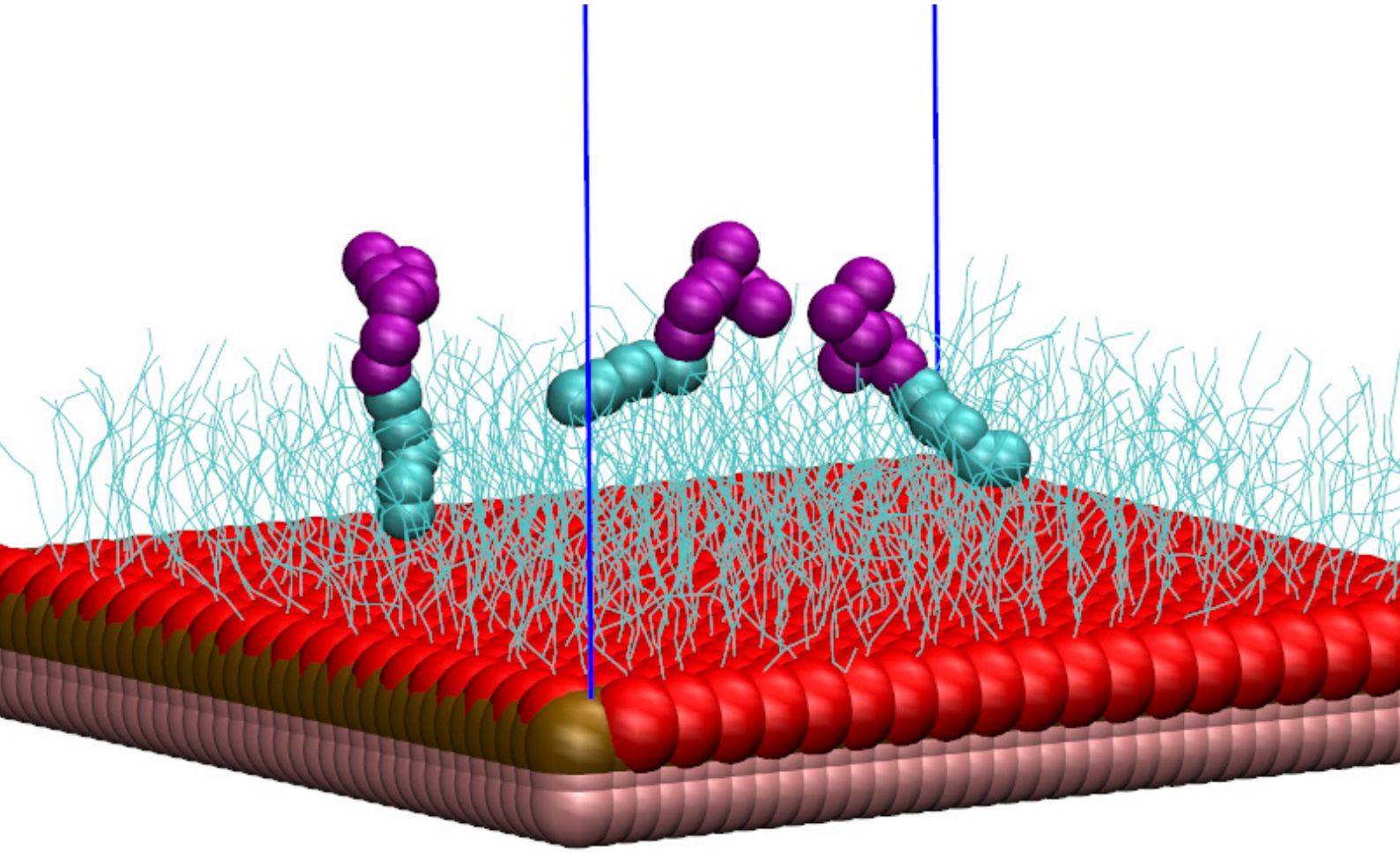
Science Highlights 2013 - 2014

Scientific Computer Department (SCD)



Science & Technology Facilities Council

Contents

Designed by Andrew Collins STFC Media Services
Produced by RCUK's internal service provider.

Welcome

I am privileged to write the introduction to the Scientific Computing Department's annual report for 2013/14, not least because I was only in post as acting director for 1 month during that year. That being said, I am deeply impressed with the breadth and quality of work that has gone on in the department during that time, and much more so since then, as my understanding of the department and its operations across all divisions at local, national, and international level has become significantly clearer.

Reading through the report I cannot help but be impressed by the quality of the work that has been carried out, but before I draw attention to a few of the highlights, I would also add that I am also struck by so much of the work of the department which is not included in this report, such as the SCARF computing service for the STFC facilities, or the cloud deployment building from the Tier-1 automated configuration system, or the establishment of the Research Data Alliance (RDA) incorporated as a UK company with its address at RAL. (RDA now has 2300 members from 72 countries). The highlights within this report should therefore be appreciated and welcomed, but also be viewed as the tips of the icebergs: rightly getting our attention, but aware of the bulk, and other success just beneath the surface.

So particular highlights for me include the article by Rik Anderson about the CAF project (Computer Aided Formulation), where scientists from the computational chemistry group within SCD have teamed up with Unilever, Syngenta and Infineum, to secure a Technology Strategy Board backed £1m project to advance the development of formulated products using CAF. The benefits of this are expected to be taken up by both industry and academic partners alike. An excellent result.

I also greatly appreciated the article by Karl Richardson, describing his experiences with the High Performance Systems group at Daresbury, a group I worked with very closely during the year, as they set up the energy efficient computing and data analytic systems within the Hartree phase 2 program. It is rewarding to understand how much Karl has learned from his year, and also to acknowledge how much he has given to the department.

I would also like to thank all the staff of SCD for their hard work and for their achievements during the year, during a time of considerable change. Without their efforts, none of these achievements would happen. Our staff are the Scientific Computing Department.

Finally, whether you are an existing stakeholder, or if you are considering talking to us about what we may be able to provide for you, and how we might be able to work with each other into the future, I hope you enjoy this report, and recognise it for what it is – a significant flag marking the progress of a dynamic and adaptive department which is looking forward to a rapidly developing future.

Thank you and regards,

David Corney

Acting Director of the
Scientific Computing Department

BIOMOLECULAR SIMULATION ADVANCES

Our understanding of the structural basis of biology is advancing year-on-year. Macromolecular crystallography (MX) continues to provide more atomic structures of individual proteins and small complexes, while Nuclear Magnetic Resonance spectroscopy (NMR) provides structures of smaller proteins as well as structural restraints and dynamic information for larger macromolecules. These structures are publically accessible via the Protein Data Bank, which released the 100,000th structure on 14th May 2014.

Lower resolution information on macromolecular complexes, viruses and molecular machines comes from electron microscopy and small angle X-ray and neutron scattering. Single particle electron cryo-microscopy (cryoEM) in particular is advancing rapidly, with developments such as direct electron detectors, and is providing an increasing number of atomic resolution structures.

Biomolecular simulation provides added value to this experimental structural information. Most experimental information is averaged over time and over multiple copies. While single particle cryoEM does image individual particles, 3D reconstructions are built from projections of thousands of particles. Simulation allows one to follow a single particle as it evolves through time. It also allows in silico experimentation, for example investigating the effect of amino acid mutations. With robust force fields and faster computers, experimentalists are increasingly turning to simulation to provide additional insight into their structures.

Within this context, the Computational Biology Group in SCD contributes to two consortia aimed at supporting biomolecular simulation. The collaborative computational project CCPBioSim organises a programme of training courses in the methods and software available, which are attended by experimentalists as well as computational scientists. CCPBioSim is also developing a number of software tools, within the general area of multiscale modelling. Quantum mechanical calculations of active sites can be linked with classical mechanics of the protein scaffold. Coarse-grain modelling then allows larger systems and timescales to be accessed.

Our contribution to this community effort is the tool FESetup for the automatic setup of protein–ligand free energy simulations, using techniques such as Thermodynamic Integration and MM–PBSA. The tool will automatically parameterise a (possibly large) set of ligands with AM1/BCC, and combine these with a protein in a complex. Common substructures between pairs of ligands are determined, and used for alchemical simulations to determine free energy differences between the complexes. FESetup supports the molecular simulation packages Sire, AMBER and Gromacs. Supported force fields are all modern AMBER force fields including GAFF. We particularly aim at ease of use and robustness of the code.

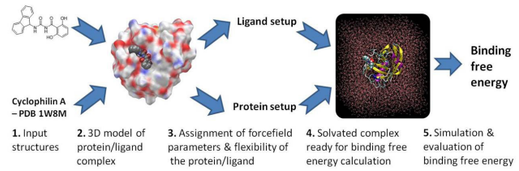


Figure 1: Protein-ligand free energy simulations in FESetup.

The code is written in Python 2.7 and comes in a self-extracting bzip2 archive script.

The work is carried out in collaboration with Julien Michel (Edinburgh) and Christopher Woods (Bristol). The first official version (1.0) was released in December 2013. Future plans include extending the code to support other popular biomolecular simulation software, additional force fields and parameterisation schemes.

The second consortium that we support is for High End Computing in biomolecular simulation (HECBioSim). One major role of the consortium is to make compute time on the national supercomputer ARCHER available to members, and so far we have distributed around 150 million AUs. Projects are judged on scientific quality and on their suitability for high end resources. A key part of the remit is to widen the take-up of HPC resources by the community, in particular for those with an experimental background who may be put off by traditional access mechanisms.

Figure 2: HECBioSim provides access for the community to the national supercomputer ARCHER.

HECBioSim is also developing software tools suitable to high end computing. In the field of biomolecular simulation, this often means running ensemble simulations (e.g. Replica Exchange) in order to increase sampling. Submitting multiple simulations to a remote platform such as ARCHER can benefit from automated tools, and we are developing ProxyApp in collaboration with Charlie Laughton (Nottingham) to address this area. This tool will simplify the transfer of job parameters and data files from a local system to a remote compute resource such as ARCHER, so that use of HPC is no more difficult than using the local cluster.

Analysis of the resulting MD trajectories is also challenging, and future work will focus on tools for retrieving, storing and interpreting the output data.

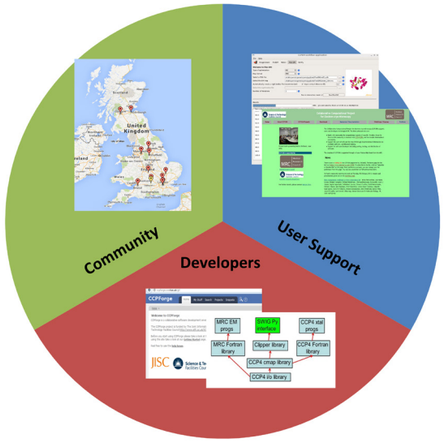


Figure 3: Overview of CCP-EM activities. The green sector refers to community building activities, and is represented by a map of current groups performing high resolution cryoEM. The blue sector refers to user support activities, which are illustrated by the CCP-EM GUI currently under development and a snapshot of the CCP-EM web site. Finally, the red sector refers to support of software developers, and is illustrated by the front page of CCPForge, which hosts the CCP-EM software project, and a schematic of the developing CCP-EM library.

As stated earlier, these simulations are of increasing interest to structural biologists, but equally the simulation community depends heavily on the structures determined by the latter. Our group have coordinated the CCP4 project for macromolecular crystallography for many years. A new 5 year grant, funded by BBSRC, MRC and Diamond, means that CCP4 will continue to develop novel software and support training. The application of cryoEM in structural biology has matured in recent years, and we have initiated a new project CCP-EM to support computational analysis in this field. The aim is to follow the open and collaborative ethos of CCP4, although it is early days yet.

CCP-EM is beginning to put together a software suite for single particle cryoEM, including programs for particle picking, 3D reconstruction, and flexible fitting of atomic models. A particular selling point of CCP-EM is its close links to CCP4, which is timely given the increasing overlap between “high resolution” cryoEM and “low resolution” crystallography. We are investigating the use of standard crystallographic software for building and refining atomic models into high resolution cryoEM maps. CCP-EM is also re-using many software libraries from CCP4, which as well as being efficient from a developers’ point of view, will also help scientists move between the two fields.

As an indication of the maturity of the field, a national facility for cryoEM is being built on the Harwell campus. Diamond will provide infrastructure and support for data acquisition, while CCP-EM will provide support for data analysis and scientific interpretation. This arrangement mimics the support that CCP4 gives to the MX beamlines at Diamond.

In conclusion, challenges in biology are being addressed through a combination of experimental and computational techniques. The Computational Biology Group in SCD contributes to the development of several of these techniques, and thus is central to the push for multi-disciplinary studies.

Author
M. Winn, STFC Daresbury Laboratory

References

[1] http://www.ccpbiosim.ac.uk/ [2] http://www.hecbiosim.ac.uk/ [3] http://www.ccp4.ac.uk/ [4] http://www.ccpem.ac.uk/

TANK SLOSHING SIMULATION USING SMOOTHED PARTICLE HYDRODYNAMICS

The *sloshing* of fluid induced by the movement of its enclosure is an important phenomenon. How the fluid responds can provide an indication of its underlying properties and any engineering solution involving the containment of a fluid that may slosh needs to be understood to ensure stability. Practical examples include fluid containment in moving vehicles, where rapid bulk motion of a large volume of fluid provides significant engineering challenges, such as the effects on a motorsport vehicle as it races, movement of fuel in containment systems on aeroplanes, and the potential for large tanker vessels to capsize while transporting large fluid volumes across water.

Sloshing is a problem that can be difficult to capture using traditional computational fluid dynamics (CFD) as, while appropriate resolution of the Navier-Stokes equations will successfully capture bulk fluid motion, most methods rely on a fixed mesh. This fixed nature can restrict the formation of important hydrodynamic structures, such as breaking waves; therefore an approach that does not rely on a fixed mesh is attractive.

Smoothed particle hydrodynamics (SPH) was initially defined to solve unbounded problems within astrophysics such as simulating galaxy formation. It is conceptually simple as, rather than using an Eulerian approach involving a fixed mesh, it takes a Lagrangian approach and uses a set of computational points (or particles). Each particle has an associated distance, or smoothing length, and any particles that fall within this contribute to its current physical state according to a distance based smoothing function known as a kernel. A number of these functions exist; however, all are designed to satisfy the Dirac delta function. Each particle in the simulation has a number of physical quantities associated with it, such as density, velocity and pressure.

When solving the incompressible Navier-Stokes equations for fluid flow using SPH, each particle in the system has an associated mass and as the simulation evolves each particle is able to move. This occurs according to Newton's second law and due to conservation of momentum derived according to the weighted contribution of its neighbouring particles. The mass of each particle remains constant, so conservation for this is performed differently depending upon whether the method is considered weakly-compressible or incompressible; the latter being computationally more intensive.

One of the key SPH communities is the SPH European Research Interest Community (SPHERIC). This group is driving the development of software packages to solve SPH based problems, both for academia and industry. One open-source example is DualSPHysics [1], which aims to allow weakly-compressible SPH simulations to be performed on commodity workstations and computing hardware. The code is written in C++ and can use either multiple CPU cores within a shared memory system via OpenMP or a single NVIDIA graphics processing unit (GPU) using CUDA. Future versions of DualSPHysics will use multiple GPUs [2] to increase the scale of the simulations that can be achieved in order to overcome the performance and memory limitations of a single accelerator.

DualSPHysics has been used to simulate a sloshing benchmark case provided by SPHERIC [3,4]. The SPH implementation of the experiment initially discretises the geometry involved using a cubic layout of particles, with initial spacing between each particle at 1 mm. This produces a simulation involving approximately 100,000 particles. Acceleration of the fluid due to the tank oscillation is achieved using a set of external forces applied to the fluid particles.

Figure 1 provides a snap-shot of two points in time from the simulation with camera footage provided on the left showing the real fluid in green and the SPH simulation on the right. Particle colour is showing pressure in both cases. It can be seen that the kinematics of the SPH fluid is well matched to the experiment.

The computed pressure history can be seen in Figure 2, the general form of the results follow the experimental data well, with the peak pressure value matching almost exactly. The pressure values are relatively noisy, especially following the initial peak impact at around 2.36 s; however, this is to be expected given the weakly-compressible nature of the simulation and does not detract from the model's general ability to capture the bulk fluid motion. The use of a different boundary solution may improve this, as would use of incompressible SPH.

As DualSPHysics can currently be run using multiple CPUs or a single GPU, Figure 3 provides a comparison of the time needed to reach the solution using various hardware configurations. The CPU-based runs were performed using an Intel Core i7-4770 (3.4Ghz, 4 cores, 8 threads) and the GPU-based runs using a GeForce GTX 680 and 660 (Kepler GK104), both hosted in the workstation from which the CPU results were obtained. SPH is rapidly gaining credibility as a CFD tool complimentary to traditional techniques; it can provide significant advantages in scenarios that involve complex or moving components. As the software improves to take advantage of new computational hardware, it is likely SPH will find new and interesting applications, as well as improved accuracy through the use of better refined methods; smaller scale discretisation's and models that better mimic reality.

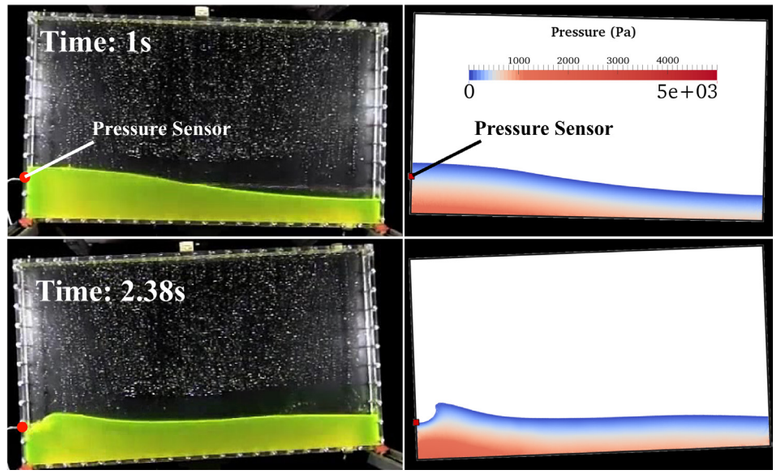


Figure 1: Two snap-shots comparing results from the SPH sloshing model and experimental data.

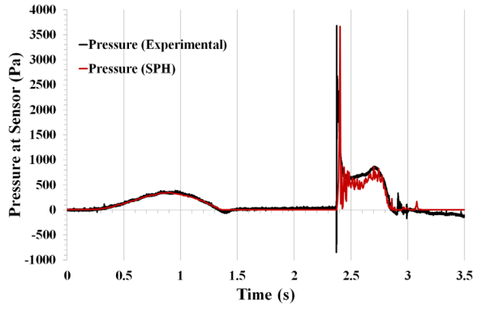


Figure 2: SPH and experimental pressure history at sensor.

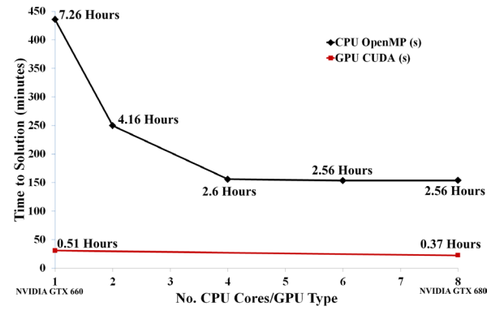


Figure 3: Time to solution results for the same simulation run on various numbers of threads on an Intel Core i7-4770 and two NVIDIA GPUs.

Authors
S. M. Longshaw, D. Emerson, R. Barber, C. Moulinec, STFC Daresbury Laboratory

References

[1] Crespo A.J.C, Domínguez J.M., Rogers B.D., Gómez-Gesteira M., Longshaw S.M., Canelas R., Vacondio R., Barreiro A. and García-Feal O., DualSPHysics: open-source CFD solver based on Smoothed Particle Hydrodynamics (SPH), Computer Physics Communications, 2014.
[2] Domínguez J.M., Crespo A.J.C., Valdez-Balderas D., Rogers B.D. and Gómez-Gesteira M., New multi-GPU implementation for Smoothed Particle Hydrodynamics on heterogeneous clusters, Computer Physics Communications, 184: 1848-1860, 2013.
[3] Souto-Iglesias A., Botia-Vera E., Martín A., Pérez-Arribas F., A set of canonical problems in sloshing Part 0: Experimental setup and data processing, Ocean Engineering, 38(16): 1823-1830, 2011.
[4] Delorme L., Colagrossi A., Souto-Iglesias A., Zamora-Rodríguez R. and Botia-Vera E., A set of canonical problems in sloshing Part I: Pressure field in forced roll. Comparison between experimental results and SPH, Ocean Engineering, 36(2): 168-178, 2009

f-ELECTRONS UNDER PRESSURE

Lanthanide compounds exhibit a remarkable variety in electronic properties, driven by a complex interplay of ligand chemistry and the dual character of *f*-electrons. Of particular interest are compounds situated at the boundary of an *f*-electron localization-delocalization transition, where a change in valence can be brought about by changes in the external parameters. Apart from providing added insight into the fundamental physics of *f*-electron systems and the associated valence fluctuations, understanding the coupling between the structural, electronic, and magnetic degrees of freedom can lead to the developments of novel multifunctional materials and the design of improved stimuli responsive devices. In order to correctly predict valence transitions under pressure, first principles calculations need to address electron-electron correlations and band formation on an equal footing, and without introducing materials dependent parameters. The self-interaction corrected (SIC) local spin-density (LSD) approximation implemented in our LMTO and KKR all electron codes distinguishes between localized (core-like) and delocalized (itinerant) *f*-electron states based on total energy considerations. The ground state can then be determined from the global energy minimum.

The rare earth mono tellurides crystallize in the NaCl (B1) structure at ambient conditions. The corresponding ground state electronic structures are determined by the degree of ***f***-electron localization, with SIC-LSD predicting the localized divalent configuration ($B1^{2+}$) to become energetically favourable for SmTe, EuTe, DyTe, TmTe and YbTe. Under pressure these compounds undergo a structural NaCl (B1) to CsCl (B2) transition either followed or preceded by a valence (2+ to 3+) transition. For TmTe, [1] from the calculated enthalpies as a function of pressure (shown in Figure 1a) we predict a $B1^{2+}$ to $B1^{3+}$ insulator to metal transition around 3.0 GPa, in good agreement with experiment as can be seen in Fig. 1b. With further increasing pressure, around 15 GPa a structural transition from $B1^{3+}$ to $B2^{3+}$ is predicted. A similar behaviour is observed in SmTe, with the valence transition setting in around 10 GPa followed by the structural transition around 12 GPa. In Figure 2 a schematic density of states (DOS) for a rare earth mono chalcogenide at both low (a) and high (b) pressure is shown. The red, blue, and black areas represent respectively the chalcogenide *p*, and the lanthanide *d* and *f* states.

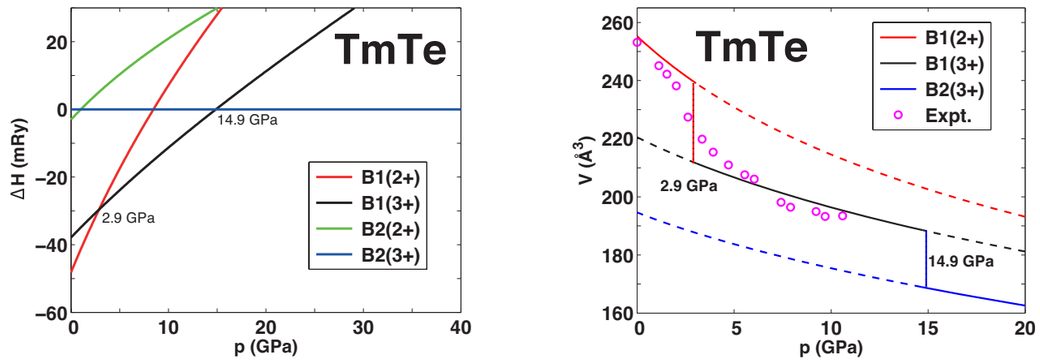


Figure 1: Behaviour of enthapies (a) and volume (b) for TmTe under pressure.

Here, in the trivalent configuration, one of the *f*-states is treated as delocalized, and the *f*-manifold separates into an integer number of localized states (situated below the anion *p*-states) and a fractional number n_f, of delocalized *f*-states, partially occupying the narrow peak at the Fermi level (indicated by the vertical green line). From the SIC-LSD total energy point of view, the trivalent configuration depicted in Fig. 2a is unfavorable, since filling the narrow *f*–peak (n_f) results in a marginal increase in binding energy, and instead the corresponding state prefers to localize, gaining the associated SI energy. The actual ground state will be divalent and insulating instead. Under pressure due to increased *f*–*d* charge transfer n_f becomes smaller, as a result of the *f*-peak emptying in favour of the broader *d*-band, and band formation becomes increasingly favourable. Therefore, at sufficiently high pressures, delocalization occurs, with the resulting ground state being the trivalent configuration depicted in Fig 2b.

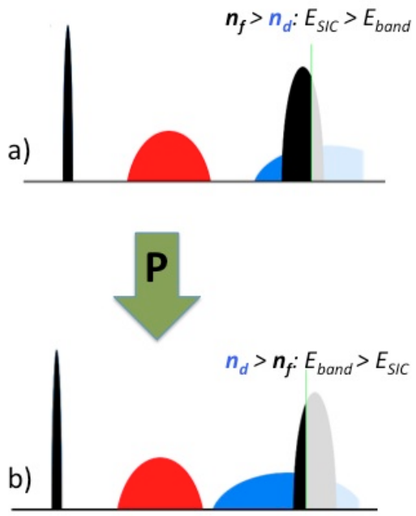


Figure 2: DOS of rare earth mono chalcogenide under low (a) and high (b) pressures.

It emerges that quite generally, for rare earth compounds, the *f*-electron delocalization transition is energetically favourable for $n_f \leq 0.65$-0.70 electrons. Depending on the size of n_f (i.e. degree of localization) at ambient pressure, the structural transition occurs before or after the delocalization transition. The increasing trend towards *f*-electron localization, from SmTe to EuTe, and TmTe to YbTe is reflected in considerably increased pressures being required to bring about the *f*-electron delocalization. As a result the valence transition only occurs after the structural transition to the CsCl phase at pressures of 47 GPa in EuTe and 31 GPa in YbTe. [1]

We find that, for a given volume, the number of delocalized *f*-electrons, n_f, is roughly 0.1 electrons higher in the B2 compared to the B1 structure. In EuO this interplay of structural and valence degrees of freedom leads to a very complex phase diagram. [2] Starting from the insulating divalent NaCl phase at ambient pressure, an isostructural insulator to metal transition occurs at around 49 GPa. In the pressure range between 49 and 60 GPa, the observed near degeneracy between $B1^{3+}$ and $B2^{2+}$ configurations indicates a possible coexistence of NaCl and CsCl phases with relative proportions that will be highly sensitive to the experimental conditions. At around 60 GPa, a full-fledged structural transition to the CsCl phase is accompanied by a localization transition to the divalent configuration ($B2^{2+}$) due to the increase in n_f. At around 70 GPa a further 2+ to 3+ valence transition occurs, now in the CsCl phase. The sequence of valence changes under pressure is shown in Figure 3. The occurrence of a divalent CsCl phase between 60 and 70 GPa, as well as the overall decrease in valence from the nominally trivalent NaCl phase (effective valence 2.37) to nominally trivalent CsCl phase (effective valence 2.28) are strongly reminiscent of the reentrant valence behavior proposed in the pressure experiments by Souza-Neto et al. [3]

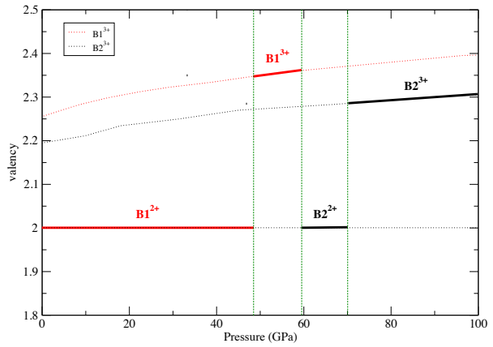


Figure 3: Effective valency of EuO under pressure.

Authors
L. Petit, M. Lüders, Z. Szotek, W. M. Temmerman: STFC Daresbury Laboratory. A. Svane: Aarhus University

References

[1] L. Petit, A. Svane, M. Lüders, Z. Szotek, G. Vaitheeswaran, V. Kanchana, and W. M. Temmerman, J. Phys.: Condens. Matter 26, 274213 (2014).
[2] L. Petit, Z. Szotek, M. Lüders, W. M. Temmerman, and A. Svane, Phys. Rev. B 90, 035110 (2014).
[3] N. M. Souza-Neto, J. Zhao, E. E. Alp, G. Shen, S. V. Sinogeikin, G. Lapertot, and D. Haskel, Phys. Rev. Lett. 109, 026403 (2012).

ACCELERATING FORMULATED PRODUCT DESIGN BY COMPUTER-AIDED APPROACHES

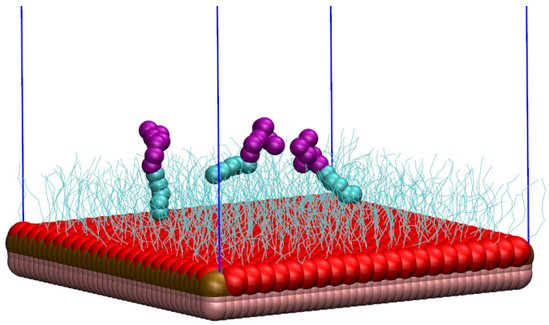
Complex formulated products, such as shampoos, detergent powders and liquids, processed foods, paints, adhesives, lubricants and pesticide granules, are ubiquitous in everyday and industrial life. The UK market for formulated products is worth around £180bn a year, with further potential in emerging overseas markets of around £1,000bn (Chemistry Innovation KTN Strategy Report 2010). Whilst in civil and mechanical engineering, the design process is done almost entirely by computer with reduced physical prototyping formulated product design is still predominantly an ad hoc labour-intensive process. Scientists from the Computational Chemistry Group in SCD have teamed up with Unilever, Syngenta and Infineum, to secure a TSB backed £1 million project to advance the development of formulated products using Computer Aided Formulation (CAF).

Active Research

The CAF project will build upon the potential of mesoscale modelling to drive a radical change in speed of formulated product design for manufacturability and in-use performance. The technical challenge is to develop predictive models that are simple yet accurate enough to enable predictive product design. This will reduce the time to market and development costs of a new or reformulated product. Initial results within the consortium across an extended formulation space (detergents, speciality chemicals, agro-technical formulations) are encouraging, to the point where the solution behaviour, formulation stability and surface interaction properties are accurately predicted for small subsets of chemical entities. The research team from the SCD will expand the capabilities of the established mesoscale method, dissipative particle dynamics (DPD), by incorporating appropriate physical models required to reproduce important experimental behaviour. Developments made to the DPD methodology during this project will address a number of deficiencies within the current framework and will broaden the range of applicability of this method. These developments will permit the simulation of more realistic and complex systems and will provide greater understanding into the mode of action of many different formulated products.

Derived Benefit

It is expected that there will be a number of positive outcomes from this project beyond the immediate commercial partners, who stand to develop better products faster and more sustainably. Academic beneficiaries from the field of condensed matter will be able to make the most of the advances to the DPD methodology and it is anticipated that the developments will be of use to researchers in other areas such as biophysics, e.g. the simulation of biomembranes and lipid bilayers. Furthermore, it is expected that DPD will play an ever-increasing role in multi-scale modelling approaches through bridging of the atomistic and continuum scales. In such approaches, atomistic simulations are performed to build the DPD models, followed by DPD simulations that provide the necessary input to the continuum codes. Implementations of such approaches can circumvent assumptions in the continuum codes since the mesoscale simulations can provide more accurate estimates of the thermodynamic state within the localised regions compared to a constitutive equation. With an improved DPD framework resulting from this work even more accurate estimates can be achieved. DPD methodology developments occurring during the project will be fully incorporated into the UK academic DPD simulation code, DL_MESO. This simulation code is freely available to academic users world-wide under license.



Surfactant molecules adsorbed onto a silica surface.

Author
R. Anderson, STFC Daresbury Laboratory

References

M A Seaton, R L Anderson, S Metz and W Smith, DL_MESO: Highly scalable mesoscale simulation, Molecular Simulation 39, 2013, 796 – 821.

NOT JUST BELLS AND WHISTLES: DELIVERING NEXT GENERATION DATA SERVICES

Data services are important for scientific data: research is increasingly "data driven", and tools and services are needed to store data, make it available for analysis, archiving, discovery and referencing. Growth in STFC's data holdings is broadly exponential; traditionally the Large Hadron Collider (LHC) has been the biggest user, but CEDA (climate) and Diamond are catching up. The challenge is to provide services for STFC's wide range of users covering all the data lifecycle, enabling owners to comply with STFC's data policies, and yet not have to provide a different tool to each.

The backend data storage services are based on CASTOR, a storage system also run (and developed) by CERN, and Oracle and MySQL databases. For convenience we provide higher level interfaces such as StorageD, "Elastic Tape," the Preservica data preservation service, iRODS, and backup services.

The CASTOR team is preparing for next year's LHC run two, expected March '15: all services need their operating systems upgraded, and there may also be another CASTOR upgrade, depending on CERN's release schedule for CASTOR. We have developed "Elasticsearch", a log analysis engine. CASTOR's logging is very verbose, and sifting through logs for relevant information – debugging, troubleshooting, analytics – can be very time-consuming, and the new search system should alleviate this problem.

Following the deployment of new disk servers (and retiring old ones), LHC CASTOR currently manages 13.6 Petabytes of disk, but facilities only about ¼ Petabyte – the difference is due to the different usage, where LHC is working directly with the data, whereas facilities are mostly using CASTOR as an archive. For facilities, however, we run a 360 TB disk store based on CEPH, with plans to extend probably up to a Petabyte this year. CEPH offers improved resilience, lower administrative overhead, and is widely used.

SCD is participating in EUDAT, a flagship FP7 project to provide data management services across disciplinary, technical, and geographic boundaries. For example, in collaboration with CSC in Finland and the European Bioinformatics Institute, we hold replicas of data from the diXa project (itself an FP7 project, managing chemical safety data), currently totalling 169 Gigabytes in 15,933 files. With iRODS as a front end to CASTOR, the storage system also holds an online archive of data from the Synchrotron Radiation Source once based at Daresbury Labs. This dataset extends to more than 600 Gigabytes in 2.65 million files and is part of the Open Access data published by STFC. A web interface to the data can be found at http://eudat-node1.esc.rl.ac.uk/irods/auto.php. We also contribute a service registry, scalability studies based on linked open data and object storage systems, as well as federated authentication and authorisation.

The EUDAT project brings together data services providers and user communities from across Europe.

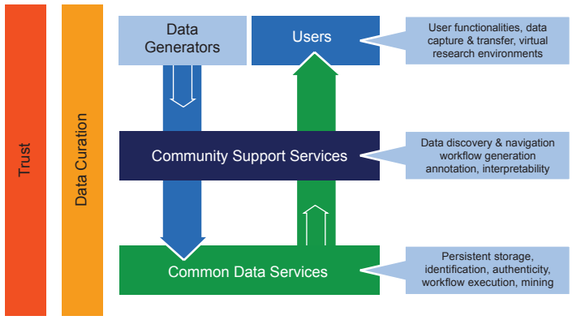


Figure 1: The EUDAT project brings together data services providers and user communities from across Europe.

Managing small files and metadata

Like EUDAT, facilities rarely interact with CASTOR directly; they use StorageD as a friendlier front end to manage large numbers of small files. In StorageD, CEDA and DLS together currently hold about 4 Petabytes in more than 650 million files, an average size of only 6 Megabytes. New data is added at rates exceeding 20 Terabytes per day, but, as in many other areas, the rate of increase is increasing.

Elastic Tape ("ET", and not to be confused with Elasticsearch) is broadly similar to StorageD but adds the concept of workspaces with quota management. Users send data from their workspace to ET to free up disk space, and recall it later when needed. Unlike StorageD, users can delete data in ET (StorageD archives data, so deletion requires special privileges), and can also manage file aggregation in UNIX tar archives which users are likely to be familiar with already. Future work includes support for external metadata checkers as input to the registration system and creation of reporting and statistics tools.

Big Data Backups

Backing up a few tens of terabytes can take weeks using traditional approaches. We are exploring alternative approaches that bring this down to a practical level. One that has already been successful is to break the backup into lots of small backups. These run more efficiently - the daily backup completes within one day! We are also experimenting with systems to store relatively small sets of data from desktop PCs. It is very similar to cloud storage products such as Skydrive but allows us to keep the data on-site, which is important when handling commercial or sensitive data. It can be used on an ad-hoc basis through the browser-based frontend, or can run in the background through a synchronising client. Of course we also check regularly that we can restore from the backups.

Digital preservation

ISIS need to preserve experiment data for at least 10 years so we provide a preservation service based on Tessella Safety Deposit Box (SDB). It ingests data packages supplied by ISIS, analyses them to ensure they are complete, generates additional archive metadata, and writes the data to tape-backed storage in triplicate. SDB can also be configured to regularly "wake up" data and check integrity.

One question is whether we should pursue accreditation for the service, e.g. Trustworthy Repositories Audit & Certification? In the case of ISIS, we work so closely that they know us better than any auditor, but it might be worth considering it for new users of the service.

Databases

The Database Team manages a total of 13 production Oracle databases and 2 MySQL; as well as 3 development Oracle databases and 1 MySQL. The databases support applications with high availability requirements, with over 100 different schemas. The size of each database ranges from a few hundreds gigabyte up to 3TB. Together, they serve over hundred thousand requests per day, with the busiest peaking at 9K operations per second.

Growth is also an important factor: during 2014, we've had to deploy three new Oracle and one new MySQL database to support new applications, and the total volume of the databases has grown over the year by 60%, to 5TB. The biggest growth rate has been in the StorageD and ICAT databases for Diamond.

Among other activities, the Database Team has been investigating Oracle's "Golden Gate" for synchronisation and High-Availability, as well as Oracle Database 12c.

Authors
R. Appleyard, C. Cioffi, J. Jensen, C. Ketley, C. Prosser,
J. Sierra, S. de Witt, STFC Rutherford Appleton Laboratory
R. Downing, STFC Daresbury Laboratory

SCAPE: SCALABLE DIGITAL PRESERVATION

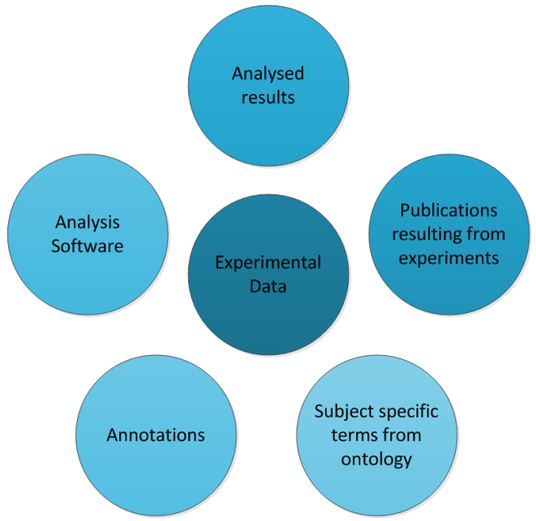
The Research Data Group in the Data Division have been working with partners across Europe in a project which aimed to investigate how digital preservation tools and techniques could be scaled up to heterogeneous collections of complex digital objects at the Terabyte scale. The scale of these digital collections implies that preservation activities will require automation through the use of workflows and high-performance systems. STFC were involved in three main areas of work: data preservation; organisational preservation policy and adoption of parallel techniques.

The SCAPE project was co-funded by the European Union under FP7 ICT-2009.4.1 (Grant Agreement number 270137) and ran from February 2011 to September 2014. It was led by the Austrian Institute of Technology and STFC led the Research Data Workpackage.

Data preservation of Research Objects

A research project produces many outputs during its lifespan; some are formally published, some relate to the administration of the project and some will relate to the stages in the process. Changes in culture are encouraging the re-use of existing data which means that data should be kept; discoverable and useable for the long term. For example, a scientist wishing to reuse data may have discovered the information about the data from a journal article; but to be able to reuse this data they will also need to understand information on the analysis done to produce the data behind the publication. This activity may happen years after the original experiment has been undertaken and to achieve this, the data digital object and its context must be preserved from the start.

Having ascertained that there is a need not only to capture context for a given digital object but also to preserve that context over the long term for effective preservation and re-use, SCAPE built on work from other projects on the notion of a Research Object which enables the aggregation of information about research artefacts. These are usually represented as Linked Data; thus RDF is used as the underlying model and representation, with a URI used to uniquely identify artefacts. For the SCAPE project, the focus of the research lifecycle is the experiment undertaken at the ISIS Neutron Spallation Facility. By following the lifecycle of a successful beam line application, we can collect all the artefacts and objects related to it, with their appropriate relationships. As this is strongly related to allocation of the resources of the facility, this is a highly appropriate intellectual unit for the facility; the facility want to record and evaluate the scientific results arising from the allocation of its scarce resources.



Some of the types of context for research data.

Preservation Policy

Preservation Policies are important as they support decision making, help to choose the right preservation actions and will support the quality assurance of these actions. The SCAPE policy framework consists of three preservation policy levels going from a high level abstract view of preservation within an organisation to more defined description of policy intent through to concrete applicable statements which can support automated workflow. One of the major outputs of this work is the Catalogue of Policy Elements [1] which describes topics which policy for preservation may be made.

Each policy element is described by a standard set of characteristics. These include a description, risks, stakeholders for the policy decision, some real-life examples and some questions to foster discussion. Together the characteristics should give the reader enough information about why the policy element is important. Feedback on this work is welcome on the wiki.

Hadoop as a parallel technology

We investigated the use of Hadoop and Taverna workflows to enable large scale migration of data from a local data format to a domain standard format (NeXus format) using existing tools. The results showed that although there are overheads in using Hadoop the parallelisation of the workflow shows an increase in performance compared to a non-parallelised version for a 1.1TB dataset containing small files, but there were I/O performance issues as the size of files increased. An important observation is consideration of the time taken to transfer the data onto and off the Hadoop filesystem. Hadooop expects the data to stay within the file system to get the most benefits from executing the application near the data.

What next?

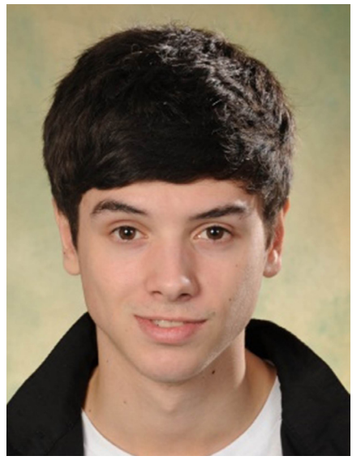
Work on the preservation of complex data objects will be taken forward within the STFC programme to deliver benefit to the Facilities we provide and we will also be exploring preservation policy in our own context.

Authors
C. Jones, A. Duncan, B. Matthews, A. Wilson ,
STFC Rutherford Appleton Laboratory

References
[1] http://wiki.opf-labs.org/display/SP/Policy+Elements

A YEAR WITH THE HIGH PERFORMANCE SYSTEMS GROUP

In 2014, for the first time in SCD at Daresbury, the High Performance Systems Group (HPSG) offered a 12 month Year in Industry studentship . The scheme is run by the Engineering Development Trust , and has worked very successfully in other groups at RAL. It was felt that there was an opportunity for a promising student to do some useful project work around collating and improving the internal documentation used by HPSG. An outline job description was drawn up and authority to recruit was both sought and granted. Following a number of interviews, pre-university student Karl Richardson was appointed to the role. This is the story of his year with HPSG in his own words.



Karl's Story

My name is Karl Richardson and I am currently on a 12 month placement with the Year in Industry programme at the Science and Technology Facilities Council (STFC), based in Daresbury. Prior to my placement I completed two years at sixth form studying IT, Business, Philosophy and Ethics. During my time at sixth form I enrolled into the Year in Industry programme as I knew it offered great opportunities that could give me vital work experience to improve my career prospects.

EDT offered me a number of opportunities to work with companies in various industry sectors all over the country. However, when I was offered the position at STFC Daresbury I could not refuse as I knew it would offer me great experience and a chance to gain a wide range of skills in a major high profile research council. The placement at Daresbury was also convenient and practical as I live within walking distance of the lab.

Following interview, and after being offered the post by my line manager Dave Cable, I was daunted by the role at first as I knew from my research that STFC is a huge research council specialising in a wide range of sciences, from laser technologies to computational chemistry. I was stationed within the Scientific Computing Department in the High Performance Systems Group, and I found that the terminology used by my colleagues was almost alien to me. The learning curve was steep. However, it didn't take long to settle in as both my line manager and colleagues in the group were incredibly helpful and approachable.

After a few weeks, when I had become accustomed to the building's layout and I had got to know some friendly faces, the job turned out to be not so daunting, but interesting, as I found myself involved in projects that I hadn't expected to be involved with. For example the implementation of the Hartree Centre phase 2 and energy efficient system, in which I was responsible for taking minutes, maintaining the actions list and then circulating the documents, round the project team.

Once I had settled in I spent a lot of time shadowing members of the team in order to become familiar with the various job roles (and skills) within the group. I wrote short biographies of my colleagues for use in grant applications and to improve my own understanding. I learnt how to use Linux as a graphical interface, as well as learning some basic navigation and useful commands.

During my 12 months at STFC I have also been involved with some public engagement work which was an unexpected bonus. This involved attending primary schools with other members of staff, and assisting with presentations. I took part in a teacher training event at Daresbury, based around using the Raspberry Pi computer. My public engagement experience also involved helping at the national Skills Show at Aintree, where I met up with a former mentor from the Career Academies programme that I had been involved with during sixth form. My mentor was intrigued by my academic experiences over the past two years and asked me to write an article on them for his Career Academies newsletter and website.

My role with HPSG is to seek out and improve internal group documentation and to enhance sharing of knowledge within the group. This has included tasks such as creating Standard Operating Procedures (SOPs) for the systems team, an important aid to ensure routine systems work is carried out effectively and consistently. I also act as secretary to a number of project and group meetings. This involves the minute-taking, maintaining action lists and generally sharing unwritten knowledge amongst relevant people.

I have been involved in various projects during the year. For example, the group-specific induction pack that I created for newcomers to the group. This is an idea that is now being copied by other groups in the department. The induction pack consists of a number of helpful documents for new starters such as relevant project codes, phone numbers, room names, site layout, departmental and group organograms, the short biographies of group members, induction checklists etc. The induction pack also contains a document I created which lists, explains and describes all the document repositories used by HPSG, along with information on how to access them. The induction pack proved to be an ongoing piece of work which involved a large degree of research and attention.

Another aspect of my job role is to maintain the HPSG webpage on the departmental website; this requires regular updates using a content management tool. My line manager often sets additional specific tasks for me to do, such as the repositories document. More recently I was invited to observe interviews for my successor. This was a valuable experience for me and by being there I was able to offer candidates the opportunity to ask me questions about my placement so far.

As a result of working with such highly skilled colleagues and undertaking many varied tasks, I have developed and learnt new skills which I believe are of great value to me. I now have more confidence in my ability to communicate effectively with other team members, and my listening skills have also greatly developed. I largely attribute this to my work as secretary to meetings, as I believe being able to listen and capture unwritten knowledge during a meeting and reproduce it accurately is a great skill that can only benefit me in the future. I also believe my adaptability to new and different working environments has been increased, as I was able to adapt within an alien working environment, and be able to work effectively.

A skill I have learnt and am currently developing is my understanding of the operating system Linux. As I previously mentioned I have been working through a Linux Essentials course, and reinforcing my knowledge with practical sessions on a Linux OS via VMware. This is something completely new to me and I believe could potentially open up many more opportunities.

After my 12 month placement here at Daresbury I plan to continue working and gaining more experience in different working environments, preferably business or sales orientated as I believe I have the personal drive and understanding to succeed in this line of work. Gaining more working experience will equip me with knowledge and understanding to improve my career prospects or put me in good stead for university.

Next steps

It's not only Karl who has benefitted from this experience. HPSG's somewhat ad hoc trial of the Year in Industry programme has led, through Karl's success, to a significantly more ambitious programme of work for his successor. It may also influence our future recruitment strategy.

We wish Karl all the best for the future, and hope that our next student is every bit as effective.

Authors
D. Cable and K. Richardson, STFC Daresbury Laboratory

STORING AND MANIPULATING ENVIRONMENTAL BIG DATA WITH JASMIN

JASMIN is a "super-data-cluster" designed to provide a high-performance high-volume data analysis environment for the UK environmental science community. Thus far JASMIN has been used primarily by the atmospheric science and earth observation communities, both to support their direct scientific workflow, and the curation of data products in the STFC Centre for Environmental Data Archival (CEDA). As a result of the early experiences of using JASMIN, discussed here and the improvements in scientific workflow that developed, during 2013-14 the Research Infrastructure Group of SCD completed the first part of a £7M upgrade to JASMIN for NERC, in collaboration with our colleagues in CEDA. Here we discuss examples of the scientific work flow based on JASMIN before its upgrade to highlight the experiences that drove the requirements for the upgrade.

The JASMIN super-data-cluster is the central node of a geographically spread environmental e-infrastructure. It deploys petascale fast disk connected via low latency networks to a range of computing services initially designed to serve a range of clients, but primarily those with "big data handling needs" from UK atmospheric and earth observation science.

The JASMIN core systems were installed at the STFC Rutherford Appleton Laboratory (RAL) in early 2012. The technical architecture was chosen both to deliver ease of management and to provide a very flexible high performance storage and analysis environment.

The JASMIN architecture before upgrade is depicted in figure 1. The system before upgrade essentially consisted of five major components:

1) The low latency core network (based on Gnodal switches but now migrated to Mellanox);
2) The Panasas storage sub-system;
3) The batch compute system ("Lotus HPC");
4) The data compute systems providing both bare metal compute and the hypervisors for virtual machines;
5) Two image stores to support the private disks of the virtual machines.

Before upgrade the system provided 600 cores and 5 PB of usable disk. The upgrade has added another 7PB of disk, ~3,000 cores plus a substantially redesigned network and a significant investment in Cloud Computing software from VMware

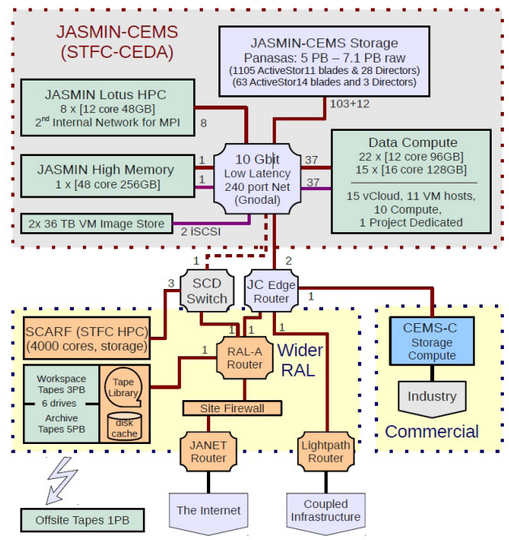


Figure 1: Key JASMIN components.

A. Data Migration Experience

While selected science users had access to JASMIN directly after installation, the major task within the first year had been to migrate the CEDA archive into the JASMIN archive.

The migration of 1.2 PB of existing archive data from legacy NAS storage to the new Panasas storage was nontrivial and took the best part of a year although the bulk was completed in 6 months. The bulk of the copy operation was performed by a batch queue based "copy service" that the Research Infrastructure group built for CEDA. It used two "worker nodes" using the Panasas pan_pcopy utility to perform parallelised copy from the legacy NAS. Performance (and number of copy nodes and threads used) during this step were limited by legacy network and filesystem issues. Bandwidths varied between 2 to 300 MB/s, leading to initial transfer times for some dataset/filesystem combinations of days to weeks. Once this operation completed for each partition, further rsync and bitwise comparison to tape was performed.

These checks were to detect (and fix) any instances of data corruption (before or after transfer). Approaching 100 old NAS servers were decommissioned after the migration to the single Panasas multi PB data store, saving power and management overheads.

B. Lotus Usage

Lotus was only made incrementally available to users through the fourth quarter of 2012 but even so completed its 1,000,000th job earlier in 2014. Users routinely submit suites of 10,000-30,000 processing jobs at once, usually representing one job per dataset file. Before the JASMIN upgrade this block of processing could take 4-5 days elapsed to complete, already a step change in capability, whereas after the upgrade this is already down to 2 days and we expect this to be less than 1 day when all the compute capability is brought online.

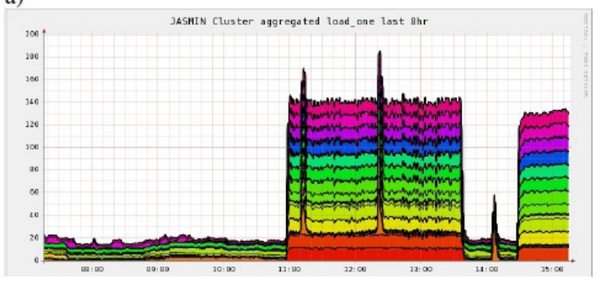
This highlights the difference in work loads between a super-data-cluster and a traditional HPC compute cluster. The super data cluster is working on 10,000's or millions of files whereas traditional workloads for an HPC cluster may consist of reading a small number of start up files, then expending significant computational time to calculate new results from that.

Typical Lotus jobs consume the entire small cluster, and would stress traditional NAS/NFS storage systems because of the number of files being accessed in parallel. Figure 2 shows an early small scale example of Lotus performance during a satellite reprocessing job. It can be seen that the I/O load from just one host is sustaining around 3 Gb/s read, so all 8 nodes would have been demanding 24 Gb/s from the same file server.

After upgrade JASMIN/LOTUS has around 200 hosts to sustain this per host data rate and our benchmarking in 2014 has shown the new part of the Panasas file system capable of delivering 140Gbytes/s, to keep up with this IO demand. The JASMIN upgrade also included a new non-blocking low latency network to allow these massive data flows, capable of currently ~3Terabit/sec. When the original JASMIN 5PB of Panasas storage is integrated into this network, our estimates show a combined IO capability of ~250Gbytes/sec, arguably making JASMIN one of the top ten systems in the world for IO capability.

The reprocessing job in Fig2 appears to have been compute bound on JASMIN, where it would have been I/O bound in a traditional data analysis environment. This observation clearly shows that if more compute capacity had been available, the job could have completed faster — however, even with the computing limit at that time, the suite of jobs completed approximately one hundred times faster than the previous analysis environment. In this case it has meant that whole mission reanalysis tasks that previously took months can be completed in days, completely changing the nature of what can be done with new analysis algorithm testing and evaluation.

a)



b)

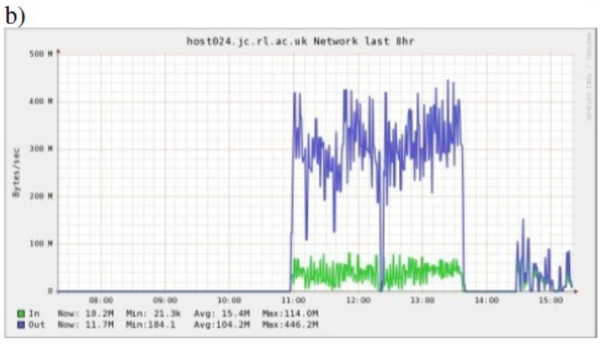


Figure 2: Lotus CPU and network performance during a small satellite reprocessing job.

C. The impact of highly parallel disk

In the original JASMIN architecture design we assumed that scientific users of the service would use virtual machines (VMs) with archive and group workspace access as replacements for servers or desktop computers in their own institutions. This has been the pattern, but as we have observed typical users, we have seen them exploiting workflows based on relatively low expectations for input/output.

Traditional NFS file servers are only able to serve a handful of processing threads, so many scripted work flows have grown up that serially feed these threads with new work. A common pattern is one machine or thread per temporal unit (often a calendar month) of a dataset, leading to utilisation of o(10) concurrent processing tasks. Access to a massively parallel file system allows orders of magnitude scale up in the number of parallel processing threads. To migrate users from traditionally scripted processing to use massively parallel I/O capability we needed a method of migrating users from a traditional server/VM environment to a higher throughput environment.
We achieved this by the use of extensions to a "virtual head node" concept first developed for the SCARF cluster by the Research Infrastructure Group.

Traditional high performance/throughput clusters have shared machines called "head" nodes. These allow users to login and submit jobs into the cluster but don't run jobs themselves. They provide users with editors, compilers and analysis tools. For resiliency the head node for the SCARF cluster had been virtualised several years ago, so in JASMIN we were able to take advantage of that to provide communities (or even users) their own virtual head node machine running a dedicated/customised full GUI desktop.

Unlike a shared use head node, users can install software and run code on it, both directly and via a batch queue. This allows users to test code and batch submission scripting without affecting the main physical cluster (or being limited by waiting times in the wider cluster queue). The VM batch queue is managed remotely by the physical cluster scheduler but initially only provides the internal VM resources. The VM batch queue can be easily extended to have access to the full computational cluster, providing significant increases in resources. Any additional system software requirements will have already been tested in the same environment as the rest of the cluster, so can be quickly rolled out.

Using this approach we have found that the sub-projects maintain ownership of their workflow in their VM, still have significant flexibility, but have that coupled with the ability to scale out to the wider cluster. This means that some workflows have already moved from o(10) to o(104) threads.

However before upgrade, Lotus and JASMIN had relatively small compute capacity, so users could not gain the full benefit of massive parallelisation. In the 2012-13 SCD report the RI group reported on work using the SCD SCARF cluster and the JASMIN archive, which demonstrated the need for the very high performance IO that JASMIN provides but with the compute capability that SCARF provides. This directly fed into JASMIN compute upgrade plans where the number of cores are now similar to SCARF but where the host IO and network capabilities come from LOTUS.

Conclusion

Some of the JASMIN communities have already built workflows suitable for clouds, but many have not. It is not yet clear whether the uptake of our batch compute reflects that balance, or that we simply did not have enough cloud resource, and so batch computing is more efficient from a user perspective.

Currently we also have an issue with our internal cloud in that we are unable to export our high performance storage into the JASMIN cloud since we cannot yet adequately constrain access. While we are working on solutions to this, in the interim this too will affect the balance of use.

Most of us believe that managed clusters like LOTUS or SCARF allow users to get the best performance and to get on with science rather than worrying about how to get a virtual cluster to work for them, but they are not universal views. There is a tension here between the constraints of working with a shared massively parallel file system and massively scalable compute. We believe the scale of the JASMIN upgrades will both allow us to serve the user communities, and to explore both sides of this debate.

Author
J. Churchill, STFC Rutherford Appleton Laboratory
This article is heavily based on the IEEE Big Data Conference 2013 paper of the same title: "Lawrence, B.N. , V.L. Bennett, J. Churchill, M. Juckes, P. Kershaw, S. Pascoe, S. Pepler, M. Pritchard, and A. Stephens. Storing and manipulating environmental big data with JASMIN. Proceedings of IEEE Big Data 2013, p68-75 doi:10.1109/BigData.2013.6691556" and presented by this author at the conference.

LHC DATA PROCESSING AT THE GRIDPP TIER-1

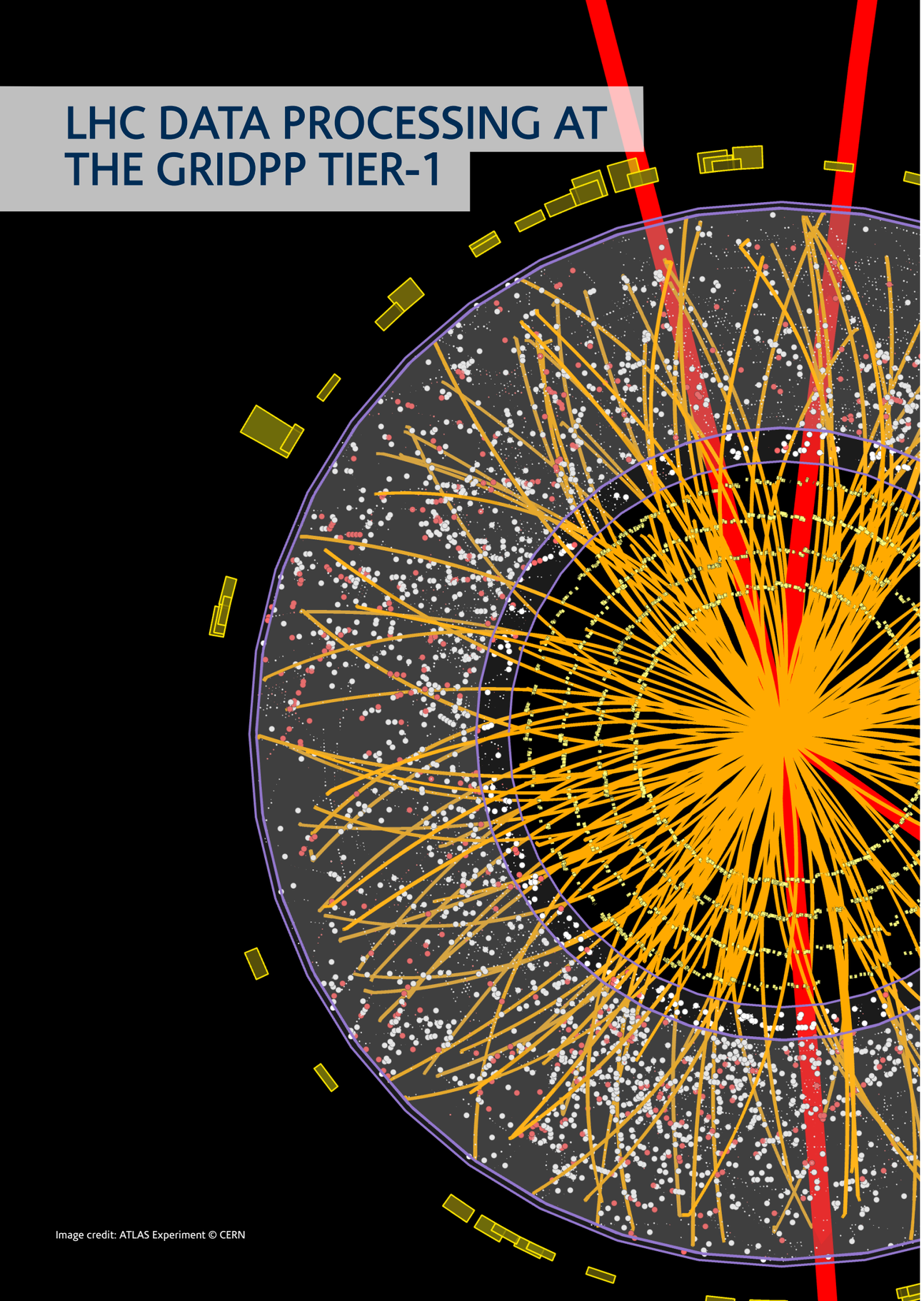


Image credit: ATLAS Experiment © CERN

The Discovery of the Higgs Boson and the 2013 Nobel Prize for Physics

The CERN Large Hadron Collider (LHC) switched off at the end of "Run 1" on the 16th February 2013. The many Petabytes of data taken were rapidly processed and analysed on the Global LHC Computing Grid of which the GRIDPP Tier-1 Centre at RAL is a major partner.

The Tier-1 provides over 10 Petabytes of disk storage spread over 500 disk servers and another 10 Petabytes of tape storage in an SL8500 tape robot. The data is received from CERN on a dedicated 10Gb/s Optical Private Network (OPN) and is then available to be processed on a compute cluster of 10,000 cores. The data processing challenge was huge, but on the 4th July 2013 the ATLAS and CMS experiments announced that they had found the long sought after Higgs Boson predicted by the Standard Model of particle Physics as necessary to explain why some elementary particles have mass.

On the 8th October 2013 the Nobel Prize for Physics was awarded jointly to François Englert and Peter W. Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider". The Tier-1 team at RAL were very proud to have played their part in the work of the experiments which helped confirm this discovery.

Preparing for LHC Run 2

The LHC is being upgraded as it prepares to restart data taking for "Run 2" in early 2015. When it restarts it will deliver higher energy beams than ever before and data taking will then continue at much higher rates than we have seen to date, until the next long shutdown in 2017. Although the LHC is currently switched off, the experiments have not been idle as they not only continue analysing existing data but also begin the challenging task of modelling the behaviour of the LHC and experiments at the higher beam energies planned for Run 2. Data rates on the Tier-1 have recently far exceeded those seen during the big 2013 push for the Higgs discovery, peaking at over 1 Petabyte a day in recent weeks.

In order to meet the challenges of Run 2 the Tier-1 is pushing forward essential developments. In order to meet the challenges that higher data rates will bring the network backbone was recently upgrade to 40Gb/s. By 2018 over 70 petabytes of tape storage will be needed and so the robot's older T10000B tape drives (having a capacity of 1TB per tape) have been replaced with the latest generation (T10000E) which have an 8TB tape capacity. The disk system is planned to grow annually until by the end of Run 2 in 2018 it will exceed 20 petabytes. Not only are capacities increasing considerably but also capabilities and the Tier-1 has development projects underway to deploy a private cloud interface and high capacity object store in order to meet the needs of a wider range of emerging high capacity science projects expected in the coming years.

Author
A. Sansum, STFC Rutherford Appleton Laboratory

HARNESSING PHYSICS: THE CHALLENGE OF EFFICIENTLY SOLVING OPTIMIZATION PROBLEMS WITH PDE CONSTRAINTS

Many important engineering problems can be formulated as optimization problems where one of the constraints is a partial differential equation. One of the main bottlenecks in current algorithms is the solution of a large linear system of equations. By understanding and using the underlying structure of the problem researchers in the Numerical Analysis group have developed efficient fast iterative solvers for such systems, enabling problems to be solved which are larger, more accurate, or with more realistic models.

The laws of nature are written in the language of differential equations, and our desire to understand the world requires us to solve ever more complex partial differential equations (PDEs) over complicated geometries. This need has been the driving force behind much progress in the field of numerical analysis since the early 20th century. Today, developments in algorithm design and advances in scientific computing mean that problems that were intractable thirty years ago are now routinely solved using a laptop. However, especially in the field of engineering, the aim is to harness our understanding of the laws of physics in order to shape the world around us to meet our desires. The field of the optimal control of PDEs provides a framework for doing this algorithmically. Here we formulate the problem as minimising a cost functional subject to constraints, namely the PDEs that describe the physics we wish to control.

For example, suppose we wish to apply some ideal heat profile over a space; a specific instance where this occurs is in some modern cars, which permit individual passenger temperature control. The state of the system, namely the temperature distribution, is a solution of the heat equation. We can affect the solution by changing the boundary conditions, for example by adjusting the warmth of the air blown into the system from vents distributed around the edge of the space. The optimal (boundary) control problem here is to find what temperature to blow into the system - i.e., what boundary conditions to apply to the PDE - that gives a state closest to our ideal. An additional complication is that we need to solve the problem subject to additional constraints on the control (since there is only a finite range of temperatures the blowers can handle) and on the state (as the ambient temperature should never be too extreme at any point in the space). The latter can be particularly challenging to satisfy numerically.

The “all-at-once” or “one-shot” approach has recently become popular for solving such problems. Here we discretize the PDE, typically using finite elements, and the solution of the control problem reduces to solving a linear system - or a series of similar linear systems if we have bound constraints or a non-linear PDE - with a block structure. Such a system is very large; there are typically three times the number of unknowns as in a forward PDE solve. However, the underlying grid structure ensures that the component blocks are all sparse, and we avoid explicitly forming this matrix.

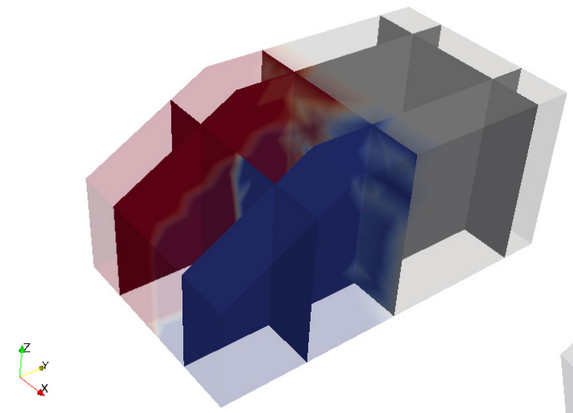


Figure 1: Desired state.

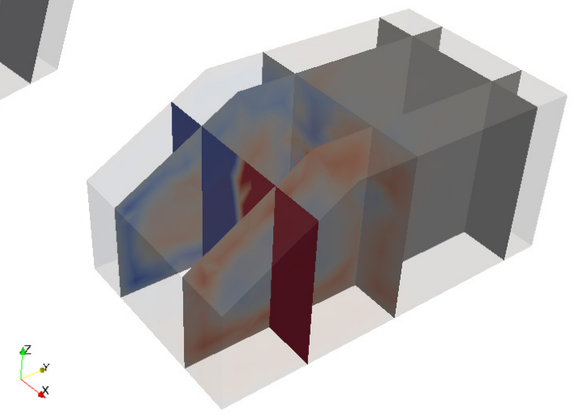


Figure 2: Optimal state.

Due to the size and properties of the linear systems involved, we are unable to take advantage of the sparse direct solvers that the HSL library is renowned for, particularly when solving three dimensional problems. We therefore turn to iterative methods -- in particular, the class of methods known as Krylov subspace methods. Such methods approximate the solution by finding the best candidate in a space built up by repeatedly applying matrix-vector products with the coefficient matrix.

In order for Krylov subspace methods to be efficient they need to be applied together with an appropriate preconditioner, that is, a linear transformation which in some sense reduces the conditioning of the system, and hence accelerates convergence. Finding an efficient preconditioner is highly problem-specific, and some members of the Numerical Analysis group have been have been working on the development of a new preconditioning paradigm for problems in the optimal control of PDEs.

A foundational component of this approach is to treat the linear system in terms of its component blocks, not as a whole. This allows us to utilize techniques such as multigrid, which have been highly successful in the solution of the forward problem. By combining these tried and tested techniques with novel ways of approximating mass matrices we have developed preconditioners which are optimal, in the sense that they scale linearly with the number of degrees of freedom and are independent of modelling terms such as the regularization parameter.

A distinguishing feature of our approach compared to the reduced-space methods that are commonly used is that we never have to solve a PDE (or its adjoint) exactly; all that is needed is the application of a preconditioner for the PDE, e.g., a multigrid V-cycle. This allows us to solve PDE-constrained optimization problems very efficiently; it typically takes only ten to twenty times longer to solve the much harder control problem than it does to solve the equivalent forward problem.

These preconditioning techniques have been developed in conjunction with a full theoretical study of their behaviour, including a characterization of the eigenvalues of the preconditioned system. An analysis of their behaviour has now been done for a range of PDEs, both stationary and time-dependent, and for a number of different regularization terms. There is more work to be done in this field, particularly solving problems with non-linear PDEs or coupled PDEs describing multi-physics simulations, but we have shown that by understanding and respecting the underlying mathematics we are able to get faster linear system solves. Such solves are the bottleneck in PDE-constrained optimization solvers, and by reducing the time needed here we get more efficient ways of using technology, driven by algorithms, to harness our understanding of physics in order to shape the world around us.

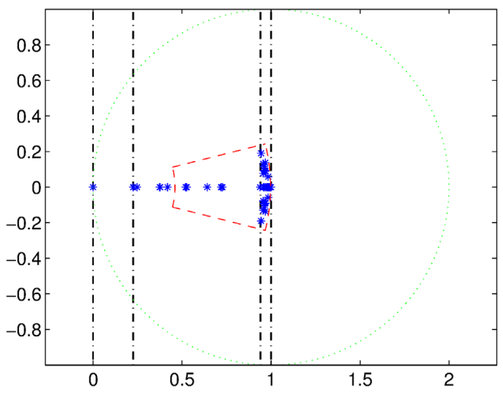


Figure 3: Eigenvalues and bounds for a flow control problem.

Author
T. Rees, STFC Rutherford Appleton Laboratory

References
The work of the Numerical Analysis Group is largely funded by a research grant from EPSRC (EP/I013067/1).
Tyrone Rees, H. Sue Dollar, and Andrew J. Wathen. Optimal solvers for PDE-constrained optimization.
SIAM Journal on Scientific Computing, 32(1):271–298, 2010.
Tyrone Rees and Andrew J. Wathen. Preconditioning iterative methods for the optimal control of the Stokes equation.
SIAM Journal on Scientific Computing, 33(5):2903–2926, 2011.
H. Sue Thorne. Distributed control and constraint preconditioners.
Computers & Fluids, Issue 1, 461-466, 2011.
Andrew T. Barker, Tyrone Rees, and Martin Stoll. A fast solver for an H1 regularized optimal control problem. 2013 (under review).

NEXT GENERATION ATMOSPHERE AND OCEAN MODELS

The Met Office's numerical weather prediction and climate model code, the Unified Model (UM), is almost 25 years old. Up to the present day the UM has been able to be 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, with performance increases from each new generation of computers now being primarily provided by an increase in the amount of parallelism rather than an increase in the clock-speed of the processors themselves, running higher resolutions of the UM now faces the double challenge of code scalability and numerical accuracy.

The UM's atmospheric dynamical core makes use of a finite-difference scheme on a regular latitude-longitude mesh. The regular latitude-longitude mesh results in an increasingly disparate grid resolution as the mesh resolution increases due to lines of longitude converging at the poles, see the left hand image in Figure 1. For example, a 10km resolution at mid-latitudes would result in a 12m resolution at the poles. The difference in resolution leads to increased communication at the poles and load balance issues which are known to impair scalability; it also leads to issues with numerical accuracy due to the difference in scale.

The GungHo Project

To address this problem the Met Office, NERC and STFC initiated the GungHo project. The primary aim of this project is to deliver a scalable, numerically accurate dynamical core. This dynamical core is scheduled to become operational around the year 2020. The project is currently investigating the use of quasi-uniform grids, such as the icosahedral and cubed-sphere grids (see the right hand image in Figure 1) and finite element methods.

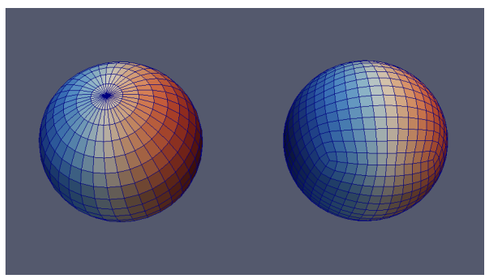


Figure 1: Illustration of a regular lat/lon grid and a cubed sphere grid.

The associated GungHo software infrastructure is being developed to support multiple meshes and element types thus allowing for future model development. GungHo is also proposing a novel separation of concerns for the software implementation of the dynamical core. 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.

Rather than writing the PSy layer manually, we 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), or alternatively, generate the PSy layer automatically.

PSyclone is currently able to generate correct PSy code for version 0.1 of Dynamo - the name given to GungHo's implementation of the dynamical core. PSyclone also provides loop fusion and OpenMP loop parallelisation transformations to help optimise Dynamo. We are continuing to work on PSyclone as the Dynamo software matures, to support any updates to the Dynamo API and to increase the number of transformations available.

The GOcean Project

The GOcean project, a NERC-funded collaboration between STFC and the National Oceanographic Centre (NOC), was initiated to determine whether the PSyKAl approach would also be useful for Ocean Models. This is particularly timely since NEMO (the ocean model used by the Met Office and many other organisations around the world) is not well-placed to take advantage of the proliferation of light-weight cores found on many current accelerators. Therefore, the results of this project will be important in helping the NEMO community decide on how to develop the code in the future.

Ocean models can be implemented using either a regular or irregular mesh. However, they do not suffer with the polar problems faced by atmosphere models since meshes can be constructed where the poles are placed over land. This means that a regular latitude-longitude mesh with its attractive numerical properties can still be used, even at high resolution. Since the Gungho project is looking at a finite-element approach on a quasi-uniform mesh, GOcean decided to investigate the PSyKAl approach for the case of finite-difference on a regular mesh.

To this end, two existing codes are being re-structured following the PSyKAl rules: the first, a shallow water model developed at the National Centre for Atmospheric Research in the US and the second, a code written by NOC which captures some of the main features of NEMO. An example of the output of the latter for a simplistic model of a tidally-forced estuary is shown in Figure 2.

Although relatively short and therefore simple to re-mould, these codes share many characteristics (mesh setup, numerical kernels, application of boundary conditions) with their larger, more complex cousins. So far, the PSyKAl approach appears to fit well. It has been possible to break up both of the models into kernels which operate on a single grid-point while making the top, Algorithm layers cleaner and simpler to understand. Further, the PSyclone system has been extended to support the generation of the PSy layer for an early implementation of the shallow water model.

The codes have also served to demonstrate the science-neutral choices that a developer can make when implementing a model of this type. We are working to extend the Gungho kernel meta-data specification in order to capture this information. Once the choices that have been made are explicitly specified, it will then become possible for the PSyclone code-generation system to also perform a number of correctness checks.

While the re-structuring of these two codes has gone well, we are just beginning to investigate the implications for performance. Oceanographers are unlikely to adopt a new approach to writing their models if there is a performance penalty (as is the case for Atmospheric modellers). The aim now is to ensure the Psyclone system supports any changes to the API (such as changes to the kernel meta-data specification) and to further develop the PSyclone system in order to obtain the same, or possibly even better, performance as the original codes on a range of parallel systems.

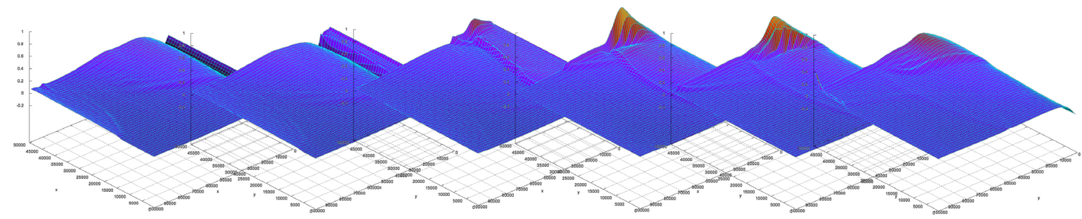
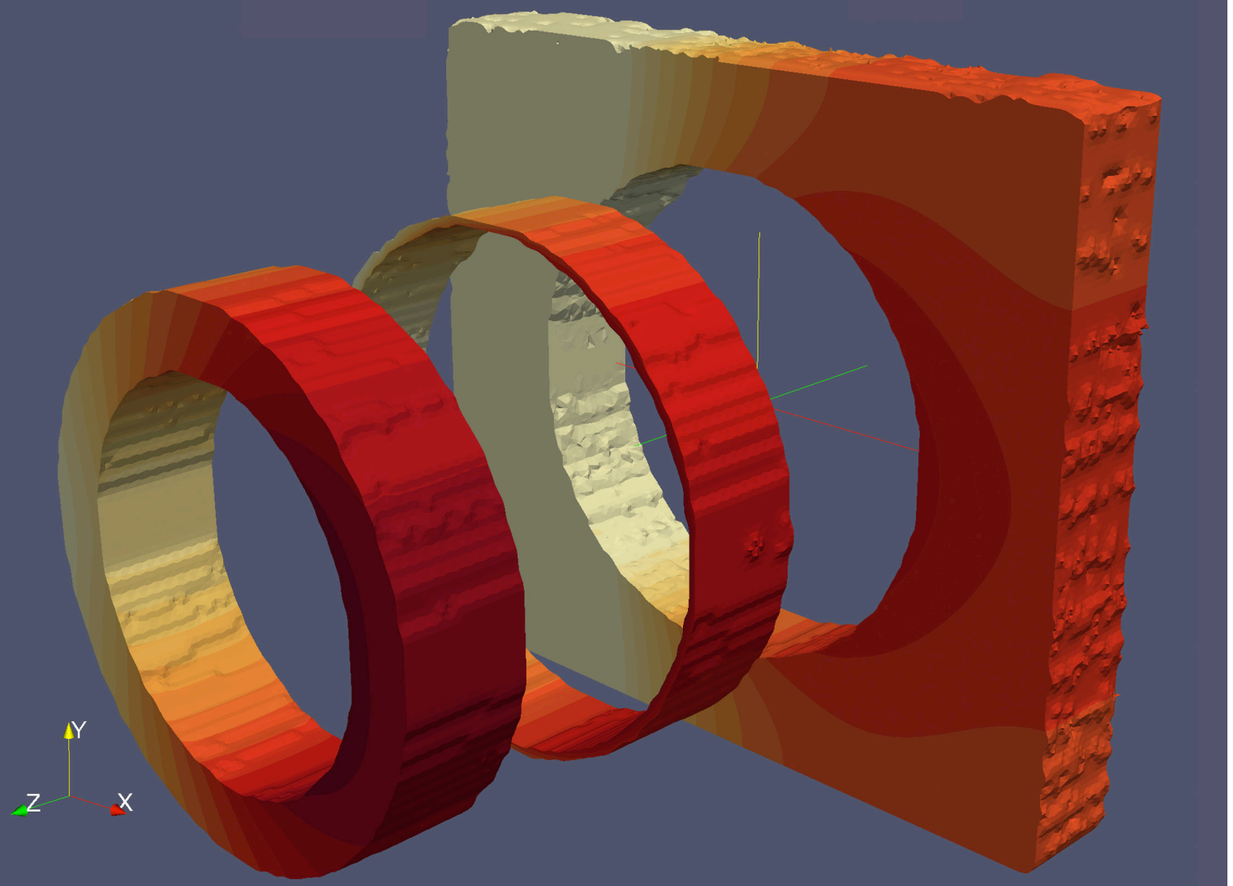
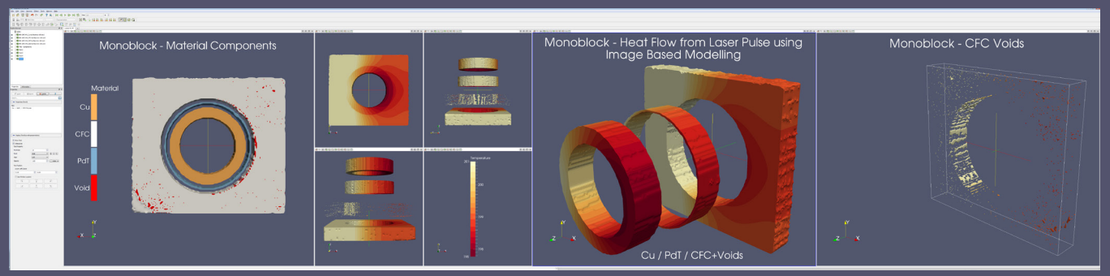


Figure 2: Output of the NOC test code.

Authors
M. Ashworth, R. W. Ford, A. R. Porter and G. D. Riley, STFC Daresbury Laboratory

PARAVIEW EVERYWHERE: DIFFERENT WAYS TO TRANSMIT AND VISUALISE SCIENCE





SCD contracted Dr Lee Margetts and Ms Louise Lever at the University of Manchester to help add transient thermal modelling capability to the open source parallel finite element software ParaFEM. This enabled Dr Llion Evans at the Culham Centre for Fusion Energy and his team to use it to study plasma facing wall components. Videos of this work are available on YouTube; www.llion.co.uk/youtube

Visualisation Matters – the need to put the human back in the computational loop

Survey Time

Over the last year the visualisation group, part of the Technology Division within SCD, has reconsidered the real visualisation needs of the computational sciences community. A series of informal and formal surveys are underway and the first ones have tackled the Tomographic Imaging and CFD (Computational Fluid Dynamics) communities. These wanted to find out which tools were actually being used and the best methods to support them. https://www.surveymonkey.com/s/97KD2V5

Surprisingly, open source was not always the most important issue but the easy creation of plug-ins, new readers and writers, as well as analysis tools have been requested. There was also indicated a strong growth in the use of the ParaView visualisation system (http://www.paraview.org/) that is an open source, multi-platform data analysis and visualisation application where users can build systems including adding qualitative and quantitative techniques.

We will consider some usage over the past year and future cross-facility projects being developed:

Use Case 1: Hartree development

One of the best way to understand and explain complex high resolution data, is to show it on a very high resolution display and walk through and around a problem; ParaView's stereo display modes, efficiently allow such immersion with the data to be made possible. Within the Hartree visualisation suites we have explored multiple modes of ParaView usage including; remote interaction using an IPad on a client system; kiwi viewer and ParaViewWeb for mobile devices; and the new Catalyst method for HPC specific network visualisation parallelism.

The ParaView code base is designed in such a way that all of its components can be reused to develop vertical applications and it has been successfully deployed on an IBM Blue Gene system. It is also being used within development projects and user communities; within the ISIS 'Mantid' data flow system and for volume visualisation within the RCaH community (new tools include http://www.tomviz.org/).

Figure 1: Different users using ParaView; clockwise from top left, Code_Saturne monitoring of remote visualisation for HPC data runs; LiDaR data evaluation and visual focussed qualification for smoothed particle hydrodynamics for Stephen Longshaw (School of Mechanical, Aerospace & Civil Engineering at the University of Manchester); Remote server from client end for IPad like control panels for simple interactions; multi-modality and stereoscopic display for Matthieu Chavent (University of Leeds Researcher in Molecular Visualisation), Biophysical Modelling, Biological Complexity and Jose Arregui-Mena's (University of Manchester) ParaFEM simulation of a brick from a Nuclear Reactor core, showing tens of realisations at the same time (rfem.blogspot.co.uk).

The other two popular tools for tomography users were Avizo and ImageJ that we run licences and training courses for and now create wrappers for the reconstruction, image filters and some segmentation methods as part of the CCPi core activities: http://visualisationmatters.com/ http://www.ccpi.ac.uk/

Use Case 2: The human-in-the-computational-loop: still important for supercomputing

An important role for ParaView is to allow crude computational steering within large computational batch runs. Although systems like the IBM Blue Gene/Q are highly ranked on the green computing list (STFC system is #26 on the June 2014 Top500 Green List at 2,178 mflops per watt) the computational cycles are not free and still limited to a certain size of cores. Therefore users have to be careful in selecting the right parameters when running a set of jobs. Currently, the Hartree system's GPFS large file store is connected by fibre to the visualisation suites that has enabled a dynamic monitoring method for looking at long computational jobs easily to be developed.

Figure 2: Training and steering example: Stefano Rolfo and Charles Moulinec presenting and running CFD code in the Hartree training room, and then in the Hartree Visualisation Facility.

So for example in the CFD (Code_Saturne) example there were multiple very long jobs submitted, with each job taking many hours to run but every five to ten minutes an intermediate file was produced that could be viewed by ParaView. This enabled a scenario where users could return to the visualisation suite with a cup of tea, and monitor the outputs, interacting and investigating the data sets. If everything was going well then the jobs could be left to complete, otherwise selective jobs could then be terminated and restarted with different parameters. This ability to put the human back in the computational-loop can be incorporated very simply, but can easily save a huge amount of computational cycles and energy.

Use Case 3: Software for the Future: integrating ParaView and OpenFOAM

The open source CFD packages OpenFOAM and foam-extend are being used as the basis for developments in a SoFT (Software for the Future) project for new marine structure simulations. These large scale simulations require visualisation of big data sets generated on parallel systems and ParaView provides many of the important visualisation techniques needed for CFD simulations and also has features that address large scale parallel data. One of the most useful aspects of ParaView is its ability to run in client-server mode so that the data of a simulation can remain on the parallel file system it was generated on and only the required visualisation geometry is sent to the client. This avoids the need to transfer many GigaBytes of data to the client, which can be slow and inconvenient.

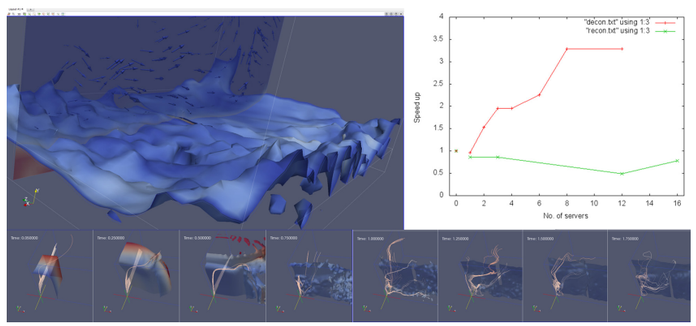


Figure 3: Top left: water-air interface of 3D dam break simulation, showing one time step of a simulation using a mesh of 204943 nodes and creating about 5GB of data. Flow vectors are shown on a slice. This transient case has been used for early benchmarking for small sets of visualisation server nodes. Client server mode using the distributed data (top right red) gives better visualisation performance than reconstructed case (green). Bottom figure shows surface and stream lines through time.

Benchmarking visualisation processes is more difficult than benchmarking computational processes since there are many more factors to consider such as the speed of network connection, the local graphics system and the level of detail required in the rendered image. In the rest of this project ParaView work will continue to be incorporated including the Catalyst extension that will generate visualisations of simulations without having to store complete data from every time step of a simulation.

Future Use Case 4: IMAT beamline: linking the human-visualisation to the facility imaging capture process

The IMAT (Imaging and Materials Science & Engineering) beamline, due for commissioning in 2015, is going to be the first neutron imaging instrument at ISIS that will offer unique time-of-flight tomography-driven diffraction techniques. This will capitalise on the latest image reconstruction procedures and event mode data collection schemes involving a strong visualisation component. ParaView has been tested and will incorporate fast parallel image reconstruction algorithms with on-the-fly image processing and visualisation to inform and guide experiments.

This project will not only bring the human into the computational loop but also bring the human and visualisation system within the neutron data capture and computation loop in a semi-real-time process.
This is an exciting advance as its aim is to make efficient use of both computation time as well as experimental neutron beamline time; with visualisation being key.

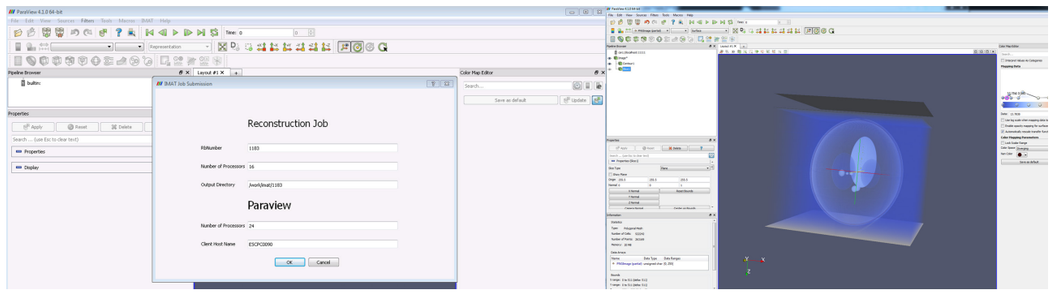


Figure 4: Phantom data generated for benchmarking the IMAT project reconstruction and visualisation (parallel rendering). Portal shows how to submit a reconstruction job and ParaView server job from ParaView client.

Conclusions and Alternatives

ParaView is a powerful visualisation tool that is well adapted to HPC applications and is under ongoing development. There are software alternatives available, and some are being actively used within the visualisation group; but looking at one product and its uses across a year gives us an interesting snapshot of highlights. This has just been the start of our exploration and as other opportunities arise we are willing to rise to the challenge.

This project is bringing together staff from the Technology, Data Services and Systems Divisions within SCD, as well as staff from the IMAT beamline and Mantid (http://www.mantidproject.org/) software development team in ISIS; project managed by Erica Yang.

Authors
M. Turner, STFC Daresbury Laboratory, R. Fowler, S. Nagella, STFC Rutherford Appleton Laboratory

References

Survey main site: http://visualisationmatters.com/ https://www.surveymonkey.com/s/97KD2V5
Hartree Centre: http://www.stfc.ac.uk/hartree/default.aspx
Visualisation Group webpage: http://www.stfc.ac.uk/SCD/research/tech/vis/43937.aspx

THE HARTREE CENTRE - BEING ESSENTIAL TO INDUSTRY AND BEYOND

This year marked our first full year of operation since the Chancellor of the Exchequer George Osborne opened the Hartree Centre officially in February, 2013. Much has been achieved in that time and the Centre is already being recognized as a new force for delivery and change within our industry.

Shortly after the opening of the Hartree Centre, we were delighted to welcome the children of the late Douglas Hartree to our Centre and share with them our exciting plans, that are now very much coming to fruition.

Phase 1 of our activities covers two specific areas:

- Modelling & Simulation
- Software development for the future

The Centre's original capital funding model was based on providing economic impact through engagement with industry. The Hartree teams have worked with over 30 industrial clients large and small to deliver on projects where we have helped them design better products and reduced the time and cost of their development cycle, by using our compute assets, software and talent. There is a huge diversity in the products we have worked on and the following are just some of the examples:

- Detergents
- Vacuum fans
- Battery technologies
- Cars
- New Polymers
- Transportation systems
- Insurance risk management

In recognition of the value we have added to their design wokflows, a number of our clients are now undertaking their 2nd, 3rd or 4th projects with us.

Through the Hartree machine access programs, we are delighted to have also been able to work with over 30 UK based Universities, helping researchers in those to get better outcomes from their science using Hartree assets. Amongst these users, a Nobel Laureate.

The evolution of software used in computational science and HPC, is essential to the opportunities and benefits we seek to deliver to our clients and partners. At the same time there are pressures for change in how this software is architected:

- New demanding challenges e.g. Big Data
- New power efficient architectures

Our software development activities currently include:

- Software for next generation Radio Telescopes (SKA)
- Next generation Weather Modelling (Met Office)
- Meso Scale formulation software (for Intel Phi)

We were delighted to be granted an additional £19M of capital investment from Government, specifically for research and outputs in the area of Energy Efficient Research Computing architectures and software. After extensive review, we now have live one of the most architecturally diverse hardware estates anywhere in the world. Work on these platforms by Hartree Researchers and others from Industry, Research and academe, will help define the architectures of the future making the challenges of 'scale' essential to the development of future supercomputers sustainable. This alongside, our new work with Big Data Analytics, covers Phase 2 of the Hartree Centre evolution.

The possibilities for Industry to work with Simulation & Modelling are almost limitless and across all industry sectors. At the end of 2013 we defined 4 clear commercial service offerings that embraced what we can do for our clients in this area.

- Collaborative R&D – Combining our domain skilled experts with, software and machines, to create an outcome.
- Software & Algorithms – Code development for new workloads and platforms
- Training & Skills – Intermediate and expert training for client research staff
- Platform as a Service – Timeshare access to our compute assets

Everything is going to change

There are 3 forces providing opportunities and challenges for what might be known as the Supercomputing sector.

- Power – Both availability and affordability
- Big Data – Extracting insights from increasingly large and available data sources
- Democratisation – Providing broader access to the capabilities delivered through Supercomputing

Hartree is playing a major role in how these forces of change are harnessed, exploited and ultimately delivering value to Industry and Society.

Massimo Noro from Unilever.

Our activities addressing the challenges of Power have been mentioned already.

Earlier in 2014, Hartree installed its first and one of the most sophisticated Big Data Analytics Clusters in Europe. This cluster can look at 'rested' data as well as real-time data from sensor networks., analysing trends and providing predictive feedback. We aim to learn significantly from the project work we undertake on this cluster, which will help provide us with the know-how needed to define some of the future architectures that will be demanded in this area, as data size and availability of data grows exponentially.

Broadening the use of these systems requires a sufficient skill pool to exist. Numerous summaries point towards skill shortages in most western economies in the areas of HPC and Big Data. We are excited to launch in collaboration with the University of Liverpool in October 2014, an MSc focussed at the applied use of HPC and Big Data, and we believe this will be helpful in addressing the skills issue, but not the entire answer. Building on the work we have done in conjunction with Unilever and IBM, we intend to user interface design approach, addressing the skills issue with simpler user interfaces for better accessibility.

Delivering on our objectives for Hartree requires greatly on an eco-system of great collaborative partners. Our partnerships mentioned in this report last year with IBM and Unilever have grown and are delivering real value. We also have new partners.

We were delighted to be chosen as an Intel Parallel Computing Centre earlier in the year. Continuing with our partnership theme of adding 'scale' to our own compute capacities, Hartree's Michael Gleaves travelled and to China and was accompanied by the Minister of State for Business Innovation and Skills, Dr Vince Cable, when signing an agreement with Guangzhou who operate the World's largest SuperComputer Tianhe-2.

Vision

We have made tremendous strides in our first full year of operation, achieving or exceeding all the goals and aspirations that were set for the Hartree Centre. We remain very focussed on delivering the vision that we have set out for ourselves covering the next 5 years.

- Through our capabilities, know-how and talent, be viewed by Industry as 'essential' and the 'go to' partner for their Research & Development activities, by adding significantly to their competitive advantage.
- To be acknowledged as being at the very pinnacle of thought leadership, defining, developing and delivering 'Intense-Computing', allowing our partners and clients to seize the profound opportunities emerging from a data enabled World.

Author
C. Brereton, Hartree Centre

Contacts and Further Information

Mr David Corney	Director of the Scientific Computing Department	david.corney@stfc.ac.uk
Mrs Laura Johnston	Personal Assistant	laura.johnston@stfc.ac.uk
Mrs Shirley Miller	Personal Assistant	shirley.miller@stfc.ac.uk
Mrs Jean Pearce	Personal Assistant	jean.pearce@stfc.ac.uk
Mr Damian Jones	Administrative Assistant	damian.jones@stfc.ac.uk
Mrs Karen McIntyre	Administrative Assistant	karen.mcIntyre@stfc.ac.uk
Mrs Esme Williams	Administrative Assistant	esme.williams@stfc.ac.uk

Division Heads

Dr Juan Bicarregui	Data Division	juan.bicarregui@stfc.ac.uk
Mr Cliff Brereton	Director, Hartree Centre	cliff.brereton@stfc.ac.uk
Dr Peter Oliver	Technology Division	peter.oliver@stfc.ac.uk
Dr Andrew Sansum	Systems Division	andrew.sansum@stfc.ac.uk
Dr Paul Sherwood	Applications Division	paul.sherwood@stfc.ac.uk

Group Leaders

Dr Mike Ashworth	Applications Performance Engineering	mike.ashworth@stfc.ac.uk
Mr David Cable	High Performance Systems Group	david.cable@stfc.ac.uk
Prof Dave Emerson	Engineering and Environment	david.emerson@stfc.ac.uk
Mrs Debbie Franks	Library Services	debbie.franks@stfc.ac.uk
Prof Chris Greenough	Software Engineering Support Centre	christopher.greenough@stfc.ac.uk
Mr Nick Hill	Research Infrastructure	nick.hill@stfc.ac.uk
Dr Jens Jensen	Data Services	jens.jensen@stfc.ac.uk
Dr Brian Matthews	Research Data	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
Dr Andrew Sansum	Petascale Computing and Storage	andrew.sansum@stfc.ac.uk
Dr Jennifer Scott	Numerical Analysis	jennifer.scott@stfc.ac.uk
Dr Martin Turner	Visualisation	martin.turner@stfc.ac.uk
Dr Martyn Winn	Biology and Life Sciences	martyn.winn@stfc.ac.uk

Websites

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

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

Sci-Tech Daresbury, Daresbury, Warrington, Cheshire WA4 4AD, UK
T: +44 (0) 1925 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: Isaac Newton Group, La Palma:
Joint Astronomy Centre, Hawaii.

Science & Technology
Facilities Council