Science and Technology Facilities Council

Scientific Computing Department

Annual Review 2020

Contents

Scientific computing is fundamental to modern research. This broad and rapidly-advancing field involves exploiting advanced computing capabilities to understand and solve complex problems in science. UKRI-STFC's Scientific Computing Department (SCD) is one of the UK's leading centres of expertise in data-intensive science, and home to sophisticated high-performance hardware.

It is one of the largest departments of its kind with over 200 computational scientists, software engineers and project support staff – and it is steadily growing as the demand for more research data management, storage and analytics increases at pace.

Our staff are located at two UK campuses – the Rutherford Appleton Laboratory in Oxfordshire, and the Daresbury Laboratory in Cheshire. They have cutting-edge skills and expertise in scientific software research and development, and world-leading capabilities in 'big data' storage and analysis, visualisation and simulation, and scientific information management.

SCD supports some of the UK's most advanced scientific facilities and provides the tools that enable the scientific community to discover and deliver vital research. We are advocates of Open Science, making research outputs available to encourage a swifter route from research to innovation, and new products and services that benefit people.

"Everything we touch and know and use can be attributed to somebody who can likely call themselves a scientist and, for the rare few, it offers a chance to leave a legacy that can last through centuries."

Stephen Longshaw,
Principal Computational Scientist, STFC

Foreword

Welcome to the 2020 annual review of the UKRI-STFC Scientific Computing Department.

To say that this year was unusual and challenging is perhaps stating the obvious but, with the appearance of the COVID-19 pandemic and subsequent restraints, some of our priorities and ways of working had to change at very short notice. In spite of this our work to support research had to continue so, for most of us, it was a case of carrying on with business as usual, albeit from a distance.

Our staff are very versatile and keen to share their skills, so at the start of the pandemic we joined the international research community efforts to find inhibitor drugs and vaccines by providing access to computing resources and expertise through some of the long-running Collaborative Computational Projects. We were also able to provide substantial computing capability and software to support the UK National Facilities in their COVID work and, like so many others, the highly successful Folding@Home project. Some of this effort is highlighted within this Annual Review.

We are addressing the issue of climate change, not only through the types of research we carry out and support, but also by driving forward decisions to actively support UKRI's Environmental Sustainability strategy. Our work this year has included using computer modelling to aid development of new, sustainable materials that will help the environment by using less power and reducing the need to consume natural resources; we have started to explore how the DAFNI platform can begin to support the National Digital Twin agenda, and we have introduced a programme of continuous improvement in our work with information and operational security management.

Encouraging the next generation of potential scientists remains a high priority and many of our activities for young people and school groups have been adapted for online workshops. In doing this we have been able to extend our reach to include students from much further afield, so the online sessions are likely to remain a fundamental part of our public engagement activities.

Remote working arrangements were extended to the many conferences and workshops we run. The success of these – with many more people attending than we would usually have seen – has really made us think about how we might handle future in-person events. Although we hope to return to some physical events in 2021, adding a 'virtual' option will enable people to take part who normally would have been excluded due to caring responsibilities, cost or time constraints.

We look forward to whatever new challenges 2021 brings and we hope to see some of you in our travels, either virtually or in person.

Tom Griffin,
Director

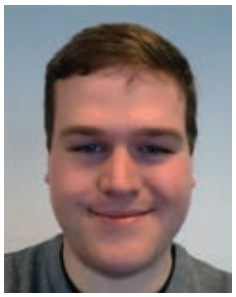
Skilled people

Dedicated people, skills and careers

The dedicated and talented staff working in STFC's Scientific Computing Department (SCD) come from very diverse backgrounds. Some are mathematicians, or started their careers in physics or biology before switching to a role supporting computational science. Others are communicators, administrators, or caretakers and managers of research data.

They all have a wealth of expertise and knowledge which they share with colleagues, and with the graduate students and apprentices they train and mentor each year. This knowledge is a valuable asset for the many projects and experiments being carried out within some of the UK's national facilities – such as the ISIS Neutron and Muon Source, the Diamond Light Source and the Central Laser Facility.

Here we introduce you to just a few of our very skilful staff, graduates and apprentices.

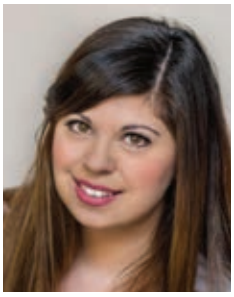


Stephen Longshaw is a Principal Computational Scientist who splits his time across a number of responsibilities, including: general project management; reaching out to scientific collaborators and industrial partners; line management responsibilities; finding and applying for new funding for his group; running and maintaining a number of research communities, helping to organise conferences and technical workshops and also speaking at national and international events.

He says, "I work on, or lead, a number of the projects that fund the Computational Engineering Group. This group uses some of the world's fastest computers to better understand nearly all aspects that engineering touches. This might be understanding how fluid sloshes in the tanks in a wing of a commercial aeroplane, how to increase the output of a wind turbine by optimising its blades, or even modelling what happens during catastrophic natural disasters that cannot be studied in a laboratory due to their destructive nature."

"I help to understand how specific types of modelling techniques can be applied to particular problems and also how our computer models can be made bigger, better quality or be produced more quickly by harnessing the very latest in supercomputing technology."

Stephen is also a project lead and technical contributor for the CoSeC (Computational Science Centre for Research Communities) support provided to the Collaborative Computational Project on Wave Structure Interaction (CCP-WSI+) community. As a part of the CoSeC management office, he also looks at various aspects of the programme, helping to establish and achieve their general scientific and technical goals and outputs.



Layla Mirmalek is a Junior Project Manager who helps to make innovative ideas become a reality. Her role sits within SCD's Project Management Office where, among other things, she encourages staff to share ideas for improvement of processes and services. She then develops feasibility studies to see if they can be brought to fruition.

"I would say the best part of my role is the variety of work and the innovative people I get to liaise with. There's an atmosphere of fostering ideas and encouraging innovation, so I always feel welcomed when bringing solutions forward. I appreciate that STFC is focused on progressing your development as an individual, and already I have been able to successfully complete the Association for Project Management (APM) Chartered Institute's Project Fundamentals Qualification course."

A crucial element of her job is to provide support for Horizon 2020 Projects, which form part of the biggest ever EU Research and Innovation programme. Going forward she will be supporting our continuing research in EU projects through Horizon Europe, a new research framework with slightly different rules.

Peter Oliver is Head of Operations for SCD and responsible for the running of the department. This includes: tracking and making sure the departmental strategy is implemented; financial management of the £12M of resource spend across 90 active projects; management of the SCD graduate scheme and much more. He says the best part of his job is "working with great people with fantastic ideas!"

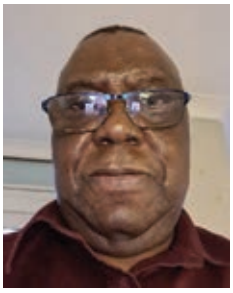
"There are so many great opportunities and projects in SCD to get involved with that I have to limit myself to not get drawn into too much."

As well as being Head of Operations, Peter is project sponsor for the Data and Analytics Facility for National Infrastructure (DAFNI), the national platform to satisfy computational needs in support of data analysis, infrastructure research, and strategic thinking for the UK's long term planning and investment needs.

Bethan Perkins is a Senior Software Engineer in SCD's DAFNI team and says, "I work to build a computer platform called DAFNI which gives scientists a place where they can run their models. Those models look at things like where train stations should be built and how flooding occurs in a city."

Bethan's role primarily focuses on understanding how scientists use the system and what they will require in the future.

"DAFNI is built to serve the infrastructure modelling community and I use my past experience as a modeller to talk to those in the community and find out about their work – which tells you a lot about what they want from the DAFNI platform. I'm also a co-tech lead on one of the two DAFNI teams, run demos of our data service and maintain the release cycle of the platform."



Feckson Kayumba obtained his Bachelor of Arts with Library Studies in 1982 from the University of Zambia and started his library career with the Mindolo Ecumenical Foundation Library, Kitwe, as Assistant Librarian.

His career has taken him on a journey with a variety of organisations before studying for his Masters in Library and Information Management at the University of West London. Feckson joined STFC in 2014, where he took a role at the Rutherford Appleton Laboratory Library as an Assistant Librarian.

"Working with an institutional repository, open science, and data management is a unique aspect of library work that I had never encountered before in my professional journey. The institution too is unique with advanced research programmes that require the professionalism of library staff to keep up with the provision of information resources and services for its users", explains Feckson, understandably delighted with the diversity of his role.

Mark Thorley joined SCD three years ago to take on a role with the European Open Science Cloud (EOSC). He is a Senior Research Project Manager for the EOSC Secretariat projects. These projects have sought to create an environment for hosting and processing research data to support EU science, leading to new insights and innovations, higher research productivity and improved reproducibility in science.

Mark describes his role as, "A 'behind the scenes' fixer who ensures things happen and run smoothly, and who is working to make scientific results much more open and accessible."

On a day-to-day basis he helps to support the work of the EOSC Governance Board and EOSC Steering Board. He is also working on developing and implementing open data and open science policy and setting up a UK open data coordination group.

James Gebbie has been with SCD for 7 years, now working as a Senior Computational Scientist. His current role sees him working for the collaborative HECBIOSim/CCPBioSim consortia under CoSeC (Computational Science Centre for Research Communities). James helps the scientists working within this consortia to access and make use of huge supercomputers so they can tackle enormously complicated biological problems.

He says, "I spend all my time supporting the large number of UK based scientists working in biomolecular simulation with their use of supercomputers. This includes writing new software, optimising existing software, compiling on new hardware platforms, benchmarking and profiling and training in existing and new methodologies. I also designed and run the new advanced training platform that allows a large number of scientists to train online in how to use the bleeding edge techniques, all within a web browser."

James is also upgrading Longbow, SCD's generic simulation automation tool which enables scientists to fire a large number of simulations to any supercomputer with one simple command and it will send inputs, run the simulation and bring back the results as if it had run on their own laptop or PC.

Graduates and apprentices, past and present

Each year we have an intake of new graduates and apprentices starting their careers with us. They have an opportunity to pursue their interests, joining one of our teams and gaining immediate exposure to exciting projects and challenges that will help them to develop new skills and technical abilities. Many of them remain with us once they have completed their training.

Louise Davies is a former graduate scheme member, joining STFC in 2016 after studying Computer Science at the University of Cambridge. She is now a permanent staff member, working in the Software Engineering Group, part of the Scientific Computing Department (SCD). Louise explains her day-to-day job is "developing software to support scientists in their work at STFC".

Louise worked on a diverse range of projects during her time in the graduate scheme. "The most rewarding was my external placement. I got to learn their systems and was able to introduce some new ideas and improvements. I also really enjoy my current project; this is due to the project itself and the people I work with".

Louise's current project focuses on the redesigning of web applications that enable scientists to access both metadata and the data from experiments they do in facilities such as ISIS Neutron and Muon Source, the Diamond Light Source and the Central Laser Facility.

Since starting at STFC, Louise has progressed well in her career. "I'm really pleased with the way my career is going. I am now a technical lead for the project I am currently working on, I have managed and supported other graduates and also provided technical support to colleagues whilst working on my projects".

Aidan McComb completed his apprenticeship in 2018 and is now a permanent member of the Data Services Group. His day-to-day work involves being a systems administrator for the group, working on a variety of projects and dealing with system issues that may arise.

He is currently working with ECHO, a ceph-backed storage service developed at STFC to meet the UK's data storage commitments to the Large Hadron Collider (LHC) experiments at CERN, and which stores over 23 petabytes of LHC data.

Describing his role Aidan says, "I help to store large quantities of data, assist people to access that data and fix issues with the system it's stored on." He says one of the most useful skills he has gained is an understanding of server administration, which allows him to do his job effectively.

During 2020 Aidan has been working mainly from home but, prior to the Covid-19 restrictions, he had the opportunity to travel to international conferences. One of these was the Supercomputing conference in Denver at the end of 2019, which he found exciting. "I was visiting a country I'd never been to before and also seeing and talking to people about the latest technologies available."

7 apprentices working in SCD in 2020

Laura Murgatroyd joined the graduate scheme in 2019 after completing an integrated masters in physics at the University of Birmingham.

The computational biology group, where Laura is working, supports CCPi, the Collaborative Computational Project in Tomographic Imaging. She is currently working on developing an app for processing and reconstructing tomographic X-ray images which will be produced at EPAC (the Extreme Photonics Application Centre), which is currently under construction and due to be operational in 2024.

The rapidly changing landscape of research at SCD provides an opportunity for graduates to work on projects that are coming to fruition. "It was very rewarding releasing an institutional login system for PSDS (a website providing access to chemical databases), and seeing this being used every day by scientists," Laura says of a previous project she contributed to. In fact, this project has had useful applications to the recent transition to remote working due to the coronavirus pandemic.

The challenges Laura has faced when learning about new systems and software has pushed her to work towards a chartered status with the Institute of Physics, "It is still early days but I feel that I have already learnt a lot about software development," she says.

5 graduates completed their training in 2020

James Acris started as an SCD Graduate in 2020. He is working as a software developer for the powerful DAFNI computing platform, that allows researchers to upload models and datasets of national infrastructure systems such as energy demand models and flood prediction forecasts, with the potential to cover a huge range of research.

He has been writing code to not only maintain DAFNI systems but also improve them. He says, "I find it exciting to be contributing towards a system that could help us better plan for the long term future as a country."

Kieran Howlett is in his first year of the apprenticeship scheme and is working in the Research Infrastructure Group (RIG). The team supports many projects within STFC, the UK and Europe by managing computer resources and providing e-infrastructure.

He says, "I am currently setting up a server for use within SCD and creating a new piece of software to simplify server control. I also deal with the helpdesk side of the SCARF supercomputer and for PEARL, a powerful deep learning system, to assist users with issues and allow smooth operation."

His decision to join us was based on the commitment STFC gives to all of our apprentices, providing opportunities to grow and improve, and also because "the overall job looked enticing to be a part of."

17 We had 17 students at various stages of our Graduate Development Scheme

Inspiring and involving

Encouraging and teaching the next generation

We have a particularly active Public Engagement Committee within SCD which has continued to meet regularly throughout 2020 and, although three face to face events were held in the first part of the year, it has been a year of remote public engagement for us and for STFC as a whole. We were forced to abandon an extensive planned programme including the Daresbury Open Week, for which work was well under way. Alongside altering our day-to-day working methods, we very quickly identified some activities we could do online and have continued to develop more through the year. We now have a portfolio of remote activities which we'll combine with our existing face-to-face activities once we're able to bring people back into the lab. This all contributes to STFC's long held strategic aim to extend the reach of its public engagement to audiences geographically remote from an STFC site.

Whilst this year was initially an exercise in improvisation, we have created new resources specifically for a remote audience and adapted existing activities where appropriate. Throughout the process, we have relied on the STFC Evaluation Framework and the concept of "Generic Learning Outcomes" to ensure our new remote events continue to meet the same objectives as our face-to-face portfolio.

Moving forward, we will learn from our pilot events this year to deliver an improved programme in 2021, which we expect to be another year of remote engagement. And whilst remote public engagement brings both benefits and challenges, we firmly believe the remote activities we have developed and will develop during the pandemic will remain a key part of our public engagement programme even when we can start face-to-face engagement again.

Work Experience Webinars and Python Workshop

Whilst the STFC Work Experience Programme was unable to proceed in its usual face-to-face format, we still wanted to be able to deliver something to the students who would otherwise have been part of the programme at an STFC site. As part of the wider STFC remote work experience programme, our staff delivered a webinar to 40 work experience students covering the work SCD does.

As well as this we reworked our Python Workshop into a remote format; Mika Shearwood (a computing apprentice) and Jacob Ward (SCD Cloud Team) who reformatted the content into a series of Jupyter notebooks hosted in a JupyterHub instance within the STFC Cloud. Mika also led the workshop, supported by four members of SCD staff, who were on hand to provide assistance and answer any questions the students had.

The new online workshop was extremely successful, allowing the team to reach a much larger audience from a greater geographic range than in previous years. Typically the face-to-face workshop would only reach 20 students, while the remote version was able to reach over 70 students. The workshop received good feedback from students as well with an average feedback score of 4.5/5 and 93% of students saying they are more confident with python following the workshop.

Virtual Python Workshop

93% of students increased their confidence in using python!

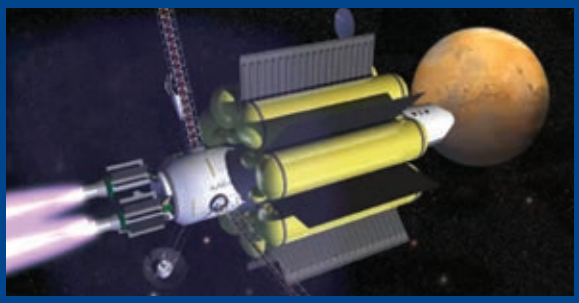
Virtual Saturday Coding Club

reached children as far as the **United States**

Ada Lovelace Arduino Challenge Day

In October, software engineer Greg Corbett led a virtual computing themed challenge day for over 300 children in Years 7, 8 and 9. This was the first virtual coding Public Engagement event of this scale, and the first to be aimed directly at school groups after they returned to the classroom in September. The aim of the day was to repair the Ada Lovelace, the first ship to Mars carrying a crew, using Arduinos to build and programme replacements for mission-critical hardware and software systems that were damaged by a solar flare.

The event was held over video conferencing and started with an introduction to physical and virtual Arduino microcontrollers, as well as block-based programming languages. Each school was then assigned a system to repair and a breakout room with two or more members of STFC staff from across our sites and departments, who offered advice to the groups and help to debug hardware and software issues when they occurred.



To facilitate this event, each school was sent physical Arduinos and a set of electronic components, as well as an electronic resources pack enabling them to replicate the day in the future on their own. The day also included a virtual tour of the Rutherford Appleton Laboratory Data Centre, provided by Jacob Ward, and a talk from Sarah Byrne, one of SCD's newest graduates, about her path from GCSEs to STFC – for this part of the day we were joined by additional schools.

One of the benefits of the remote nature meant the event was able to reach over 350 schoolchildren, from a more geographically diverse area, instead of approximately 40 local schoolchildren that the physical event usually caters for.

Author: Greg Corbett, Distributed Computing Infrastructure Group

Infographics (pages 8-11) designed by Rebecca Humble, Impact Team.

Building influence and fostering collaboration

International Support and Collaboration

SCD continues to work internationally with researchers and computational experts to support global scientific advancement. Although we were unable to travel to conferences and meetings in this year, we were still able to run our own events – some of which had surprising numbers of attendees – and also attend other international forums. New challenges and opportunities presented themselves with the emergence of Covid-19. Our staff and facilities were an active part of the international research community response to the pandemic, offering much needed expertise and resources.

Support for global COVID-19 efforts

Collaborative Computational Projects (CCPs)

Immediately following the release of the genetic sequence by the Chinese, the structural biology community swung into action to determine the three dimensional structures of the proteins produced by the virus. These form the basis for efforts to find inhibitor drugs and vaccines.

The Collaborative Computational Projects CCP4 and CCP-EM develop software for determining protein structures by crystallography and electron microscopy respectively, and form part of the toolkit being used against Covid-19.

Martyn Winn, leader of SCD's Computational Biology group (which co-ordinates the two CCPs) said, “Many years of work by ourselves and our collaborators have gone into producing and distributing the best software. We are proud that our software can help in the fight against this disease. We are continuing to look for ways to contribute to the international effort."

Leading rapid responses

Our staff within the CCPBioSim and the HECBioSim consortia coordinated efforts to avoid duplication, paired groups doing Covid research with others who had relevant expertise and available capacity, provided rapid responses to software availability and development requests, and benchmarked new and existing methodologies and regimes along with new hardware to allocate High Performance Computing (HPC) resources more efficiently.

Juan Bicarregui, who heads our Data Division, co-chaired the Research Data Alliance Covid-19 fast track working group. This was set up to clearly define detailed guidelines on data sharing and re-use to help researchers follow best practices and maximize the efficiency of their work. Their recommendations and guidelines were published in May 2020.

Donating GPU and CPU capacity

The Diamond Light Source benefitted from resources provided by the STFC Cloud for use in a very large screening experiment and in crystallography work requiring substantial computing capability and software. Researchers working with Diamond received additional Central Processing Unit (CPU) resources.

We were able to donate large numbers of Graphics Processing Unit (GPU) capacity and CPU cores from the STFC Cloud for the Folding@Home distributed computing project for their efforts in analysing and understanding the virus. Our UK Tier-1 supercomputing facility, which stores and processes vast amounts of data from the Large Hadron Collider at CERN, also donated thousands of CPU cores to Folding@Home.

International Conferences and workshops

Despite the challenge of organising and attending mainly virtual events in 2020, our staff have engaged in a wide range of far-reaching conferences and events. Our own conferences and workshops have been extremely successful, with delegates being able to attend from much further afield than might have been possible for a physical event. As Tom Griffin mentioned in his foreword to this review, it has really made us think about how we will run events in the future as a remote element is likely to be advantageous for some people who might otherwise find themselves excluded.

Computing Insight UK – our own Supercomputing Conference

After 30 years of physical conferences, the 2020 edition became the first-ever virtual Computing Insight UK (CIUK) conference, held in December. The move online opened up some new opportunities, including an extended programme of events that ran throughout the month of November in the build-up to the conference.

CIUK 2020 included the usual programme of high quality presentations, with the keynote talk delivered by Torsten Hoefler from ETH Zurich who presented on "A Data-Centric Approach to Performance Portability". And CIUK 2020 would not have been complete without hearing from the ever-popular Martyn Guest, a founding member who has spoken at every conference to date. Something missing this year was the exhibitor arena, which gives people the opportunity to speak to vendors about the latest hardware available and discuss any operational problems. This is something that delegates always enjoy and we hope that the next CIUK will once again be an in-person event, but still include a virtual option.

One of the highlights was the introduction of the CIUK Student Cluster Challenge. This was an online competition, with teams from the Universities of Liverpool, Durham, Bristol and York, who completed a series of ten individual challenges over an eight week period, using various different systems provided by supporting companies. It was a very close-run competition with each team winning different challenges, but it was Durham University who were ultimately crowned CIUK 2020 Cluster Challenge Champions. The team will now go forward to represent CIUK in the International Supercomputing (ISC'21) Cluster Challenge in the summer of 2021.

Designed by Andy Collins

Impact posters displayed on STFC's 'virtual exhibition' website.

Supercomputing 2020

Every year, a delegation from Scientific Computing travels to the US for the biggest and most heavily attended annual Supercomputing conference and exhibition. In 2020, with the global pandemic bringing a halt to international travel, SC20 became an online conference and STFC responded with a strong virtual presence.

We created a 'virtual exhibition' – a website which showcased a selection of the High Performance Computing (HPC) research STFC is involved in internationally, and the benefits this brings to science, industry and society. This virtual platform displayed posters that highlighted STFC research within the HPC community, publications, and a number of educational and scientific videos.

Members of SCD and its close relative, the STFC Hartree Centre, led and participated in workshops covering topics from 'scalable algorithms for large-scale systems' to the 'challenges faced by software engineers'. As always, it was good to catch up with colleagues from across the HPC community and around the world, and it was exciting to see once again how far high performance computing has advanced since the last conference.

Electron cryo-Microscopy Spring Symposium

For the first time ever, in 2020 the CCP-EM Spring Symposium for the Electron cryo-Microscopy (EM) community was held entirely online in a virtual format due to travel and social restrictions surrounding Covid-19. This was one of the first conferences affected by these restrictions but the response to the new format from the worldwide cryoEM community was remarkable. The number of participants soared to over 2,700 people, almost 10 times the usual number of people taking part!

Tom Burnley, who leads the CCP-EM core team said, "With the conference season cancelled we wanted to bring the EM community together in these challenging times and were amazed at the response. As cryoEM provides a tool that can help understand the molecular mechanisms of COVID-19 and other diseases, it's more important than ever that the EM community can share inspirational findings and exciting new methods."

This year's conference focused on cryoEM methods to solve protein structures – vital for understanding how the cells in our bodies respond to diseases. It had a full agenda of exciting and thought–provoking talks and presentations from a good variety of people from within the global EM community.

Chain of amino acids or biomolecules called protein - 3D illustration.

CCP5 40th Conference and AGM

Now in its 40th year, the Collaborative Computational Project no. 5 (Computer Simulation of Condensed Phases) conference became an online event, packed with engaging international speakers and presentations. The talks covered computational research on a wide range of topics, from the performance of tyres, through iron deficiency anaemia, to quantum computing for scientific calculations.

Attendance was high and showed a four-fold increase in registrations and viewers of the live-streamed webinars, including people from the US, India, Japan and Europe.

This year's conference introduced a biennial CCP5 Prize and Lecture Award, and announced the inaugural winner as Professor Kostya Trachenko, Queen Mary University of London. The award was given for his 'outstanding contributions to modelling and theory of condensed matter phases including liquid state theory, radiation effects and DL_POLY development'. He received £1000, a medal, and expenses to cover attendance at next year's 40th Anniversary celebrations.

Prof. Kostya Trachenko

International Research Software Engineering Leaders Workshops

The research community is realising the importance of the role of a research software engineer (RSE) because researchers are not only looking for new approaches and improved quality of research software, but also for software that enhances the research it supports.

Dr Alejandra Gonzalez-Beltran is a Fellow of the Software Sustainability Institute (SSI) and leader of Scientific Computing's Data and Software Engineering Group. She was a leading voice at the 'RSE-London-South-East-2020' workshop in which participants discussed four areas that they believe cover the full breadth of the RSE domain – policy, training, community, and software development. They have made recommendations to support and grow RSE communities and for collaborative projects with a more global scope.

Later in the year Alejandra joined colleagues from 13 different countries for the 2nd International RSE Leaders Workshop 2020, to discuss other aspects of RSE challenges. They focused on bringing diversity, inclusivity and equity to RSE communities, ideas to produce new resources and create communication channels along with a platform to facilitate them, and ways to grow and consolidate the international community. A particular highlight was hearing that participants in Argentina, Colombia, France, Canada and Spain plan to start new RSE communities in their respective regions.

Testing and tracing in epidemiological models

Back in March 2020, Dr. Tim Colbourn from University College London was assembling a team of epidemiologists, economists and simulation experts to quickly develop new ways to model the nascent COVID-19 pandemic, with the goal of producing cost-benefit analyses on possible strategies for its suppression or management. He was particularly interested in modelling the power of an extensive program of contact tracing that could identify and isolate most people who had been close to a known patient, and limit the spread of the disease. Dr Simone Sturniolo explains how he got involved.

While I had no previous expertise with epidemiology, I volunteered to help as someone with extensive experience in mathematical modelling in Python. During the course of the next month, together with other members of the team, we developed a modified version of a very well-known model that we have then gone on to use for simulations of a number of scenarios involving the COVID-19 epidemic in the UK.

While they may sound like they should be the purvey of experts of disease and biology, the mathematics of epidemiological models share a lot with that of chemical reactions. In the simplest possible model, the *Susceptible-Infected-Recovered* model (or SIR for short), there are essentially two types of reactions. If a person who is *Susceptible* meets one who is Infected, they have a certain probability of becoming infected themselves. On the other hand, after some time in that state, an Infected person is *Recovered*, having overcome the disease or succumbed to it; either way, they will be unable to contract it again or spread it any further.

These two processes alone are already enough to predict in broad strokes the dynamics of the 'waves' of infection we've all sadly grown used to; the cases rise at first as the infection spreads, then fall as the number of *Susceptible* individuals dwindles and thus the fire runs out of fuel, so to speak. What characterises this model are the specific numbers used in it – the infectiousness and the duration of the infection, which are properties specific to each

disease and determine how much of the population is likely to get infected, as well as how quickly. The effect of some non-pharmaceutical interventions in this sense is relatively easy to model. For example, widespread mask use will effectively reduce the infectiousness of the disease. The hard part, in this case, is estimating exactly by how much, which can only be done with empirical observation.

Including the effect of testing and tracing, however, is trickier. Our work focused on a *Susceptible-Exposed-Infected-Recovered* model, very similar to the above except that it includes a category of Exposed individuals who are incubating the disease but still do not display the symptoms and cannot spread it, but will progress to being Infected in due course. This appeared to be closer to the reality of COVID-19, with its relatively long incubation times. Self-isolation was included by adding the possibility of marking an individual as *Quarantined*, which prevented them from contacts (whether they were healthy or infected), thus removing them for a time from the pool. Testing was simulated by discovering and isolating Infected individuals at a rate that depended on how intensive the program was.

The trickier part was contact tracing. In a model in which every individual is simulated independently and has a personal history, it is possible to keep track of meetings and thus identify everyone who had encountered someone who has tested positive to the disease during the last few days, so that they can be isolated as a precaution – just like in real life. But models that track every individual, called Agent-Based Models, can be expensive to simulate, especially for large populations.

We focused instead on creating a Compartment-Based Model that only keeps track of the numbers of individuals in each group without defining personal histories. This, however, makes it impossible to trace back contacts explicitly. To solve this problem, we introduced an innovation to the model, creating additional compartments that acted as a 'memory' to keep track, for example, of how many encounters between *Infected* and *Susceptible* people had happened over the last two weeks. The number of people who could be traced was then taken to be a fraction of these.

This approach is just approximated, but it turned out to work extremely well in all realistic scenarios, and to fail its predictions only in cases where tracing was so slow and rare it mattered very little anyway. Thanks to this new model, it was possible to conduct a further study that explored a lot of different strategies in short times and with very little computational resources.

Author: Simone Sturniolo,
Theoretical Computational Physics Group

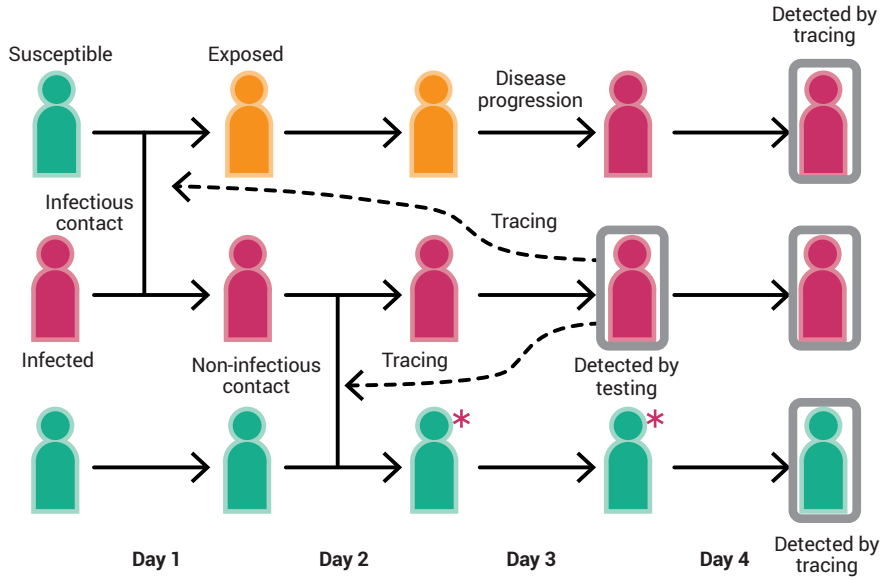


Illustration of a test and trace model.

The code for this work has been made available for free at **https://github.com/ptti/ptti/tree/ptti-theory-paper**, and a paper detailing the model has been published:

S. Sturniolo, W. Waites, T. Colbourn, D. Manheim, J. Panovska-Griffiths, "Testing, tracing and isolation in compartmental models", PLoS Comput Biol 17(3): e1008633. **https://doi.org/10.1371/journal.pcbi.1008633**

DAFNI Champions focus on growth

This year DAFNI (the Data and Analytics Facility for National Infrastructure) has introduced a new expert focus to extend priority areas for its user communities.

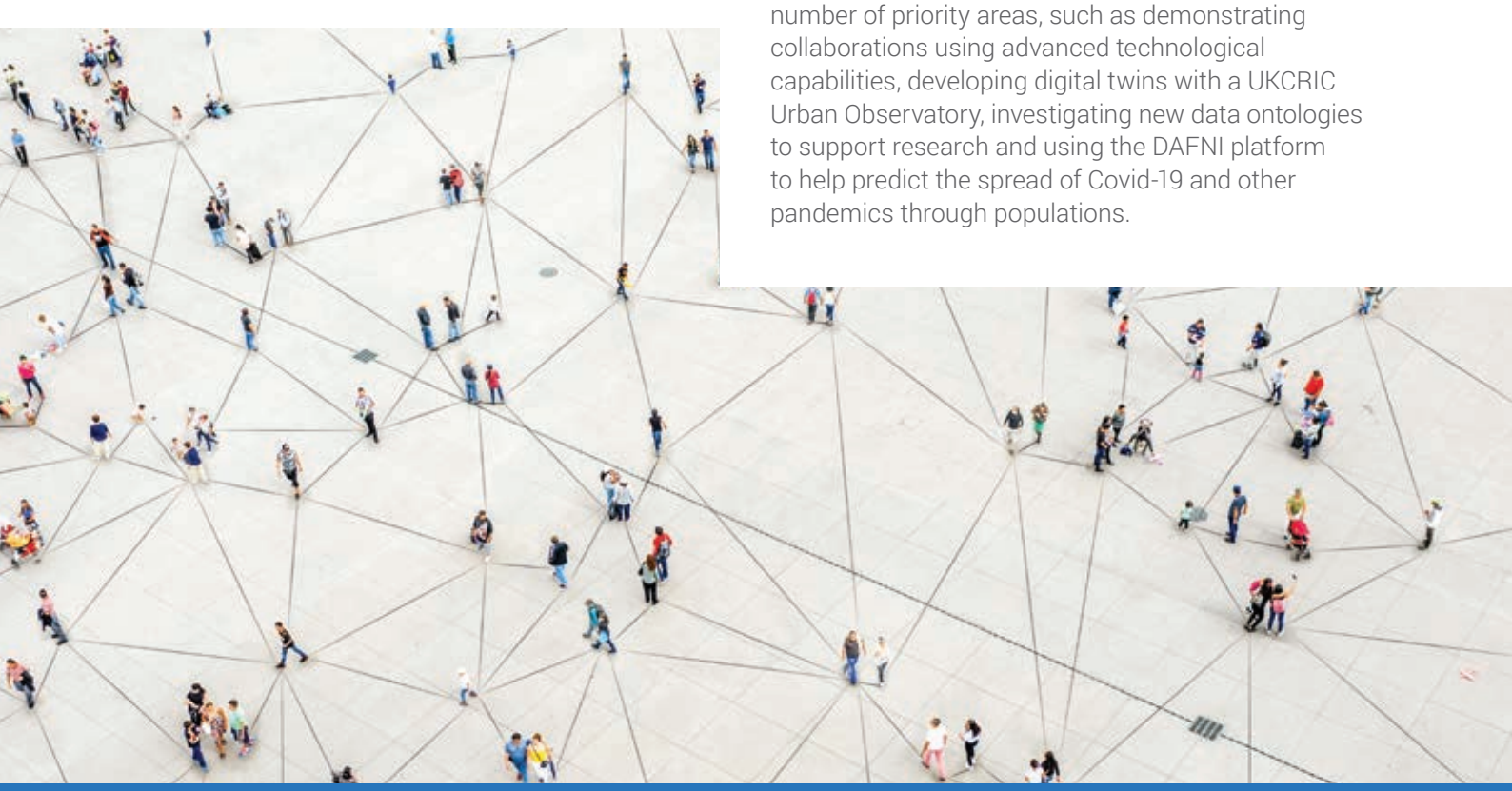
The DAFNI vision is to allow researchers to use state-of-the-art modelling, simulation and visualisation to better inform and develop strategic thinking. DAFNI has been developed so researchers can run models on a common computing system with a central repository of data. Using cutting edge computers, models and simulations can be scaled to greater coverage and resolution. In fact collaborators are not only able to use the platform as a shared workspace to integrate models and data from different sectors, they are also provided with a suite of tools and capability to reduce many of the interoperability issues researchers face when accessing information from disparate sources.

In this third year of its development, the DAFNI team has started to explore how the platform can begin to support the National Digital Twin agenda, an opportunity to transform the efficiency and reliability of national infrastructure. In a digital twin, data from real systems are fed into a running computer simulation, to predict outcomes that can then be fed back to the real system to steer its course. DAFNI will provide a platform to bridge these virtual and physical worlds, coupling diverse simulation models that will together contribute to digital twins that are national in scope.

In 2020 we have provided an extensive programme of user interaction in the form of DAFNI Roadshow demonstrator events, demonstrator videos and the development of help guides to support users to fully engage with the platform and develop their research. DAFNI has also been involved in a number of new projects which have meant we are working more closely with Government departments such as Defra, Climate Change Committee, Environment Agency, Office of National Statics and England's Economic Heartland as well as extending our reach to academic researchers beyond our originating partners.

As a result DAFNI's community is steadily growing and now has more than 100 users. This number rises steadily with each interaction the DAFNI team has with the research community. This growth has been boosted by a number of Champions from within our own Governance Board members. Their involvement has not only supported the development of the DAFNI platform but is extending and stretching its capability within a growing community of users. This extensive work demonstrates the demand for the platform and provides a significant stepping stone to launch the platform operationally in 2021.

Our Champions are spearheading projects in a number of priority areas, such as demonstrating collaborations using advanced technological capabilities, developing digital twins with a UKCRIC Urban Observatory, investigating new data ontologies to support research and using the DAFNI platform to help predict the spread of Covid-19 and other pandemics through populations.



DAFNI Champions and focus areas:

Dr Cristian Genes, University of Sheffield:
Cristian's focus is on demonstrating the unique, advanced technological capabilities that make DAFNI an ideal platform for collaborative development, validation and implementation of National Digital Twins for large-scale complex systems.

Mr Adrian Hickford, University of Southampton:
Adrian is to carry out advocacy within the transport research community, to develop and strengthen DAFNI's relationships with transport bodies and researchers regionally and nationally.

Dr Simon Jude, Cranfield University:
Simon is providing interaction with the Cranfield Living Laboratory and Urban Observatory as well as developing a pilot Digital Twin using DAFNI to link real-time and near real-time water quality data from Cranfield Urban Observatory and the National Water and Water Treatment Test Facility.

Dr Juste Raimbault, University College London:
Juste is extending the capability of the platform to bring on the MATSim multi-agent transport simulation framework and apply it to the current Covid crisis, using the EpiSim model to add pandemic indicators to travel indicators and to model them on DAFNI.

Professor Liz Varga, University College London:
Liz is leading a team of researchers from her own institution as well as Newcastle, Cranfield, Leeds, and Oxford which will culminate in the development of an ontology framework for infrastructure research, the absence of which is seen as a challenge to DAFNI's sustainability.

We are looking towards 2021 as a year whereby we seek sustainability funding for DAFNI and continue to grow and extend the DAFNI platform support for the infrastructure research community.

For further information on the work of the Champions and the many other projects DAFNI is becoming involved in, please check out our website at: www.dafni.ac.uk

DAFNI is funded through an £8 million capital investment from EPSRC as part of the UK Collaboratorium for Research on Infrastructure and Cities (UKCRIC).

Author: Marion Samler, DAFNI

Developing advanced and sustainable technologies

Advanced electronic structure methods for spintronics

Magnetism and magnetic materials remain a central topic in our understanding of materials and are crucial to many of our established technologies including data storage, power transmission and also many newer technologies, such as the superconducting circuits that make magnetic resonance imaging possible. Magnetic materials are now being investigated as a route toward computing devices with significantly lower power consumption than today's technology and as practicable building blocks of quantum computers. These technologies are loosely termed "spintronics" because the fundamental operations are done with spin instead of the charge, as in conventional computers.

Magnetism is both a fundamentally quantum mechanical effect and also an effect that emerges due to the complex interactions between each and all of the atoms that make up a piece of material. This means that our theoretical methods must be able to treat systems with many interacting particles while faithfully describing the subtleties of their quantum nature. The interacting particles are the electrons which form the "glue" bonding atoms together in the solid and our understanding of them is based on electronic structure theory.

Recent investigations by the Theoretical and Computational Physics group have included an application for the first time of an advanced method for predicting the electronic and magnetic properties of materials to perhaps the most important material in spintronics, yttrium iron garnet (YIG). This material has a complicated structure containing magnetic iron atoms in different positions, interacting via oxygen atoms, and has previously been too complicated to be treated using accurate methods. In their recent paper[1], they showed how modern computational approaches, in particular employing many powerful graphic-card processing units (on the Marconi M100 supercomputer at Cineca, Italy), can be exploited to show good scaling for taking advanced electronic structure methods to larger and larger systems.

The significance of this achievement is that, for the first time, a truly predictive theoretical method has become available that is able to treat materials of the kind of complexity that is interesting for experiments and applications. This accuracy depends on no assumptions other than the positions of the atoms in the material. The method used in the study is called "quasiparticle self-consistent GW".

In their YIG study, the team were able to show that the description provided by this method was sufficiently accurate to be able to predict detailed magnetic properties such as the spin wave spectrum (the energies of magnetic excitations seen in neutron scattering experiments, an important technique for probing magnetism at the atomic level) and how the magnetisation depends upon temperature.

In order to incorporate temperature effects, the authors quantified the magnetic interaction between the iron atoms, known as the Heisenberg exchange integral. They used this in a novel simulation method for the magnetism at finite temperature that describes correctly the fluctuations in the spin system that give rise to the loss of magnetisation at the critical temperature. Conventionally, this modelling has used approximations that mean the spins act as classical objects and it has not been clear when this approximation is justifiable. Based on the precise evaluation of the Heisenberg exchange made possible by the quasiparticle self-consistent GW method, the team were able to demonstrate the effect of this assumption, and how using the correct statistics for finite temperature magnetism is crucial. This is an important conclusion for all theoretical work in finite temperature magnetism.

The YIG calculations, with 80 atoms, are the biggest yet using this advanced method – until now systems of up to about 10 atoms only have been possible. This demonstration shows that parameter free, accurate methods are now accessible for large, complicated materials problems. In the past theoretical methods often needed to be tuned to match experimental data: instead the new methods are truly predictive. This predictive power is crucial to understanding materials and developing new technologies. We must be able to trust that interesting effects predicted by theory are real, and not merely artefacts of our limited modelling capability.

The project is a collaboration between STFC's Scientific Computing Department, Dr Joe Barker, University of Leeds and Dr Dimitar Pashov, King's College London.

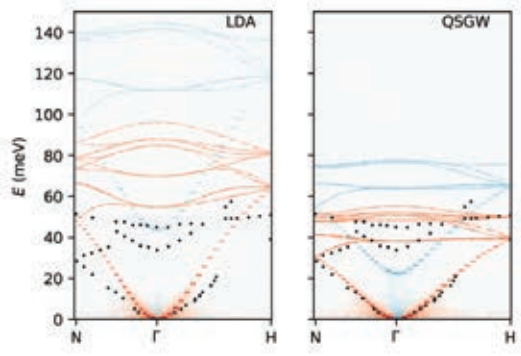


Figure 1, spin wave spectrum calculated using conventional theoretical method, left, (LDA- the local density approximation in density functional theory) and, right, the new calculations. Black dots are experimental data (see the paper[1] for references).

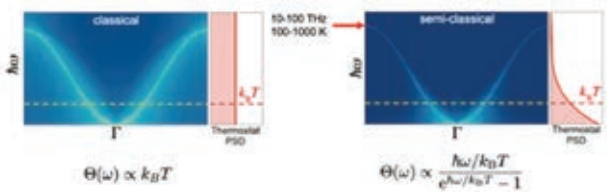


Figure 2, occupation of spin waves at finite temperature: conventional methods, left, over populate high energy spin waves, resulting in a poor description of thermal effects. The new semi-quantum method, right, shows the correct occupation of spin waves.

Author: Jerome Jackson,
Theoretical and Computational Physics Group

[1] For the full paper see:
https://iopscience.iop.org/article/10.1088/2516-1075/abd097

Dissipative particle dynamics: adding attraction to interface problems

From the many challenges that our society is facing today, environmental problems are probably the most pressing ones. Humankind has been consuming natural resources at a higher rate than the environment could regenerate. To address these challenges a broad programme of research is being undertaken to develop new sustainable materials and to advance our understanding of their structure-property relationships at the nano- and mesoscales (atomic scales, not visible to the naked eye).

Computer modelling, which is a prime theoretical tool to study matter from first principles (from fundamental laws of physics and chemistry), meets fundamental difficulties for biofuels, sustainable materials in fast moving consumer goods formulations, functional organic, food science and printing technology materials because of the vast gap between their molecular and functional properties scales. In order to bridge this gap a range of multiscale and coarse-grained methods – grouping together atoms or molecules to form interacting beads – have been developed.

When it comes to describing the fluid motion at mesoscale, dissipative particle dynamics (DPD) is arguably the most widely used particle-based method describing various aspects of multiphase flow (eg. various combinations of liquids and gases). It is fast, simple to implement, and extends the range of accessible time and spatial (length) scales by several orders of magnitude, extending thus the range of accessible problems. However, in its standard form DPD employs only the repulsive force to describe the interaction between the beads, representing small fluid volumes. It is therefore unable to capture

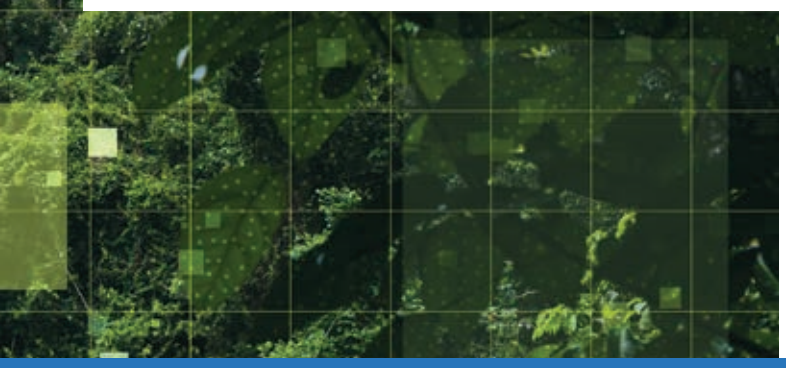
the liquid interface phenomena including wetting, spreading, and many capillary effects, such as seeping into small gaps.

The currently adopted way to overcome this shortcoming is to include many-body interactions in the model. This is not a strictly necessary condition, as the wealth of capillary phenomena within the liquid state theory is represented by simple particle –particle interactions, known as the Lennard-Jones potential. However, modelling these interactions at the coarse-grained level would negate the advantages of using DPD; ie. by slowing down its speed and scale. Fig.1 illustrates the standard DPD potential, our new potential, as well as the Lennard-Jones potential in the inset graph.

When two states of matter, liquid and gas, meet, an interface is formed (see Fig.2) and this isn't easy to model because the liquid and gas phases have different densities, and transferring particles from one phase to another requires doing some work. Our new method accurately predicts these phases as well as the interface, which has a structure different from either phase (see illustration at Fig.3). We know it is accurate because the simulated densities match predictions. The critical point is where a particle can freely move from the gas phase into the liquid phase and vice-versa.

This exploratory work, done using the SCD-developed code DL_MESO, demonstrated high versatility of the modified DPD potential and its suitability to tackle the interface problems. Further work is required to derive appropriate parameters to solve practical cases. What happens at the interface affects all aspects of our living environment. For example, the manufacture of biofuels or the safe storage and transport of food depends on the interface where the liquid and gas meet. This could determine the shape of the pipe needed for the biofuel, or the type of material used to contain the foodstuff.

Authors: Michael Seaton, Vlad Sokhan and Ilian Todorov, Computational Chemistry Group



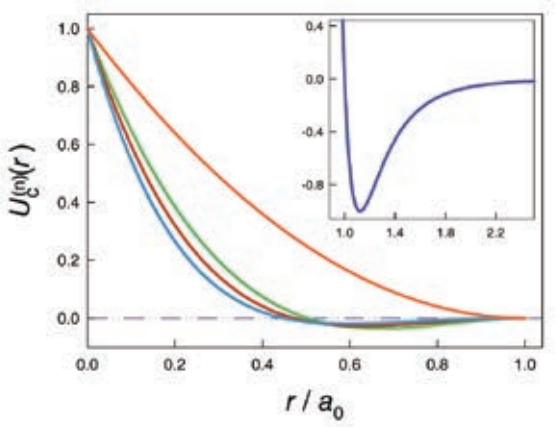


Figure 1. Standard DPD (orange line) and new proposed (green, red, and blue lines for different test cases) highlighting the difference with the standard DPD (~0.5a0 – 1.0a0). In the inset, the standard Lennard-Jones potential, widely used in atomistic simulations, is shown in indigo.

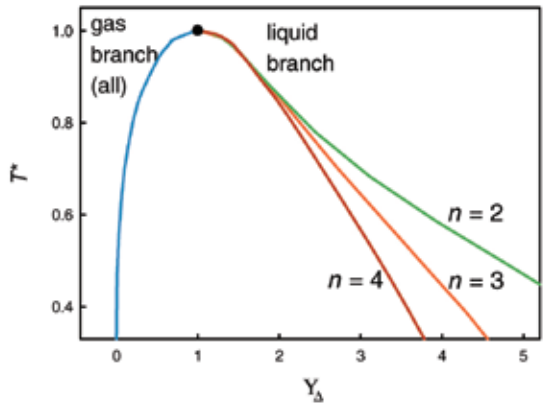


Figure 2. Coexistence curves for the new potential, calculated for three test cases – quadratic, cubic, and quartic repulsive forces. Blue line denotes the gas branch (common for all powers) and green, orange, and red lines are for the liquid branches. Black dot denotes the critical point. Reduced units have been used.

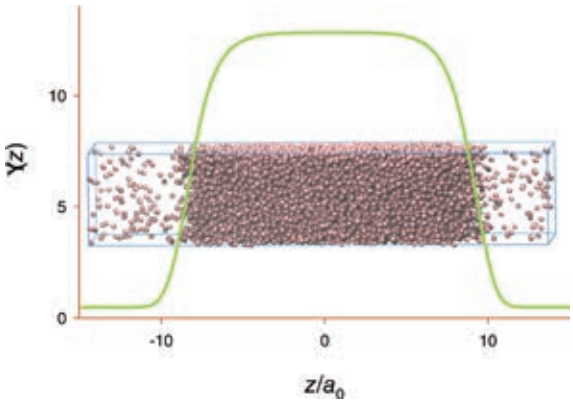


Figure 3. A sketch of the system in slab geometry used to estimate the coexisting densities. The graph presents the density profile of the DPD particles (green line) giving the values of two coexisting densities, where the profile is flat and showing two liquid-vapour interfaces at z≅±10.

For more information about this work please see the associated paper **https://www.tandfonline.com/doi/abs/10.1080/08927022.2019.1578353**

Interscale modelling of soft matter: The Cheshire Cat's approach to accelerated multiscale molecular dynamics

Soft matter is a collective term for condensed matter systems that exhibit complex cooperative molecular behaviours that make soft materials distinct from simple liquids and solids. Typical examples are found all around us: detergents – soap, shampoo, washing liquids; polymers – putty, glue, plaster; colloids – paints, dairy products, mayonnaise, ketchup, spreads and ice cream; everything biological – bacteria, blood, soft tissues.

At the molecular level, a profound property of soft matter is the ability of its constituents to change their shape and consistency. This enables a vast diversity of relatively long-lived metastable molecular assemblies which often coexist, transmute and morph, be it fluffy conglomerates of surfactants and/or other surface-active agents (in detergents, food and healthcare products), highly-functional complexes of lipids, proteins and DNA (in building blocks of a living cell), specially designed block-copolymer gels (e.g. in drug-delivery systems), etc. The associated rich spectrum of intricate thermodynamic and kinetic phenomena is attributed to cooperative processes concurrently occurring at a range of length and time scales: atomic vibrations – picoseconds/Angstroms, molecular motion – nanoseconds/nanometres, and mesoscale phenomena – (micro-) seconds/microns.

Standard classical approaches to molecular simulation focus on a prechosen level of detail, using either pure atomistic (where every atom is represented) or simplified coarse-grain (CG) models (where groups of atoms are represented as a single cluster). As a result, full-atom simulations of condensed soft matter suffer from long, often intractable, relaxation times at scales larger than a few tens of nanometres, whereas most of the computational effort is spent on motion of the smallest species such as solvent molecules (most often water). In contrast, upscale (coarse-grain) simulations, although relieving the slow relaxation issues, lack sufficient detail and at best, deliver only qualitative results that can rarely be directly compared with experiment.

Coarse-graining is done by clumping groups of atoms into larger clusters, also known as CG beads, so a molecule of 100 atoms is represented by far fewer connected clusters so the numbers, and hence the computational effort required, drastically reduce, see Figure 1.

There exist a number of multiscale methods for systematic optimisation of CG models (force-fields) against reference data obtained from either full-atom simulations of relatively small target systems, or (rarely) experiments providing structural information on the target aggregated states. However, such methodologies inevitably involve tedious iterative procedures and produce CG models too specific for the input data used as reference, thus suffering from overfitting and making the obtained CG models non-transferable between different state points (in terms of temperature and/or concentration). Furthermore, in order to study the self-assembling and possible restructuring processes, i.e. different aggregate shapes and structures emerging at the same time, one has to either introduce some sort of transformation mechanism between a few specific CG models, or employ an agile, self-adjusting force-field.

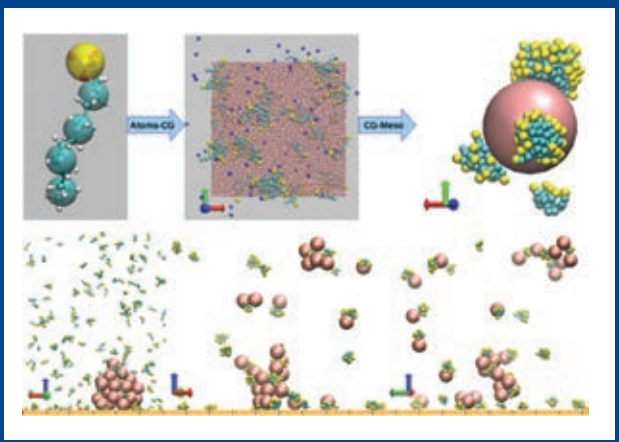


Figure 1. Illustration of coarse-graining and the CG-to-mesoscale workflow in the case of SDS surfactant (the most abundant detergent). Upper row: CG model with 5 beads per molecule is, first, designed and the CG forces are optimised by directly linking the atomistic and CG dynamics. Lower row: mesoscale simulation, mimicking the washing liquid action, becomes feasible for systems 10 times larger than the original atomistic sample. Additional nano-scale ingredients such as 'dirt' grains can be added too. The 'dirt' aggregate (pinkish) is initially stuck (adsorbed) on the 'cloth' surface (orange). As the simulation progresses, the detergent molecules permeate the dirt, breaks it up into smaller grains and thereby dissolves it into solution.

With this in mind, we develop a novel approach to multiscale simulation that effectively combines the full-atom and coarse-grain representations for the same system, thereby allowing for accumulating statistics at the fine-detail level while accelerating the fine-detail kinetics due to coupling with the coarse-grain dynamics. Our method is also enhanced by a robust on-the-fly (re-)optimization of the CG force-field in the course of simulation. Therefore, due to the agile nature of the CG interaction model, its 'transferability' is automatically ensured by the simulation protocol. The approach is illustrated in Figure 2 with a snapshot from a combined atomistic-CG simulation of phospholipid membrane (a bio-cell exterior barrier).

The 'interscale' main concept is to gradually, and repeatedly, switch on and off the CG force-field – similar to the appearance and disappearance of the Cheshire Cat's Grin in Alice's Wonderland. This way we allow for the two representations, fine-detail and coarse-grain, to concurrently evolve, while, from time to time, to also effectively speed-up the fine-detail dynamics. This is illustrated with a principal diagram given in Figure 3.

The simulation is started with a fine-detail initial configuration, i.e. at the lowest level in the diagram. Coarse-graining is then done on-the-fly during simulation, and Boltzmann-weighted 'hops' up and down in the strength of the CG force-field are carried out: the further up in the levels the stronger the influence of CG forces becomes. At the top level the fine-detail kinetics is essentially driven by the CG dynamics, whereas at the bottom the pure fine-detail dynamics is restored and statistics are accumulated.

The interscale simulation protocol has been recently implemented in the development branch of DL_POLY (a software suite), being tested now in the case of self-assembly and phase behaviour in surfactant and lipid systems.

This new combined method can probe phenomena that either is not captured at all, or as realistically, by either method used on its own. Our tests reveal at least 10-fold speedup in effective diffusion of the solute species, which facilitates long time relaxation and self-assembly processes. In certain cases, like with detergent molecules arranging in spherical nano-aggregates (micelles), we observe micelle formation within 5-10 nanoseconds, while in a 'brute-force' pure fine-detail simulation micellisation occurs on the time scale of microseconds.

Authors: Andrey Brukhno and Ilian Todorov, Computational Chemistry Group

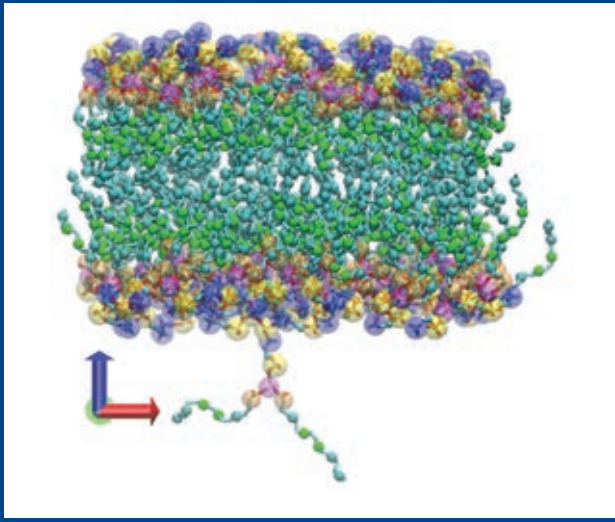


Figure 2. A snapshot from an interscale MD simulation of a phospholipid bilayer. The semi-transparent CG model representation is overlaid with the underlying atomistic representation placed in the background. The dynamics of atoms and CG beads are coupled in the course of simulation.

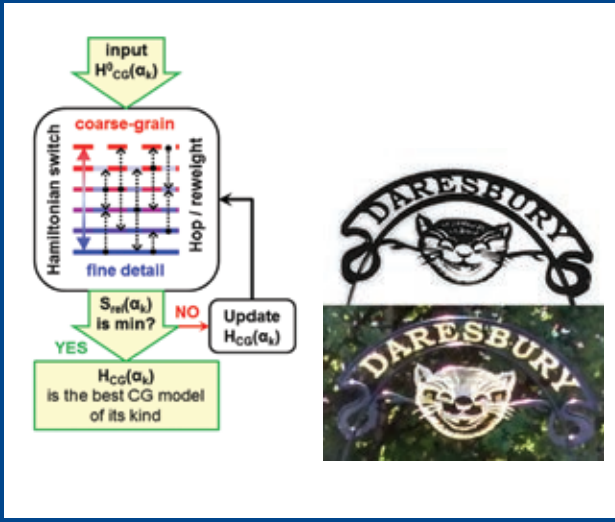


Figure 3. Schematic diagram illustrating the working principles of the interscale simulation protocol. The CG optimisation is performed on-the-fly by means of iterative updates of the CG force-field parameters. This iterative process allows the CG model to accommodate with respect to the changes in the molecular structure(s).

Developing algorithms to solve real-world problems

Mathematical models are used to solve a whole host of problems, from being able to better predict changes in the economy to the design of new drugs to fight diseases. In each case, complex numerical models are designed to fit real-world observations.

These models affect us all across our daily lives, so they must be designed accurately and efficiently.

Tyrone Rees, Nick Gould and Jennifer Scott, all part of Scientific Computing's Computational Mathematics group, have developed an algorithm that is capable of solving nonlinear least-squares problems. Least squares is a statistical method that is used to find the regression line (aka 'line of best fit') for a data-set collected from your real-world observations. This is a critical step in model design.

The algorithm has now been incorporated into the Numerical Algorithms Group's library, an Oxford-based software company, where it has been used to develop a novel data fitting solver that is faster (solves 60% of problems in less time), more robust (solves 25% more problems) and requires fewer callbacks than its predecessor.

> **"One of the central problems in computational mathematics is being able to fit a suitable model to observed data."**
>
> **Tyrone Rees, SCD Computational Mathematics Group**

Significance of least squares in the real world

Models are used to represent, understand and simulate data that has been collected in an objective manner. When designed well, models allow us to form predictions and make logical decisions about the best course of action, according to the data.

Model design begins with a problem. The first step to designing a successful model is to understand what the fundamental nature of the problem is. Once this has been determined, we can use our current understanding of the situation to theorize a model that fits these defined first principles.

The next step is to establish the parameters – or the factors to be considered – for the model.

For example, in civil engineering, models are used to design roads in cities that reduce traffic. The parameters for this may include:

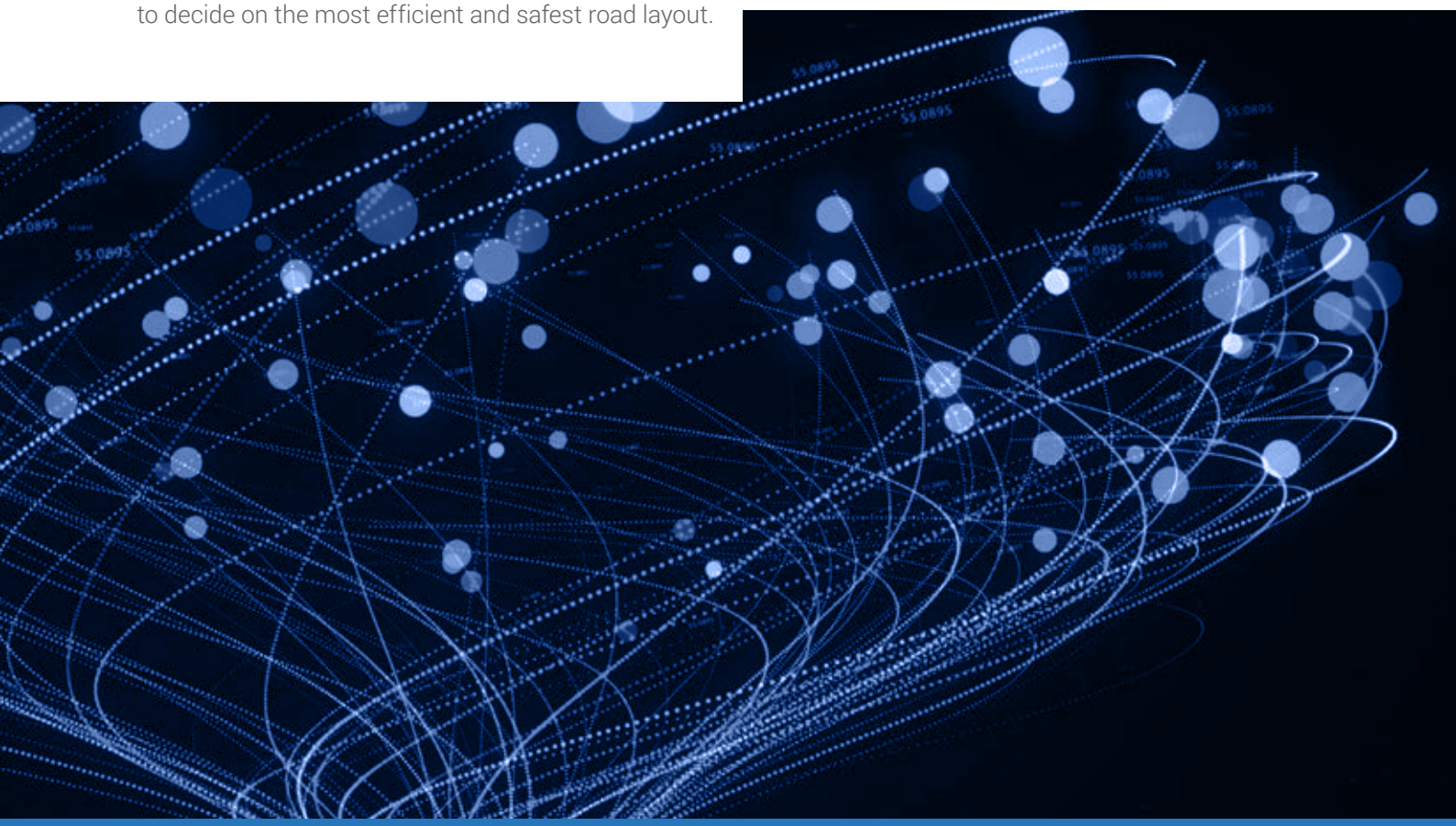
- How many people drive?
- Where are the busiest driving spots?
- How many people commute from outside the city?
- Where do pedestrians walk?
- Where do pedestrians need to cross the road?
- Where are local business, shops and schools?
- Easy access for emergency services?

All of these factors must be taken into consideration to decide on the most efficient and safest road layout.

Once the parameters have been decided, they can be translated into mathematical symbols which can be solved before being implemented into an algorithm. This is achieved by collecting the appropriate data. Once collected, it is important to establish if a relationship exists between the variables, by finding the regression line with minimal error. Least squares helps us do this.

If a relationship is found, it can be either linear or nonlinear. The difference between the two types of models is that the parameters of a linear model can be estimated in a linear manner, whilst the nonlinear models cannot. The independent variables of linear models do not have to follow this rule to still be linear. This means that linear regression lines do not necessarily have to be straight lines. They can express different curves if the parameters are linear in nature, but an independent variable is raised to an exponent function, or it contains a logged or inversed term.

In real-world scenarios, most relationships are nonlinear in nature. For example, with regard to road design, how many people drive is not a constant variable. We see that there are peak traffic times each day, with more cars on the road at commuter times when driving to and from work or school. Even these daily peaks and troughs are not constant as we also have increases during holiday seasons, such as when people are returning home for Christmas.



The least squares method

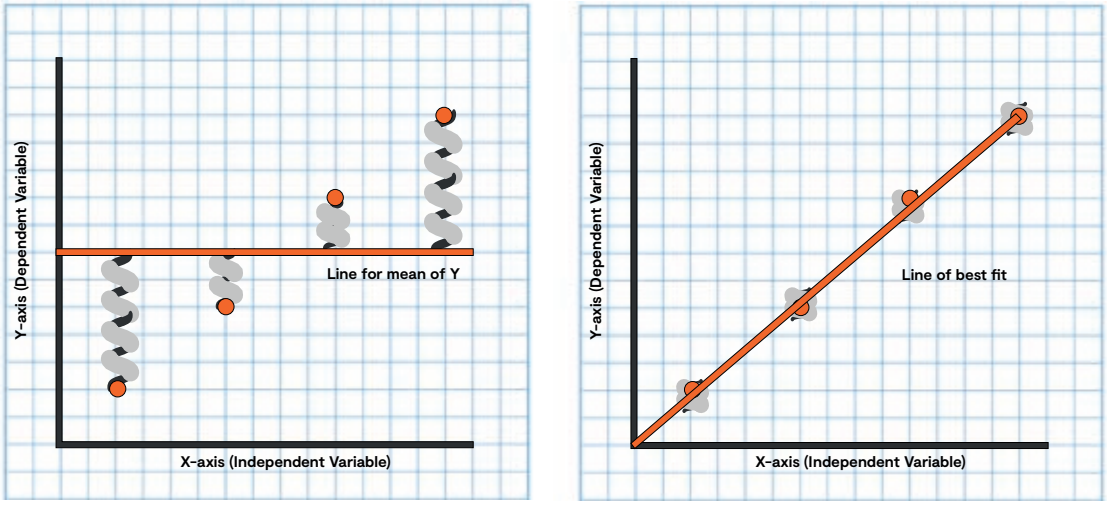
To visualise this method of regression fitting, imagine you have collected your data and plotted it on a graph. On this graph, you draw a line through the mean of Y (the average result for your dependent variable) and fix it in place. Then you attach springs to the individual data points and the line you have drawn. You notice that some springs are more extended than others. Now, if you allowed that line to move, the springs would exert a force on the line, rotating or bending it, and shortening the length of the springs.

By shortening the springs, we have found the regression line that has the least amount of error between our real-world observations and the model we have designed.

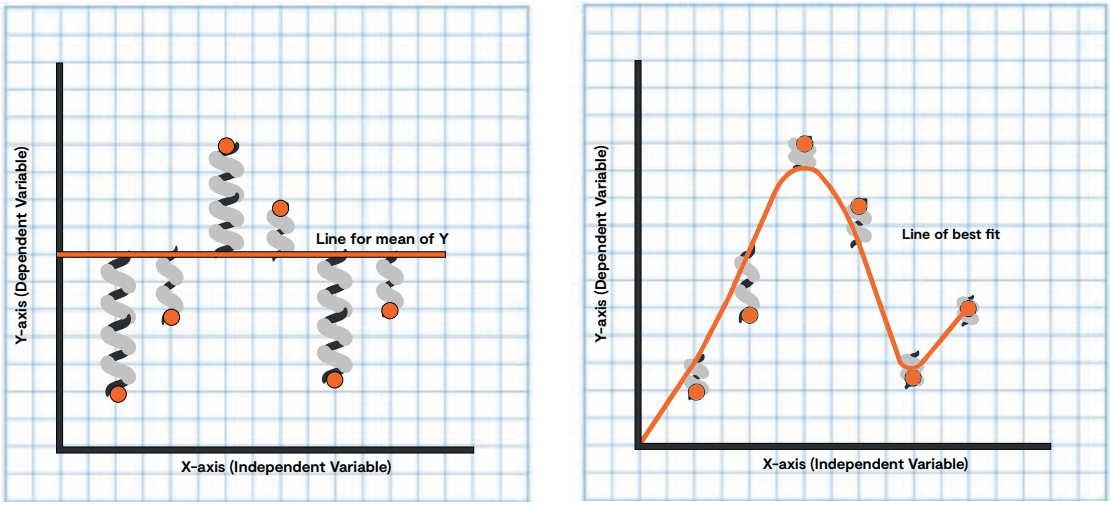
In essence, the shorter the springs, the less error there is, and the better the model.

So, why use an algorithm?

Most models have a huge number of parameters to consider. Mathematicians consider a model fitting problem to be small to medium in size if it includes up to the thousands of parameters. In such cases, attempts at drawing a graph and regression line would be futile. This is made even more challenging when the relationship between the parameters is nonlinear in nature. So, to solve this issue, we use algorithms and software to determine the regression line for us. This allows us to design better models and make more informed decisions.



Linear Regression fitting



Nonlinear Regression fitting

Authors: Tyrone Rees, Computational Mathematics Group, and Hal Dodman, Impact Team

Delivering solutions for continuous improvement

Scalable Cloud infrastructure provides vital services for Research Facilities

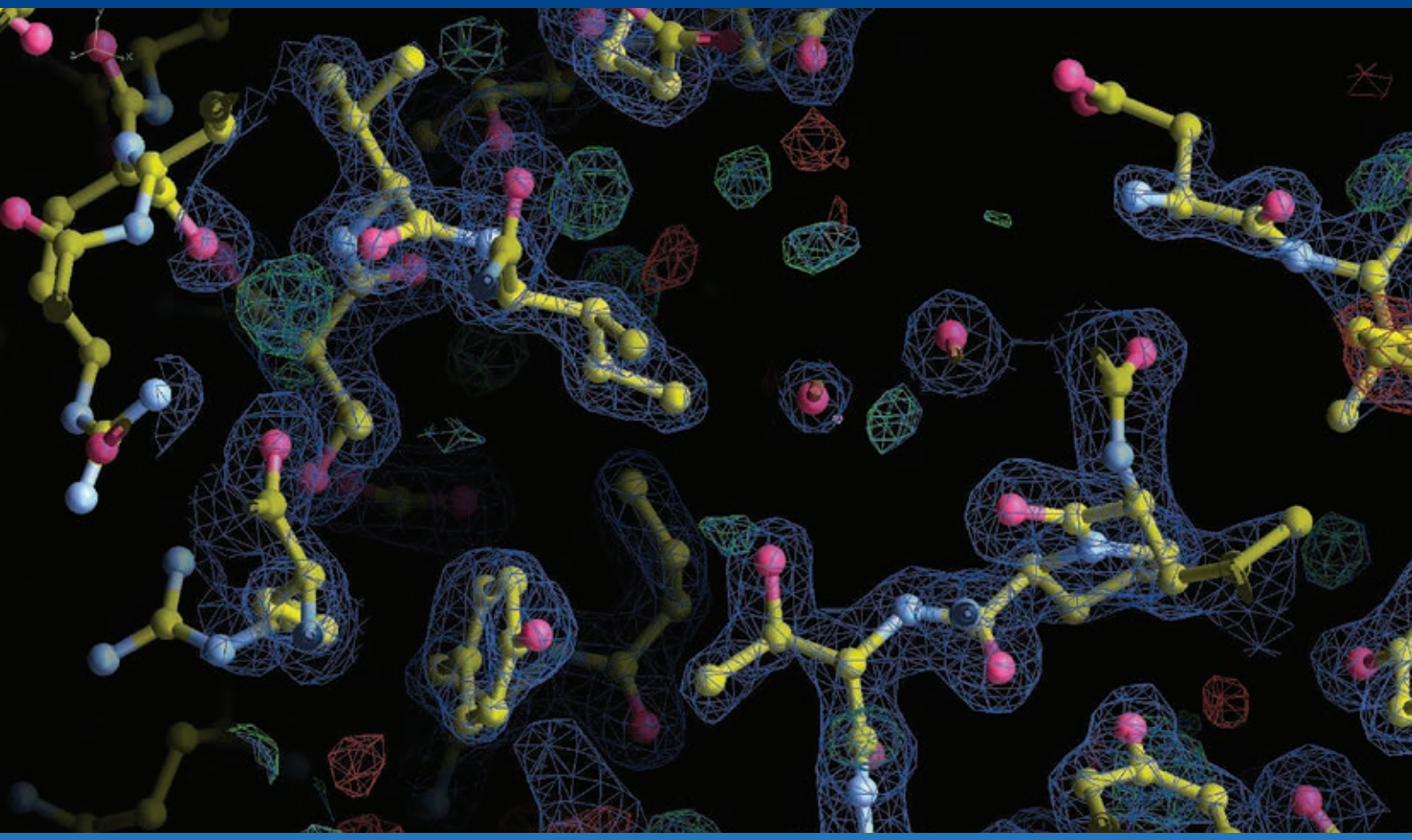
The STFC Cloud is a dedicated cloud infrastructure which enables users to access compute resources across the research facilities provided by STFC and partner organisations. Run by the Scientific Computing Department, it is designed to be flexible to allow as many different use cases as possible.

The aim of the Cloud is to allow users to perform complex data analysis as and when required, without the overheads of running their own computing infrastructure – an approach often referred to as Infrastructure as a Service. This allows for the rapid deployment of a highly scalable, dynamic computing environment which can be tailored to the needs of the user, easily adding or subtracting resources such as RAM, CPU, GPU and storage to the allocation to mirror the changing needs of the use case/s.

Virtual Machines (VMs) are supported on Hypervisors (physical hosts) and are available in a number of pre-configured setups which dictate the compute resource available. These VMs and Hypervisors are orchestrated in the STFC Cloud with the use of OpenStack technology. OpenStack manages the creation of VMs and which Hypervisor they will reside on, enabling migration between different hosts, attaches and allocates storage volumes to VMs and records the usage of the resources.

In 2021, the STFC Cloud configuration will include over 30k CPU cores, 200 GPU cores and 1.5PB of disk.

Crystallography visualisation



Due to the scalability provided by the Cloud computing environment at STFC, when the ISIS Neutron and Muon Source and the Diamond Light Source were seeking to increase the resources available to them during the Covid-19 global pandemic, the STFC Cloud team was very pleased to be able to reallocate additional processing power to those projects to aid the scientific research required to understand the properties of Covid-19. Additionally, STFC Cloud resource contributed to the Folding@ Home initiative, reaching a ranking within the top 400 of some 250,000 worldwide contributors.

The STFC Cloud also the hosts the Data Analysis as a Service (DAaaS) platform.

The DAaaS Platform

DAaaS provides an environment for STFC Facility users to analyse their data, using the appropriate software, data and processing on the STFC Cloud. Accessed via web browser, the platform allows the user to select an environment suitable for their analysis, and is available on the STFC Cloud as a VM for as long as they need.

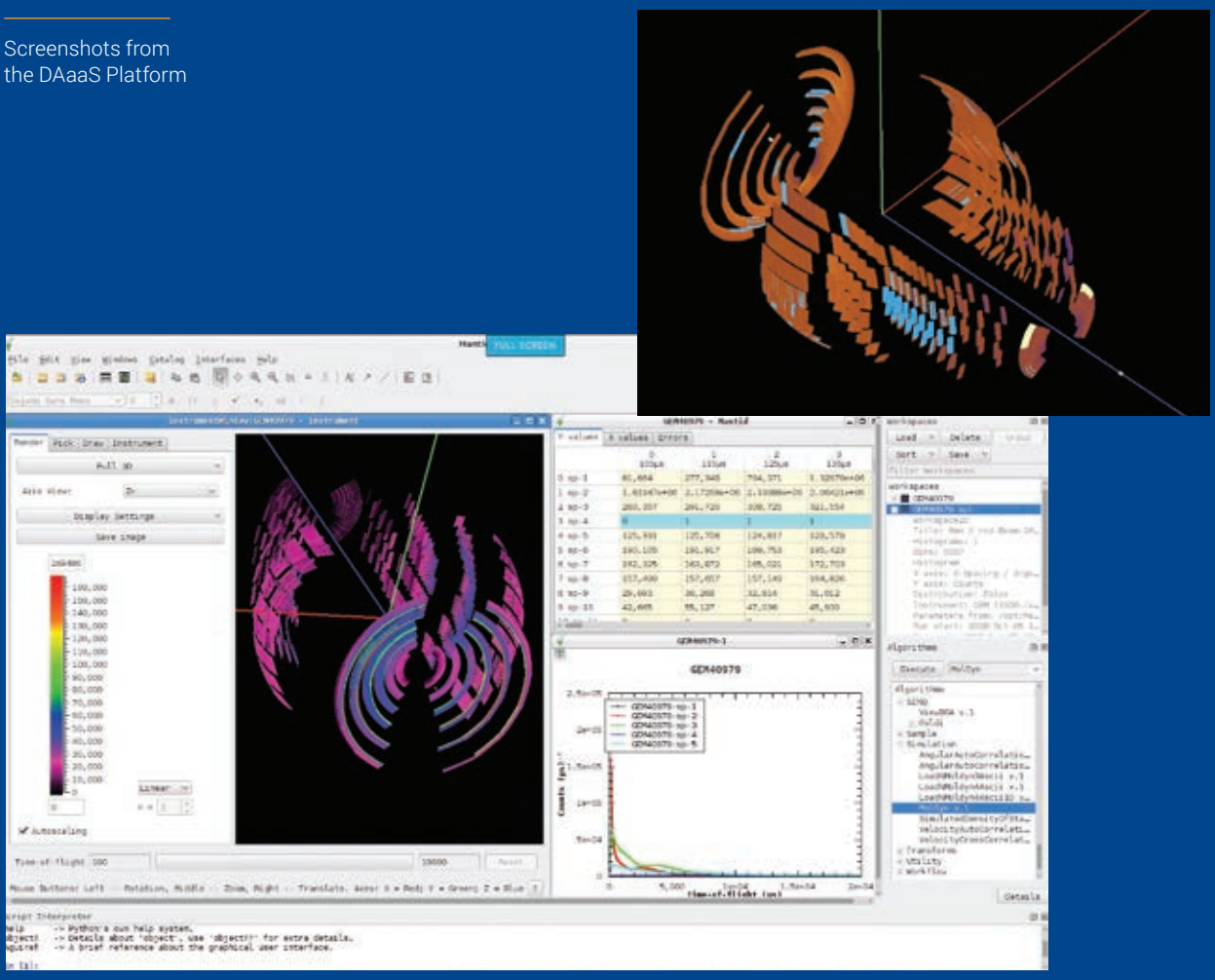
ISIS Instrument Scientists and visiting scientists who have come to use the facility can then use the DAaaS platform remotely. DAaaS hosts specific environments for Excitations, WISH, Reflectometry, Muon, Disordered Materials, IMAT, SANS, Engin-X and Crystallography groups, supporting 34 instruments.

The contribution of DAaaS during the pandemic was significant, with the ISIS User Cycle going ahead in the autumn of 2020 without any users present, and yet still able to analyse and visualise large datasets without any inconvenience.

The DAaaS team are passionate about their platform and making people's lives easier. It is rewarding when DAaaS users describe the service as a "life saver tool" or say "DAaaS has come to the rescue" because you know that what you do every day at work makes a significant impact for people in the scientific community.

Authors: Julimar Romero, Dynamic Infrastructure Group, and John Good, Cloud Services Group

Screenshots from the DAaaS Platform



Distributed Research Trust and Security

Distributed Digital Research Infrastructures continue to grow in importance following the research needs of scientific communities. Consequently, there is also an increasing need to provide assurance that these infrastructures operate in a secure and trustworthy fashion – both within the infrastructures themselves and as partners in the global research community. Staff within SCD and the Particle Physics Department (PPD) have extensive experience leading the development of this capability in distributed research security across Information Security Management, Trust and Security Policy, Identity Management and Operational Security. The combination of this expertise provides the basis for a new team, Distributed Research Trust and Security (DRTS).

Information Security Management

Information security management frameworks define the processes needed to carry out security activities. These activities are coordinated by staff in PPD for the GridPP project – the UK contribution to the Worldwide LHC Computing Grid (WLCG) – and the EU Horizon 2020 EGI-ACE and recently completed EOSC-hub (European Open Science Cloud) projects. In addition, staff in SCD and PPD led the activity that defined the Security for Collaborating Infrastructures (SCI) Trust Framework, which has been endorsed by many of the global eInfrastructures including GridPP, WLCG and EGI. This framework enables the management of security risk across collaborating infrastructures by specifying a number of requirements that an infrastructure should address.

Trust and Security Policy

The foundation of a secure and trustworthy infrastructure is a clear set of expectations of the behaviour of its users, its services, and of the infrastructure itself. In 2003 staff in PPD led the creation of the WLCG Security Policy Group, which expanded later to also cover the EU project Enabling Grids for E-sciencE (EGEE). This work was continued through 2010 with the creation of the EGI Foundation to maintain a set of policies in place across WLCG. Staff from SCD and PPD have built on this work by taking a leading role in developing policy templates for a Policy Development Kit (PDK) as part of the European AARC (Authentication and Authorisation for Research Communities) project. This PDK has been used recently as the basis of a Trust Framework for

IRIS, the eInfrastructure for Research and Innovation for STFC, a collaboration of STFC's science activities and its national computing centres at universities and STFC's own sites, driven by the physics communities supported by UKRI-STFC. This work is in turn contributing to the wider development of distributed infrastructure security policies as part of a mutually beneficial cycle.

Identity Management

Identity management is a vital aspect of distributed trust, as having an appropriate level of assurance of the identity of users is essential. 20 years ago, the EU DataGrid Project "CA Coordination Group" was created and subsequently led by a Rutherford Appleton Laboratory PPD staff member. This group was charged with managing the creation of the X.509 certificate-based Public Key Infrastructure used for Single Sign-on (authentication) to GridPP, WLCG and other Grids. The UK eScience Certificate Authority, managed by staff in SCD, provides such certificates to UK researchers, with around 3000 certificates in current circulation of which a fifth are user certificates. Authentication using these certificates subsequently enables their access to resources both within the UK and globally via appropriate authorisation mechanisms.

More recently, token-based federated identity mechanisms (already widely used by social media platforms and other internet services) are playing a vital role in the evolution of Authentication and Authorisation Infrastructure, both in the WLCG and across the research and education sector. The IRIS IAM (Identity Access Management) federated identity

proxy, also managed by staff in SCD, plays a critical role both in providing coherent access to the IRIS infrastructure and as an exemplar of the use of this technology to enable researchers to access computing resources in a secure, unified manner. This work proceeds as part of an international collaboration: SCD staff organised and ran a very successful IAM Users Workshop in January 2020 involving key members of the identity management community.

Operational Security

After the creation in 2005 of operational security teams to help prevent security incidents in WLCG, EGEE and other related Grids, with the creation of the EGI Foundation in Amsterdam in 2010 a strong operational security team was formed – and has since been in continuous operation - as a collaboration between staff from CERN, CESNET (Czech Republic), GRNET (Greece), Nikhef (Netherlands), and STFC. Staff from SCD and PPD have played significant leadership roles within this activity and continue to do so today, including the chairing by staff in PPD of the EGI Software Vulnerability Group (SVG). The SVG plays a vital role in assessing the impact of vulnerabilities in software used in the EGI infrastructure and advising sites in actions to be taken – sometimes urgently – as a result.

Since 2020 a new IRIS Security Team has been operating to develop and support the operational security capability within IRIS. This new team, made up of staff from across IRIS and coordinated by staff in SCD, builds on and enhances the vital processes developed by an existing team (now part of the larger team) designed to support GridPP sites. A particular focus of this work is the essential collaboration between security teams. Corporate (or campus) security teams often have a focus on core services including email and collaboration tools, while network security teams focus on the underlying network infrastructure and typically work at a national level. Research security teams have a focus on domain specific requirements including "custom" software stacks, distributed user communities and facilities, and large data transfers. Collaboration between these types of teams is essential, both within the UK and internationally. Finally, staff in SCD are leading and enabling collaboration in the sharing and active use of threat intelligence to enhance distributed incident response, which has been identified as a cornerstone of the WLCG security strategy in the next 5 years.

Leading future security development

Information security management, trust and security policy, identity management and operational security are all vital elements in the operation of a secure, trustworthy distributed infrastructure. Staff in SCD and PPD have been involved for many years in the creation and leadership of activities in these areas on the international stage. We are building on our experience to support security activities in Digital Research Infrastructures, both those in operation today and those that develop in the future.

For more information, please contact the Distributed Research Trust and Security Team at drts@stfc.ac.uk

Author: David Crooks,
Distributed Computing Infrastructure Group

Contacts

Mr Tom Griffin	**Director, Scientific Computing Department**	**tom.griffin@stfc.ac.uk**
Mrs Shirley Miller	*Personal Assistant*	*shirley.miller@stfc.ac.uk*
Prof Tony Hey	**Chief Data Scientist**	**tony.hey@stfc.ac.uk**
Mrs Sally Prydderch	*Personal Assistant*	*sally.prydderch@stfc.ac.uk*

Division Heads and Group Leaders

Dr Peter Oliver	**Head of Operations**	**peter.oliver@stfc.ac.uk**
Mrs Sally Prydderch	*Personal Assistant*	*sally.prydderch@stfc.ac.uk*

Group Leaders

Mr David Macaree	Project Management Office	david.macaree@stfc.ac.uk

Dr Juan Bicarregui	**Head of Data**	**juan.bicarregui@stfc.ac.uk**
Mrs Amanda Chapman	*Personal Assistant*	*amanda.chapman@stfc.ac.uk*

Group Leaders

Dr Brian Matthews	Data Science & Technology/ DAFNI	brian.matthews@stfc.ac.uk
Ms Elizabeth Newbold	Open Science Services	elizabeth.newbold@stfc.ac.uk

Mr Gordon Brown	**Head of Software Infrastructure**	**gordon.brown@stfc.ac.uk**
Mrs Amanda Chapman	*Personal Assistant*	*amanda.chapman@stfc.ac.uk*

Group Leaders

Mr Gordon Brown	Dynamic Infrastructure	gordon.brown@stfc.ac.uk
Dr Alejandra Gonzalez-Beltran	Data and Software Engineering	alejandra.gonzalez-beltran@stfc.ac.uk
Dr Jeyan Thiyagalingam	Scientific Machine Learning	t.jeyan@stfc.ac.uk
Dr Tyrone Rees	Computational Mathematics	tyrone.rees@stfc.ac.uk
Dr Fan Yang-Turner	Data Analysis as a Service	fan.yang-turner@stfc.ac.uk

Dr Andrew Sansum	**Head of Systems**	**andrew.sansum@stfc.ac.uk**
Mrs Amanda Chapman	*Personal Assistant*	*amanda.chapman@stfc.ac.uk*

Group Leaders

Mr Nick Hill	Research Infrastructure	nick.hill@stfc.ac.uk
Ms Alison Packer	Data Services	alison.packer@stfc.ac.uk
Mr Ian Collier	Distributed Computing Infrastructure	ian.collier@stfc.ac.uk
Mr Alex Dibbo	Cloud Operations	alexander.dibbo@stfc.ac.uk

Dr Barbara Montanari	**Head of Computational Science and Engineering/ CoSeC**	**barbara.montanari@stfc.ac.uk**
Ms Esme Williams	*Personal Assistant*	*esme.williams@stfc.ac.uk*

Group Leaders

Prof David Emerson	Engineering and Environment	david.emerson@stfc.ac.uk
Dr Gilberto Teobaldi	Theoretical and Computational Physics	gilberto.teobaldi@stfc.ac.uk
Prof Ilian Todorov	Computational Chemistry	ilian.todorov@stfc.ac.uk
Dr Martyn Winn	Biology and Life Sciences	martyn.winn@stfc.ac.uk

Administrative Team

Mrs Karen McIntyre	karen.mcintyre@stfc.ac.uk
Ms Georgia Lomas	georgia.lomas@stfc.ac.uk
Ms India Reeves	india.reeves@stfc.ac.uk
Miss Helen Walker	h.walker@stfc.ac.uk

Further information

Scientific Computing Department	**www.scd.stfc.ac.uk**
CoSeC (Computational Science Centre for Research Communities)	**www.cosec.stfc.ac.uk**
DAFNI (Data and Analytics Facility for National Infrastructure)	**www.dafni.ac.uk**
Science and Technology Facilities Council (STFC)	**www.stfc.ukri.org**
UK Research and Innovation	**www.ukri.org**

Notes

Nick Hill,
Research Infrastructure Group Leader,
in the Scientific Data Centre at the
Rutherford Appleton Laboratory.

Head Office
Science and Technology Facilities Council,
Polaris House, North Star Avenue, Swindon SN2 1SZ, UK

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

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

Establishments at:
Rutherford Appleton Laboratory, Oxfordshire
Daresbury Laboratory, Cheshire
UK Astronomy Centre, Edinburgh
Chilbolton Observatory, Hampshire
Boulby Underground Science Facility, Boulby Mine, Cleveland