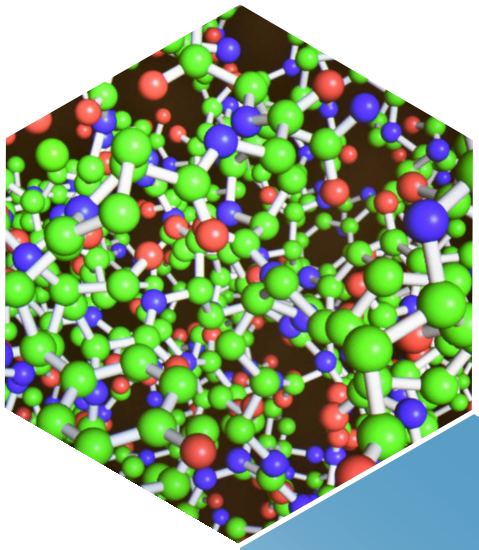
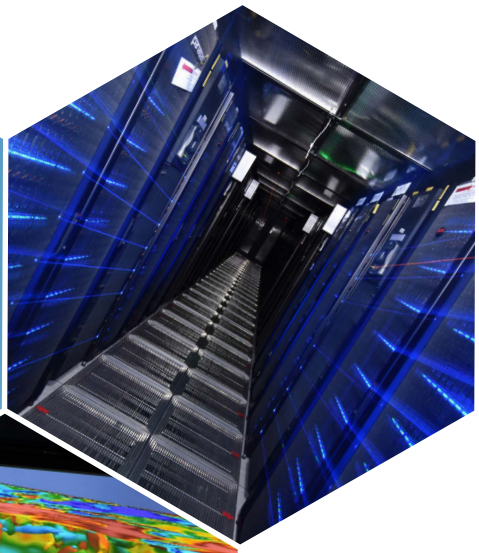
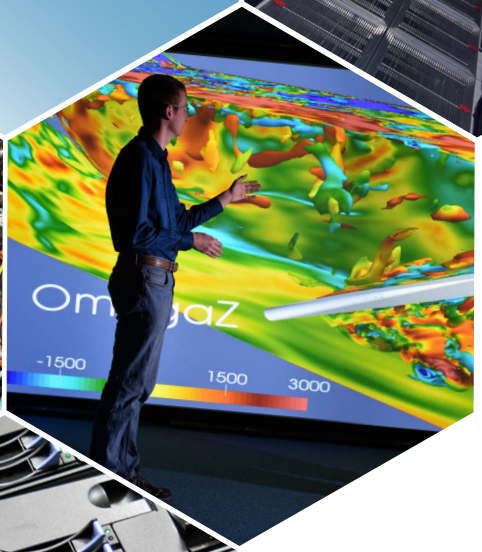
Science & Technology Facilities Council



SCIENCE HIGHLIGHTS 2015-16





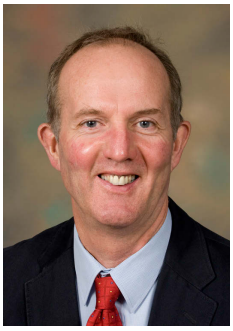
SCIENTIFIC COMPUTING DEPARTMENT

Contents

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

Note: All images are courtesy of STFC unless otherwise stated.

Welcome



Welcome to this year's Science highlights from the Science and Technology Facilities Council's Scientific Computing Department. A lot can happen in a year, and indeed a lot has happened; including a new Conservative government, the outcome of the Comprehensive Spending Review, the development towards UKRI, and the Brexit vote. The landscape has changed considerably, and as a department, we have to change with it to ensure we continue to maximise our output for STFC, UK and international science.

As a result, this year we have been evolving our Scientific Computing Department Strategy, incorporating all the goals from the original strategy developed in 2015, but building a stronger approach to specific areas as they have evolved and emerged as serious concept in line with supporting our mission and vision, and in line with strengthening the UK e-infrastructure.

One of these specific areas is the "STFC UK-T0" association, which involves a bottom up collaboration across the STFC facilities, the Culham Centre for Fusion Energy, together with most of the international programs that STFC supports in the UK, including:

- The LHC Experiments (ATLAS, CMS, LHCb and ALICE)
- Other current and future HEP Experiments (ILC, NA62, MICE, T2K, SNO+, HyperK, DUNE)
- The Square Kilometer Array (SKA)
- The Large Synoptic Survey Telescope (LSST)
- Dark Universe space mission (EUCLID)
- Low Frequency Radio Astronomy (LOFAR)
- Advanced-LIGO
- Cerenkov Telescope Array (CTA)
- Dark Matter (Lux-Zeplin)

The UKT0 initiative then, is an association of activities from within the STFC science programmes and facilities, with the aim of coordinating computing provision more effectively and efficiently. By sharing expertise and computing infrastructure, the domain specific expertise, vital for all experiments is minimised. The "STFC UK-T0" association directly addresses many of the recommendations of the STFC computing strategy review.

The second area emerging as a key development area within the SCD strategic plan, is the development of the "Ada Lovelace Centre." Experimental and observational science is undergoing a revolution driven by the new generation of facilities and instruments, and by dramatic advances in detector technology which generate ever greater volumes of data. In addition, the experiments now being performed at STFC and CCFE's large-scale facilities are becoming increasingly complex, often requiring advanced computational modelling to interpret the results. There is also an increasing requirement for the facilities to provide near real-time feedback on the progress of an experiment as the data is being collected. All of these trends are requiring a closer coupling between software, data and compute resources. The Ada Lovelace Centre will generate a step change in the quality and quantity of the science delivered through our Facilities by building capacity in advanced software development for the handling, analysis, visualisation, integration, modelling and interpretation of experimental data. It will also deliver a new generation of data scientists and software developers, who will be trained at the frontier of their disciplines and become pioneers in the emerging field of data intensive science.

STFC are committed to the ALC. In the updated STFC strategy (due out in a few weeks), a new Data Intensive Science theme is identified which includes the following:

- High level Objectives:

"Transform the efficiency of campus research through the Ada Lovelace Centre, a multidisciplinary approach to real-time data processing, computer simulation and data analytics for our national facilities"

- Deliverables:

"Establish the Ada Lovelace Centre, an integrated, cross-disciplinary data intensive science centre for capitalizing on research at the national facilities including Diamond, ISIS, CLF and CCFE"

So, please enjoy the many offerings from this year. I hope that this year's highlights provide you with a good understanding of some of the real strengths of the STFC's Scientific Computing Department, and some of the ways that we intend to use the departments' skills and expertise to develop and strengthen SCD into the future, for the benefit of STFC, UK, and international science.

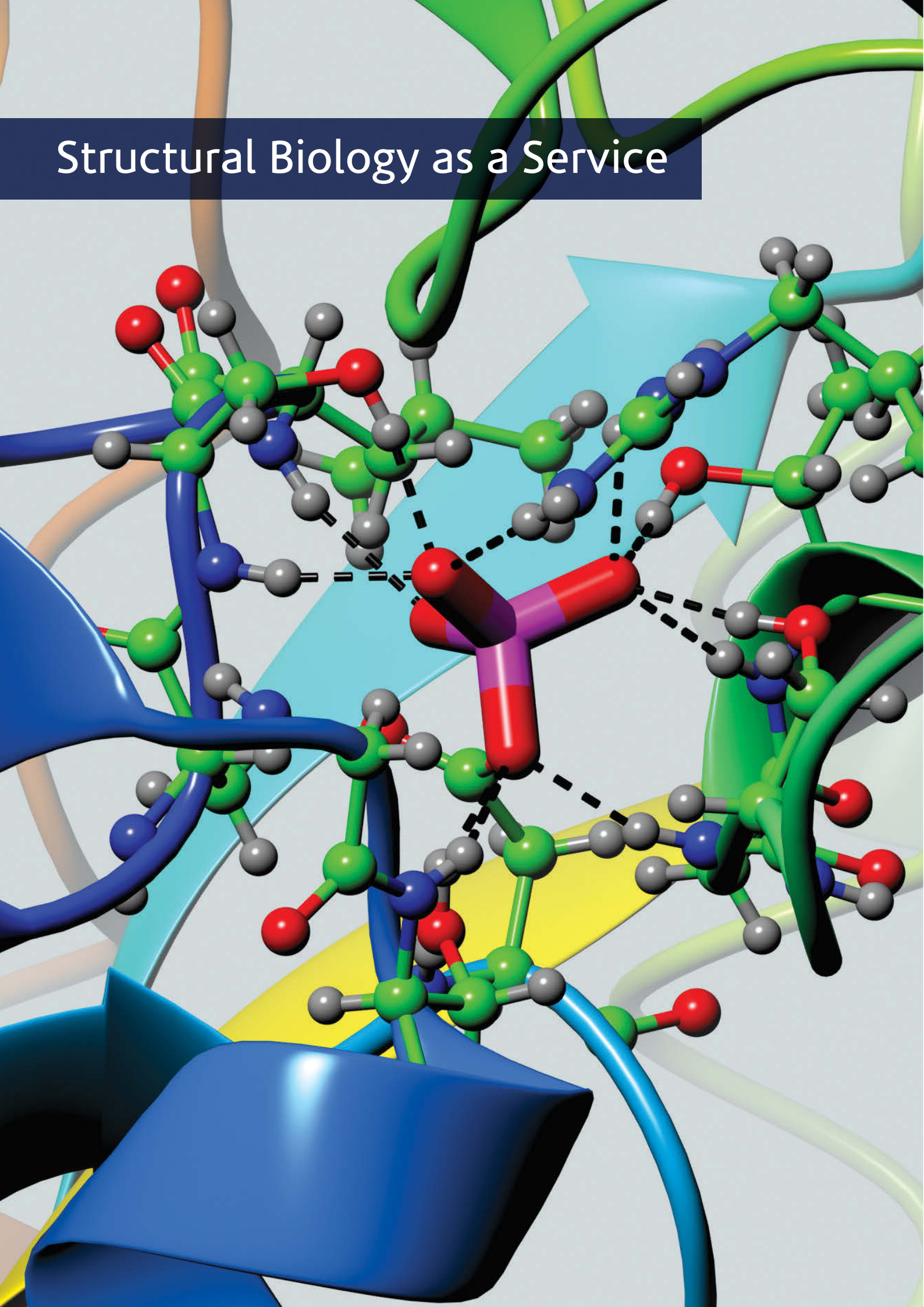
As usual, I would like to finish by thanking the staff for their contribution to making the department, a success. I said last year that we have become used to rapid change and evolving quickly to adapt to the developing needs of our scientific user communities. This year has proved again, that our staff are able to adapt rapidly to challenging and rapidly changing circumstances, to help ensure the growing success of the STFC's Scientific Computing Department. Thank you.

Thank you and regards,

David Corney

Director, SCD

Structural Biology as a Service



There are two themes characterising modern structural biology research. Firstly, there is diversification in the range of experimental techniques being brought to bear on a particular biological problem, including crystallography, electron microscopy, NMR, mass spectrometry, SAXS, optical microscopy, as well as a range of genetic and biochemical manipulations. Secondly, improvements and automation for each technique are leading to a rapid increase in the volume of data being generated. Thus, a researcher needs the e-infrastructure to handle the storage and processing of large quantities of data, as well as an appreciation of the value and reliability of different kinds of data.

These pressures are leading to an increasing role for computational services in structural biology. These are not simply concerned with the provision of compute resources, but also encapsulate the accumulated knowledge of a particular scientific technique. These services are often tightly coupled with experimental facilities, which generate many of the initial datasets. The Computational Biology Group in the Scientific Computing Department is closely involved in the development of a number of these services, and we describe here some recent examples.

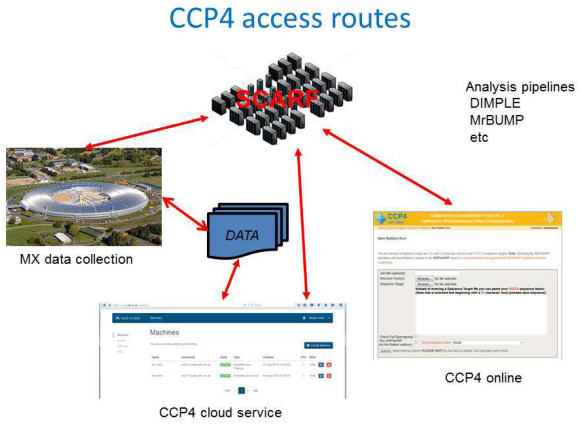


Figure 1: Modes of online access to the CCP4 suite, with particular relevance to data collected at Diamond.

The CCP4 suite for macromolecular crystallography, whose development is coordinated by the Computational Biology Group, has been traditionally deployed as a downloadable software package. Many of the individual programs can be run comfortably on a laptop, and much effort has gone into making installation and software updates straightforward for users. CCP4 has also developed several software pipelines for automated structure solution. These pipelines generally work by testing different solution strategies or parameter choices, and thus require larger CPU resources than human-directed structure solution. CCP4 online (https://www.ccp4.ac.uk/ccp4online/) was set up many years ago to provide user access to web services running these pipelines, supported by a back-end compute cluster. As well as supporting around 1000 unique users in the last year, these services provide developers with many test cases for improving the pipelines.

There are, however, several limitations to the static implementation of CCP4 online. It is hard to integrate further compute resources or data sources, and it is hard to replicate the service at different locations. In the CCP4-DaaS project, we are trying to address these issues through a solution based on the SCD Facilities Cloud (see Figure 1). Registered users can launch a Virtual Machine (VM) running Scientific Linux, and with the CCP4 suite pre-installed. While many programs can be run locally in the VM, compute-intensive pipelines can be staged to the departmental cluster SCARF. A permanent storage area is mounted in the VM, and visible as a folder My_Data in the user's home directory. Crystallographic data collected at the Diamond facility is also made available via the ICAT data service. CCP4-DaaS is in beta testing, and will soon be made available to users of STFC facilities. Users will be authenticated via their STFC federal ID, issued as part of their facility access.

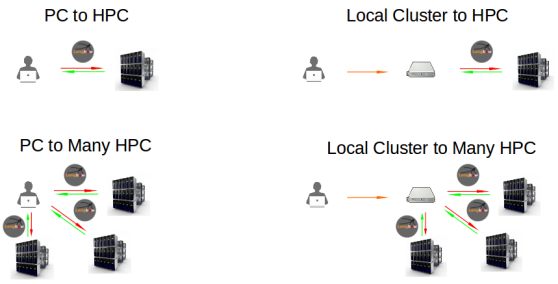


Figure 2: Modes of operation of the Longbow job submission tool. Longbow can connect to one or more remote HPC resources, and can be run on a laptop/desktop or as part of a longer workflow on a local cluster.

Staging of computations to remote resources is a concern for many communities. While Grid and Cloud solutions try to abstract out resource issues, there are many situations in which a specific HPC resource

is required. This may be for funding reasons, for example when time has been obtained on a national service such as the UK's ARCHER supercomputer. The Longbow job submission tool, developed jointly by the Computational Biology Group and the University of Nottingham, aims to simplify the process of running biomolecular simulations on remote HPC resources (see Figure 2). From the user point of view, all input files and results are viewed on their local desktop. By prepending "longbow" to a command line (e.g. running the namd2 executable), the necessary files are staged to the HPC resource, a queue submission script is written and submitted, progress of the job monitored, and results files returned to the desktop. Longbow needs to be configured with details of the remote resource, and preferred options, but then runs invisibly in the background.

Longbow has been implemented in Python, with a core library and a command line executable. Plugins have been created for specific schedulers and specific application programs, and others can be added easily. While the initial focus has been on biomolecular simulation, Longbow has also been integrated into the CCP-EM software suite for electron cryo-microscopy.

While CCP4-DaaS and Longbow represent developments in specific areas of structural biology, we are also involved in efforts to provide higher level integration. At the European level, we coordinate the West-Life project, funded by Horizon 2020 to develop a Virtual Research Environment (VRE) for structural biology. Partners in the consortium include providers of structural biology web services, developers of analysis software, and experts in European e-infrastructure. The VRE aims to give a scientist unified access to a wide range of software, services and data sources. Behind the scenes, e-infrastructure developments enable mounting of remote data sources, and running of jobs on Grid and Cloud resources.

In the current design, two frameworks (Cloudify and Infrastructure Manager) are being evaluated for portal virtualisation, DIRAC will be used as the common job dispatcher for both Grid and Cloud environments, and EUDAT solutions will be leveraged for data handling. Portal virtualisation allows deployment of the different components of a structural biology service in the cloud, e.g. the web front end, database servers, and compute resources. As so-called cloud orchestrators, virtualisation frameworks need to connect with a variety of standards. For example, Cloudify implements the TOSCA standard for defining cloud components, while Infrastructure Manager supports OpenNebula providers such as the STFC Facilities Cloud used by CCP4-DaaS (see above).

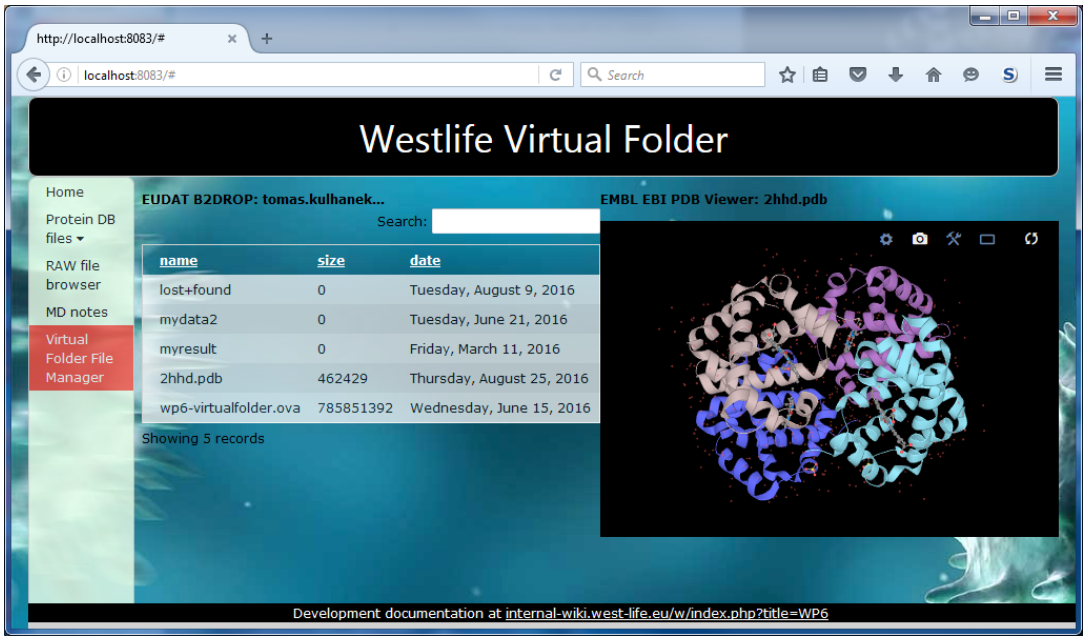


Figure 3: West-Life Virtual Folder, integrating numerous data sources including user folders on the EUDAT B2DROP service and public data at the Protein Data Bank.

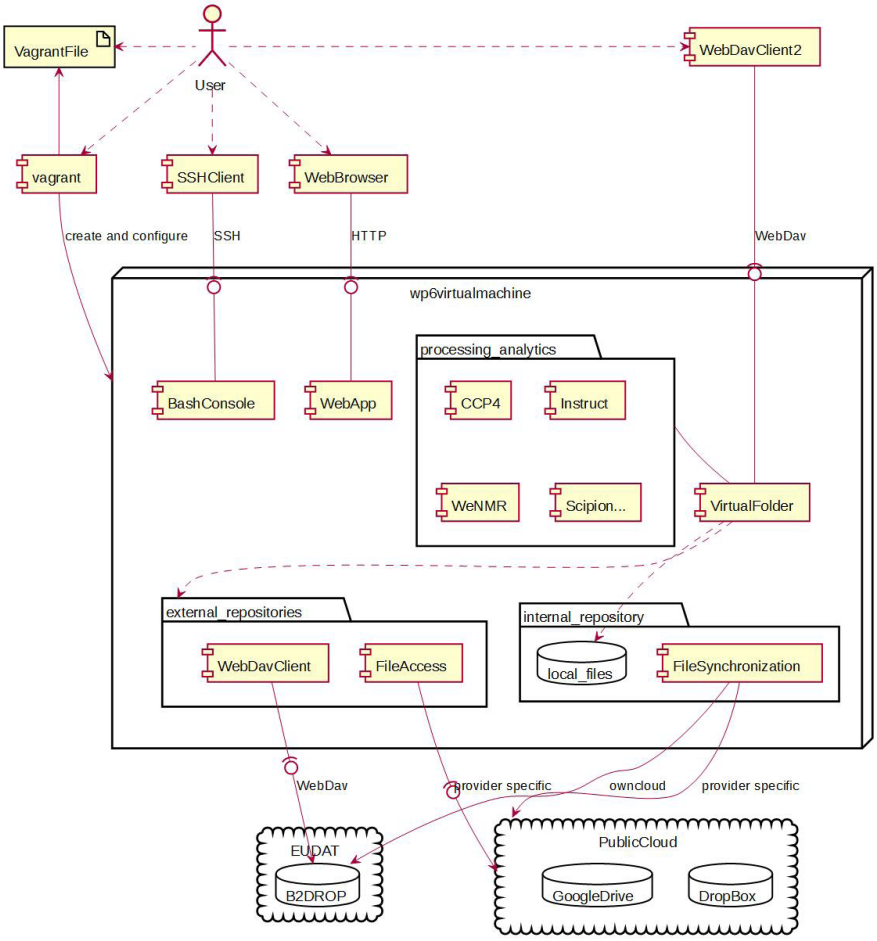


Figure 4: UML diagram showing the prototype architecture for a single deployment scenario.

One component of the structural biology portal, being developed by STFC, is a virtual folder integrating local disk space (local disk, USB) with numerous online data sources (EUDAT,DropBox, GoogleDrive), see Figure 3. The virtual folder is deployed as a headless VM which can be accessed via a web browser or via a WebDav client (Figure 4), and thus can be integrated into the West-Life VRE. The virtual folder can connect to a user's account on data services provided by EUDAT (the European e-infrastructure providing data storage for research communities), for example the B2DROP service which provides a Dropbox-like working area for a collaboration.

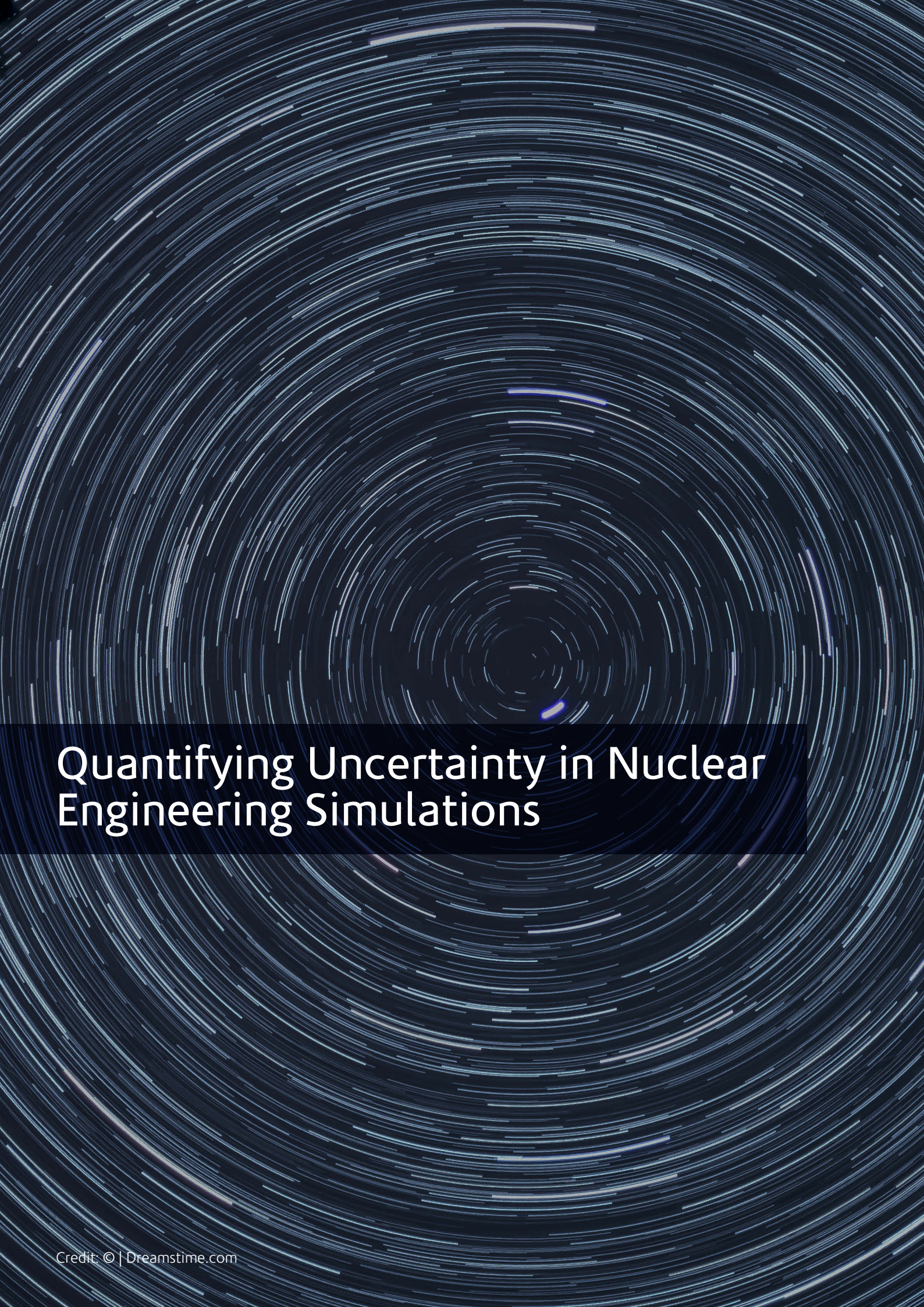
In summary, these developments will enable structural biologists to tackle increasingly complex projects, by making a wide variety of tools and data available within a user-friendly environment. These efforts are central to SCD's remit to support user communities and facilities.

Authors
T. Kulhanek, J. Gebbie and M. Winn, STFC Daresbury Laboratory

References

[1] M. D. Winn et al. (2011) "Overview of the CCP4 suite and current developments" Acta. Cryst. D67, 235-242 DOI:10.1107/S0907444910045749

[2] Gebbie-Rayet, J, Shannon, G, Loeffler, H H and Laughton, C A (2016) "Longbow: A Lightweight Remote Job Submission Tool" Journal of Open Research Software, 4: e1, DOI:10.5334/jors.95



Quantifying Uncertainty in Nuclear Engineering Simulations

The numerical modelling of fluid motion via computational fluid dynamics (CFD) of nuclear power plants gives insight into complex phenomena that affect plant safety; however, uncertainties in the operating conditions and the physical models applied, as well as the numerical description of the fluid space, all influence the mixing of fluids. Understanding how all these factors affect the solutions obtained is in the interest of researchers and regulators as well as designers and operators at utilities and nuclear engineering companies. A benchmark exercise organised by the Paul Scherrer Institute for the Working Group on Analysis and Management of Accidents of the Nuclear Energy Agency examined these factors on the mixing of two fluids in a square duct. The work presented here was performed in a collaboration between EDF and STFC.

Introduction

Currently nuclear power contributes approximately 20% of the UK energy mix with half this capacity expected to be lost by 2025 [1]. This has led to a renaissance in nuclear power in the UK, as it is seen as a secure source of low carbon baseline energy, which has led to the recent approval of the first nuclear power plant (Hinkley Point C) to be built since 1995 [2]. A research cooperation agreement between China has also changed the research landscape for nuclear power in the UK [3]. Both these developments will lead to opportunities for high performance computing centres in carrying out modelling activities on all parts of the nuclear fuel cycle in order to gain understanding of physical phenomena occurring at the plant scale.

Identifying safety issues

The safe start-up, operation and shutdown of nuclear power plants is affected by the mixing of fluids passed through the reactor core. The fluid or coolant used to cool Pressurized Water Reactors (PWR) such as Sizewell B or the proposed reactors at Hinkley Point C, Moorside and Sizewell C is an aqueous solution of sodium borate. The coolant moderates fast neutrons (neutrons with a kinetic energy of ~1MeV) to thermal neutrons (0.025eV) capable of sustaining a nuclear chain reaction of fuel containing fissile materials such as Uranium-235.

During abnormal conditions (i.e. startup, shutdown or design basis accidents), the coolant may become locally diluted and this portion of fluid may then be transported to the reactor core and cause a local increase in the number of fast neutrons, which may as a consequence result in exponential power increase from which damage to the reactor core may ensue. Current and future reactor designs have used numerical simulations of the core physics and the fluid mixing, as well as systems codes to demonstrate physical phenomena and aid the development of reactor safety systems [4]; however, there are uncertainties in the size of the diluted coolant, the impact of mixing and the impact of the dilution on fast neutron moderation.

Boron dilution

Many experimental studies have been performed examining the mixing of diluted coolant at the lab and pilot scale, while reactor scale studies have been limited to commissioning runs of reactors built twenty or more years ago [4]. Therefore, numerical models have been developed, verified and validated on smaller scale single effect and integral effect studies (IES), which examine the effect of competing phenomena arising from the resistance to the fluid flow (viscous and inertial forces) and the fluid density (gravitational and inertial forces) [5]. The effect of fluid flow and density are characterised by the Reynolds and Froude numbers, respectively, and they are both dependent on the characteristic length of the system concerned. However, it is difficult to scale experiments in accordance with both the Reynolds and Froude numbers, as the proportionality of the characteristic length is not the same. Large changes in scale will result in significant changes in the Reynolds number, from high Reynolds number at reactor scale to lower Reynolds at IES scale, while the Froude number may remain the same through appropriate selection of the conditions applied.

GEMIX Benchmark

There is limited understanding on the scale-up of numerical models from IES up to the reactor scale with respect to the Reynolds number. This is particularly important as both the experimental and numerical studies have inherent uncertainties in their accuracy, which implies a region of validity that both datasets should occupy. Consequently, when scaling up to reactor scale, it is the corresponding region of validity which should be explored in any fluid dynamic simulation at reactor scale to ensure the integrity of any resultant safety case upon which these simulations are based. Therefore, a benchmark exercise was proposed to examine the effect of the uncertainties in the conditions applied and physical models of the mixing between fluids with a small difference in density [6-7]. The fluids were mixed in a square duct, with the lower density fluid above a higher density fluid (Figure 1) [6-7]. To simulate the scale-up process,

three experimental datasets containing velocity and concentration profiles were released to all participants to verify and validate the physical models. One dataset remained blind for the participants to produce a range of the flow simulated, which would be submitted for assessment (Table 1). The datasets were defined by the adjustment of the operating conditions to change the Reynolds number and the Froude number. The velocity and fluid density were changed (Table 1). The fluid density was adjusted by adding sucrose to the fluid used, which was water. Particle Image Velocimetry (PIV) and Laser Induced Fluorescence (LIF) were used to capture the mixing of the fluids in the duct.

Table 1: Conditions applied to the experimental datasets considered.

	Low Velocity	High Velocity
Water	Open	Open
Sucrose Solution	Open	Blind

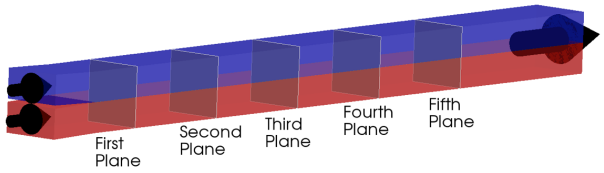


Figure 1: Duct Geometry. The planes indicate the locations of the PIV and LIF measurements released by the benchmark organisers. Red = Higher Density; Blue = Lower Density.

Our approach

The open source CFD software, Code_Saturne (http://www.code- saturne.org), is used to perform all simulations considered here. It uses the Finite Volume method to numerically solve the Navier-Stokes equations and the equations for the motion of sucrose and a fluorescent dye. Turbulence is modelled by use of the Smagorinsky form of Large Eddy Simulation (LES). OpenTURNS (http://www.openturns.org) was used to provide a range of values to which the velocity conditions could be adjusted according to the 95% confidence interval of the experimental data.

Results

Time-averaged profiles of the velocity and the concentration were obtained from LES and compared against the blind data in Figure 2, while field plots of the velocity and the concentration of the LIF dye are depicted in Figure 3 along with the variance in these values. In each plot in Figure 2, experimental data is presented with error bars fitted to the 95% confidence interval, while two simulation results are presented at the upper and lower range of values applied to the inlet conditions.

The profiles of the simulations in Figure 2 show broad agreement with the experimental data; however, the magnitude of the peak values of the velocity under-predict the experimental profiles and there is significant dissipation or smoothing of the peaks as the fluid travels along the length of the duct. The experimental profiles become sharper along the duct length. This suggests that some phenomena is not being captured by these simulations, particularly as the flow in the duct is a developing flow and the turbulence used is not able to correctly assess the boundary layer phenomena. The uncertainty bands in the simulated data were insignificant. Further work is required to fully explore these bands when we can better capture the physical phenomena. Nevertheless, the concentration profiles and fields presented in Figures 2 and 3 are consistent with the experimental data released by the benchmark organisers and those reported in [6-7].

Conclusions and future work

The simulation of mixing in the GEMIX duct using LES showed broad agreement with the experimental studies; however, the experimental results indicate that the flow is developing and we are not capturing all the phenomena observed in the duct. Further simulations are being performed to try to capture all the physical processes occurring in the duct before further uncertainty analyses can be performed.

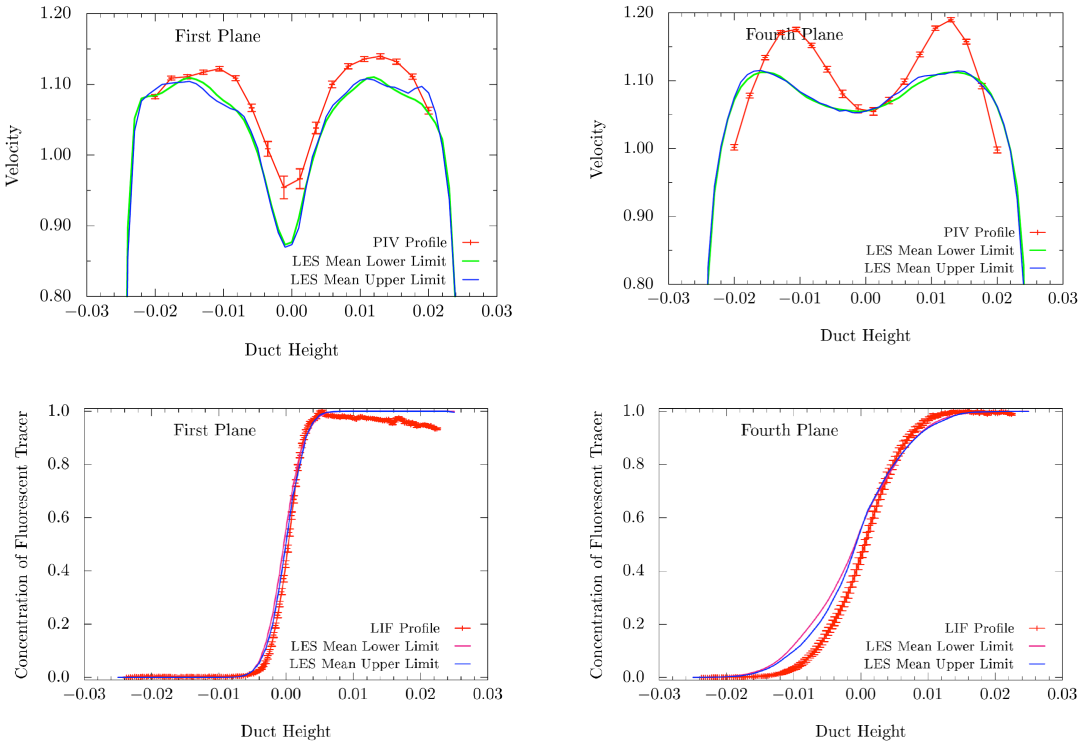


Figure 2: Profiles of velocity and concentration at first and fourth measurement planes.

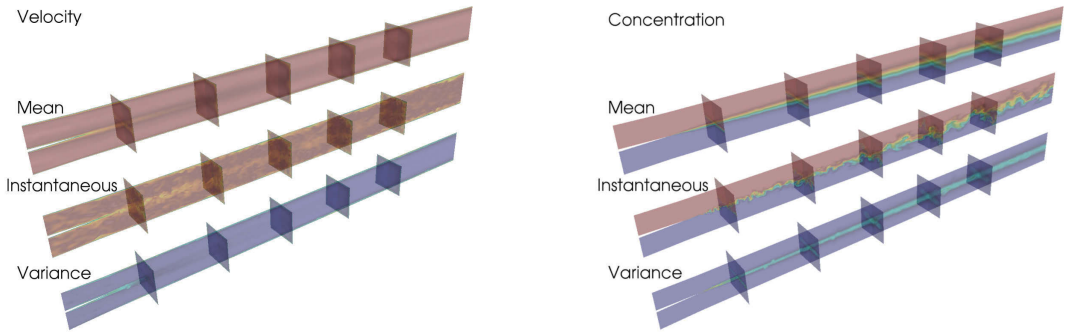


Figure 3: Fields of mean velocity and mean concentration with the corresponding instantaneous and variance fields. Warmer colours indicate larger values.

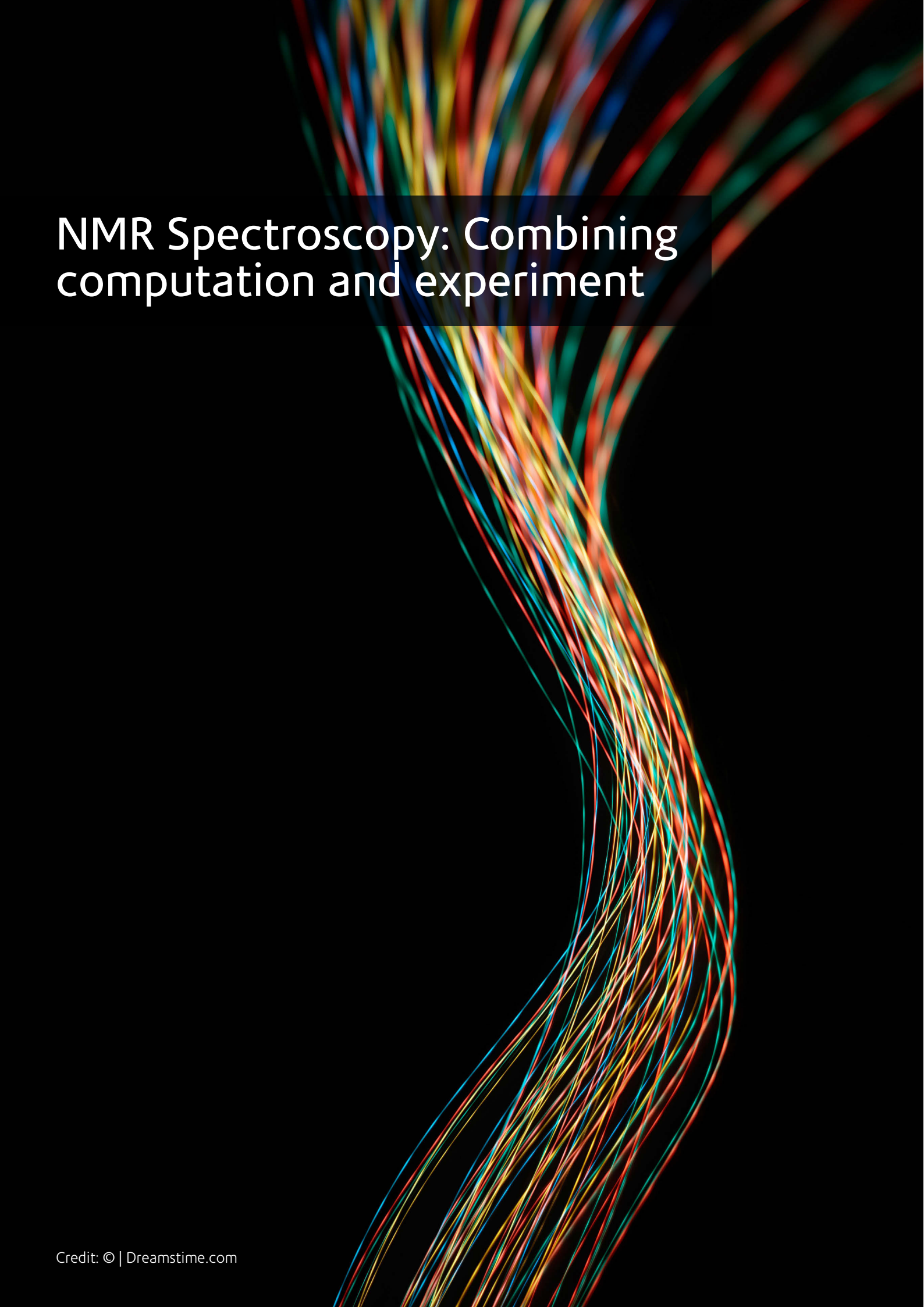
Authors
G. Cartland Glover, S. Rolfo, C. Moulinec, R. Barber, D. R. Emerson, STFC Daresbury Laboratory

References

[1] World Nuclear Association, Nuclear Power in the UK, http://www.world-nuclear.org/information-library/country-profiles/countries-t-z/united-kingdom.aspx, Downloaded 05/09/2016.
[2] Hinkley Point C media team (2016), Press Release, https://www.edfenergy.com/energy/nuclear-new-build-projects/hinkley-point-c/news-views/FID, EDF Energy, Downloaded 05/09/2016.
[3] NNL media team (2015), Press Release, http://www.nnl.co.uk/news-media-centre/news-archive/uk-china-joint-research-and-innovation-centre/, National Nuclear Laboratory, Downloaded 05/09/2016.
[4] Office for Nuclear Regulation (2012), Plant modelling, http://www.onr.org.uk/research/2012/section-f.pdf, Office for Nuclear Regulation, Downloaded 05/09/2016.
[5] T. Hoehne, E. Krepper, and U. Rohde (2010), Application of CFD Codes in Nuclear Reactor Safety Analysis , Science and Technology of Nuclear Installations , Volume 2010, Article ID 198758.
[6] Nuclear Energy Agency (2016), Computational Fluid Dynamics in Nuclear Reactor Safety, https://www.oecd-nea.org/nsd/csni/cfd/, Organisation for Economic Co-operation and Development, Downloaded 07/09/2016.
[7] J. Fokken (2015), Velocity and Concentration Field Measurements of Unstratified and Stably Stratified Confined Wake Flows Past a Plate, Doctor of Science Dissertation ETH Nr. 22988, ETH- Zuerich.

The authors would like to thank Romain Camy of EDF for his contribution and the Hartree Centre for providing time on their machines.

NMR Spectroscopy: Combining computation and experiment



Credit: © | Dreamstime.com

Nuclear Magnetic Resonance (NMR) has historically been a fundamental technique to explore the molecular structure of new substances. Today, modern powerful ab-initio simulations allow to predict NMR results with unprecedented accuracy, making this technique able for the first time to give even deeper insights into crystalline order in both organic and inorganic compounds. This new potential, however, can be somehow stifled by the practical problems posed by the usage barrier between experimental NMR scientists and the running and interpretation of computational simulations.

The CCP for NMR Crystallography, or CCP-NC, aims at abating this barrier by releasing software tools designed to make it easier and more intuitive to interpret ab-initio NMR results and compare them with experiment, as well as automate as much as possible of the process. Its first product, MagresView, has now been fully released as a live web application and provides the experimental community with the possibility to visualise and post-process NMR data in the "magres" file format. This format is now being pushed as a new standard for the field, and MagresView is contributing to its diffusion by providing a standard tool that employs it.

Since its inception, Nuclear Magnetic Resonance has been a very precious tool in unraveling the details of chemical structure in many substances, compensating for the weaknesses of other techniques such as X-Ray Diffraction. Fundamentally, NMR is a technique that allows us to measure the magnetic field experienced by specific atomic nuclei inside a sample. This magnetic field is strongly influenced by the arrangement of the electronic cloud surrounding the atom itself, and thus can be a precious insight in what kinds of bonds the atom itself is forming and with which species. In addition, since the signal has such a short ranged origin (a few Angstroms at best), NMR does not require fully grown single crystals like some other spectroscopic techniques and works well with powders or even liquids

However, the relationship between chemical structure and local magnetic fields is all but trivial. Historically, empirical observation and basic theoretical considerations allowed for interpretation of the strongest contributions to the signal, but the smaller details would be too hard to predict, and the information they conveyed lost. Recently, though, new advancements in both algorithms and sheer computational power have made it possible to obtain far more accurate predictions of NMR parameters starting with chemical structure using ab-initio simulations. This paved the road to proper NMR crystallography, a previously unthinkable enterprise, extending the amount of information that we can extract from the same data about the overall molecular and crystalline structure of a substance. Such a technique has great potential in those fields, like pharmaceutical research, that work with organic molecular compounds that organise in many subtly different crystalline forms but can be hard to investigate with diffraction techniques due to the predominance of light atoms or the difficulty to grow large single crystals.

One of the bigger obstacles to this approach being adopted by the community at large lies in the difficulties of connecting properly the computational and the experimental sides in terms of skills, know-how, conventions and language. As it is often the case, these two different fields have developed different ways of describing the same things, and it does not help that the measured quantities in NMR experiments are often pretty far removed from the most natural physical descriptions of the parameters as an ab-initio simulation can derive them. For example, a typical product of an ab-initio NMR calculation could be the "shielding tensor" at a given site, a matrix returning the magnetic response of the electronic wavefunction to an external field; however, chemical NMR most often deals with "chemical shift", namely the trace of that tensor referenced to the same value in some standard chemical compound (since absolute values are easily calculated but fundamentally nigh impossible to measure in an experiment). The vast amount of data processing required just to parse the results of an ab-initio calculation in a form that is easily digestible and comparable to experiment can in itself be a huge barrier to usage of this novel approach. For that reason CCP-NC has funded the creation of MagresView, a software tool dedicated exactly to solving this problem in a user-friendly way.

MagresVies is a web application relying on JMol, a commonly used open source molecular visualisation software developed by Robert Hanson. JMol works also as a web applet (both Java and Javascript), and through a close collaboration with its creator in the last years it has implemented both the ability to parse magres files, the CCP-NC format of choice for storing NMR data, and process basic NMR quantities. In addition to that, the magres file format has been adopted as output both by the CASTEP and the Quantum Espresso ab-initio software packages, making it a possible candidate for a future general standard. What MagresView adds on top

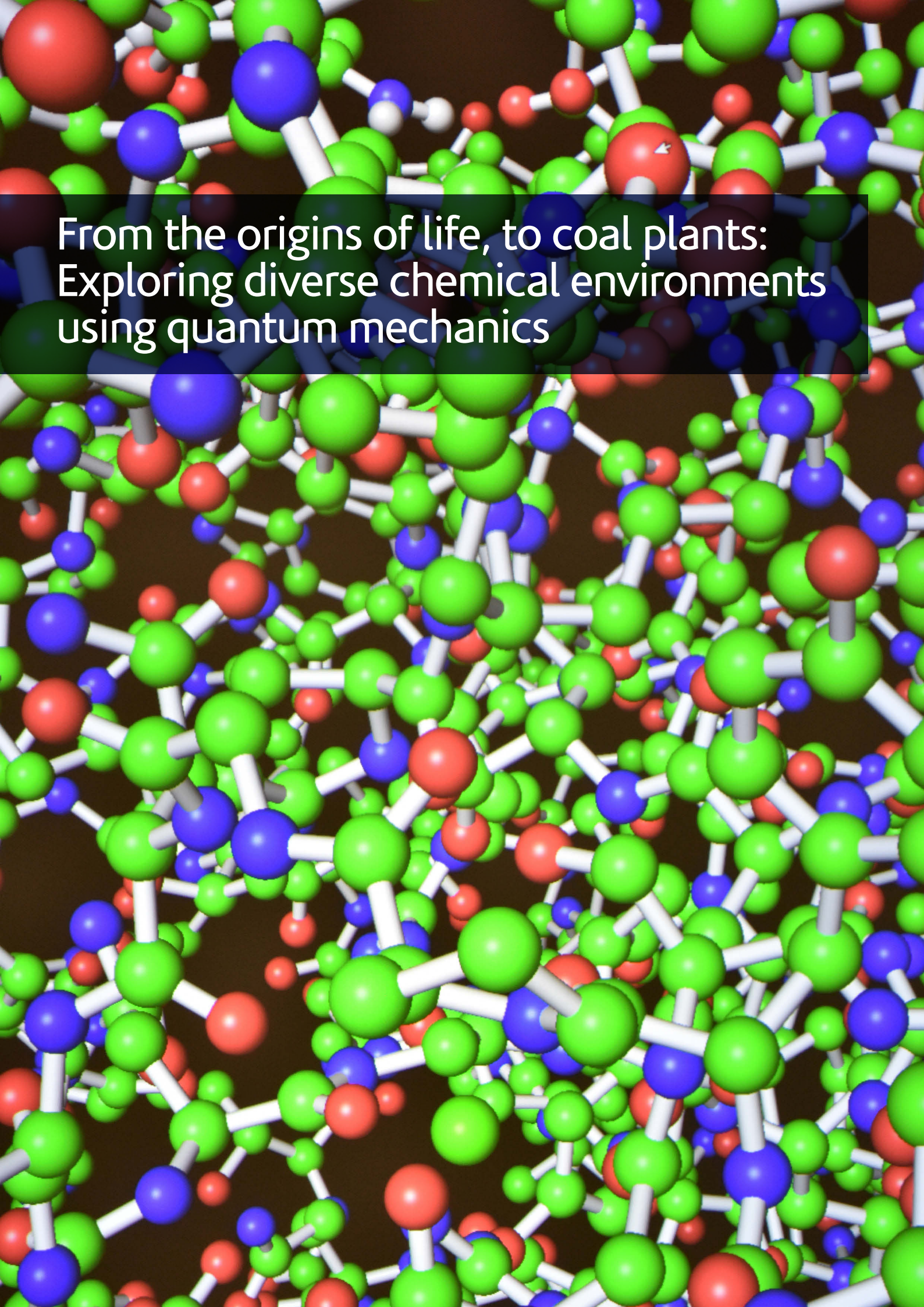
of the basic JMol capabilities is a Javascript powered user interface that allows to execute all the most common visualisation and post-processing operations in a simple, intuitive manner.

For example, tensorial quantities like the shieldings and the electric field gradients are visualised in MagresView in ellipsoid form. This provides an easy way to convey the information contained in the tensors' eigenvalues and eigenvectors, which is far more meaningful than their explicit form. In addition, it provides text and color scale labeling based on all the most commonly defined ways of expressing them: chemical shift, anisotropy and asymmetry, but also span and skew, used by different conventions. Similarly, visual links are employed to display atom-atom coupling effects, limiting the number of links visualised at a time by user selection to avoid overcrowding the plot.

In addition to these capabilities, MagresView also provides some basic computational ones. Not only are tensors diagonalised, but the relative rotation of their principal frames can be expressed in the form of Euler angles (a commonly used quantity in quadrupolar NMR), dipolar couplings can be calculated from scratch, quadrupolar shifts include up to second-order corrections, and quantities that are affected by the choice of isotope or magnetic field properly account for them. All this can be saved into textual output that can then serve as input for other software. An important target of MagresView in fact is to ease the use of spin dynamics software like SIMPSON, which are necessary to predict the results of the most sophisticated NMR experiments, even given the basic parameters. MagresView provides formatted output that can be directly copy-pasted into a SIMPSON input file, thus reducing the distance from ab-initio simulation to NMR experiment simulation to one simple step.

MagresView has been available to the public for more than a year now at http://www.ccpnc.ac.uk/magresview as a live web version and on CCPForge as stand-alone source code. The success of the software across the NMR community, especially the UK one, is very satisfying, with Google Analytics reporting often more than 100 page hits per week. We hope that it will be the first stepping stone to build a better bridge between the computational and experimental communities in this sector and prelude to a vaster collaboration leading to systematic understanding and classification of chemical structures by their computed NMR parameters, which would all-around greatly benefit future crystallography efforts.

Author
S. Sturniolo, STFC Rutherford Appleton Laboratory



From the origins of life, to coal plants: Exploring diverse chemical environments using quantum mechanics

All chemical reactions (interactions involving the making or breaking of chemical bonds) occur as a result of the activity of electrons; the behaviour of the local electrons determine whether a chemical bond changes. The small size of electrons (approximately 2.818 x 10^{-15}m) belies their influence and role in the properties of matter, both perfectly encapsulated by the periodic table where from element-to-element across the rows, a single electron separates neighbouring elements and their various attributes. Although it is also true that each element has a different number of protons, these do not contribute to chemical reactivity in the same way as do electrons.

Exploring electron behaviour in-depth would enable us to understand the whys and wherefores of chemical reactivity, eventually leading to the design and chemical engineering of material systems. Although such in-depth understanding of electron behaviour at the experimental level is difficult to achieve, it is relatively straightforward to do it theoretically using a branch of mathematical physics called quantum mechanics (QM). Originating from Max Planck in 1900, augmented by Einstein in 1905 and re-vamped in the 1920s, QM has been around now for over 100 years. Pierre Hohenberg and Walters Kohn and Sham made QM tractable to implementation in theoretical chemistry and physics in the 1960s.

With the subsequent development of various flavours of applied QM, the growth in computational power and development of high performance computing (HPC) systems together with the refinement of matrix algorithms, we arrive at today's capabilities whereby material systems containing hundreds to millions of electrons can be explored and their behaviour understood. In this way, we gain much-needed insight to inform experimental results, plus we are able to make predictions about environments that are difficult or impossible to reproduce (consider the centre of the Earth for example).

In this report we present two examples of how we have used a particular category of QM called density functional theory, (DFT) to explore such diverse chemical environments as the early Earth, and the material expelled from the chimney stacks of coal plants.

General QM Methods

Basically, the computational implementation of DFT comprises a series of minimisation routines, with the aim of locating the lowest energy minimum or ground state for a given atomic arrangement and imposed convergence criteria. The initial configuration provides the initial atomic arrangement within which the interaction energy of the electrons is minimized. This is followed by another minimisation method to move the atoms incrementally in the direction of decreasing forces (between the atoms). When the increment is complete the electron energy is again minimised, and the dual-cycle routine continues until the convergence criteria – imposed by a human – are met.
A positive feature of all modelling at all scales is that we can work with perfect examples of material systems, which provide a baseline against which the physical samples can be measured. This can help experimentalists to distinguish between genuine results and experimental noise.

All of the calculations for the following two case studies were carried out using the plane wave, pseudopotential code, CASTEP[1] using 8 to 64 cores on a HPC system.

Origins of life

For a detailed account of this work see the full article[2]. What do we mean by the phrase 'Origins of life'? In the current context we mean the precursors of nucleic acids and peptides, where the former hypothetically progressed to form RNA and DNA, and the latter to proteins. The building blocks of both nucleic acid and peptides are amino acids, and these and their constituent building blocks were thought to be abundant on the early Earth. The formation of peptides from amino acids is enabled if they are brought together in a confined environment, within proximity of a catalysing mechanism that promotes peptide bond formation. One such environment is believed to be layered clay minerals, where the charged layers are sandwiched together with oppositely charged, hydrated ions as the 'filling', see Figure 1.

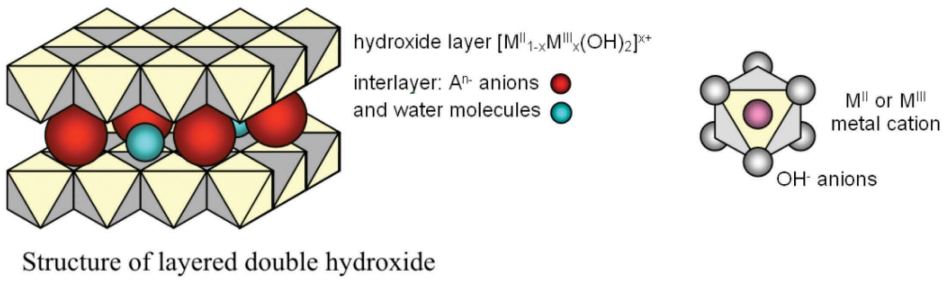


Figure 1. Schematic of layered double hydroxide (LDH)[3]. LDH comprise a class of layered clay minerals where the layers are positively charged and the interlayer 'guests' are negatively charged

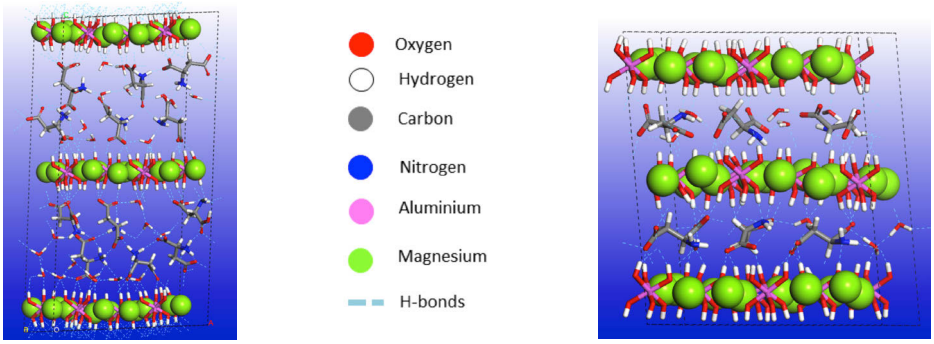


Figure 2. Higher charge density LDH models showing two interlayer spaces where, on the left each contains 6 x aspartate (-1) and on the right 3 x aspartate (-2). Comparing the DFT-computed interlayer spacing to experimental results confirmed aspartate (-2) is present in the LDH interlayers.

In this joint experimental/theoretical study we explored two similar LDH: $[Mg^2Al(OH)_6]^+$ and $[Mg_3Al(OH)_8]^+$ and two forms of amino acid: aspartate with a charge of -1 and -2. The different ratios of Mg:Al in the LDH determines the relative effects of layer charge on the amino acids in the interlayer space, where $[Mg_2Al(OH)_6]^+$ has the higher layer charge density. The two charges of aspartate represent the effect of different pH environments, where at pH 7 aspartate is present as a zwitterion with a charge of -1 and at higher pH it has a charge of -2. The effect of hydration was also considered, with the number of water molecules per amino acid ranging from zero to five. An example of a LDH/aspartate model is shown in Figure 2.

Comparing the DFT-computed interlayer spacing to experimental results confirmed aspartate (-2) is present in the LDH interlayers.
The QM computational results provided geometrical information of the amino acids, an aspect of the intercalated LDH/aspartates that was difficult to analyse by experimental methods alone. This was in part due to the lack of perfection of the experimental crystal structure, making the resulting analytical spectra difficult to interpret. However, in combination with the DFT results, we were able to determine that

(i) Aspartate exists as a dianion i.e. aspartate (-2) within the LDH interlayers.
(ii) The layer charge density is the main influence on the orientation of asparate within the LDH interlayers.
(iii) To a lesser extent, the degree of aspartate hydration influences aspartate orientation within the LDH interlayers.

These results support the hypothesis that the interlayer environment of LDHs structurally orients the amino acids to be favourable for peptide formation, and furthermore, that these layered, clay minerals could have played a vital role in the origins of life on Earth.

Coal plants

For a detailed account of this work see the full article[4]. Human activity causes the release of toxic mercury (Hg) into the environment, up to 42% of this is due to coal combustion and unintentional release through by-products. With recent regulations enforcing reduced Hg emissions, and new coal-fired plants held to still lower emissions, there is significant research into Hg-sorbents to prevent Hg emission. One such sorbent is coal fly ash, a waste material of burning coal and is present in the flue gas stream of the chimney stack alongside volatile Hg. Although 99.95% of this fly ash is captured, approximately one ton per plant is released into the atmosphere every day. What is not yet known is the form of Hg adsorbed to coal fly ash and how much Hg is released along with it. The adsorbed form of Hg determines its bioavailability and hence potential toxicity to the local environment as well as to humans. For example, Hg falls into aquatic environments, becomes adsorbed by fish that are ultimately eaten by humans. Also, industrial-scale waste products are potential sources of feedstock in various sectors such as the building industry. For example, coal fly ash could be used as a building material in concrete mixes, hence the stability or otherwise of Hg adsorbed to fly ash must be thoroughly assessed. See Figure 3 for a schematic overview of the above.

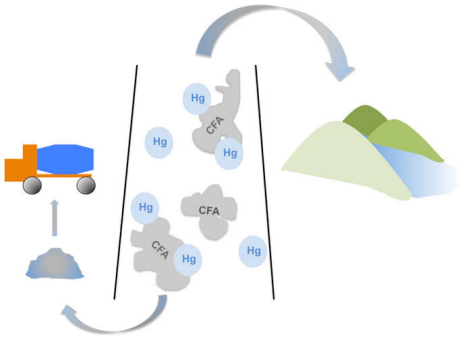


Figure 3. Schematic showing volatile mercury (Hg) adsorbing to coal fly ash (CFA) in the chimney stack of a coal plant. Some of the mercury-containing waste coal fly ash is emitted to the environment, and some is collected and used as feedstock for the building industry. Is the mercury adsorbed to the coal fly ash in a stable form and therefore unlikely to leach into its immediate environment?

In this joint experimental/theoretical study the research question was: Does the Hg adsorbed to the fly ash exist in a form that is readily released to the environment? To answer this question we first needed to understand that fly ash is composed of crystalline and amorphous organic and inorganic material, and second to identify the forms of Hg attached to these components. Our experimental colleagues took two batches of fly ash obtained from a Kentucky coal plant and reacted one batch with a simulated flue gas stream (containing Hg) while keeping the other unreacted. They then used a range of techniques to identify '…(1) the size fractions of coal fly ash with which Hg is associated, (2) the solid phases and elements associated with Hg, and (3) the molecular-level speciation of Hg in the coal fly ash'[4].

They needed help from DFT to answer (3) because from their analytical spectra obtained by Fourier transform infra-red (FTIR) they observed a shift in peaks around 2900 cm^{-1} between the reacted and unreacted fly ash. Crucially, their experience enabled them to associate this wavenumber with carboxylic acid functional groups, leading to the hypothesis that within the flue gas stream, Hg was binding to a carboxylic acid functional group (i.e. COOH) within the fly ash. This functional group is present as one of the amorphous organic components comprising fly ash.

To test this hypothesis we carried out DFT calculations on (gas-phase) $OHCH_2COOH$ and $OHCH_2COOHgOH$, relaxing the electrons and atomic geometry as previously described. We then carried out vibrational frequency calculations to simulate infra-red (IR) spectra to compare with the experimental spectra. The carboxylic acid molecules and their simulated IR spectra are shown in Figure 4.

The shift in peaks between the unreacted and reacted fly ash samples was exactly matched by the DFT results, and to a lesser extent the change in intensity on Hg uptake by carboxylic acid functional groups was also reproduced, thus confirming Hg binds to carboxylic acid functional groups in coal fly ash. This type of bound Hg is relatively stable and therefore unlikely to leach into its local environment. Together with the experimentalists' results to the previously mentioned questions (1) and (2), we were able to identify the likely stability of the Hg phases found in fly ash emitted from a coal plant. We concluded that low Hg exposure would result from waste coal fly ash emitted from coal plants as well as that used in the building industry.

Summary

As in any modelling simulation it is important to identify the essential elements of the environment you want to explore, as well as the information required from the data. In the previous example of the 'Origins of life' the focus was peptide formation in the aspartate-filled interlayers of LDH, and the information sought from DFT primarily structural and geometrical. In the case of 'Coal plants', the focus was the type of Hg species adsorbed to coal fly ash, and the information sought from DFT was simulated IR spectra from vibrational frequency calculations. Both of these studies illustrate the depth of insight obtained from DFT calculations, shown to its best advantage when working with our experimental partners.

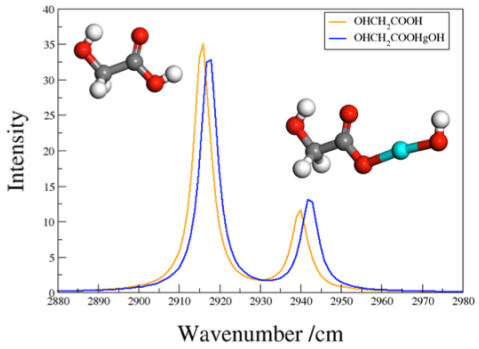


Figure 4. Colour scheme as shown in Figure 2. DFT computed IR spectra illustrating the shift in wavenumbers induced on the addition of mercury (light blue) to a (gas phase) molecule containing a carboxylic acid functional group (i.e. COOH).

Authors
D. Geatches, STFC Daresbury Laboratory

References

[1] Clark S. J., Segall M. D., Pickard C. J., Hasnip P. J., Probert M. I. J., Refson K., Payne M. C. (2005) First principles methods using CASTEP. Z. Kristallogr. 220, 567-570.

[2] Brian Grégoire, Valentina Erastova, Dawn L. Geatches, Stewart J. Clark, H. Christopher Greenwell and Donald G. Fraser. Insights into the behaviour of biomolecules on the early Earth: The concentration of aspartate by layered double hydroxide minerals; Geochimca et Cosmochimica Acta, (176), 239-258, 2016.

[3] http://tresen.vscht.cz/min/en/research-team-mineralogy

[4] Adam D. Jew, Erik C. Rupp, Dawn L. Geatches, Ji-Eun Jung, Gabriela Farfan, Louisa Bahet, James C. Hower, Gordon E. Brown, Jr., and Jennifer Wilcox, Mercury Interaction with the Fine Fraction of Coal- Combustion Fly Ash in a Simulated Coal Power Plant Flue Gas Stream, Energy Fuels, (29), 6025-6038, 2015.


Data Services

The Data Services group runs data stores for storing, archiving, preserving, analysing, and backing up scientific data; most of the data is from the Large Hadron Collider (LHC), but there are also significant holdings of climate modelling and STFC facilities data. The Data Services Group also participates in projects and other research in order to increase our knowledge and capability to support data management, to improve the services we run by making use of research and to increase the economic and societal impact of our data by providing expertise and facilities for using open data.

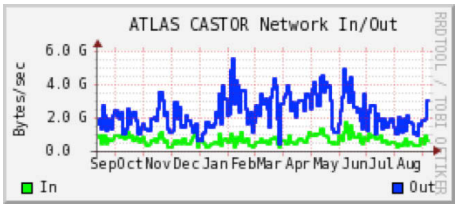
The group is involved in a number of Horizon 2020 projects, including ESiWACE, investigating the future of data storage for climate data, EUDAT which continues work on a collaborative data infrastructure for European researchers, SAGE, a project to provide a next-generation multi-tiered object-based data storage system and AARC, Authentication and Authorisation for Research and Collaboration.

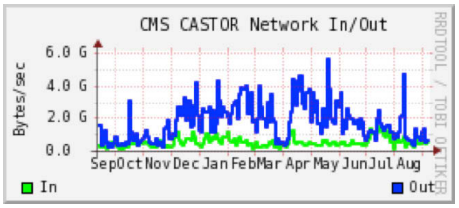
AARC

Single sign-on (SSO) is important as it enables users to easily access several infrastructures using identities they already have (their home organisation, social media) as long as these identities have a level of assurance appropriate to the resource being accessed. It removes the insecure and long obsolete infrastructure-specific username/password. For STFC, SSO traditionally meant that users of STFC's facilities had a single account with which they would do all their work; today this would be linked to their Umbrella identity so users would have a single identity across a wide range of neutron and lightsource facilities. The Horizon2020 funded AARC project (with contributions from STFC's Particle Physics and Scientific Computing Departments) is developing a unified SSO architecture based on "federated identity management" (bridging national academic identity management federations) and policy harmonisation. This work has recently led to the funding of the successor project, imaginatively named AARC2, which will continue the work on harmonisation and develop and support a framework for authentication and authorisation in research infrastructures. A good example of how this works in practice can be seen in the H2020-funded data infrastructure project EUDAT. EUDAT's B2ACCESS component provides SSO to EUDAT resources, enabling researchers to use their home institute or social media ID, or ORCID, to authenticate to EUDAT resources regardless of the type of resource or where it is running. EUDAT's "collaborative data infrastructure" provides data and metadata services to a wide variety of research disciplines and enables researchers to work together across disciplines.

CASTOR

The LHC's Run 2 has led to increased load on the CASTOR (Cern Advanced STORage Manager) system running here at RAL. CASTOR has coped well despite I/O loads peaking around 10 gigabytes per second from all LHC experiments combined. Below are network throughput plots for the two most heavily-loaded instances of the four we run:





We continue to receive large amounts of data, not only from the LHC experiments but also Facilities data, primarily from Diamond and CEDA, which is archived onto tape. In total this year we have received around 10PB of data from all users.

A major software upgrade to the CASTOR system is currently in an advanced testing stage and, in addition, we plan for a migration of Tier-1 disk storage to the new system, Echo, a Ceph object-based storage system. Once Echo is in production we will be able to greatly reduce CASTOR's hardware footprint, both due to the disk storage capability moved to Echo and the work being carried out to integrate the two systems into one combined storage system.

The data archival service for the Diamond Light Source is based around CASTOR and the ICAT metadata catalogue. This year we have refreshed the whole ICAT software stack providing much improved functionality and availability for Diamond Archive users. In addition, we now also provide a service with the option to download data via Globus, allowing the reliable recall of TB datasets to users' home institutions. With more and higher resolution instruments coming online at Diamond, data volumes being archived continue to accelerate; we now hold 1.2 billion files, over 5PB of Diamond data.

Authors
J. Jensen, R. Appleyard, C. Prosser, A. Packer, STFC Rutherford Appleton Laboratory

Managing data from the Central Laser Facility's High Power Lasers

Credit: © | Dreamstime.com

Each time the Central Laser Facility (CLF) fires one of its lasers it generates diagnostic data. The analysis of variations in the laser beam, shot trends, behaviour, energy delivered and other diagnostic information is critical in understanding experimental results.

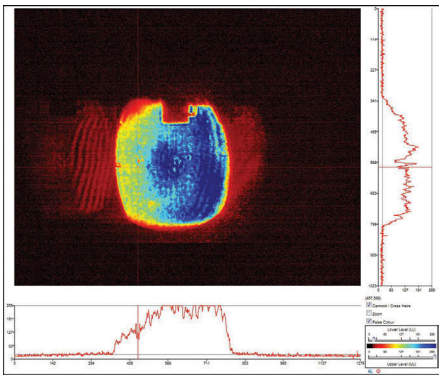
For nearly 10 years the Scientific Computing's (SCD) Research Data Group (RDG) has been working with staff supporting CLF's Astra Gemini laser to catalogue all of the diagnostic data coming from sensors, cameras and other equipment such as oscilloscopes. For the last five years the data has been stored in ICAT (www.icatproject.org), a metadata catalogue developed by RDG specifically to support Large Facility experimental data. ICAT is also in use at ISIS, Diamond, the European Synchrotron Radiation Facility in Grenoble and the Spallation Neutron Source at Oak Ridge, Tennessee.

The raw data is made available to laser operators and users who are doing experiments via a web-based application known as eCAT, which RDG developed specifically for this purpose. Operators use this data to confirm that the laser is set up and operating correctly.

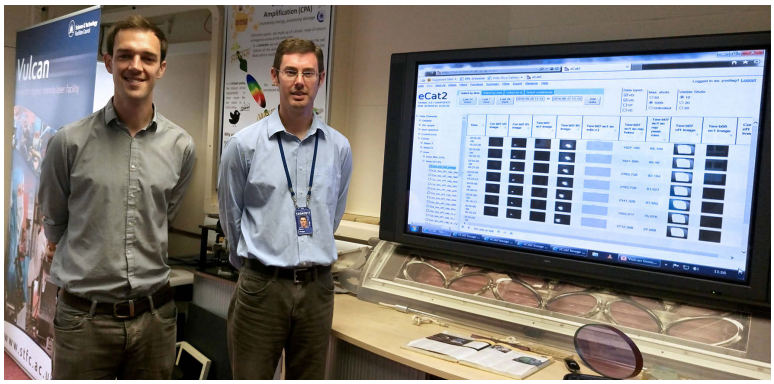
The ICAT system for Astra Gemini has proved useful primarily for allowing Operators to easily get data for each shot and combine it with other data sources to give an overview of the state of the entire beam line. Over the last year RDG has been collaborating with CLF to adapt the system to support the Vulcan laser. This involved CLF making changes to the way they save diagnostic data, mainly in the area of the file names and formats, such that the ingestion tool run by RDG can read the data and catalogue it correctly in ICAT. The users can now access the Vulcan laser data using eCAT.

Over the past year the CLF has also bought in a tool named DARB - an automated system for Diagnostics, Analysis, Review and Backup - developed at the University of Strathclyde for automatically saving data from additional diagnostics such as specialist, scientific cameras that the users set up during their experiments. This has been adapted to save the data in the format required for ingestion into ICAT, and was successfully tested for the first time on a Vulcan experiment in March 2016. The plans for the coming year are to put the Vulcan ICAT system into full production and to deploy the DARB software fully for both Vulcan and Astra Gemini.

From the long term collaboration between CLF and SCD for the Astra Gemini laser, it is gratifying to see that the system has become increasingly valued by CLF and continues to evolve to meet their requirements. Being able to take a system that has had years of development effort invested in it and, for a relatively small amount of effort, adapt it for use by another laser has been cost effective as well as allowing a complex system to be put in place within a relatively short period of time. With the two lasers now using common data management software, it also means that future updates to the software potentially benefit both lasers. The addition of the DARB tool has proved to be a successful collaboration with Strathclyde University, and again, with a relatively small amount of effort, it has been possible to customise it so that it integrates with the ICAT tools.



False colour image analysis using eCAT



SCD developers Frazer Barnsley and Kevin Phipps next to eCAT on the large screen in the Vulcan Main Control Room

Author
K.Phipps, STFC Rutherford Appleton Laboratory

Monitoring in the petascale era

The Research Infrastructure group uses advanced monitoring tools to gain unprecedented insights into the services it provides and clusters it runs.

The batch clusters operated by the group process millions of jobs each year, consuming hundreds of millions of CPU hours. This year, RIG deployed the XDMod[1] software developed by the University at Buffalo. XDMod parses the log files generated by the batch system and then extracts and reprocesses the information into a database. The software allows retrieval of detailed information from the batch system logs, for example, the size of submitted jobs, the total CPU hours used and the number of active users. This facilitates forward planning as the required capacity and the change in user base can be extrapolated over time.

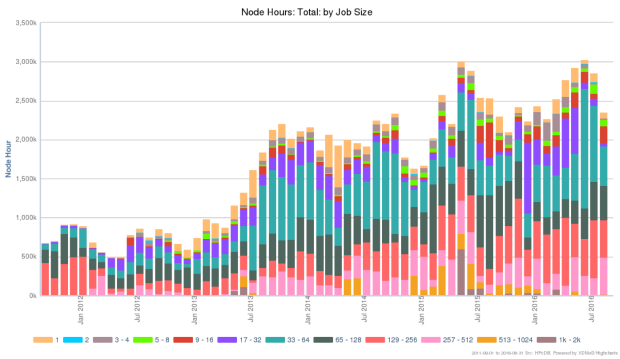


Figure 1 CPU Usage on SCARF broken down by job core count

Data ingest is a critical activity for RIG, for example JASMIN supports the NERC FAAM research aircraft [2] performing experiments while flying over the Arctic, the WLCG Tier 1 must be able to accept data from CERN when the LHC is running and projects such as ULTRA will use SCARF to support experiment users running on ISIS, who will take data from the IMAT instrument and then transfer them to SCARF to perform tomographic reconstruction during their allocated beamtime. Therefore ensuring that the network has no problems is vital: even a tiny number of dropped packets on a network link can prevent the necessary data rates being maintained due to the increased number of retries, data being held in buffers and the transmission protocols attempting to compensate for packet loss by reducing the rate of packets sent. The network that RIG maintains is especially complicated with direct optical connections to CERN and other major particle physics laboratories, the Met Office and the UK HPC facility ARCHER.

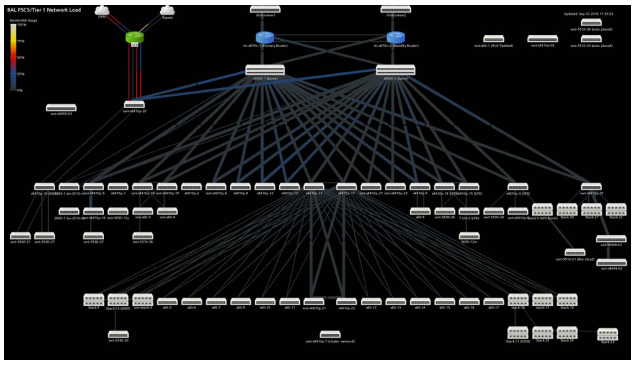


Figure 2 The WLCG Tier 1 hierarchy of the RIG network infrastructure

The open source tools Observium [3] and Cacti[4] are used to visualise the overall traffic flows and identify congestion, whereas tools such as Solarwinds [5] are used to delve deeply into problematic connections

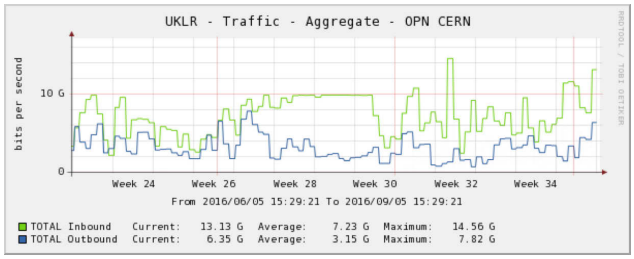


Figure 3 WLCG Tier 1 inbound traffic from CERN reaching link capacity in weeks 28 and 29, subsequently the link capacity was increased as can be seen in week 32

RIG stores a vast quantity of data on a wide range of equipment, ranging from the world's largest public installation of Panasas high-speed parallel storage that forms the basis of the JASMIN infrastructure for NERC, to two 10,000-slot tape robots providing archival storage to both JASMIN and the WLCG Tier 1. In total RIG has thirty petabytes of disk based storage and eighty petabytes of tape.

Some of the data stored are of global importance: 60% of the data gathered for the most recent IPCC climate change report are stored on JASMIN and the data stored on the WLCG Tier 1 were used in the discovery of the Higgs Boson.

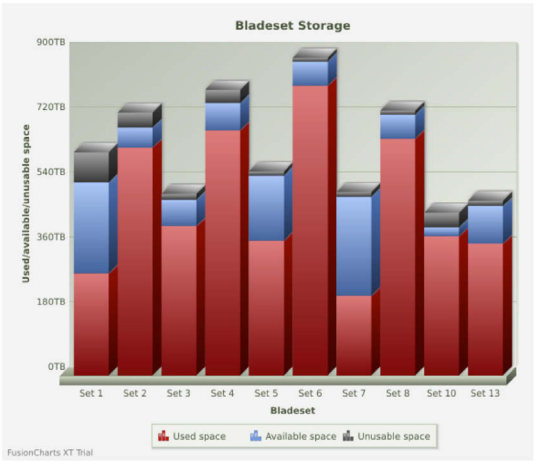


Figure 4 Breakdown of JASMIN Disk Storage

In total, RIG manages over two thousand individual computer systems, each of which has multiple temperature and other sensors and metrics. RIG uses Ganglia [6] to extract, summarise and aggregate this information. Each host gathers it own data and then reports that data back to a central aggregation server, the data on the central server is stored inside round robin database files allowing the data to be automatically summarised as it ages. This data can be extracted and re-formulated as input to dashboards allowing at glance overviews of the state of the physical fabric of the clusters.

RIG is transitioning from the Nagios monitoring system to Icinga [7]. Nagios and Icinga are both used for alerting - raising alarms about problems that have been detected on systems. RIG have an extensive suite of tests - over thirty thousand checks, ranging from simple checks that a host is accessible on the network to detailed checks for memory problems on GPGPU cards, providing a comprehensive overview of the state of the systems and guiding support staff to problems. Out of hours, significant problems can initiate a callout to alert an on-call staff member to attempt to resolve the issue.

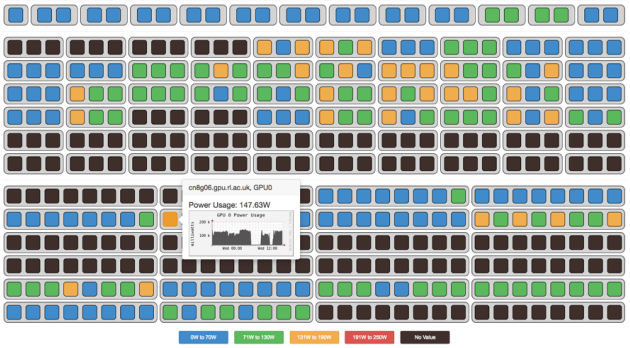


Figure 5 Emerald Cluster Dashboard showing GPU Power consumption, derived from Ganglia data

Author
D. Ross, N. Hill, STFC Rutherford Appleton Laboratory

References

[1] http://xdmod.sourceforge.net/
[2] https://www.ncas.ac.uk/index.php/en/faam
[3] http://www.observium.org/
[4] http://www.cacti.net/
[5] http://www.solarwinds.com/
[6] http://ganglia.info/
[7] https://www.icinga.org/

e-Infrastructure for Data Intensive Science

Credit: © | Dreamstime.com

The Distributed Computing Infrastructure Group was formed in May 2016 – bringing together the teams running grid services for RAL Teir-1 Centre as part of Worldwide LHC Computing Grid (WLCG), the teams developing and running the APEL accounting and the GOCDB information services used by EGI and WLCG, as well as team developing the & supporting the UKCA together and the team developing SCD's cloud platform.

Enabling Science with the SCD Cloud

The SCD Cloud, launched in 2015, provides dynamic resources to the department to support H2020 projects as well as a platform of testing and development. The SCD Cloud also underpins the Data Analysis As A Service (DAAAS) project for ISIS and other STFC facilities. DAAAS will allow Facilities users to access and analyse their data using STFC resources.

The SCD Cloud has had around 160 unique users and around 250 VMs running at any given time and almost 35,000 VMs have been created since launch. Users have come from SCD, ISIS, Diamond, and the STFC Virtualisation Working Group. There is also a pilot project for the Low Frequency Array (LOFAR).

We are close to launching an OpenStack deployment to enable greater flexibility for our use cases than the existing OpenNebula based cloud. Over the coming year we will be migrating existing users to the OpenStack deployment and providing external access to selected projects via native APIs (INDIGO Datacloud) and via the EGI Federated Cloud (West-life).

e-Infrastructure - EGI CernVM-FS Service

The EGI CernVM-FS service run by the Tier-1 team has expanded significantly since its launch in 2013. The CernVM File System (CernVM-FS) [1] is a distributed read-only file system delivering scientific software onto physical and virtual machines in a fast, scalable, and reliable way. It has not only been adopted by the CERN LHC experiments, but also by other scientific user communities.

The service we provide to the EGI community consists of a Stratum-0 server that hosts and publishes over 25 repositories comprising of 9.6 million files, 600GB in size. The repositories are maintained using the CernVM-FS Uploader. This facilitates access, via a GSI interface, to users responsible for updating the repositories. The provisioning of the EGI CernVM-FS service, its daily running and availability are defined within an Operational Level Agreement between EGI.eu [2], and STFC.

Reinforcing STFC's role in the wider community of CernVM-FS users, the second CernVM Users Workshop [3] took place at RAL in June 2016 with more than 40 worldwide participants attending the event. The technical talks and discussions allowed developers, infrastructure maintainers and users to exchange ideas, and setup new collaborations.

International e-Infrastructure Services

The Distributed Computing Infrastructure Group has taken over the development of a number of production services that are integral to several national and international scale compute infrastructures. These services (GOCDB, APEL, UKCA) form part of the core operational services for many compute sites across the UK and EU. The Grid Operations Centre Database (GOCDB) is used to record computing sites, services and other e-infrastructure components that participate in the EGI [2] and/or the Worldwide LHC Computing Grid (WLCG). The Accounting Processor for Event Logs (APEL) is an accounting system and allows individual scientists, site administrators, and VO (Virtual Organisation) managers to monitor compute resources used across large grid infrastructures. The UK eScience Certification Authority service (UKCA) [4] is essential for the UK's participation within multiple e-infrastructure projects, including WLCG and EGI. 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 [5]. This ensures our certificates can be used to authenticate users with a variety of Grid, Data, Compute and Cloud resources around the world.

Computing Apprenticeships in SCD

2016 marks the second year of SCD and ISIS recruiting computing apprentices. After a very successful first year, all three of last year's apprentices have continued into their second year. This year two new apprentices have joined each department and will spend two days a week studying at Abingdon & Witney College, working towards a foundation degree and three embedded in teams at STFC. Apprentices rotate between teams from both departments, completing four three-month placements in their first year and two six-month placements in the second year.

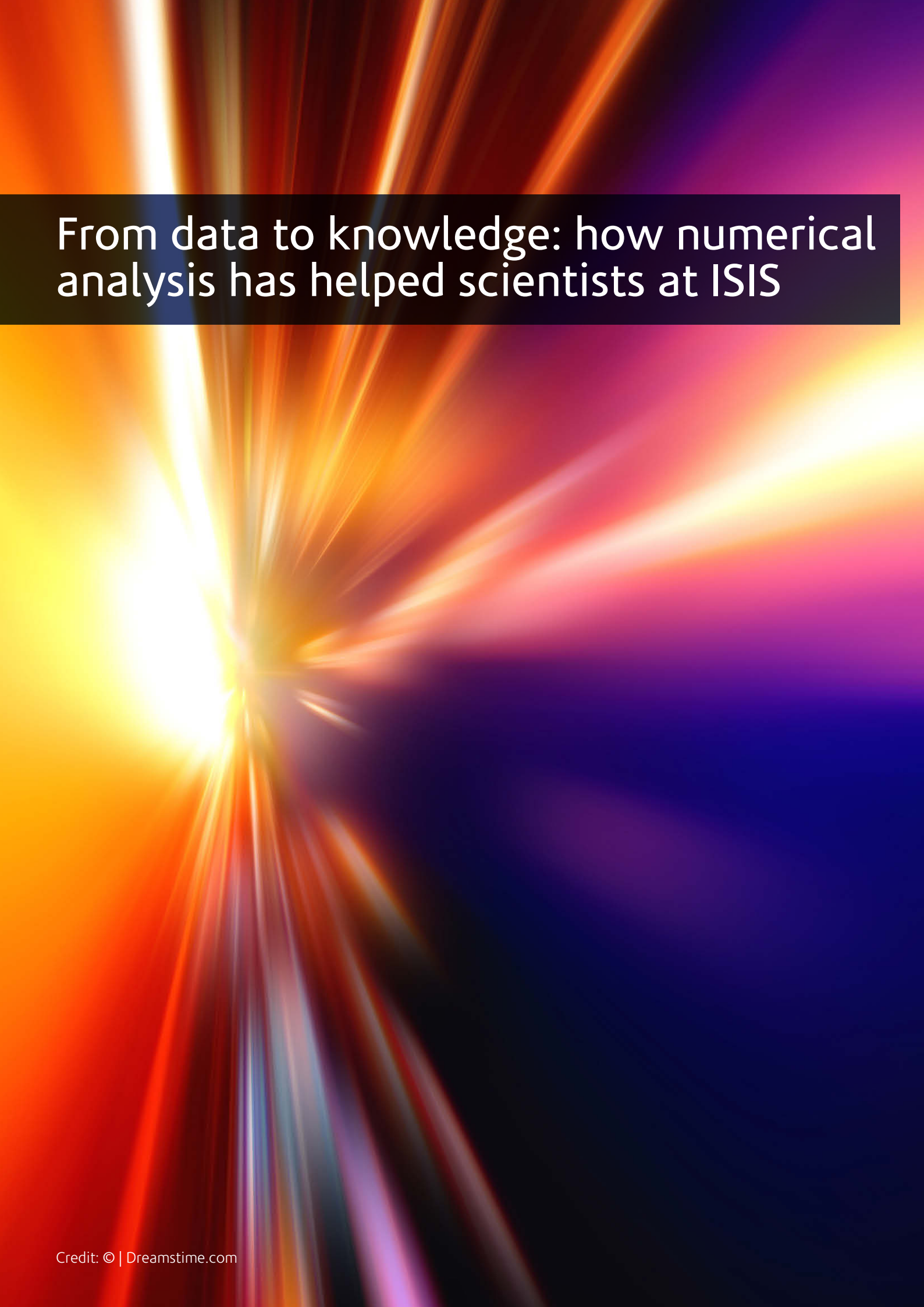
Both the apprentices and STFC get a lot out of the placements, last year's apprentices having contributed to work as diverse as hardware infrastructure, analysis of our storage systems, and cloud deployment.

Author
I. Collier, STFC Rutherford Appleton Laboratory

References

[1] CernVM File System (CernVM-FS) - http://cernvm.cern.ch/portal/filesystem
[2] EGI - http://www.egi.eu/
[3] CernVM Users Workshop - https://indico.cern.ch/event/469775/
[4] UKCA http://www.ngs.ac.uk/ukca)
[5] Interoperable Global Trust Federation - https://www.igtf.net

From data to knowledge: how numerical analysis has helped scientists at ISIS



Credit: © | Dreamstime.com

The scientists at the ISIS facility at RAL rely on fast and accurate data analysis tools to perform world-leading research, and the software package Mantid provides this functionality. They found that for a small, but significant, number of problems, the algorithms they were using are not giving satisfactory results.

A team from SCD's numerical analysis group has helped them by developing a new algorithm, which has proved to be more robust and faster than those previously used. This improved algorithm will be incorporated into the next release of Mantid, and is also available for anyone to use through the software package RALFit.

The ISIS facility at RAL is a world-leading centre for research in fields as diverse as physics, chemistry, materials science, engineering, biology and geology. A synchrotron produces beams of neutrons and muons, which scientists use to study materials at the atomic level. The instruments at ISIS output massive amounts of data. Looking at the noisy raw data alone tells you little, but scientific breakthroughs are made by learning about the atomic structure of the sample, which is revealed by data analysis. However, the scientific results produced are only as good as the tools that are used to deliver them, and to this end a team at ISIS, together with international project partners, have developed widely used data analysis and visualization package Mantid.

To solidify ideas, we consider the VESUVIO beamline at ISIS. This instrument was designed to measure the atomic momentum distribution of atoms in condensed matter systems. The data produced will consist of a number of spectra, which in turn are made of a series of peaks. The mass of an atom within the sample can be inferred from the centre of the peak, the amplitude of the peak is a function of an atom's mass and scattering cross section, and the momentum distribution of the mass can be inferred from the width of the peaks. An example of data from this instrument is given in Figure 1.

In order to get numerical values for the centre, amplitude, and width of the spectra, we must first fit a smooth curve through the noisy data. To do this, the scientist selects a function with these three values as variables, which we call $f(x;t)$, and she seeks to find the values of the variables in x that minimize the difference between the observed value and the smooth value in a least-squares sense. In particular, if (t_i, y_i) are the observations (the data points in Figure 1), then we want find the value of x that minimizes the squares of the errors, namely:

$$\min_{x} \frac{1}{2} \sum_{i=0}^{m} \left(y_i - f(x; t_i)\right)^2 = \min_{x} \frac{1}{2} ||r(x)||^2$$

Mathematically, this is a nonlinear least-squares problem. Under the hood, the default method Mantid uses to solve this problem is the implementation of the Levenberg-Marquardt algorithm found in the widely used GNU scientific library (GSL). This usually works well, but occasionally gives incorrect results.

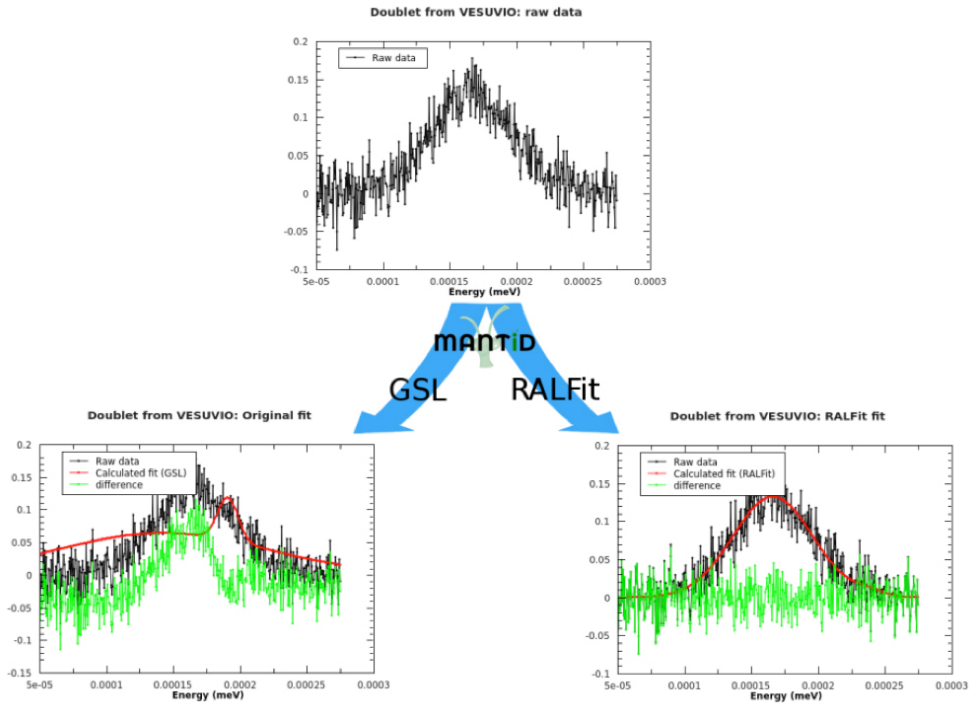


Figure 1: Raw data (top), processed with the default solver from GSL (bottom left), and the new RALFit solver (bottom right)

For example, when we fit the data above, the `best' fit was calculated as being the red curve in the bottom left image of Figure 1, and therefore incorrect values will be recorded for the centre, amplitude, and width of the peak. If this result was included in an automated run, then this error may not have been caught, and incorrect conclusions may have been drawn from the experiment. It is therefore very important that the data analysis tools used robustly give the right answer, and that any failures are flagged to the user. In addition, thousands of such problems may need to be solved for each sample, so we need the fitting routine to be as fast as possible.

The Mantid team wanted to include a more robust solver, and approached SCD's Numerical Analysis group for help. In order to improve the algorithm, we need to understand what it's doing.

The Levenberg-Marquardt algorithm used by GSL is iterative, and from some arbitrary starting point it produces successively better and better approximations to the solution. From any point, x_k, we calculate an update step, s_k, so that $x_k + s_k$ is closer to the solution. The method essentially approximates the residual, $r(x_k+s)$, by its first order Taylor expansion, $r(x_k+s) \cong r(x_k) + J_k s$, where J_k is the Jacobian matrix, i.e., the matrix whose columns are made up of the derivatives of $r(\cdot)$ evaluated at the point x_k. The Levenberg-Marquardt method minimizes this approximation, while also including a regularization term, making the sub-problem solved at each step

$$\boldsymbol{s_k} = \arg\min_{\boldsymbol{s}} \; 0.5||\, \boldsymbol{r(x_k)} \; + \boldsymbol{J_k s}\,||^2 \; + \frac{\sigma}{2}||\, \boldsymbol{s}\,||^2. \qquad (1)$$

The regularization term, σ, grows or shrinks as the algorithm progresses in line with the calculated accuracy of the approximation.

The first improvement is made by observing that the Levenberg-Marquardt sub-problem is, in certain cases, mathematically equivalent to finding the minimum within a region of the current point; instead of (1), we solve the problem

$$\boldsymbol{s_k} = \arg\min_{\boldsymbol{s}} 0.5||\, \boldsymbol{r(x_k)} \; + \boldsymbol{J_k s}\,||^2 \;\; s.t. \;\; \min_{\boldsymbol{s}} ||\, \boldsymbol{s}\,|| \;\leq \Delta, \qquad (2)$$

again growing or shrinking Δ in line with the quality of the solution.

To illustrate this idea, we look at a problem in one dimension in Figures 2 and 3; in each case, suppose we're at the red point, and we want to find the minimum of the blue function. This problem is hard, so we approximate the blue curve by the red curve, which is easier to minimize. Since the red curve only matches the blue curve close to the current point, we find the minimum within some region of x_k (shaded in pink). Usually we will have the situation in Figure 2, where the next iterate falls on the boundary of the trust region. In this case there is a direct equivalence between trust region methods and the regularization (in that, for each σ in (1), there is a corresponding Δ in (2) that will give the same answer). However, the minimum value of the model (the red curve) sometimes lies within the trust region when then solution is nearby – this is the case in Figure 3. This would be equivalent to setting a regularization parameter of zero in (1), but there is no mechanism to find when this would be appropriate in the standard Levenberg-Marquardt algorithm. Solving sub-problem (2) will therefore, in general, converge to the solution in fewer iterations than solving the regularized problem (1).

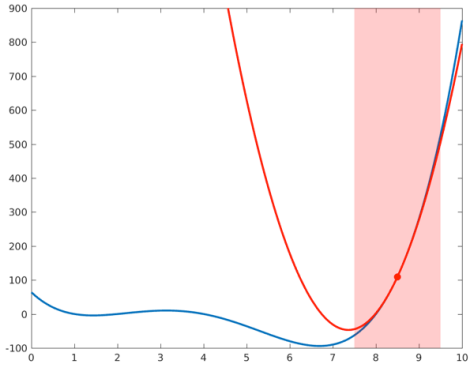


Figure 2: A trust region problem with solution on the boundary

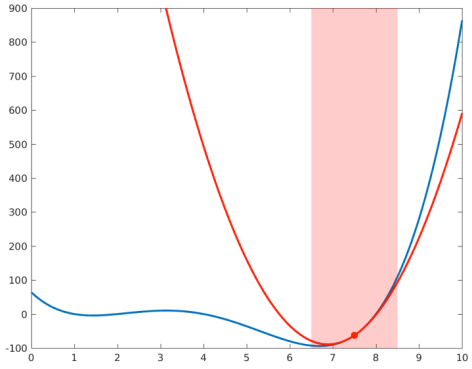


Figure 3: A trust region problem with solution within the region

A further improvement can often be made by using a better local model of the function (in our toy 1D example in Figures 2 and 3, this corresponds to choosing the red curve so that it better approximates the blue curve locally). Recall that the Levenberg-Marquardt method uses a first order Taylor approximation to the function $r(x)$, and this approximates the function we want to minimize by

$$F(\boldsymbol{x}_k + \boldsymbol{s}) \cong m_{gn}(\boldsymbol{s}) := 0.5||\, \boldsymbol{r}(\boldsymbol{x}_k) + \boldsymbol{J}_k\boldsymbol{s} \,||\text{^}2 = F(\boldsymbol{x}_k) + \boldsymbol{g}^T\boldsymbol{s} + \boldsymbol{s}^T\boldsymbol{J}_k{}^T\boldsymbol{J}_k\boldsymbol{s}$$

where $g=J_k x_k$ is the gradient. This is a second order approximation of the function we want to minimize (known as the Gauss-Newton approximation). If we allow ourselves to use second derivatives, the second order Taylor expansion of $F(\cdot)$ around x_k is

$$F(\boldsymbol{x}_k + \boldsymbol{s}) \cong m_n(\boldsymbol{s}) := F(\boldsymbol{x}_k) + \boldsymbol{g}^T\boldsymbol{s} + \boldsymbol{s}^T\boldsymbol{J}_k{}^T\boldsymbol{J}_k\boldsymbol{s} + \boldsymbol{s}^T \sum_{i=1}^{m} \boldsymbol{r}_i(\boldsymbol{x})\nabla^2 \boldsymbol{r}_i(\boldsymbol{x})\boldsymbol{s}$$

This is referred to as the Newton approximation. Note that, for the fitting problems we're interested in, if the data was not noisy then close to the solution we could expect $r_i(x)$ to be close to zero, and therefore the extra term added to $m_n(s)$ would change the value very little. However, if the data has significant noise, ignoring the second order derivatives (as done in Levenberg-Marquardt) will result in a worse approximation to the function close to the solution. If the second derivatives are not available, they can be approximated algorithmically, which will give better performance than if we use $m_{gn}(s)$ alone.

In practice, combining the two approximations, $m_n(s)$ and $m_{gn}(s)$, gives the most robust algorithm. Far from the solution $m_{gn}(s)$ performs best, as the subproblem has the same sum-of-squares structure as the true problem we wish to solve. Close to the solution it is beneficial to include the second derivative information. Our algorithm tracks progress towards the solution, and switches between $m_n(s)$ and $m_{gn}(s)$ as appropriate. Using these algorithmic improvements, we have produced a nonlinear least-squares solver, RALFit, which, in our tests, has proved to be more robust than the solver in GSL. When run on the problematic data from VESUVIO, we get the better fit given at the bottom right of Figure 1. We have also run tests on a standard set of problems (which includes all the test examples from the US National Institute of Standards), and here we were able to successfully solve more problems than the state-of-the-art GSL code currently used in ISIS.

Furthermore, our code is more efficient than the code currently used by Mantid – Figure 2 shows a performance profile comparing the number of function evaluations taken by GSL, our trust region Gauss-Newton method, and our default Hybrid method. Performance profiles are used to compare different solvers on a set of test problems: the y-axis shows the fraction of problems for which a given solver is x% of the best, and therefore the solver nearest to the top left corner performs the best, on average.

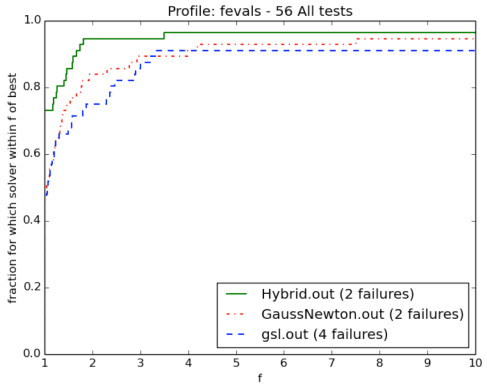


Figure 4: Performance profile

Our method is being incorporated into the forthcoming release of Mantid, and has also been used by researchers at the Diamond facility. Our code, RALFit, is written in Fortran, and C and Python interfaces are also available. It is open source and available to all (download it at https://github.com/ralna/RALFit), and some exciting new features will be added in the coming months.

Author
T. Rees, STFC Rutherford Appleton Laboratory

References

[1] VESUVIO Data Reduction and Analysis in Mantid, S. Jackson, RAL-TR-2014-009
[2] Trust region methods. A.R. Conn, N.I.M. Gould, P.L. Toint. SIAM, 2000.
[3] https://github.com/ralna/RALFit


Software Engineering Group

It has been a productive year for the Software Engineering Group and this report highlights three areas.

Launch of SESC Build Service: providing testing services for CCPForge Users

The Software Engineering Support Centre, is funded by EPSRC, to provide tools, expertise and support to the academic community with a focus on the CCPs. The latest tool launched in 2016, is the SESC Build Service which uses Jenkins to provide continuous integration and testing services to CCPForge. It has had a steady take-up from CCPForge users. Current developments include the widening of the tools available and the decoupling of the authentication and authorisation from CCPForge so that other code repository systems can be used.

Providing Training to CCP-WSI Developers

The CCP-WSI is addressing the research area of Wave Strucuture Interactions, and was formed in 2015. SEG is a partner in the CCP and leads the wokd on software engineering support and training. Gemma Poulter, together with Alan Kyffin and Steve Lamerton, ran a successful workshop for the devlopers in CCP-WSI setting standards and expectations for the code quality for the project, this includes using the SESC Build Service.

Identifying research software, persistently

The Software Reuse, Repurposing and Reproducibility project has not changed the world, but it has changed the DataCite schema! In our Guidelines for Persistently Identifying Software, we recommended a new DescriptionType property should include the value of "TechnicalInfo" to encourage the inclusion of more explicit technical information about software, when DataCite DOIs are generated. This change is now adopted and will be incorporated in the forthcoming Version 4. It will aid in the discovery, reuse and repurposing of software through the clear signposting of the technical content.

Software developers, researchers and digital preservation specialists all have a part to play in ensuring software is reusable, good quality and findable. Throughout 2016 Catherine has been out and about talking about the software identification issues and raising awareness of potential benefits at meetings and conferences that involved representatives of all three stakeholder groups. These include: Collaboration Computational Projects Steering Panel, Software Sustainability Institute's Collaborations Workshop, and the Reproducibility symposium at the Alan Turing Institute, the International Digital Curation Conference, ESIP Data Stewardship Committee (USA), the Repository Fringe and Research Software Engineers Conference. This work was also discussed at the Knowledge Exchange workshop on software sustainability and we have been included in the report.

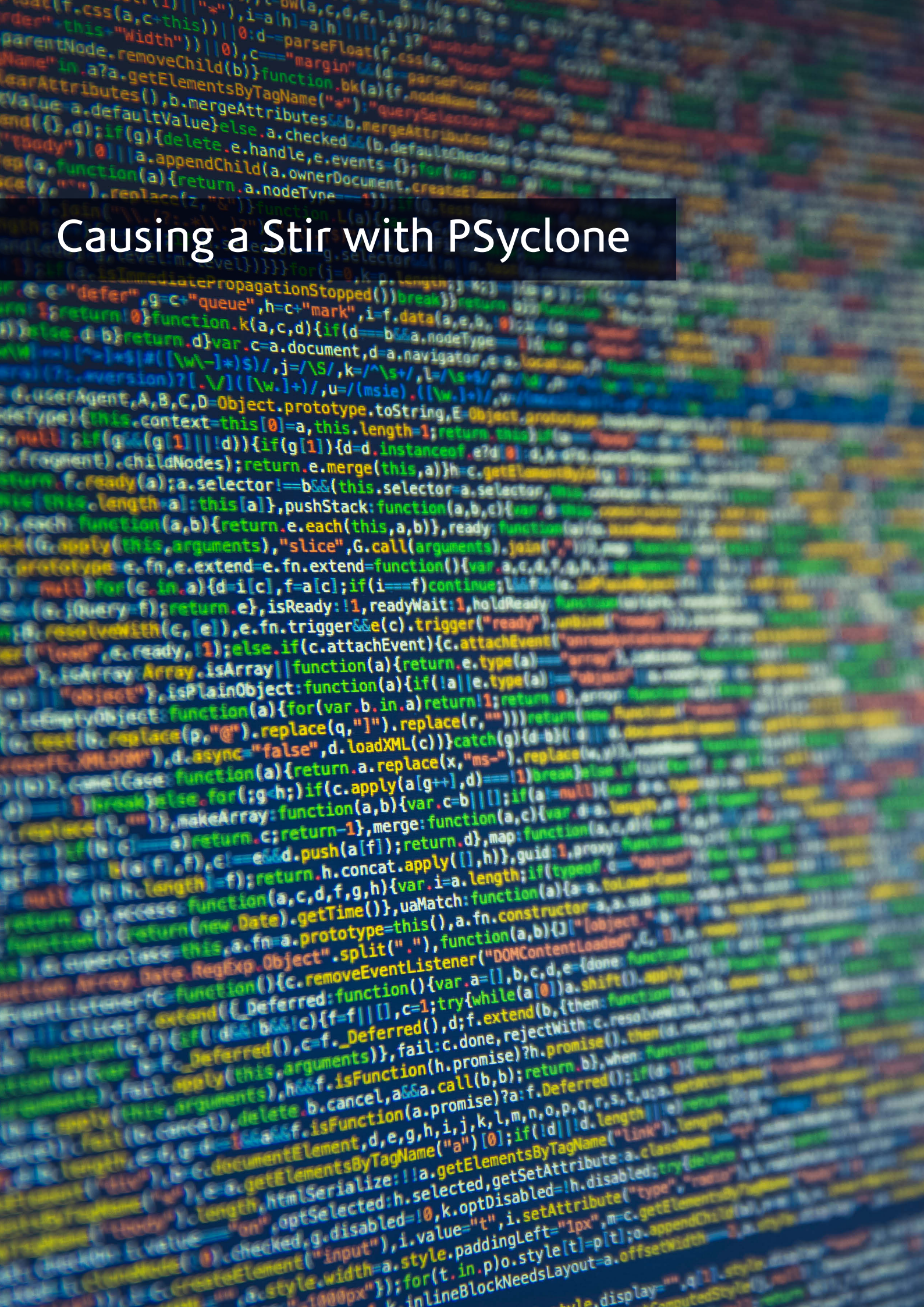
The Software Reuse, Repurposing and Reproducibility project was funded by Jisc's research data spring initiative from April to November 2015. It was led by Ian Gent, St Andrews and Catherine Jones, STFC; joined by Jonathan Tedds, University of Leicester in the second phase.

Organisational change

In June 2016 the Software Engineering Group merged with the Research Data Group to form an enlarged Software Engineering Group with three main areas: software engineering support and expertise for computational scientists, data management tools and pipelines for experimentalists and research output recording systems for STFC staff.

Author
C. Jones, STFC Rutherford Appleton Laboratory

Causing a Stir with PSyclone



When the development of a new, performance-critical application can take the best part of a decade, how do you ensure that it will run well on the hardware that is available when it goes into production and for twenty years following that? This is the challenge faced by the UK Met Office as it develops its new LFRic Atmosphere Model; the successor to the Unified Model that has been in production for some 20 years. This task would be difficult at any time, but the push towards Exascale computing is resulting in significant proliferation of competing technologies; GPUs, Intel's Xeon Phi, increasingly many-core CPUs and even FPGAs. A novel approach is being taken by the Met Office in collaboration with STFC as it tackles this problem. At the heart of the approach is the "PSyKAl" separation of concerns whereby scientific code is kept completely separate from those aspects that deal with parallel performance. A domain scientist writes the Algorithm that describes the sequence of computations and the Kernels that perform the necessary computations on a single column of data. The middle "PSy" layer then handles the business of distributing work over whatever resources are available. STFC are developing PSyclone, a domain-specific compiler, which processes the Algorithm and Kernels and then generates a suitable PSy layer for the hardware being targeted. This approach has enabled the developing LFRic code to go from being completely serial to something that runs in MPI/OpenMP hybrid mode on 50,000 cores with no change to the scientific code base - all that is required is a flag to the PSyclone compiler for MPI and a small script for OpenMP, both of which are invoked during the build process.

The LFRic Project

The Met Office's numerical weather prediction and climate model code, the Unified Model (UM), is over 25 years old. Until recently the UM has been able to run efficiently on many of the worlds most powerful computers, helping to keep the Met Office at the forefront of climate prediction and weather forecasting.

However, performance increases from each new generation of computers are now being primarily provided by an increase in the amount of parallelism (e.g. the number of cores in a processor) rather than an increase in the clock-speed of the processors themselves. This means that in order to run higher resolutions of the UM, the Met Office now faces the double challenge of code scalability (making use of additionaly parallelism) and numerical accuracy.

To address this challenge the Met Office, NERC and STFC initiated the GungHo project. The primary aim of this project was to design a new prototype scalable, numerically-accurate dynamical core (the heart of any weather prediction/climate model). This project completed earlier this year and the Met Office have set up a related project called LFRic which is building on the recommendations of GungHo in order to design and implement a full atmosphere model. This atmosphere model is scheduled to become operational around 2024. STFC are continuing to collaborate with the Met Office on the LFRic project via funding from the Hartree Centre.

PSyKAl

In the GungHo project a novel separation of concerns for the software implementation of the dynamical core was developed. This approach distinguishes between three layers: the Algorithm layer, the Kernel layer and the Parallelisation System (PSy) layer. Together this separation is termed PSyKAl.

The Algorithm layer specifies the algorithm that the scientist would like to run (in terms of calls to kernel and infrastructure routines) and logically operates on full fields.

The Kernel layer provides the implementation of the code kernels as subroutines. These subroutines operate on local fields (a set of elements, a vertical column, or a set of vertical columns, depending on the kernel).

The PSy layer sits in-between the algorithm and kernel layers and its primary role is to provide parallel performance for the target architecture. The PSy layer can be optimised for a particular hardware architecture, such as multi-core, many-core, GPU's, or some combination thereof with no change to the algorithm or kernel layer code. This approach therefore offers the potential for portable performance and single-source science code.

PSyclone

Rather than writing the PSy layer manually, STFC are developing a code generation system, called PSyclone, which can help a user to optimise the code for a particular architecture (by providing optimisations such as blocking, loop fusion, inlining etc).

In Autumn 2015 PSyclone was integrated into the LFRic build system. This meant that PSyclone generated a sequential version of the PSy-layer for the LFRic model, rather than the PSy-layer being written manually. In Spring 2016 the LFRic model went parallel, using both MPI and OpenMP, and was tested on over 15,000 cores. More recently tests have been performed on over 50,000 cores.

The switch from sequential code to parallel code took a week, and would have been much quicker had there not been a bug in the OpenMP code generation. Most importantly the science code did not change from the serial to the parallel version of the LFRic model.

From a users perspective the difference in running with or without MPI is a PSyclone command-line flag and the difference in running with or without OpenMP parallelisation is the addition of a short script which performs loop colouring on any loops that require it (loops which contain writes to a continuous function space) and then adds OpenMP parallel directives on any loops that are parallel.

Whilst PSyclone is designed to allow a code to be tuned for a particular architecture and compiler, at the time of writing little performance optimisation has been attempted for the LFRic model. The MPI and OpenMP implementations are currently very simple and generic. Most work thus far has concentrated on supporting new functionality, the latest functionality being developed is for so-called builtin's. Builtin's are Kernels which do not need to be written by the user as they are already supported by the system. Simple linear algebra routines are one example.

However, performance optimisation is now starting to be examined, primarily due to funding from STFC's Intel Parallel Computing Centre. In this case the OpenMP scalability of the model is being tested and improved on Intel architectures, particularly the Xeon Phi. This work is proving to be very useful in finding performance issues in the LFRic infrastructure and in the Kernels themselves, as well as providing feedback on the code PSyclone generates.

At the time of writing both PSyclone and the LFRic model are scheduled to be made open source by the end of the year. Once this has been agreed, PSyclone will be migrated from the Met Office repository to a more open repository, probably github.

GOcean Finite Difference API

PSyclone has been written so that it is able to support more than one API. In the case of the LFRic model, the API is for a mixed finite element scheme with an associated data model of k (the vertical) contiguous in memory and the horizontal (latitude and longitude) indirectly addressed.

A second API has been developed for finite difference codes which has a directly addressed data model. This API is currently two dimensional, and MPI parallel support is not yet implemented.

A small amount of funding has been secured in the European ISENES2 project which will examine how to best extend this finite difference API to 3D. This work will concentrate on a kernel taken from the NEMO ocean model in collaboration with the Euro-Mediterranean Centre on Climate Change in Italy.

While the finite difference API has some limitations, it has been used to test the performance of the PSyclone approach. In the first case a hand-optimised benchmark was chosen. It was shown that PSyclone would be able to get back to the performance of the original, optimised code on a range of compilers and architectures, despite splitting the code into the separate layers described above.

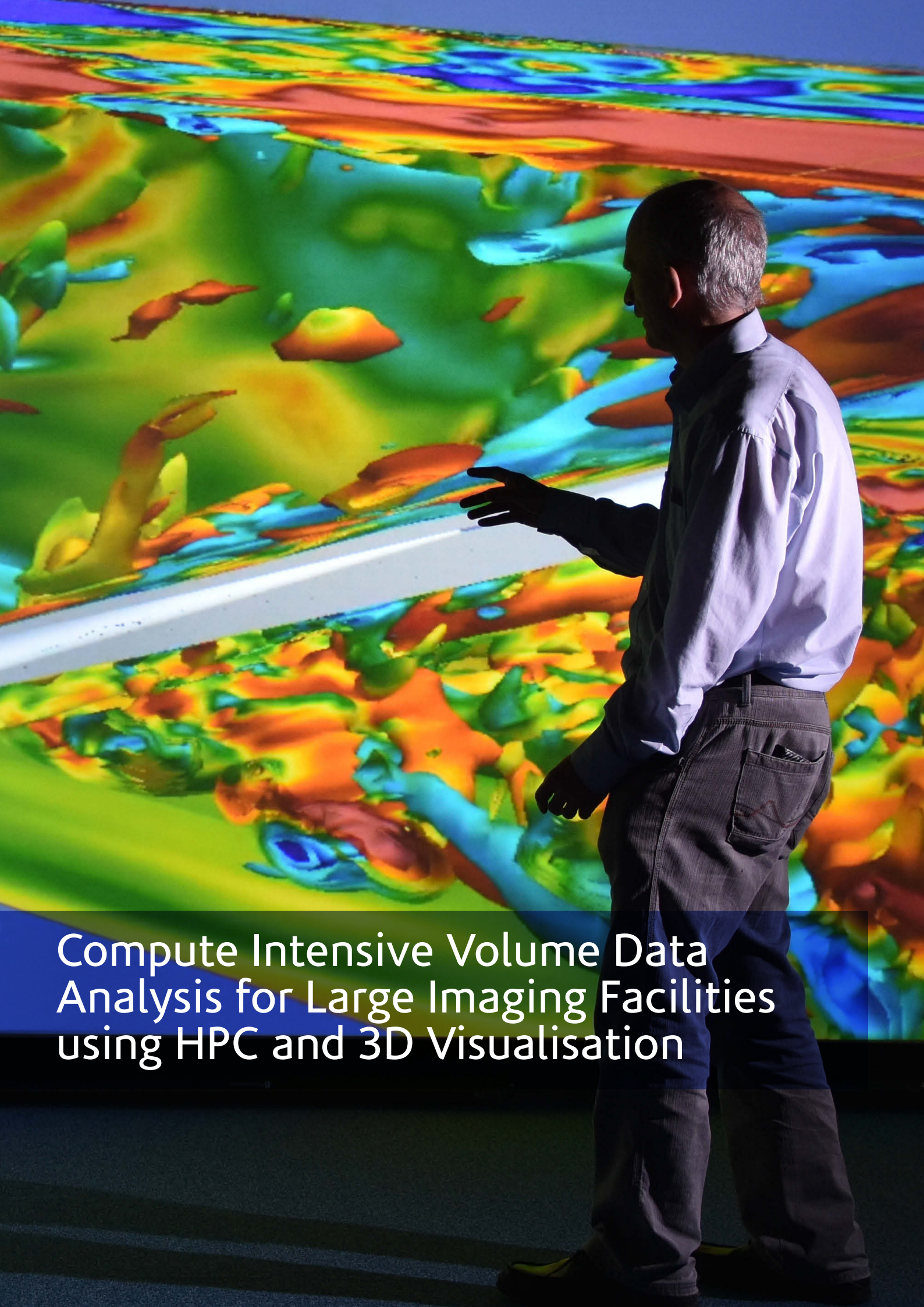
However, it was found that one architecture/compiler combination required a different hand-optimised structure to obtain the best performance. PSyclone would be able to do this with no change to the scientific source code as optimisations are applied in a script as a recipe and can be tailored for the particular architecture/compiler combination. For the hand-tuned version, two implementations would need to be supported.

In a separate benchmark, NVIDIA were asked to optimise the benchmark for GPU's. It was shown that the code could run efficiently on multi-core CPU's using OpenMP and on GPU's using OpenACC with no change to the scientific source code.

Summary

STFC are collaborating with the Met Office on their next generation Atmosphere model called LFRic. This work is funded by the Hartree Centre. The PSyclone software being developed by STFC is in use by the Met Office for their LFRic model and is integrated within their build system. The model is run in hybrid MPI/OpenMP mode using PSyclone and has been tested on over 50,000 cores. PSyclone has allowed the Met Office to move from sequential code to parallel code with no change to the underlying science code. To date little attention has been paid to the performance of the model as most effort has been directed at functionality. Funding from STFC's Intel Parallel Computing Centre is starting to change this and has already helped to improve the OpenMP scaling of the model and its performance on Intel architectures. PSyclone's GOcean API is currently not fully functional but has been used to demonstrate a case where it is not possible to have optimised single source code. The GOcean API has also been used to demonstrate the ability of PSyclone to target GPU's as well as CPU's, again with no change to the science source code.

Author
R. Ford, A. Porter, K. Sivalingam, STFC Daresbury Laboratory



Compute Intensive Volume Data Analysis for Large Imaging Facilities using HPC and 3D Visualisation

This article gives a brief introduction to the production of a video for a specific science facility located at the Rutherford Appleton Laboratory in the UK, namely the ISIS neutron source and the HPC data centre operated by the Scientific Computing Department. It illustrates the setup of a neutron imaging instrument, named IMAT (Imaging and Materials Science & Engineering) and the basic principle of an energy selective neutron imaging experiment. Particularly, the video highlights the compute intensive nature of this type of experiment and why that has motivated the use of HPC to accelerate image reconstruction process during the experiment. Finally, the video briefly demonstrates the use of volume segmentation, quantification, and visualisation techniques as an integral part of the experiment lifecycle to help deriving and determining optimal experimental strategies in the laboratory.

Background

Rapid advances in imaging detectors with over 10MP sensors have been incorporated within large science laboratories allow us to image in high-quality materials at the micro- (10^{-6}m) or nano- (10^{-9}m) scales; seeing new phenomena including cracks and deformation. A further level of complexity has recently been added that allows energy selective imaging to be performed at each pixel that is to be offered by the IMAT (Imaging and Materials Science & Engineering)[1]. instrument, currently being commissioned at the ISIS pulsed neutron source[2] in the UK.

ULTRA a compute intensive HPC platform is being built to enable high-throughput neutron tomographic image data analysis so that images can be scrutinised during an experiment rather than as a batch-mode post-process operation.

We describe the HPC-based analysis and visualisation technologies employed to enable this new mode of operating with energy selective neutron tomographic imaging. Figure 1 depicts the basic setup of an energy selective tomography experiment (left image), where a sample sits on a rotation table in between the neutron beam origin, and the detector. The right image illustrates the massive volume of 3D image stacks produced – as each energy level and combinations of energy levels can create a full 3D volume set.

Neutron Imaging Driven Scientific Discovery

Unlike normal computed tomography (CT) scans used in hospitals where one 2D image is acquired for each 'shot' (a rotation angle), in energy selective neutron imaging, an image stack comprising of potentially thousands of 2D images are collected at each 'shot'. Figure 2 illustration the data volume for the MCP camera[3], capable of collecting 3,000 images per angle where each image uniquely corresponds to one of the 3,000 energy bands; resulting in 0.3 million images during a 100 angle experiment.

The reason we do this is that neutron interactions

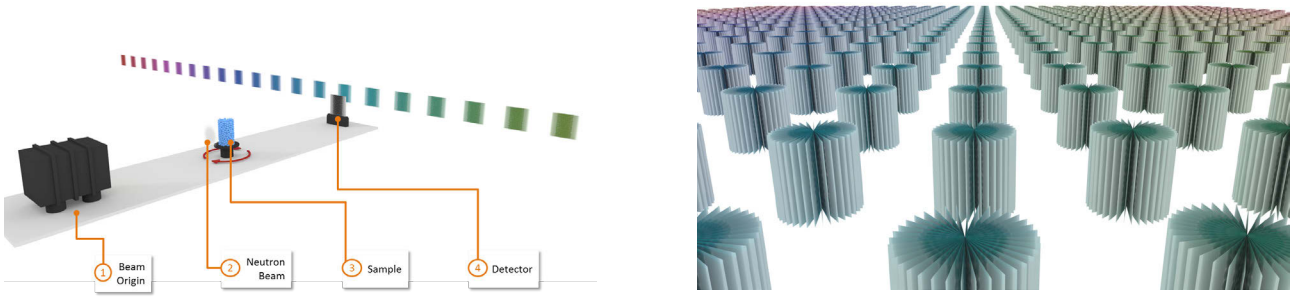


Figure 1: Left - Energy Selective Tomography Imaging (thousands of images per angle); Right: thousands of 3D image stacks, one per energy band

[1] www.isis.stfc.ac.uk/Instruments/Imat/

[2] www.isis.stfc.ac.uk

[3] www.isis.stfc.ac.uk/instruments/imat/technical/imat-imaging-and-materials-instrument8261.html

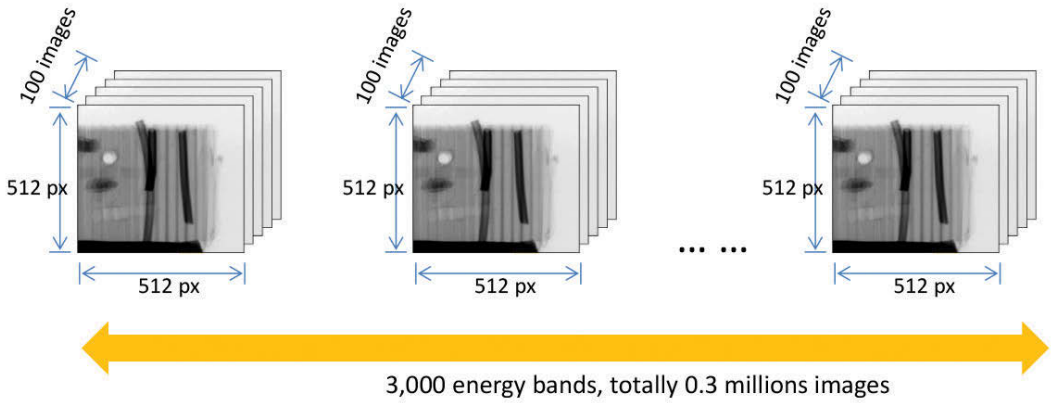


Figure 2: An Illustration of the Data Volume of Energy Selective Neutron Imaging (512px X 512px image, 100 angles – 3,000 image per angle, each image uniquely corresponding to one of 3,000 energy bands, IMAT MCP detector, totally 0.3 million images)

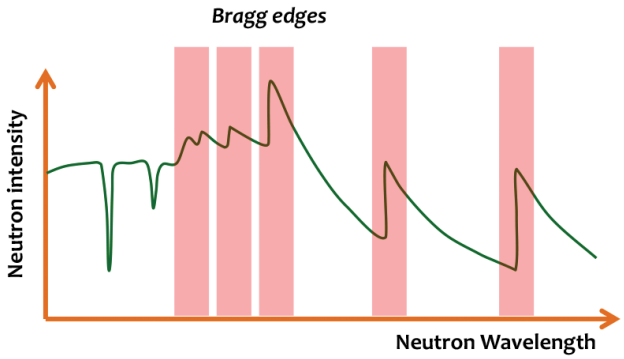


Figure 3 : Exhibition of Bragg edges in the presence of different energy wavelengths [1]

can vary drastically with neutron energy for certain materials, allowing for chemistry discriminations [2]. The amount of neutrons that is able to penetrate through a material and reach the image detectors, namely, neutron intensity, is strongly affected by the crystalline structures and microstructures of a material, exhibiting Bragg edges [1], as illustrated in Figure 3. Bragg edges show crystal structures, i.e. how atoms are arranged, thus giving information about texture, phase composition, and even strain within a sample. As illustrated in Figure 3, each Bragg peak highlighted represents a potential energy band region suitable for a reconstruction.

This information is critical in practice for many real world applications [1], including engineering, material science, geology, archaeology, and physics. For example, accurate measurements of strains inside an aircraft wing acquired under in-service conditions (e.g. high temperature & magnetic fields) are used to determine the structural integrity of the component, relevant to the type and composition of materials as well as its structural parameters (thickness, length, density) in the design. Equipping with such information, engineers are able to make better design decisions needed to balance the competing needs for safety, costs, and performance.

Image Reconstruction Using HPC

A crucial step to enable scientists to analyse energy selective image data is called tomographic image reconstruction where 2D images (radiographs) collected are used to reconstruct 3D volumetric data -- a digital replica of the sample. Scientists use the Bragg edges to guide the reconstruction process. In other words, reconstructions are performed on selective energy bands that contain interesting features; whilst at the same time, consider ways to improve the information extracted from the data, for example, to minimise image artefacts or enhance contrast between different materials within a sample. In practice, this is a complex and compute intensive process demanding significant processing power, which is not typically available on standard workstations.

Traditionally, a typical 3D reconstruction takes minutes,using the Filtered Back Projection algorithm (FBP) [3], one of the most common and fast

Figure 4: Top - thousands of image stacks ready for reconstructions based on Bragg edges using HPC; bottom: reconstructed 3D images are produced using HPC and become available to users when they are ready

algorithms. However, in energy selective imaging, the reconstructions need to be performed repetitively across selected energy ranges and as signal to noise levels are lower iterative algorithms, which are much slower than the FBP algorithm and take 100s of minutes to run, are often used.

HPC techniques are essential to speed up the process to make it useful and parallel reconstructions are possible. A range of noise and image artefact reduction and image filter algorithms are applied during the pre-processing phase of the image reconstruction also exploiting the parallism available within the HPC cluster.

Volume Image Analysis and Visualisation

In this example, we use the SophiaBeads dataset [4], a microCT dataset, to illustrate the process of reconstruction, segmentation, quantification, and visualisation. This dataset has been collected within the STFC supported research network and is an open dataset available to the public for research purpose. This dataset was reconstructed using the FBP algorithm from the TomoPy image reconstruction toolkit [6] running on a single node with 128GB RAM, 12 CPU cores using the STFC's large HPC cluster named SCARF [4]. The reconstructed volume data is then segmented using the algorithms available in a commercial software package named Avizo [5], which is popular among material science community for interactive analysis and visualisation. Simple thresholding, colormapping, and intensity range techniques have been applied to segment and isolate the beads in the volume data. Figure 5 shows the reconstructed volume and a segmented volume with colour labels corresponds to a distinct type of material. For further details of segmentation and quantification steps, readers can refer to [4], which is written by material scientists at the University of Manchester, describing in-depth insights into the rationale for using different visualisation techniques for the dataset.

[4] http://www.scarf.rl.ac.uk/

The insights obtained through this analysis process is then used to steer the next experiment step, for example, to adjust sample positions and beam alignment, or to decide whether to use different reconstruction algorithms or parameters, or image filters.

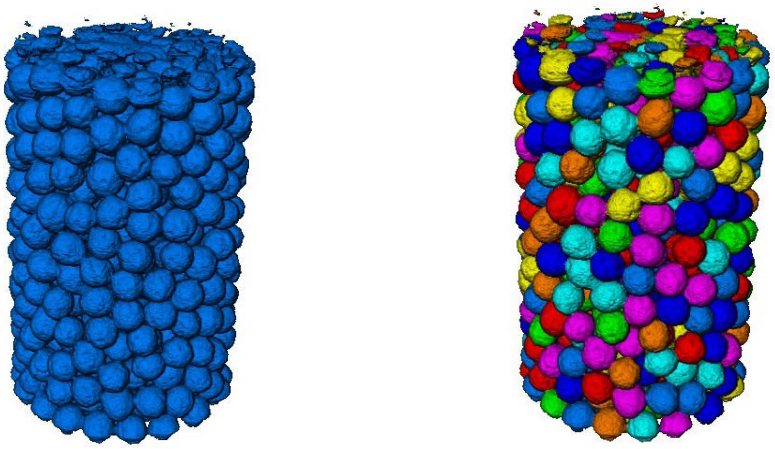


Figure 5: Left: Reconstructed 3D Volume; Right - Segmented 3D Volume Images with Isolated Beads in Different Colours/Materials

Authors
E. Yang, S. Nagella, STFC Rutherford Appleton Laboratory. M. Turner, STFC Daresbury Laboratory

Thanks also to: Winfried Kockelmann, Joe Kelleher, Genoveva Burca, Triestino Minniti - ISIS Neutron Facility - Rutherford Appleton Laboratory (RAL), Science and Technology Facilities Council.

References

[1] T. Minniti et. al., "Material analysis opportunities on the new neutron imaging facility IMAT@ISIS", Journal of Instrumentation, Volume 11, March 2016, IOP Publishing Ltd and Sissa Medialab srl.
[2] J. Santisteban et. al., "Time-of-Flight neutron transmission diffraction", J. Appl. Cryst. 34 (2001) 289.
[3] Peter Toft, "The Radon Transform - Theory and Implementation", Ph.D. thesis. Department of Mathematical Modelling, Technical University of Denmark, June 1996.
[4] Sophia Bethany Coban, "SophiaBeads Datasets Project Documentation and Tutorials", April 2015, MIMS EPrint: 2015.26.
[5] Avizo 9. "Avizo User's Guide", FEI Visualisation Sciences Group.
[6] Do a Gürsoy, Francesco De Carlo, Xianghui Xiao, and Chris Jacobsen, "TomoPy: a framework for the analysis of synchrotron tomographic data", J Synchrotron Radiat. 2014 Sep 1; 21(Pt 5): 1188–1193. DOI: 10.1107/S1600577514013939.

The SCD Facilities Programme: a Year in Transition

The science undertaken within STFC's funded facilities, such as the ISIS Neutron and Muon Source, the Diamond Light Source (DLS) and the Central Laser Facility (CLF), is becoming ever more data intensive, with more and more complex experiments being undertaken, and with ever more sophisticated instruments which take more and more varied data. As these needs grow, the volume and nature of the support needed from all aspects of scientific computing becomes ever greater for the potential from these experiments to be realised. In response to this trend, SCD has been changing the focus of its work with facilities, opening up new areas, whilst still maintaining its core services in supporting data archiving and computational clusters. As a consequence, the last year has seen significant changes to the programme which SCD undertakes with the STFC facilities, with a number of new projects getting underway, and with increased resourcing from the ISIS facility.

Two areas in particular have seen a significant expansion: Data Analysis as a Service, and the development of new algorithms and applications for scientific modelling and data analysis.

As experimental data volumes and complexity increase, the computational demands on interpreting the data increases, both in the scale of computing hardware needed, and in the sophistication of algorithms. SCD has significant computing resources, such as the SCARF multi-CPU cluster, the SCD Cloud service and the Petabyte data-store, which could make it considerably easier for users to process their data. Consequently, SCD has started a number of projects building tools and services to make it easier for users to access and use the SCD compute resources.

Users may typically want to access data analysis tools in two ways. During their experiment, they may want an on-demand analysis of data as it is collected, to give an indication of the state of the experiment, which can then be adjusted to collect better data. An example of this is the ULTRA project, which is building image reconstruction services on the SCARF cluster, accessible to users of the new IMAT neutron imaging instrument at ISIS. This allows users to reserve resources on SCARF for the time of their experiment. Data is then passed over to SCARF, reconstructed and passed back for visualisation, all controlled via ISIS's MANTID analysis tool. The second way of accessing the system is after the experiment, when users could access SCD resources remotely to reprocess the data kept in SCD's archives. The CCP4-DAaaS project is looking at supporting the CCP4 suite of macro-crystallography tools within the SCD cloud so DLS users and others can access these tools in a managed, tailored environment.

These projects are contributing to building "Data Analysis as a Service" (DAaaS) – a common platform providing data analysis service on the SCD infrastructure, which can provide a tailored scientific environment for each science community. The overall concept of the DAaaS is being developed in the ISIS supported IDAaaS project, which is building generic tools and services so different user communities and science domains can be supported within the same managed environment.

The second area which has been launched the last year has been the development of new modelling software and new analysis algorithms. SCD has extensive expertise in these areas, which has not been fully utilised. These pilot projects are changing how SCD can support facilities.

SCD has expertise in computational science, particular materials modelling, meso-scale chemistry, bio-chemical systems and engineering systems. Pilot projects to date in the facilities programme have largely concentrated on materials science, with projects on: calculating spin excitation energies for magnetic states, comparing first principles models with experiments performed at ISIS; looking for Muon stopping sites by ab initio random structure search; and ab initio molecular dynamics simulations to study the properties of organic or organometallic molecules in solution in the presence of laser light. A further project, the Materials Workbench, is looking at integrating different materials modelling tools, such CASTEP and CRYSTAL, within one environment so that a user can use these tools together, and ultimately in combination with the analysis of experimental data.

Other areas of activity are building on SCD's expertise in mathematical algorithms. For example, least-squares analysis is an important algorithm used to fit experimental data to functions built into the Mantid data analysis framework developed within ISIS. The NAG group is working with the Mantid team to improve the efficiency and applicability of these algorithms with Mantid, bringing their latest expertise to the problem. Another example is the use of image analysis algorithms to support imaging techniques in ISIS, DLS and CLF; the Visualisation group is bringing new image

analytics techniques to bear on these increasingly data rich experiments.

These new projects do not mean that the more established areas of work with the facilities have been standing still. A major effort in this year was to upgrade the systems which support data archiving for DLS. The volumes of data which DLS generate are constantly growing, with a total archive of over 5PB collected since DLS opened in 2006 – with 4PB collected in the past 3 years! An individual data set from a single experiment can collect many TB of data. The size of data means it is not easy for DLS users to pull back to their home institution. Consequently, a major effort was undertaken to upgrade the system which support users to access and download data from the DLS Data Archive, with a new data system based around the ICAT Data Service, and downloading via the Globus Online fast data transport tool introduced. This was followed by a new, easier to use data browsing interface. As a consequence the system is more responsive, reliable and easier to use.

Another area which has also seen an expansion in use is the data archiving for High Powered Lasers. This has expanded in both its functionality, with experimental data now being archived by integrating the SCD system with the GARB tool, from the University of Strathclyde, and also in its coverage; SCD's data service now works with all three main high powered lasers in CLF – Astra-Gemini, Vulcan and Artemis.

In this article, I have shown the range of the new projects and activities now being undertaken by SCD to support facilities; there are further projects which I have not had space to cover, such the CEPH Data Store and the Data Linking project which will bring publications and data together. 2015-16 has been a year of transition within the facilities programme, with much of this new programme starting, which will become established in subsequent years. Nevertheless, the breadth of the new programme shows the potential of Scientific Computing to address the challenge of a data intensive future for facilities science.

Author
B, Matthews, STFC Rutherford Appleton Laboratory

Daresbury Open Week

During 5th-9th July 2016, the SciTech Daresbury Campus opened its doors to potential customers, possible future recruits, and the general public in the first ever Daresbury Open Week (DOW). The idea was to showcase the wide range of science- and technology-related activities that happen on the campus, normally out of view. The programme included:

- A full day of activities for corporate visitors, including the opportunity to take a ride in a Tesla model S P90D electric car.
- Two full days of schools visits at key stage 2/3, and key stage 4/5. Students had the opportunity to participate in talks, tours and interactive workshops with the aim of inspiring them to take up careers in Science, Technology, Engineering or Mathematics (STEM) occupations.
- A public day on Saturday 9th July which gave open access to visitors to all areas of site. Many exciting activities were provided for all age ranges, including visits to the viewing area at the top of Daresbury Tower; tours of the ALICE/EMMA particle accelerator facility; and visits to the Engineering Technology Centre. Plus, naturally, the opportunity to walk with a Tyrannosaurus Rex!

Planning for the event started in October 2015 and involved staff from all across campus. Advertising DOW was tricky due to the period of "purdah" imposed on public bodies during the run-up to the EU referendum. However, a social media campaign proved highly successful and DOW exceeded all expectations in terms of visitor numbers:

- There were over 100 VIP guests on the Corporate Day.
- The two schools' days saw more than 1100 students enthused and engaged in STEM activities.
- Despite awful weather in the morning, and car parking difficulties, more than 7,500 members of the public attended on the Saturday.

An early estimate of visitor numbers that could be supported was 3,000, so to reach more than double this target was a remarkable achievement.

The Scientific Computing Department made a significant contribution to the overall impact of the event, especially on the public day. It did so through a wide range of activities:

- Providing demonstrations of our visualisation capabilities to corporate guests, and schoolchildren in both KS2/3 and KS4/5 age groups. Demonstrations were run continuously throughout the public day.

- KS2/3 pupils were treated to a "program a simple robot" activity.

Data centre tours team on the public day. L-R: Mark Lundie, John Whittle, Kevin Beresford, Dave Cable, Pete Oliver, Nick Hill, James Priestley. Behind the camera: John Baker.

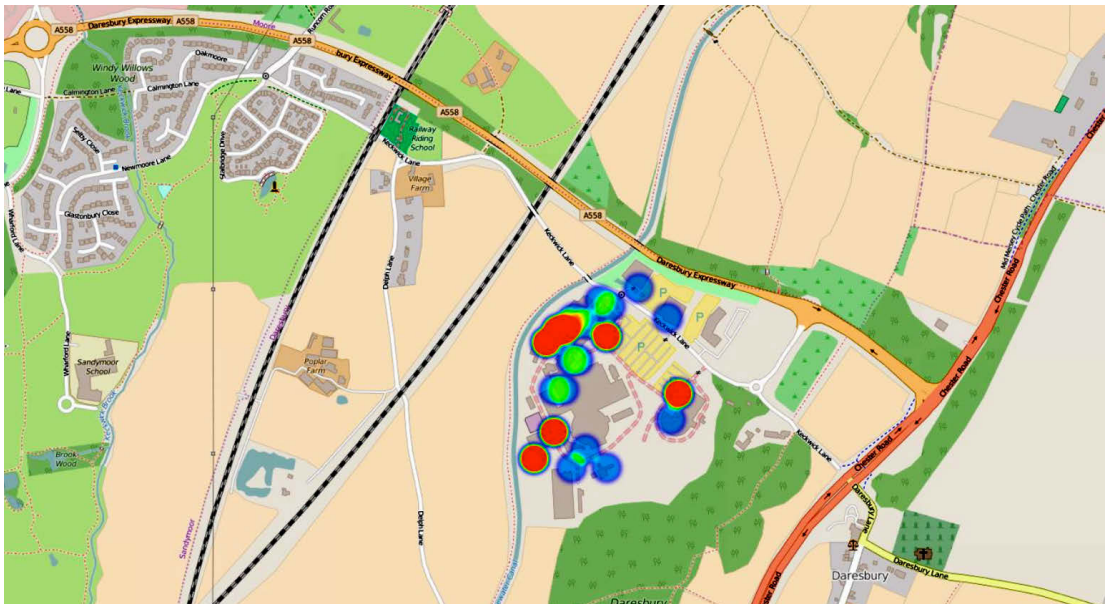


Figure 1: Heat map of visitors at a point in time during the public day. Warmer colours indicate higher numbers. Note that A Block (location of SCD) is one giant red blob! Source: Navatha Tirunagari.

- KS4/5 pupils took part in a "build a temperature sensor" activity, and were given a masterclass in the chemistry of solar cell materials.

- On the public day, SCD delivered activities including data centre tours, computer coding for beginners, Lego Mindstorms "Battle Bots", and building temperature sensors using Arduino microcontrollers.

All told, the department ran a total of eleven activities for the general public, all of which proved very popular. Queues of over 40 minutes were reported for a virtual reality tour of the ATLAS detector on the Large Hadron Collider, and the Lego Battle Bots were continuously busy for the whole day. Because of the popularity of our activities with the public, one of the busiest buildings on campus was A Block, which houses SCD.

Data centre tours team on the public day. L-R: Mark Lundie, John Whittle, Kevin Beresford, Dave Cable, Pete Oliver, Nick Hill, James Priestley. Behind the camera: John Baker.

We can actually illustrate this, because the other contribution that SCD/Hartree made to the event was to create a network of 27 scanning stations across the site, strategically located at key points or activities. Members of the public were given a plastic card, the size and shape of a credit card, with an embedded NFC (Near-Field Communications chip). They were asked to scan these cards at each station, enabling us to visualise visitor movements around site in real-time. This visualisation was shown on the large screen in the Crosfield room (see figure 1).

Figure 1: Heat map of visitors at a point in time during the public day. Warmer colours indicate higher numbers. Note that A Block (location of SCD) is one giant red blob! Source: Navatha Tirunagari.

The scanning stations themselves each comprised a Raspberry Pi connected to an NFC scanner board, housed in a plastic casing. These units were constructed and programmed by Andy Hill of the Technology department.

Navatha Tirunagari of SCD wrote the visualisation software, which will have a life after DOW as a technology demonstrator. Mark Mawson of the Hartree Centre was project manager and also provided technical support on the day, causing the step counter on his wrist to record over 20,000 steps!

We have retained the resultant data set, so that it can be analysed in different ways enabling us to discover what we can learn about visitor movements, especially how (or if) they were related to activity descriptions or locations.

It was clear on the day that the scanners were not very reliable, and this is reflected in figure 2. We found that sometimes scanners would not record a card swipe at all; on other occasions a scanner would record up to 10 scans for one swipe! There a number of reasons for this:

- The NFC scanners themselves worked only over a very short range and were sensitive to location within the plastic housing. Because they were not

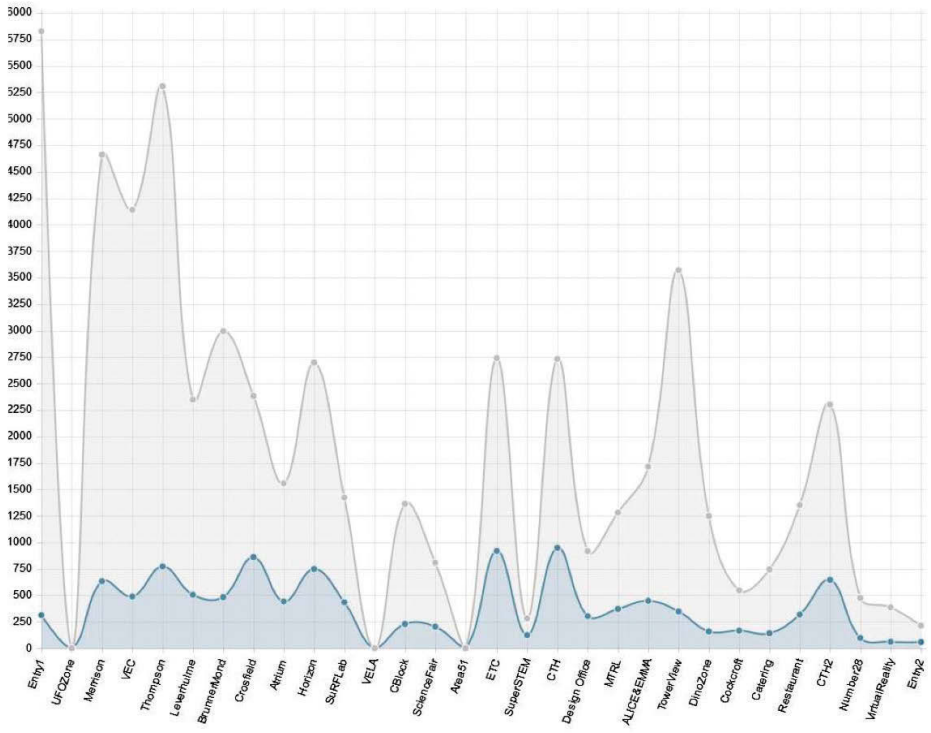


Figure 2: Count of visitor scans per location, at end of public day. Lightly-shaded areas represent raw data; darker areas showed data after post-processing to remove duplication. Source: Navatha Tirunagari.

necessarily in the same place in every unit it was easy to under-record swipes.

- This issue also meant that the best results were achieved with a specific swiping movement. This was not necessarily communicated to volunteers at each station.
- The Python script recording scans seemed to "sleep" at times. This meant that it often missed a swipe or two.

Furthermore, we had 5,000 NFC cards but over 7,500 visitors. So it was impossible to record every visitor at every location.

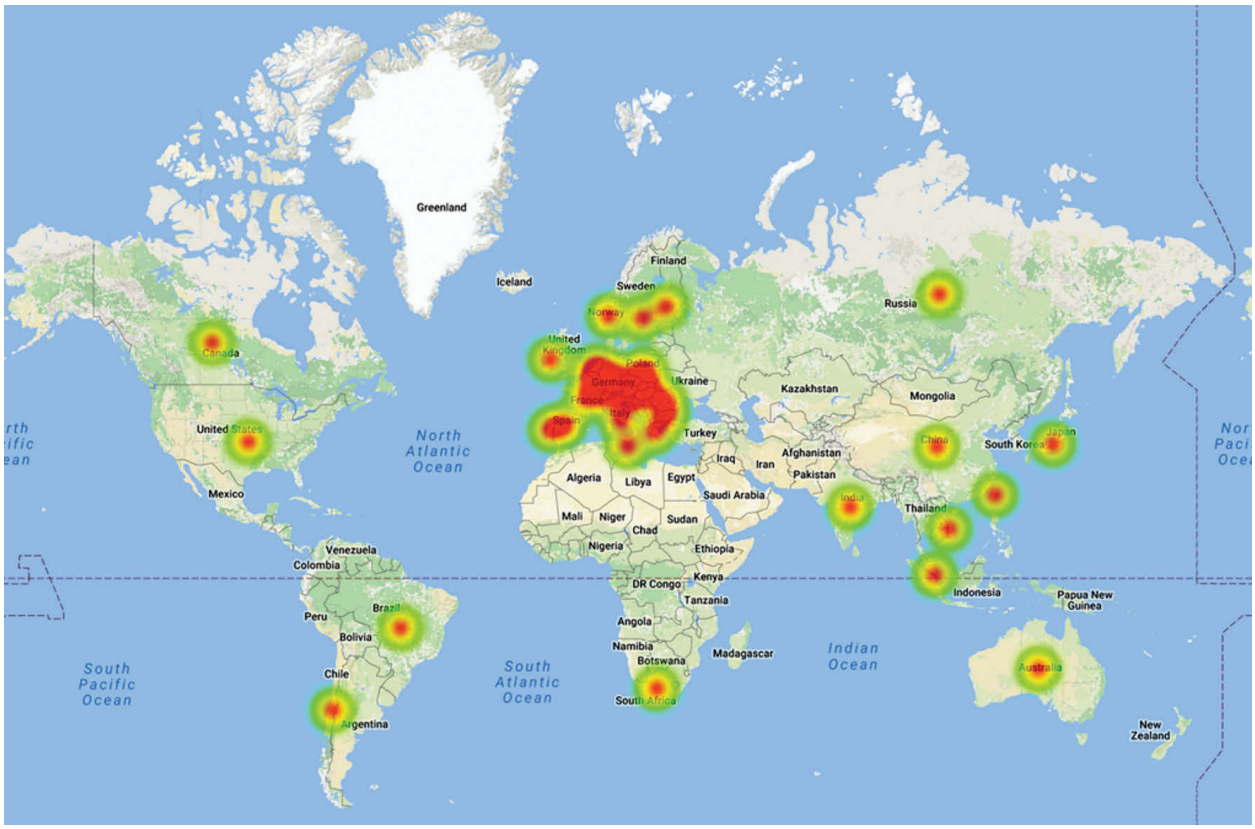
Figure 2: Count of visitor scans per location, at end of public day. Lightly-shaded areas represent raw data; darker areas showed data after post-processing to remove duplication. Source: Navatha Tirunagari.

An event like DOW cannot happen without significant levels of support from staff, which was provided voluntarily. The staff effort involved in planning and putting together the SCD contribution to DOW exceeded 600 person-hours. In terms of volunteer numbers, there was an excellent turnout from SCD and the Hartree Centre. In total we had over 70 volunteers during DOW, plus about half a dozen working with us from other departments.

Author
D.Cable, STFC Daresbury Laboratory

International Presence

During 2015-16 SCD presented, demonstrated and discussed its research and services in over 36 countries across the globe.



A few highlights are outlined below.

Scientific Computing and Hartree conquer Texas

In November 2015, Scientific Computing comrades from the STFC Rutherford Appleton Laboratory, Daresbury Laboratory and the Hartree Centre attended the biggest international supercomputing conference of the year.

SC15, held in Austin, Texas, was the place to be for anyone with a serious interest in computing and big data.

The STFC team certainly did plenty of communicating during the conference. Our booth in the exhibition hall proved to be a popular location for people to meet with Scientific Computing Department (SCD) and Hartree staff, and it attracted some new visitors who wanted to find out what we do with our supercomputing capability. It was also the venue for the presentation of an award to STFC DiRAC (Distributed Research utilising Advanced Computing) and the Stephen Hawking Centre for Theoretical Cosmology for the 'Best use of High Performance Data Analysis.'

New big data research and development opportunities with China

In January 2016 A UK delegation, which included Erica Yang from Scientific Computing, visited to China.

Organised by Innovate UK in collaboration with China's Ministry of Science and Technology (MoST) and the British Embassy, the visit enabled the 17-strong group from leading UK research organisations, Innovate UK catapults, SMEs and large corporations to deliver a series of presentations to potential Chinese partners. The aim was to encourage new collaborative big data and IOT (Internet of Things) research and development (R&D) opportunities. This trip was the first stage in the UK-China research and innovation bridges competition run by Innovate UK. It should lead to significant collaborations between leading Chinese academic institutions, companies, the Scientific Computing Department (SCD) and the Hartree Centre, either via the Newton programme or other funding sources. Erica Yang brought particular expertise and experience in working with UK PLC and the Highway sector in intelligent transport infrastructure using large scale data analytics techniques and high-throughput data system engineering methods to help develop and maintain robust and efficient transport systems in cities.

SCD staff members attending SC15

Collaboration with MoST and the British Embassy

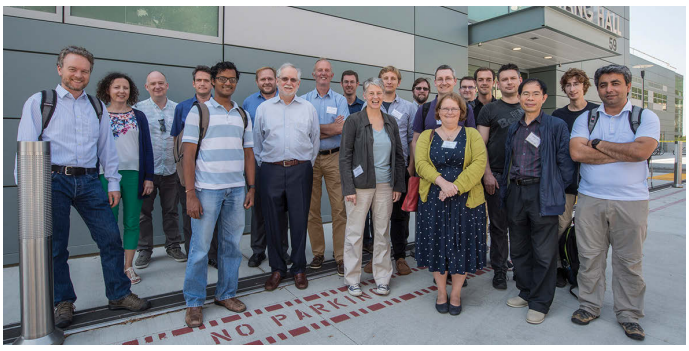
Building links with US National Laboratories

In the latest effort to actively build links with US National Laboratories, STFC Scientific Computing Department staff along with colleagues from STFC's Central Laser Facility, ISIS Neutron and Muon Source and the Diamond Light Source, went to meet their counterparts at Berkeley Lab's Computational Research Division (CRD).

Their aim was to find out more about the computational methods and software tools used at Berkeley Lab to extract information from light source data, view real-time results during beamline experiments, and analyse experimental data.

"One of the biggest bottlenecks to science from the big facilities is access to computing resources, including skilled staff," said David Corney, Director of the STFC Scientific Computing Department. "We came to Berkeley Lab because it's very similar to our organization. By visiting, we hope to have a better understanding of how Berkeley run their synchrotron and what tools they use to manage and analyse the data. We want to get some ideas that we can take back and apply, and we also hope to share some of our expertise as well."

The STFC and Berkeley Lab staff explored how their respective computing departments and science facilities work together, especially in managing data, applying mathematics and computational science to help users interpret experimental data and in supporting them to access the data remotely.



STFC and Berkeley Lab staff
(Credit: The Regents of the University of California, through the Lawrence Berkeley National Laboratory)

Authors
P. Oliver, M. O'Sulivan, STFC Rutherford Appleton Laboratory

Computational chemistry reaches out

It's been a typically busy year for the computational chemistry group with their participation in, and organisation of, a range of training sessions, presentations, summer schools and outreach activities. Training in the DL_Software codes is designed to teach materials modellers how to use DL_POLY, DL_FIELD and DL_MESO (and Chemshell) and sessions were held for over 100 participants onsite at Daresbury, Bristol, London and Guangzhou, China. The training event in China was part of a modelling school initiative that also included two extra days of research and institutional presentations to 70 attendees at two institutions. A new training feature this year included Hack Days aimed to give a more detailed introduction to the software design and methodology of the DL_Software programs, and to deepen the attendees' understanding of how the codes function.

International presentations include Michael Seaton's and Alin Elena's contributions in Sofia and Austin, Texas on modelling codes and their performance on Intel Xeon Phi, and John Purton organised both the annual CCP5 summer school and the CCP5 AGM in the UK. The former was held at the University of Lancaster and involved 70 postgraduate students, completing the 10-day school well-versed in the foundations of molecular simulations. The annual CCP5 AGM conference was held at Harper Adams University and hosted 55 delegates showcasing applications and new techniques of interest to the modelling community. All of the participants contributed to the creation of a networking environment for exchange of ideas and for future collaborative work.

The 2nd CCP5/CCPBioSim Multiscale Modelling Conference, was organised by Tom Keal (SCD), Paola Carbone and Richard Henchman (Manchester) with support from Damian Jones (DL) and Donna Taylor (DL) and was held over three days in Manchester. This event brought together over 90 researchers from around the world to discuss topics of common interest across the spectrum of electronic structure, atomistic and mesoscopic scales and to address the current and future challenges posed by multiscale modelling in biological and materials science. It featured five invited speakers from the UK and Europe and four from the US including Arun Yethiraj (Wisconsin-Madison), who went on to tour UK universities supported by CCP5. The event received very good feedback from the chairs of the CCPs and the third conference in the series is planned for April 2018.

Michael Seaton, David Bray, Annalaura Del Regno and Rick Anderson have been active in training industry users on the DL_MESO and UMMAP simulation and analysis codes as well as general principles behind coarse grained particle simulation methods. Staff from Unilever (UK and India, Bangalore), Syngenta and Infineum have adopted the use of these codes internally following the training sessions supported by the DL team. The team has been asked to create bite sized webinars that can be rolled out to future collaborators.

This year's outreach involved the whole group in a solar cell workshop held for key stage 5 students run as part of Daresbury's Open Week, and a 'Chemistry in 3D' display on the Public Day. One of the highlights was the 3D posters and 3D specs and watching members of the public sway from side-to-side in front of the posters. We told them that was the best way to see the 3D effect, although to be honest, none of us could prove it scientifically! It's been a fun and busy year for us and we're looking forward to 2017!

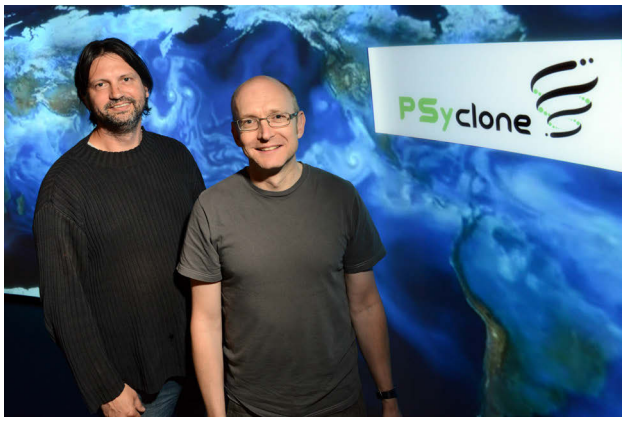
Group shot of instructors (back row) and participants at the modelling school held in Guangzhou, China.

Author
D. Geatches, STFC Daresbury Laboratory

Making an impact on science, society and the UK economy

Credit: © | Dreamstime.com

The STFC Scientific Computing Department carries out the research and development needed to drive innovation, ultimately leading to novel services and products to meet scientific challenges. It has dedicated computational specialists and support staff who are enabling research in our national facilities, collaborating with organisations worldwide and seeking technological solutions to the challenges posed by modern living. Here are just a few of the ways in which they are making a difference to science and society, and helping to boost business and the economy.



Rupert Ford and Andrew Porter, PSyclone developers

Developing new software

Our scientists have developed a truly innovative and unique software framework that automatically generates the parts of the code necessary to run on supercomputers.

PSyclone was developed for the UK Met Office and is now a part of the build system for Dynamo, the dynamical core currently in development for the Met Office's 'next generation' weather and climate model software.

By generating the complex code needed to make use of thousands of processors, PSyclone leaves the Met Office scientists free to concentrate on the science aspects of the model. This means that they will not have to change their code from something that works on a single processing unit (or core) to something that runs on many thousands of cores.

"This is quite a radical approach to the challenge of running models on different kinds of computer architectures. It significantly reduces the potential for errors and will make weather forecasting and climate modelling simpler and more efficient on the high performance computers of the future."

Dr Mike Ashworth

Bringing technology users and businesses together

The success of the UK's first High Performance Computing conference and exhibition, Computing Insight UK, demonstrates the growing importance of computational technologies to research and to businesses.

This STFC event attracted delegates from all parts of Britain as well as international exhibitors and sponsors. It brought together technology providers with a community of expert users and people responsible for the procurement and running of systems, in an environment where they could easily network and share information. Computing Insight UK will be run in December each year.

"I believe this is the only UK conference that's completely focused on High Performance Computing, Big Data Analytics and Cloud Computing."
Dave Cable

Over 200 delegates attended
50 industry exhibitors
26 UK Universities were represented

Helping to save lives

Scientists are using powerful supercomputers to visualise how blood flows through implanted blood pumps, which prolong the lives of heart patients waiting for a donor.

Their goal is to assess if using Computational Fluid Dynamics in conjunction with High Performance Computing (HPC) can increase confidence in using computer models to design these complex medical devices. The work will help to reduce the number of prototypes needed for clinical testing – so saving on time, materials and costs – and potentially lead to the earlier availability of devices for patient use.

Heart failures and diseases claim more than 17 million lives across the world each year – more than all forms of cancer combined. The number is expected to grow to 23.6 million by 2030. In the UK alone, 155,000 people died from heart diseases in 2014. It is estimated that seven million people in the UK are living with heart and circulatory diseases and healthcare costs could be as much as £11 billion.

Source: American Heart Association, British Heart Foundation

Leading large programmes

We have been selected to lead multinational programmes – for example 'West-Life'(World-wide E-infrastructure for structural biology), a €4 million EU funded programme to aid the challenge of sharing knowledge and skills needed for the development of new vaccines and drugs.

To combat strains of diseases such as meningitis and Ebola, structural biologists need to dissect the complexities of bacterial and viral interactions occurring in the human body, often having to use multiple experimental techniques and facilities. West-Life will bring the world of complex data analysis in structural biology to a simple, web browser-based, Virtual Research Environment.

Tutorials and symposiums

Every other year, hundreds of parallel computing experts convene for the Society for Industrial and Applied Mathematics (SIAM) Conference on Parallel Processing for Scientific Computing. In 2016, computer simulations to solve major offshore engineering problems – such as designing and manufacturing coastal flood defences, ships or oil-rigs that can withstand the pressure from vast bodies of rapidly-moving water and the impact of crashing waves – were on Dr Xiaohu Guo's agenda.

He and his colleague, Sergi Siso, held a mini-symposium in Paris to discuss the challenges posed by different application software and computing architectures, and find some possible solutions to improving performance and efficiency.

"The increasing scale and complexity of current and future computing systems is putting pressure on numerical algorithms and the multiple software systems we use. In other words, it can cause problems in running the software efficiently, or completely fail to run them, so we need to find ways to balance this out." Dr Xiaohu Guo

CIUK 2015

Publishing technical papers

In 2015-16 the Scientific Computing Department scientists recorded the publication of 83 papers in ePubs, STFC's repository for scientific and technical outputs. These have all been published in prestigious peer-reviewed journals and some have even made the front cover – such as Leon Petit and colleagues' work with a class of rare-earth materials called 'lanthanide-transition metal compounds'. These play an increasingly important role in the development of new materials, including the lightweight permanent magnets used for modern applications such as mobile phones, radiation detectors and air-conditioning systems. The work was showcased on the front cover of the American Physical Society's journal Physical Review Letters

Journal front cover (credit American Physical Society)

Social interaction

We are now frequently using media releases and social media outlets to let people know about our work and the impact it is having. This year we have increased our social media presence via twitter so we are building on the number of people who follow us, our work is being mentioned more often by others and we are having more direct interaction with them than ever before.

Follow us on twitter @SciComp_STFC

Author
M. O'Sulivan, STFC Rutherford Appleton Laboratory

Notes

Notes

Notes

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	jen.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
Ms 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

Group Leaders

Mr Luke Mason	Applications Performance Engineering	luke.mason@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
Mr Tom Griffin	Research Data	tom.griffin@stfc.ac.uk
Mr Nick Hill	Research Infrastructure	nick.hill@stfc.ac.uk
Mrs Catherine Jones	Software Engineering Group	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 Erica Yang	Visualisation	erica.yang@stfc.ac.uk
Dr Martyn Winn	Biology and Life Sciences	martyn.winn@stfc.ac.uk

Websites

Scientific Computing Department	www.stfc.ac.uk/scd
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