	alison.packer@stfc.ac.uk
Dr Jennifer Scott	Numerical Analysis	jennifer.scott@stfc.ac.uk
Dr Ilian Todorov	Computational Chemistry	ilian.todorov@stfc.ac.uk
Dr Martin Turner	Visualisation	martin.turner@stfc.ac.uk
Dr Martyn Winn	Biology and Life Sciences	martyn.winn@stfc.ac.uk

Websites

Scientific Computing Department	www.stfc.ac.uk/scd
Hartree Centre	www.stfc.ac.uk/hartree
Science and Technology Facilities Council	www.stfc.ac.uk

Science and Technology Facilities Council
Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxfordshire OX11 0QX, UK
T: +44 (0) 1235 445000 F: +44 (0) 1235 445808

Sci-Tech Daresbury, Daresbury, Warrington, Cheshire WA4 4AD, UK
T: +44 (0) 1925 603000 F: +44 (0) 1925 603100

www.stfc.ac.uk

Head office, Science and Technology Facilities Council, Polaris House, North Star Avenue, Swindon SN2 1SZ, UK
Establishments at: Rutherford Appleton Laboratory, Oxfordshire; Daresbury Laboratory, Cheshire:
UK Astronomy Centre, Edinburgh; Chilbolton Observatory, Hampshire: Boulby Underground Science Facility
Boulby Mine, Cleveland.

Science & Technology
Facilities Council