Science and Technology Facilities Council

Scientific Computing Department

Annual Review 2021-22

SCD Annual Review edited and produced by
Marion O'Sullivan, Impact and Communications Manager
marion.osullivan@stfc.ac.uk

Cover image: STFC Scientific Data Centre at the Rutherford Appleton Laboratory.
Inside front cover image: JASMIN Super-Data-Computer

Contents

FOREWORD

Tom Griffin, Director SCD 5

CONNECTING PEOPLE AND BUSINESSES

Computing Insight UK – Scientific Computing's Flagship Annual Conference 6

INSPIRING CAREER CHOICES AND ENHANCING SKILLS

Expanding Skills and Knowledge 8

Encouraging Open Science Practices 10

Inspiring the Next Generation 12

Remote^3 – Public Engagement in a Pandemic 14

ACCELERATING INNOVATION

A Shining Example of Collaboration to Advance Solar Energy Technologies 16

Fast-Tracking Drug Discovery using the STFC Cloud 18

A Sustainable Spin to Energy-Efficient Computers and Smart Phones 20

Nature's Aerodynamic Blueprints Inspire New Engineering Designs 22

Using Computers to Design Safe and Sustainable Nuclear Reactors 24

DELIVERING SOLUTIONS

IRIS: Addressing Security Challenges in a Digital Environment 26

The Trail from Data to Policy in the COVID-19 Pandemic and Beyond 28

OpenCLIM: Assessing the Impact of Climate Change 30

Global Optimization of Parameters for ISIS Neutron and Muon Source 32

Intelligently Deciphering Small Angle Scattering Data 34

CONTACTS **36**

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 rapidly growing to meet the ever-increasing demand for innovative software solutions, digital research infrastructure and computational expertise in a variety of disciplines.

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, with 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 provide 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.

Foreword

Welcome to the 2021-22 annual review for STFC's Scientific Computing Department.

Despite the challenges posed by the COVID-19 pandemic continuing well into 2021 we have had a very productive year. You will see from some of the articles in this review how we have been collaborating with others to understand the spread of the disease, developing new computational tools and software, and helping to compile and analyse data to improve public health policies.

Environmental sustainability and the route to net zero carbon is high on government agendas and we are keen to encourage the use of materials that are abundant and accessible, removing the need for harsh mining practices and enabling improved manufacturing processes. Our computational scientists and engineers are addressing such challenges by providing the tools and expertise that enable researchers to make scientific advances in these areas.

COP26 in November 2021 highlighted the urgency of dealing with the effects of climate change. We are partnering with other organisations to investigate, and find ways to mitigate or adapt to, some of the most pressing climate risks that will affect people and their environments, agricultural practices, and the diversity of wildlife.

Our work provides a foundation for many of the experiments carried out in the UK's large-scale national facilities, increasing efficiency and speed of results, and enabling researchers to concentrate on their science rather than the computational requirements. A major concern is the security of data and how it is accessed. Through the IRIS digital research infrastructure we are leading the development of services that enable users to access data securely and confidently using robust authorisation mechanisms.

As you will note from this review, an important and immensely rewarding aspect of our work is being able to inspire future generations of scientists. I am particularly pleased that our staff are very active in encouraging young people to get involved in coding projects – something that often helps to boost their confidence and shows that computational science isn't such a difficult subject.

I am also pleased that, with the easing of COVID restrictions, we have been able to return to holding physical events. In December we welcomed some of our research communities and industry partners to our flagship HPC conference and exhibition, Computing Insight UK. As one of the first events that people had attended for almost two years, it was very well received and particularly successful. I hope we will have the opportunity to meet some of you at our conference in December 2022.

Tom Griffin,
Director

Connecting people and businesses

Computing Insight UK – Scientific Computing's Flagship Annual Conference

In December 2021 we were able to return to hosting an in-person event – the first for us since the UK went into lockdown following the Covid-19 outbreak in early 2020. This conference was a great success as delegates and exhibitors celebrated being able to meet and network once again, albeit with some Covid precautions still in place.

Dr Barbara Montanari, Director, CoSeC

Computing Insight UK (CIUK) brings together the UK academic computing community, industry and researchers. It combines an exhibition of the latest High Performance Computing (HPC) hardware and software with a wide and varied programme of presentations and events. These are all delivered by speakers who are experts in their fields, and can provide delegates with some insight into how they use HPC in real life situations. This year's conference was held as a hybrid event, with more than 350 people attending physically and on-line.

Scientific Computing's Director, Tom Griffin, said, "CIUK was really well attended this year. We were so pleased to be back in person after a year of Covid disruption and last year's virtual event. It was really good to see the level of interaction we had between researchers, system administrators and the industry partners who were there for the event."

The industry exhibitors provide delegates with an opportunity to find out about the latest technology, discuss product roadmaps and compare pricing schedules. And having 400 plus people attending the conference also provides the vendors with a chance to build their UK customer base.

One first-time exhibitor said, "As a storage company looking to build relationships with the HPC community in the UK this event has been absolutely fantastic," and he is already looking forward to CIUK 2022.

An exciting aspect of this conference is the Student Cluster Challenge. In 2021 it became a hybrid competition with a series of online challenges carried out during the two months leading up to the conference, followed by some physical, hands-on challenges during the conference itself. Teams of six students take part and the goal is for the competing teams to build their own cluster and run software on it, and this has to be achieved within a fixed power limit. It is a great opportunity to improve the skills base within the student population, essentially training up the next generation of people who will be building and running our HPC systems.

We congratulate the winning team from the Universities of Bath and Bristol, and also the runners-up from Durham University. The challenge was supported by a number of companies, including Alces Flight, Boston, OCF and Lenovo, who made systems available and set the challenges for the teams to complete.

Running in parallel with the main CIUK programme was the inaugural conference for CoSeC, the Computational Science Centre for Research Communities, which is hosted and supported by SCD. This also proved to be a popular part of the overall event. Dr Stephen Longshaw, Principal Computational Scientist, said "Over the course of the day attendance was really good, all speakers delivered fantastic content, highlighting the breadth of interrelated computational science that happens across the CoSeC programme.

CIUK evolved from what was once a technical workshop with a focus on evaluating new computing systems for academia. It is now one of the UK's premier HPC conferences, held each December in Manchester.

More information: **www.stfc.ac.uk/ciuk**

Top: Steve Hindmarsh, Head of Scientific Computing at the Francis Crick Institute, speaking at CIUK 2021

Above: CIUK 2021 Exhibition Hall

Inspiring career choices and enhancing skills

Expanding Skills and Knowledge

DAaaS team meeting

Part of our mission is to actively contribute to the developing skills agenda in scientific computing and related disciplines. Providing training and knowledge-sharing workshops plays an important role in ensuring our own staff and the facilities and communities we support receive the essential skills they need to aid their research.

Many of our groups provide such specialist training, including the Computational Science Centre for Research Communities (CoSeC), which supports the advancement of scientific research by developing and strengthening software to analyse and solve increasingly complex problems in multiple disciplines - physics, chemistry, life sciences, engineering, and more. Its staff run training workshops, and host scientific visits to facilitate the exchange of knowledge and expertise between researchers.

CoSeC provides a hub for exchanging knowledge and expertise through training and outreach. In the last year (to March 2022) 44 members of staff have conducted a total of 2545 training days through activities such as: running hackathons where scientists come together to write code to solve a problem under the guidance and mentorship of CoSeC staff; running a 10-day summer school designed to teach modelling and simulation techniques that can be used to explore a wide range

of materials such as those used in the nuclear industry, the walls of cells found in the human body, and the impact of meteorite fragments on the surface of a rocket.

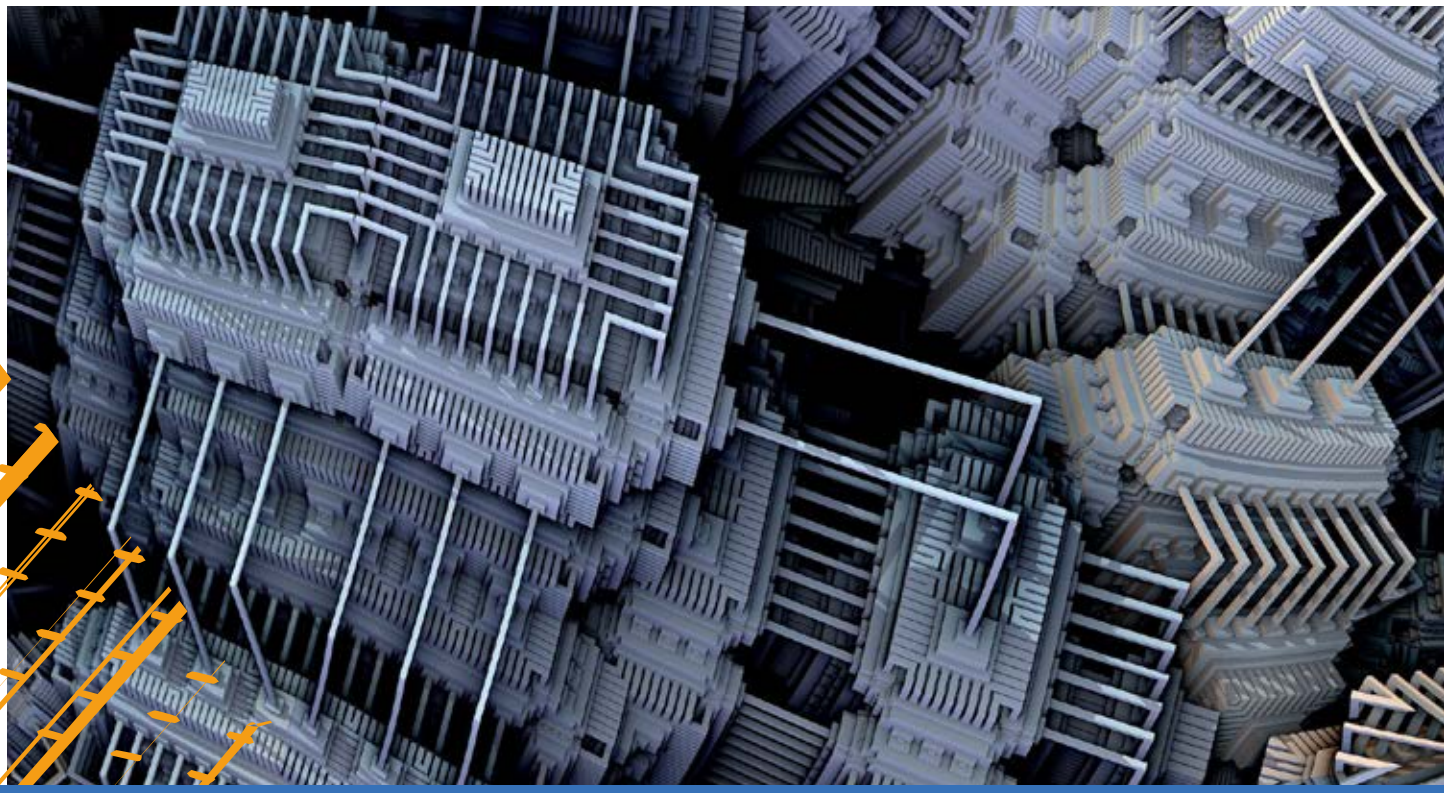
CoSeC is especially keen to recognise the work of Early Career Researchers and ran the CoSeC Impact Award again this year; the overall winner and 3 runners-up were mostly PhD students at the time of applying.

In 2021 our Data Analysis as a Service (DAaaS) Group worked with the ISIS Neutron and Muon Source to run specific training courses for users of ISIS instruments. They ran three separate training courses over 21 days and trained more than 120 users. One of these was a four day Virtual Reflectometry training school, with talks, live demonstrations for some of the tools, and hands-on guided practical sessions – all carried out in a virtual environment and attended by students from around the world. Reflectometry is a technique for measuring the chemical composition and structure of thin layers at a surface or interface, and used in applications such as medical devices or fuel cells.

And our Scientific Machine Learning Group ran four intensive training courses in AI and Machine Learning (ML). Each course ran for four to five days and trained groups of 40 – 50 people from some of our large facilities, STFC-funded programmes, and our own staff-base. They also held a webinar for much broader audiences, focusing on ML for large scale facilities which covered topics from the development of new ML-based tools to the advances in ML technology inspired by the challenge of facilities' data.

In addition to running specific training courses, we organise and participate in a number of conferences and workshops for a variety of research communities – such as the CECAM (Centre Europeén de Calcul Atomique et Moléculaire) Workshop for Data-enabled Atomistic Modelling, a hybrid event with both in-person and remote attendees, hosted by CoSeC in November. And the hybrid workshop for Integrating Quantum Computers in Condensed Matter Physics Simulations, held in London in September.

CoSeC is funded by the Engineering and Physical, Medical, and Biotechnology and Biological Sciences Research Councils, and supported with staff and computational expertise from STFC's Scientific Computing Department.



Encouraging Open Science Practices

We encourage knowledge exchange and engagement within STFC and the wider UK Research and Innovation (UKRI) communities. The desire to make research data open and accessible has never been stronger, and within the STFC National Laboratories there is also an obligation to comply with best practices in Open Science.

Open Science Café

Enabling busy STFC researchers and support staff to keep up-to-date with Open Science.

The Open Science Services Group has provided expertise and guidance through the Open Science Café initiative. Introduced in the latter part of 2021, this is a series of short, informal monthly webinars on Open Science topics. The webinars are available to staff from all STFC National Laboratories' departments and many are extended to other parts of UKRI and the communities it supports.

The group has already had some very interesting talks from both STFC staff and from organisations within the wider research community – such as how the Alan Turing Institute engages in sharing open and reproducible research practices; how STFC's Centre for Environmental Data Analysis tackles the issues surrounding crediting data sources; and an introduction to the Digital Preservation Coalition and its work.

"The talks are both informative and entertaining, providing a mixture of inspirational examples and practical advice in 'bite-size' chunks that are easy to understand and digest, and give opportunities to ask questions. And because they are short they are popular and well attended, as staff can easily fit them into their working day," says information specialist Tracy Colborne who, alongside colleague Katie Yates, organises and hosts the webinars.

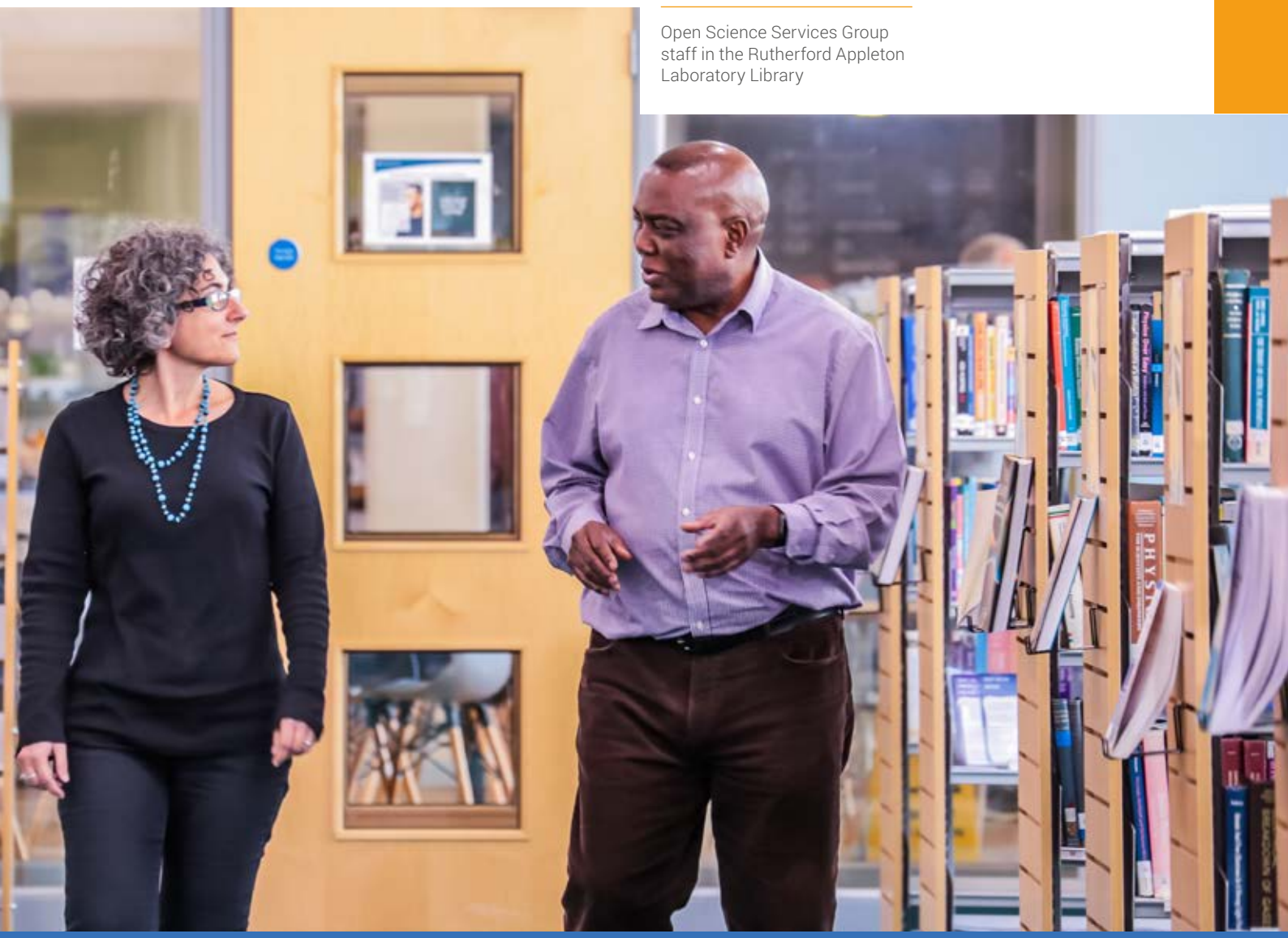
The webinars give attendees the chance to get to know and interact with the group, which is providing an invaluable resource for knowledge exchange and engagement.

Open Access Policy

In early 2022 UKRI introduced a new Open Access policy that applies to all core-funded staff, including researchers in the National Laboratories. Sara Fletcher, who leads on impact for the ISIS Neutron and Muon Source says, “The Open Science Services Group has been really proactive in developing guidance on how to comply with the new policy and supporting individual researchers on the best routes for publishing their scientific outputs.”

Monitoring and evaluation is crucial to delivering the STFC and UKRI visions. It provides an evidence base to support effective decision-making and is an important means of ensuring investments in research and innovation deliver on their aims; push the frontiers of human knowledge and understanding; and deliver economic, cultural and societal impact. For research-intensive departments like ISIS, this means capturing, analysing and reporting on the impact of their scientific publications.

The Open Science Services Group has provided the tools and expertise for departments to capture and analyse their research impact. EPubs, the open archive for STFC research publications, allows departments to capture relevant metadata that can then be imported into research evaluation tools, which the group also supports. This has provided new insights into the research impact of the facilities, informed the development of the business case for the ISIS Endeavour programme and is being used to underpin a pilot study to evaluate the impact of facility research in addressing the Carbon Net Zero challenge.



Open Science Services Group staff in the Rutherford Appleton Laboratory Library

Inspiring the Next Generation

We aim to inspire and encourage the next generation of computational scientists by engaging children and families in science coding activities. We have a very active Public Engagement Team who work hard to bring a variety of events to schools and members of the public in the UK and further afield.

In 2021 we:

HELD 61 Public Engagement Events

ENGAGED With 58 Schools

REACHED Over 3,400 Students, Teachers and Members of the Public

ENCOURAGED 99 Children Aged 8-16 with our SATURDAY CODING CLUBS

INSPIRED 383 People, Including: Scout Groups, Schools and Families with our REMOTE[3] Robotics Coding Project*

DELIVERED A Computing Workshop using LARGE HADRON COLLIDER DATA for 947 A-Level Students as part of a Particle Physics Masterclass

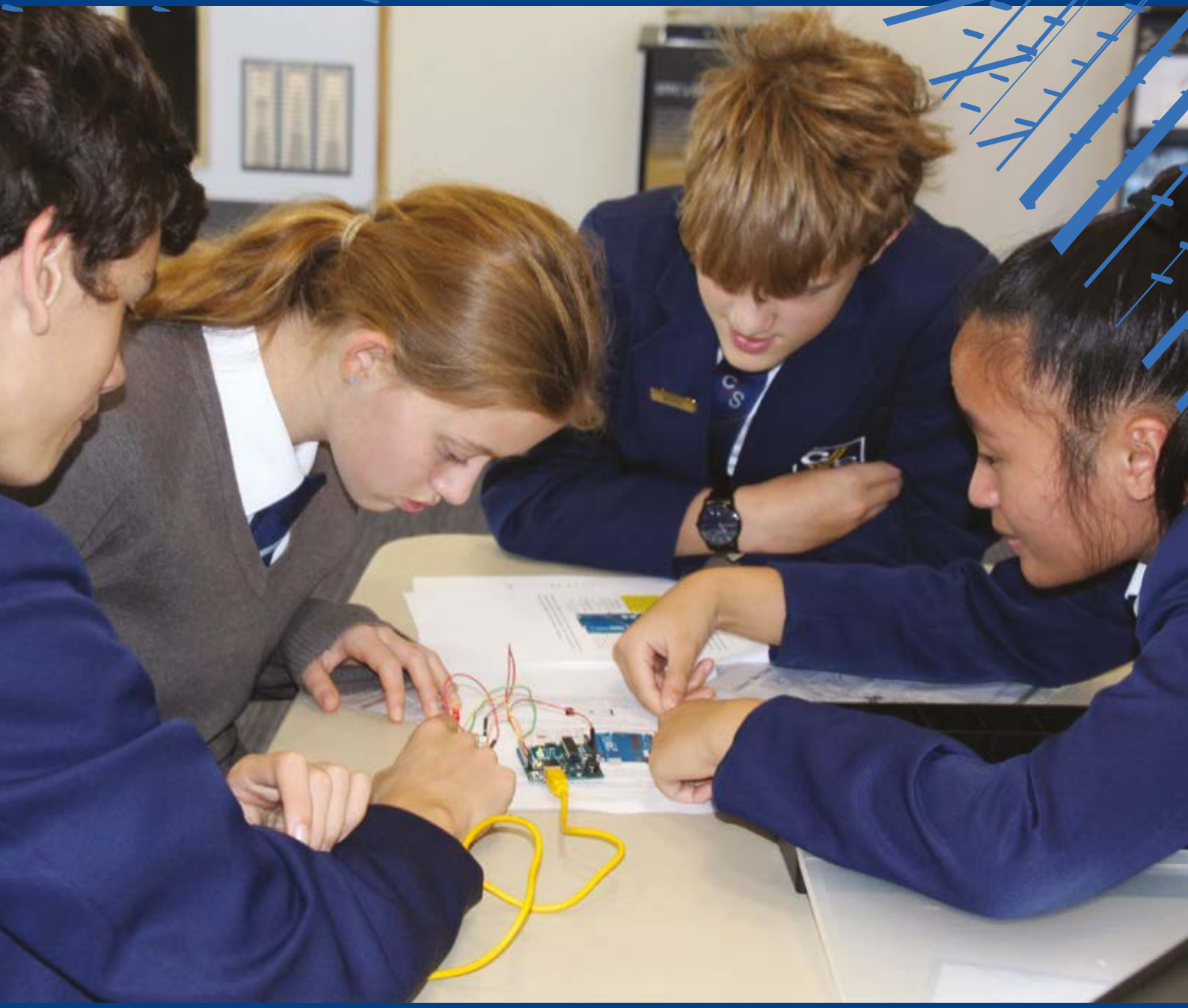
EXCITED > 850 Members of the Public with our 'GLOW YOUR OWN' Sessions**

INSPIRED 124 Students and Teachers with a Coding and Design Challenge in Similar Conditions to those Found on MARS

HOSTED A Summer School on Coding with MICRO:BITS*** for 247 Students

RAN Virtual Python Code Masterclasses For 211 Work Experience Students

Graphics designed by Orlagh Simpson, Impact Team, Scientific Computing



School students test their coding skills

* Remote[3] was originally designed to enable remote schools in the Scottish Highlands to access STEM activities and challenges, such as using remote sensing in the remote environment of Boulby Underground Laboratory. It has since been widely expanded to reach many more schools.

** Glow Your Own creates unique lanterns using Arduinos (simple electronics boards)

*** Micro:bits are pocket-sized computers that introduce you to how software and hardware work together. This activity was held in partnership with Computing at School.

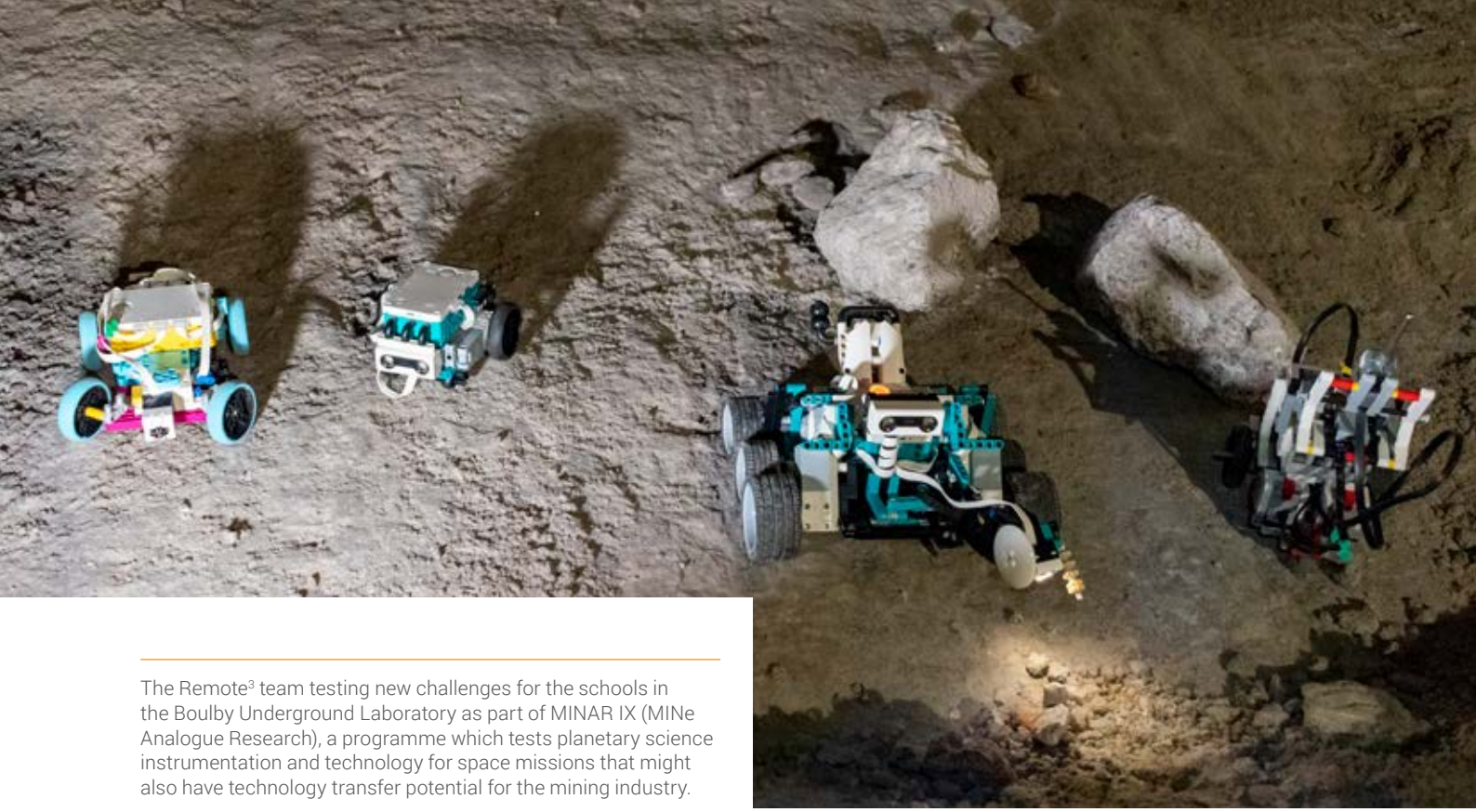
Remote^3 – Public Engagement in a Pandemic

The Scientific Computing Department (SCD) has always actively supported STFC's Public Engagement programme, with activities ranging from Data Centre tours for site visits to week-long computing workshops for work experience students. One such activity is Remote^3, which stands for Remote Sensing by Remote Schools in Remote Environments.

Remote^3 is a Spark Award funded collaboration led by the University of Edinburgh (UoE) in partnership with several STFC departments – Public Engagement, Boulby Underground Laboratory and Scientific Computing. This school robotics project has been designed to help students from schools in the Scottish Highlands access exciting STEM (science, engineering, technology, mathematics) challenges. Schools are provided with a LEGO Mindstorms Robotics kit with which they must design their own "Mars" rover, all whilst receiving support from a mentor from SCD, Boulby, or UoE. This rover will need to complete a series of announced challenges, with final testing taking place remotely at the Mars Yard at Boulby Lab.

Remote^3 was originally scheduled to launch in March of 2020; however, as with many things, even the best laid plans did not account for a global pandemic. With after-school clubs no longer happening, the original plan was no longer feasible for a 2020 launch. Despite this, the original remote engagement component of Remote^3 meant that the team was well placed to exploit existing connections with schools and pivot to a programme of fully virtual outreach.

The first steps taken for Remote^3's move away from the original plan was organising a series of weekly challenge webinars, with the audience expanded to reach more than just the original set of schools. Children all over the country were invited to attend these webinars, which introduced a new coding challenge each week, as well as providing information about the uses of robots in science, and the opportunity to learn about the research carried out at STFC's Boulby Underground Laboratory. The challenges tasked attendees to create code using languages and tools including Scratch and MakeCode, ranging from designing a space exploration game to completing simple tasks on a pre-made LEGO Mindstorms robot – all while receiving support and advice from the Remote^3 team.



The Remote^3 team testing new challenges for the schools in the Boulby Underground Laboratory as part of MINAR IX (MINe Analogue Research), a programme which tests planetary science instrumentation and technology for space missions that might also have technology transfer potential for the mining industry.

The Remote^3 team has also started to work with other audiences as well as with schools. The team has taken part in and run activities for the Muslim Scout Fellowship. This collaboration reaches a different audience than the schools, with the team working with the fellowship to run activities to support Beavers and Scouts as they work towards achieving some Space and Astronomy themed badges. Since the first collaborative event, which took place on 27th January 2021, there have been a number of further events including remote activities, Facebook Live talks, and, as of February 2022, an in person tour of Boulby for Scout Leaders. These events have been a huge success for Remote^3, allowing the team to exploit lessons learnt over the course of the COVID pandemic to engage with an entirely new audience. Over the coming months the team plans to build on the work already done and develop and launch a Remote^3 scouting badge, with the aim of sharing the activities even wider.

As lockdown restrictions have eased over the end of 2021 and start of 2022 Remote^3 has been able to relaunch the original school clubs scheme, with the launch event taking place on the 30th March 2022. Schools have been provided with their Mindstorms kits, and mentors have been introduced to their schools. Despite the return to the original plan, Remote3 plans to still utilise contacts and resources so that they continue to engage with different and new audiences – including engagement at both virtual and in-person events and festivals.

Over the past two years Remote^3 provided SCD with the opportunity to continue to participate and lead innovative engagement, even whilst the world was locked down. Through the partnership with the University of Edinburgh, we have been able to engage new audiences. Over this time, SCD has also continued to build and improve a portfolio of remote engagement including: partnering IF Oxford, Fusion Arts, Oxford Christmas Lights Festival and Blackbird Leys Community Centre to create a new interactive Arduino coding session; a python workshop for work experience students on the STFC Cloud training platform; and is continuing to run remote and hybrid training sessions for our staff.

The opportunities provided and lessons learnt over the past two years as we have adapted to virtual engagement have been invaluable and will be applied as STFC continues to develop its Public Engagement programme to a more hybrid world.

Authors: Greg Corbett & Tom Dack, Distributed Computing Infrastructure Group

Accelerating innovation

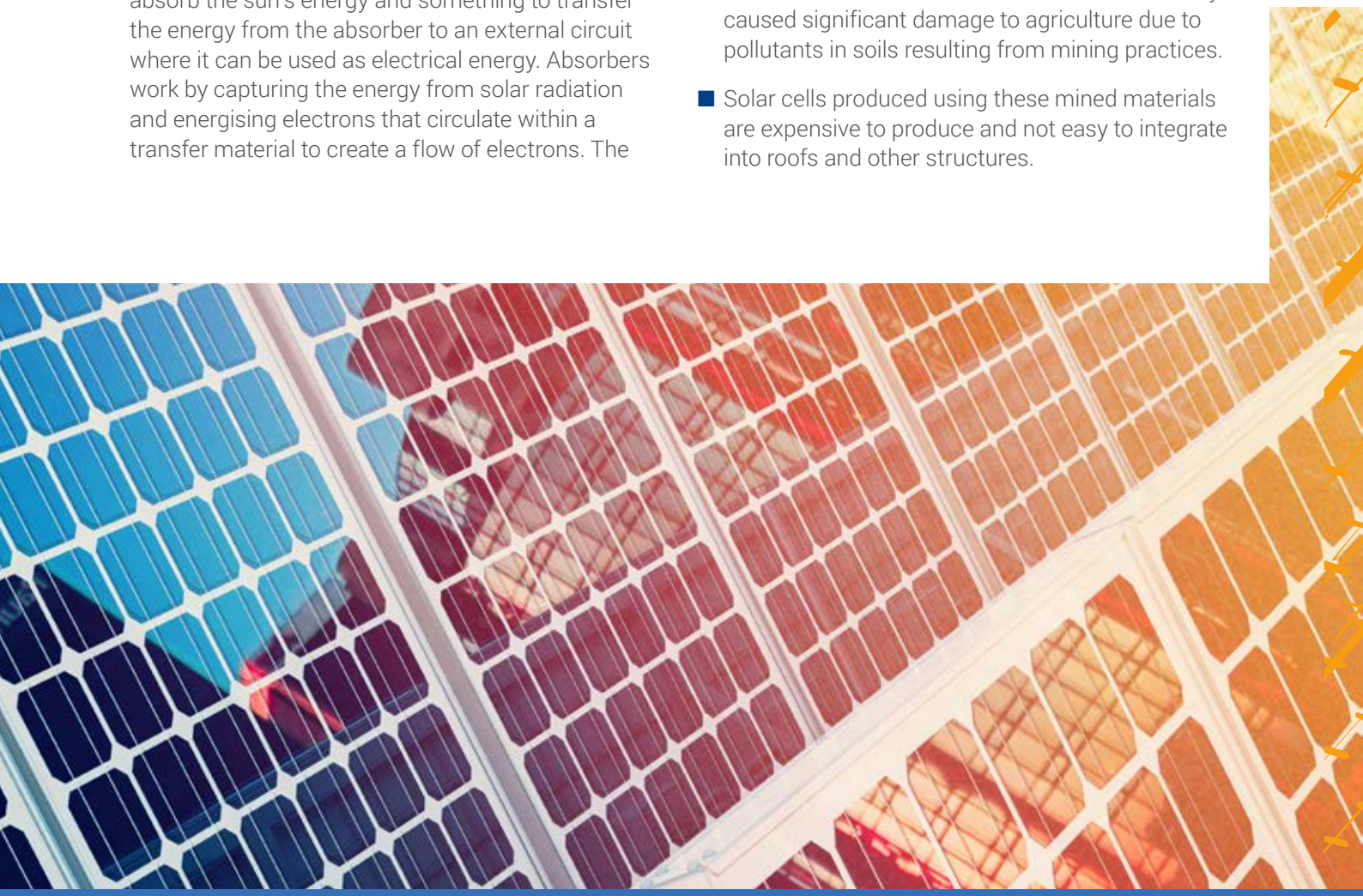
A Shining Example of Collaboration to Advance Solar Energy Technologies

Our computational scientists are working with two UK universities, using real experimentation and high quality computer modelling to accelerate solar cell technology. Improving this technology should push down the price of solar cells and encourage their use as a more accessible and affordable energy solution worldwide.

As the climate crisis intensifies, the switch to renewable energy has become a fundamental need for the future. To meet the UK's ambitious net zero targets and also keep up with modern demands, the research and development of clean energy solutions need to advance quickly.

Solar cells use radiation from the sun to generate the flow of electrons that produce electricity. The solar cell needs two main components – something to absorb the sun's energy and something to transfer the energy from the absorber to an external circuit where it can be used as electrical energy. Absorbers work by capturing the energy from solar radiation and energising electrons that circulate within a transfer material to create a flow of electrons. The transfer component is commonly made from (semi-conducting) materials such as silicon or rare earth metals, both of which have major disadvantages:

- The manufacture of semi-conductors has several negative impacts on the environment; polluting underground water - that ultimately might be used as drinking water and for crop irrigation - and it also generates toxic waste.
- The manufacture of rare earth metals has already caused significant damage to agriculture due to pollutants in soils resulting from mining practices.
- Solar cells produced using these mined materials are expensive to produce and not easy to integrate into roofs and other structures.



Our approach

Dr Dawn Geatches, Dr Kakali Sen and Dr Ya-Wen Hsiao are collaborating with researchers at the Universities of Glasgow and Swansea, using computer models to improve the component of solar cells that absorbs energy. The solar cell type used in this research uses dyes as the absorber of the sun's radiation, and these dyes are adsorbed on the surface of the transfer component – in this case titanium dioxide.

Titanium dioxide is a common, easily mined crystal that is used safely in all sorts of products, from sun cream to toothpaste. It is abundant worldwide and easy to manufacture using existing processes. The team has identified five dyes that will be good absorbers to match this transfer material, although these still need more testing and research to take full advantage of their properties and ensure they make the best possible solar cells.

The models use theory combined with experimental data to understand how the dye molecules are moving and arranging themselves on the surface of the titanium dioxide in the cell. These configurations affect how efficiently the energy is transferred and how all components interact with each other. Computer modelling helps to show what is happening and, more importantly, why it is happening, on a level that cannot be reached with experiments alone. The findings explain or predict unforeseen behaviour that provides an insight into these dyes and how to take advantage of them.

Dr Dawn Geatches said, "The project is a combination of modelling and experimental methods, with our models enhancing the experimental data and the experimental data enhancing our models. This constant feedback loop is a really satisfying and efficient way to carry out research because it leads to faster and better results than doing experiments or theory alone."

Using dyes as opposed to the other alternatives can greatly improve the manufacture of solar cells, making them cheaper to produce and, subsequently, more affordable to consumers. The dyes can be printed onto a large flexible roll instead of the rigid structures you see with solar panels today. This makes the printed sheets much more compatible with existing infrastructures, such as roof tiles, and will make them a cheaper, more favourable way to generate electricity without relying on fossil fuels.

"Taking advantage of these abundant organic dyes, which are more sustainable and less devastating to the environment than their counterparts, will push us closer and closer to realising the full potential of renewable energy, " said Dr Ya-Wen Hsiao.

The collaborative work with Glasgow and Swansea universities has extended beyond the computational models. The project has been used in outreach programs, such as workshops for A-Level students, and integrated into STFC's summer schools. A related project has been developed to extend the life expectancy of solar cell components by introducing recyclability in solar cells.

Authors: Marion O'Sullivan and Orlagh Simpson, SCD Impact Team, in collaboration with Dr Dawn Geatches and colleagues, Computational Science and Engineering Division.

Research partners are:

STFC Scientific Computing: Dr Dawn Geatches, Dr Kakali Sen and Dr Ya-Wen Hsiao; and previously Dr Sebastian Metz (now at Fraunhofer ISE, Freiburg)

Swansea University (College of Engineering): Prof. Peter J. Holliman, Dr Christopher P. Kershaw, Dr Diana Meza-Rojas, and extended team

Glasgow University (School of Chemistry): Prof Graeme Cooke and team

Fast-Tracking Drug Discovery using the STFC Cloud

Scientists at the Diamond Light Source's XChem Facility are using the STFC Cloud to enable more efficient collaboration between researchers to develop drugs to combat diseases like COVID-19.

On average, it takes at least ten years for a new medicine to complete the journey from initial discovery to the clinic. The average cost to research and develop each successful drug is estimated to be $2.6 billion (USD). The likelihood that a drug entering clinical testing will eventually be approved is estimated to be less than 12%[1]. Drugs are generally developed iteratively, starting from initial compounds which might be identified by high-throughput screening. While this can identify starting points for the generation of drug leads, they are frequently difficult to optimise, or to rationalise the drivers for their potency.

Fragment-Based Drug Discovery

Fragment-Based Drug Discovery (FBDD) uses molecule fragments, usually around three times smaller than the potential drug candidate, to determine which of its structural features are useful when it interacts with a target protein involved in causing disease. When they interact with the protein they form what is called a protein-ligand complex. We can modify ligands to improve how strongly they bind to the target protein, taking us closer and closer to a specific, drug-like molecule.

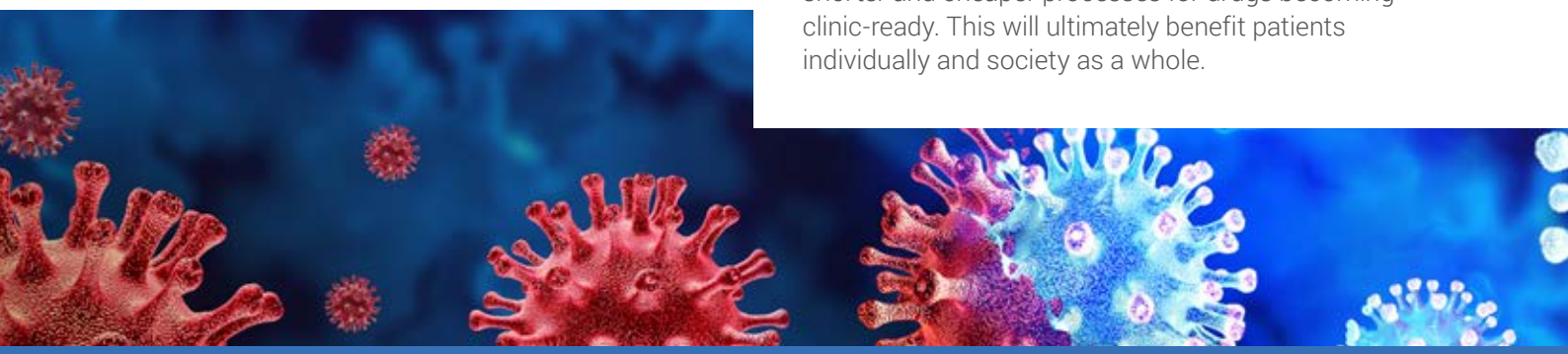
Scientists at XChem use X-ray crystallography to visualise the 3D protein structures. This technique involves introducing hundreds of fragments into crystals that contain millions of copies of the target protein, and then firing X-rays at the crystal. They can then analyse the angle and intensity of the diffracted X-ray beams to create a 3D model of the protein-ligand complex. XChem runs hundreds of these screens per day, thereby significantly accelerating the process of getting drugs into clinical trials.

How the STFC Cloud facilitates drug discovery

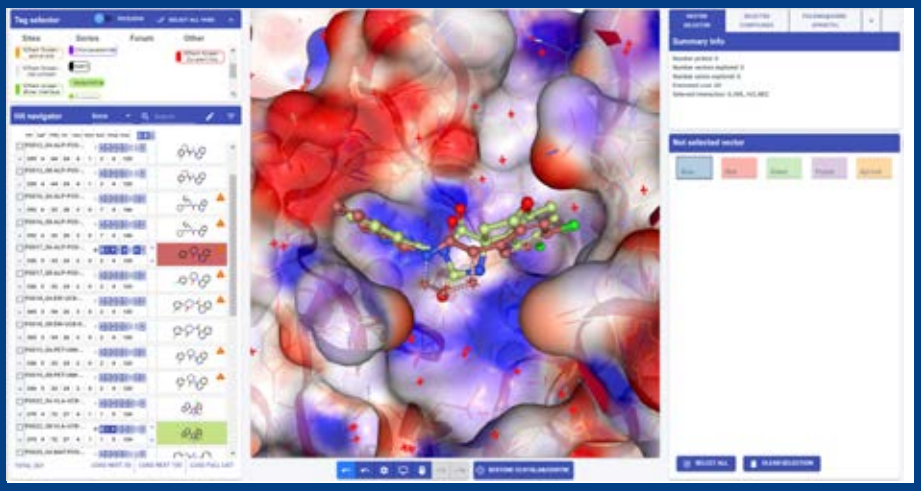
The STFC Cloud is a dedicated cloud infrastructure that provides access to compute resources for users across STFC and partner organisations. Developed and managed by the Scientific Computing Department, it's designed to be flexible, allowing many different uses. The STFC Cloud aims to enable users to perform complex data analysis as and when required, without the overheads of running their own computing infrastructure.

Facilities like XChem have given rise to an unprecedented increase in data describing potential drug molecule and protein combinations. To cope with the vast data output, and to explore it in the context of FBDD, the XChem team designed and implemented an open-source web-based application called Fragalysis[2]. Fragalysis is hosted in the STFC Cloud through the IRIS initiative, which has encouraged open science in the academic community. It has allowed multiple projects to become more efficient and collaborative. Hosting Fragalysis in the STFC Cloud vastly reduces costs and helps to ensure that it has a long-term home.

Fragalysis has become an essential tool for the team at XChem. Having the data for a given target on the platform means that anyone can access it, from wherever they are in the world. They can also share snapshots of the exact state of the page to enable better collaboration in the design of drug-like molecules. The platform has the potential to revolutionise how scientists think about FBDD, how they collaborate on projects and how they share data with the rest of the world. This could lead to much shorter and cheaper processes for drugs becoming clinic-ready. This will ultimately benefit patients individually and society as a whole.



Rational design of SARS-CoV-2 main protease inhibitor carried out using Fragalysis webtool.



"Having Fragalysis served from the STFC cloud has been nothing short of sensational in terms of what it has allowed us to do scientifically. We believe Fragalysis has the potential to provide a step-change in how scientists think about fragment-based drug discovery, and how they collaborate on projects and share data with the rest of the world. If we manage to achieve this, then this will hopefully lead to much shorter and cheaper timescales for drugs getting into the clinic. This will ultimately have individual benefits for patients and collective benefits for society."

Dr Rachael Skyner, senior software scientist for XChem

Fragalysis and COVID Moonshot

Fragalysis has already proved its value in the COVID Moonshot[3], a non-profit, open-science international consortium of scientists, biotech companies, contract research organisations and pharma companies dedicated to the discovery of affordable antiviral drugs against COVID-19 and future viral pandemics. The project was kick-started from a massive fragment screen of the SARS-CoV-2 main protease, one of the enzymes involved in the viral life cycle. It was carried out at XChem by Diamond researchers in early 2020 as the pandemic was evolving[4].

Fragalysis has been used as the primary platform for the dissemination of structural data during COVID Moonshot. The STFC Cloud has enabled the sharing of data in almost real-time, allowing the structures to be released as they are solved. This has allowed hundreds of contributors from across the globe to rapidly consider new designs for drug-like molecules and has contributed to an extremely swift turnaround for optimization cycles. They expect to progress some of the molecules designed using Fragalysis to become preclinical candidates[5]. Reaching this stage usually takes many years, whilst the candidates from Moonshot have been developed in less than 12 months[6].

The Fragalysis project benefitted from an STFC Cloud allocation through IRIS (www.iris.ac.uk). IRIS helps develop and grow the digital research infrastructure that allows STFC to continue to play a leading role in world class science. Additional funding was provided by the Ada Lovelace Centre.

Authors: Marion O'Sullivan and Pollyanna Burnet-Harris (Scientific Computing), in collaboration with Dr Rachael Skyner, Dr Frank von Delft and Daren Fearon (Diamond Light Source).

1 www.phrma.org
2 https://fragalysis.diamond.ac.uk
3 https://www.nature.com/articles/d41586-021-01571-1
4 https://pubmed.ncbi.nlm.nih.gov/33028810/
5 https://www.diamond.ac.uk/Home/News/LatestNews/2021/27-09-21.html
6 https://www.biorxiv.org/content/10.1101/2020.10.29.339317v3

A Sustainable Spin to Energy-Efficient Computers and Smart Phones

The journey to net zero carbon is complex but essential, and as such is encouraging scientists to approach research from different angles, combining techniques and tools in original ways to explore new materials for improved greener solutions.

Scientific Computing's Dr Ivan Scivetti and Dr Gilberto Teobaldi have done exactly that. Their work is exploring a new sustainable option in the science of spintronics, a commercial technology for magnetic information storage and processing, and used in electronic products such as computers, smart TVs and mobile phones.

The field of spintronics is relatively new and, so far, based on compounds comprising metals that are rare and expensive to extract from the earth, and to purify. This involves harsh and hazardous chemical processes that add additional manufacturing challenges and costs to spintronic technologies. In addition, the geopolitical issues in terms of their mining, resourcing and distribution all combine to form a strong motivation to find alternatives that are based on cheaper and more environmentally benign elements.

In a simplified picture, electrons exist in one of two spin states: 'up' or 'down'. Electrons of opposite spin states can have distinct electron transport properties – i.e. electrical conductivity. If the electrons in one spin state conduct electricity whereas the electrons in the other spin state do not, the material is half-metallic. Half-metals are capable of selectively transporting electrons of only one spin-state. This is a highly sought after property for spintronics applications that require the flowing of electron-spins – or spin-currents.

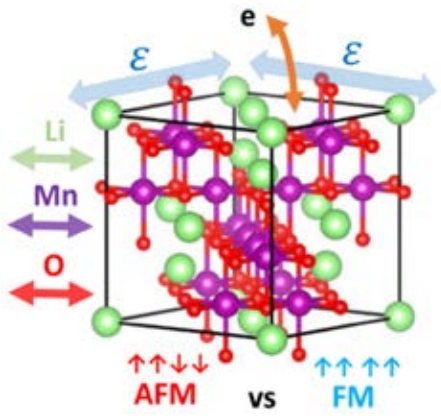
"Without these rare-earth metals, spintronic devices can't function properly. Our research, which uses advanced computer simulations to test research data and ideas, has revealed a promising alternative that is both cheaper and more abundant than the rare-earth metals currently used in computers, smart TVs and mobile phones."

Dr Ivan Scivetti, Senior Computational Scientist

Credit: Creativity103

Spin-currents do not dissipate energy as heat, resulting in devices of substantially increased energy efficiency in comparison to standard commercial solutions based on electron-currents – i.e. the flowing of electricity. Unfortunately, while relatively common in materials containing rare-earth metals, such properties are not typically present in compounds of cheap and benign transition-metals such as manganese (Mn).

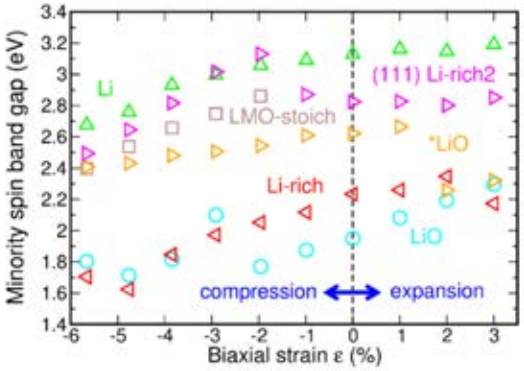
Drs Scivetti and Teobaldi approached this challenge from a different perspective, applying research strategies from surface science and computational physics to a metal oxide material (lithium-manganese-oxide, $LiMn_2O_4$) typically used for battery applications. They used quantum mechanical simulations to model thin films of $LiMn_2O_4$ under geometric strain – stretch and compression – and investigate how these factors influence the film's electronic and magnetic properties.



Schematic view of the crystal structure of $LiMn_2O_4$. Mn: violet, O: red, Li: green. The interplay between lattice strain ε, magnetic ordering and compositional changes for surface models results in the emergence of ferromagnetic half-metallic states, as required for spin-selective filtering and storage applications.

The two researchers modelled different crystallographic planes of $LiMn_2O_4$ and simulated them under varying amounts of strain. They found that, when placed under strain, the interplay between strain and different chemical compositions (the specific amount of lithium, manganese and oxygen in each plane of the layered material) determine the system's metallic and magnetic properties, and consequently their spin – and electron – conductivity.

This computational screening pointed out several unexpected half metallic solutions for different compositions of strained $LiMn_2O_4$ films. The observed half-metallic properties and predicted response to strain endows $LiMn_2O_4$ films with the capability of spin-selective electron transport in the absence of any rare-earth metal in the system, indicating more sustainable avenues to sourcing materials for spintronics applications.



Calculated minority-spin band-gap for the selected half-metallic terminations of $LiMn_2O_4$ as a function of lattice strain (ε).

Dr Teobaldi said, "This work provides unexpected insights that will inform new research avenues for the development of more sustainable, energy-efficient electronic devices than commercially used, or being considered for use, throughout the world today."

The quantum mechanical calculations require considerable computational power and were run on both national supercomputing facilities (ARCHER2) and STFC Hartree Centre's supercomputers. Approximately 1200 models (one per simulation) were built, each comprising about 200+ atoms; each simulation used on average between 96 - 512 cores* for approximately 24 - 36 hours per simulation. The electrical energy consumed to carry out this research is a small fraction of the overall potential gain in energy efficiency offered by these spin-tuned spintronic devices should they eventually be manufactured.

***Equivalent to 12 to 64 powerful laptops**

Authors: Dawn Geatches and Marion O'Sullivan in collaboration with Drs Ivan Scivetti and Gilberto Teobaldi, Theoretical and Computational Physics Group.

'Combined Role of Biaxial Strain and Nonstoichiometry for the Electronic, Magnetic, and Redox Properties of Lithiated Metal-Oxide Films: The $LiMn_2O_4$ Case', is published in ACS Applied Materials and Interfaces **https://pubs.acs.org/doi/abs/10.1021/acsami.1c18326**

Nature's Aerodynamic Blueprints Inspire New Engineering Designs

Researchers from STFC's Scientific Computing Department (SCD) and the University of Manchester are using the surface patterns found in nature to find ways to reduce fuel consumption and increase aerodynamic efficiency for aircraft, ships and cars.

When air passes over an aircraft wing the flow is disturbed by surface friction, separating the flow and leaving a gap above the surface, which creates a drag on the wing. This surface friction drag and flow separation, which are common phenomena in air, road and water vehicles, have a large impact on fuel consumption, cruising range, endurance and aerodynamic performance.

To reduce the impact, fin-like structures called vortex generators are attached to the leading edge of the wing to create a swirling mass of air that can reduce the separation gap. These are widely used for aerodynamic applications but are relatively large and can disrupt the entire field of flow if not used properly.

The research team, led by SCD's Dr Jian Fang, used high-fidelity computational simulations to study the airflow process and explore innovative ways to improve aerodynamic efficiency and reduce fuel consumption.

Inspired by the natural streamlined efficiency of birds and fish as they move through air and water, the team investigated the micro-scale pattern of ridges and grooves on bird feathers and shark skin. These tiny directional grooves, invisible to the naked eye, control the flow of fluids over their surface.

"Sharks and birds are both well adapted to moving efficiently through fluid, be it water or air. If you examine shark skin or bird feathers under a microscope, they both share a common feature: small directional grooves invisible to the naked eye that are used to control the flow of fluid over the surface."

Dr Jian Fang, Senior Research Scientist

Dr Fang and his colleagues set about mimicking these groove patterns by implementing a technique which they called 'convergent divergent riblets' (or CDRs) as an alternative to the usual vortex generators.

Using the Hawk supercomputer at the High-Performance Computing Centre in Stuttgart, Germany, the team tested how the riblets would perform when air flows over them. To do this, they adopted the direct numerical simulation approach to explore the details of the vortices induced by CDRs, using the ASTR code developed at SCD.

The researchers observed how the CDRs direct the airflow into the grooves and how the turbulent air moves along them. They looked at the influence of spacing between riblets, demonstrating that wider spacing encouraged resistance to flow separation by generating large-scale 'streamwise' vortices.

By adjusting the parameters of the CDRs they can control the scale and strength of the vortices – narrower converging and diverging strips produce small vortices, while wider strips produce larger, stronger vortices.

Dr Fang said, "As we increased the spaces between the riblets, the vortices became stronger, and through our simulations we were able to find the point at which the spacing produced the largest effect."

There is currently a trade-off, however, as reducing the flow separation can increase the overall drag on the vehicle surface. Dr Fang is confident that, with further studies, it will be possible to find ways to address that issue and achieve reductions in both.

Bird feather

Benefits:

- This work lays the foundations for optimising the aerodynamics of vehicles that have the potential for more efficient performance and lower fuel consumption.
- The ability to 'fine-tune' the size of the vortices by changing the riblet parameters provides the potential for engineers to refine the CDRs for specific applications – something that is not possible with standard vortex generators.
- A further study is now underway to explore the effects of CDRs in supersonic flows, in which high-pressure shockwaves can induce flow separation. Preliminary results are very promising.

Authors: Marion O'Sullivan and Orlagh Simpson, SCD Impact Team, in collaboration with Dr Jian Fang, Engineering and Environment Group

This project received support and resources from PRACE (Partnership for Advanced Computing in Europe) which enabled computational performance testing of the ASTR code on a large scale, using over 100,000 cores. This furnished the researchers with confidence that the code would run at scale in an optimal way. It has since enabled Dr Fang to improve the Input/Output performance and a new version of the ASTR code is soon to be released.

Using Computers to Design Safe and Sustainable Nuclear Reactors

As the UK's net-zero goals loom, Scientific Computing's Drs Stefano Rolfo, Gregory Cartland-Glover and Leon Petit are carrying out research to model and simulate designs for the next generation of safe and sustainable nuclear reactors.

With minimal greenhouse emissions, fission reactors should be a forerunner for cleaner energy production. However, the nuclear power cycle still requires the mining and transport of uranium. Only 1% of mined material is actually used in energy production while the rest is discarded as nuclear waste – and discarding it safely is far easier said than done.

Molten Salt Fast Reactors, which use a molten salt mixture to carry the radioactive material, have the potential to improve the sustainability of the nuclear cycle since they can use current nuclear waste as fuel and at the same time generate more, emission free, energy.

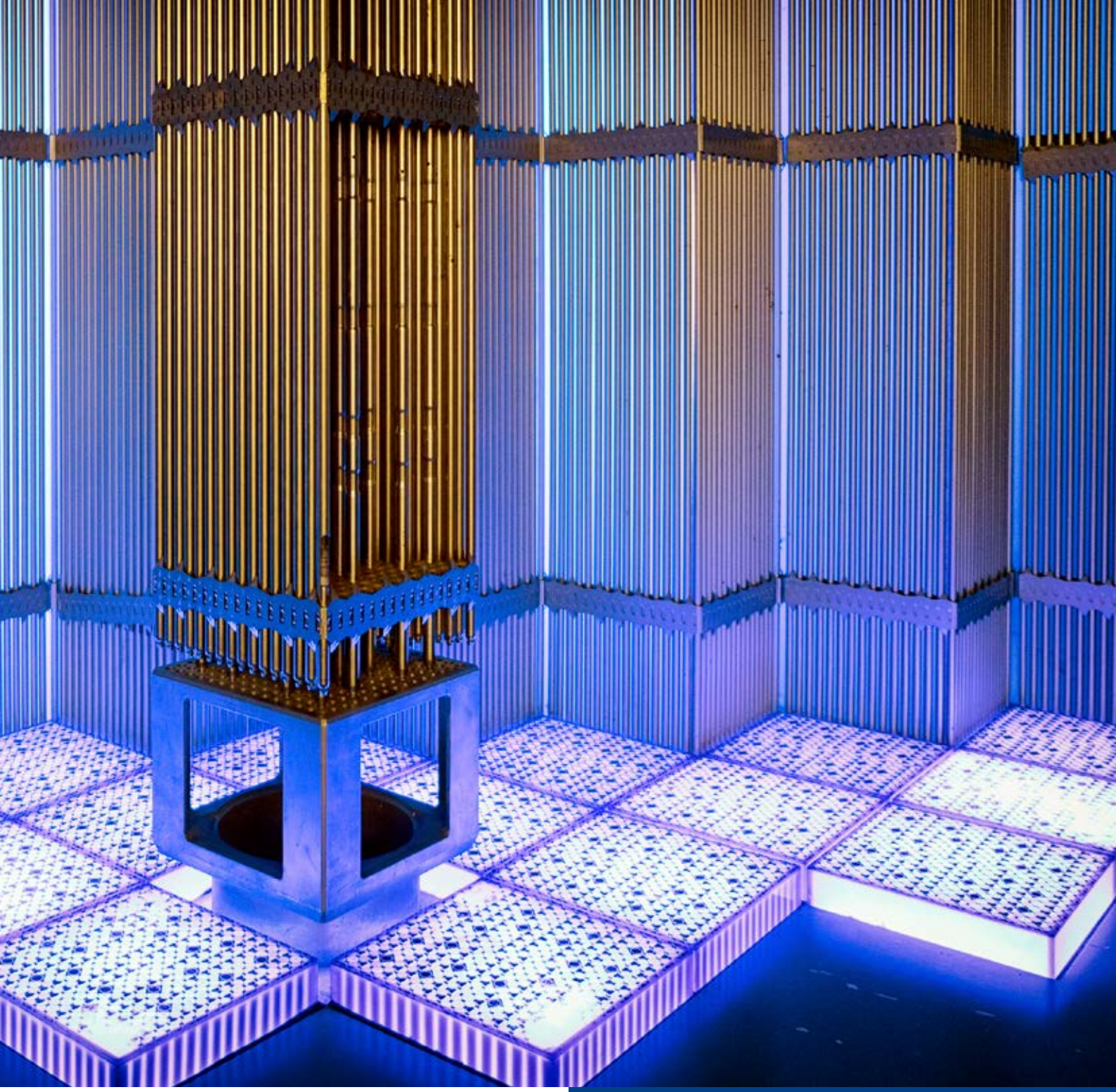
Dr Rolfo and colleagues are working alongside Professor Bruno Merk at the University of Liverpool to find ways to understand the complex chemistry of these reactors, improve their sustainability and reduce corrosion damage.

"One way to do this is by plating certain parts with less reactive metals, but doing so can affect how well the part functions and even how it reacts to heat. We must understand these effects to create the safest, most efficient reactor possible," said Dr Rolfo.

Plating of noble metals, such as platinum, rhodium or palladium, on the vessel surface will have a large impact on the reactor's life and can have a positive impact in reducing corrosion damage.

Dr Rolfo added, "We can lower our global impact even more by using the fertile and radioactive material that already exists in the above-ground storage in the UK, whilst getting rid of leftover depleted uranium, plutonium and other by-products. If the technology is successful, it could offer a secure supply of low carbon energy for several hundred years."

Making these ideas a reality still requires a lot of research. Numerical models, new relevant data



Model of a nuclear power station

stacks and datasets are still needed for us to fully understand the reactions and mechanics that will be going on in these machines.

Moving forward with this takes us a step closer to reaching Britain's net-zero goals, whilst getting rid of the nuclear waste that we expect to be an issue for generations to come.

Authors: Marion O'Sullivan and Orlagh Simpson, SCD Impact Team, in collaboration with Dr Stefano Rolfo, Engineering and Environment Group.

Delivering Solutions

IRIS: Addressing Security Challenges in a Digital Environment

IRIS is the digital research infrastructure that enables STFC-supported science and includes resources providing high-throughput, high-performance and cloud computing. The ability for this infrastructure to operate a secure and assured computing environment is essential for its continued success and development.

The STFC Distributed Research Trust and Security team coordinates all of the security development for IRIS, including Identity Management, Trust and Policy, Operational Security and Information Security Management. In this article we will discuss the different aspects of this work, how they rely on each other, and how they help us support the users of IRIS and work alongside our national and international partners.

Identity Management

A central challenge in modern computing infrastructures is Identity Management. Ensuring that an account corresponds to the correct person – authentication - and has the correct permissions – authorisation - is vital to avoid identity misuse.

This is a core challenge to address for a nationally federated digital research infrastructure such as IRIS to effectively enable the requirements of STFC's science activities. These require a well-implemented and well-supported identity management solution to act as a coherent user-facing infrastructure across the UK.

The IRIS IAM (Identity and Access Management) service, built following an architecture (Figure 1) developed by many infrastructures in Europe with leadership from STFC, allows users to access IRIS resources with their home institutional identities. This is critical to the security of IRIS, providing robust authorization mechanisms for IRIS services and resources. The experience of developing the IAM service will be crucial in the development of UKRI's national, federated digital research infrastructure.

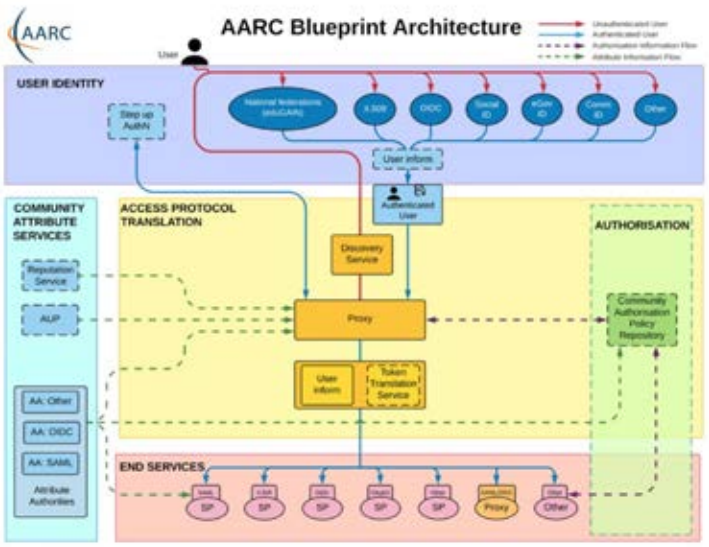


Fig1: AARC Blueprint Architecture (https://aarc-project.eu/architecture/)

Operational Security

Operational security for infrastructures includes risk management, incident prevention, security monitoring, incident response and a process of continuous improvement.

We have built a distributed security team to support the diverse high throughput, cloud and high performance computing resource providers in IRIS. This work has built on an existing team that supports GridPP, the UK contribution to the computing activities supporting the CERN Large Hadron Collider.

This team plays an essential role in sharing essential vulnerability information throughout IRIS and maintaining the capability to coordinate within IRIS in the case of a distributed information security incident. The team regularly holds Security Forums for IRIS which allow the development of an effective security culture for IRIS through regular meetings and training workshops, and is contributing to the first thematic CERN School of Computing on security held this summer.

Trust and Policy

Operational security must be supported through policy to provide assurance and allow good governance. We use a policy framework kit created alongside the architecture used by the IRIS IAM, again developed by an international body of experts with significant input from STFC. This development kit has allowed us to build a set of policies which enable a secure and assured computing environment for IRIS researchers as well as working closely with partner infrastructures.

The development of a security policy framework for IRIS has also had significant impact for our international infrastructure partners such as European Open Science Cloud, by contributing back to broader policy development activities.

Information Security Management

As the IRIS security capability continues to mature, it is important that the different aspects of this work are managed as part of a single process to make the work as effective as possible. Starting this year, all aspects of IRIS Security are now managed as part of a single activity, through which we are now planning developments for the coming months and years.

The Future

IRIS is an excellent example of a digital research infrastructure which has matured over the last four years with a security capability developing in step. We now have effective processes and services in place across all areas of distributed security and identity management, laying the groundwork for the future.

Author: Dr David Crooks, Distributed Computing Infrastructure Group

The Trail from Data to Policy in the COVID-19 Pandemic and Beyond

COVID-19 was an event that marked a 'before' and 'after' in our lives. We will all remember the days when daily statistics on positive cases, hospitalisations and fatalities were reported in each region of the world, providing metrics on the looming situation. And some of those statistics will continue to be collected for the foreseeable future.

All in this together

As the World Health Organisation was officially declaring a pandemic at the start of 2020, truly interdisciplinary teams were coming together to understand and combat the spread of the disease, determining and improving processes in healthcare and public health policy as new information was compiled and analysed.

The Scottish COVID-19 Response Consortium (SCRC)[1] was one such team, formed by a group of epidemiologists, mathematical modellers, data scientists, data policy experts, software engineers and other scientists to develop epidemiological models of COVID-19 spread. The initiative started in response to the Rapid Assistance in Modelling the Pandemic (RAMP)[2] coordinated by the Royal Society. During 2020, the consortium involved over 150 volunteers from multiple universities, research centres and industrial partners across the United Kingdom. The first activity by the consortium involved a requirements gathering exercise for a data pipeline, which could support easy annotation of any data consumed by analysis, while also tracing the provenance of the scientific outputs back to the analytical or modelling source code and to the primary data. Such a data pipeline would provide the mechanisms for scientists, as well as the public, to assess scientific evidence by inspecting its provenance, tracking every step from data source to a policy decision. Such a tool can be applied to any policy-facing research to improve public trust in data-driven science.

The requirements gathering exercise and subsequent analysis concluded that there were no existing tools satisfying all the needs of such a data pipeline. Thus, the group embarked on the design and development phases of the **FAIR data pipeline**, which later continued with a reduced group funded by STFC.

The data pipeline is **FAIR**, as per the acronym for Findable, Accessible, Interoperable and Reusable data, which refers to a set of principles to ensure that the data is machine processable and there is enough information for others to re-use.

Software and data – the basis for policy evidence

Epidemiological modellers use a variety of programming languages to describe and analyse their models. Thus, the FAIR data pipeline provides **software components** in multiple languages (R, Julia, Java, Python, C++) that support reading and writing the data that will be used by the models. The data is stored in one or more instances of a **data registry** component. Each registry includes provenance information about the data, i.e. how the data has come into existence. This reflects the entities, activities, and people involved in producing the piece of data, and can be used to assess its quality, reliability and trustworthiness. Finally, the pipeline also relies on a command line component that enables users to synchronise the information kept at multiple registries.

The main consideration in creating the FAIR data registry has been the use of data standards, which allow for the representation of information in well-established formats and common vocabularies, enabling better integration with other tools and wider re-use. The Data and Software Engineering Group in the Scientific Computing Department has been contributing specifically in this area, by making sure that the data and its provenance is represented in machine-actionable way, i.e. it can be read and interpreted by machines as well as users.

The outputs from the project are all open source and can be accessed through the dedicated website: **https://www.fairdatapipeline.org/**. More details about the pipeline can be found in the pre-print article "FAIR Data Pipeline: provenance-driven data management for traceable scientific workflows" **https://arxiv.org/abs/2110.07117**

A more trusted and verifiable future

The FAIR data pipeline is generic and can be used both for human and animal disease models, as well as across ecology and other areas of the life sciences. Indeed it has been designed and developed independently of the domain application, and we are already exploring other modelling communities that may benefit from its use. The target audience is not only modellers but also policy makers and members of the public.

By using the whole framework or some of its components, we hope to allow people to achieve our overall objective of improving trustworthiness in the science used for public policy, as scientists and members of the public alike would be able to examine and verify the trail from data to policy decisions.

Author: Dr Alejandra Gonzalez-Beltran, Data and Software Engineering Group

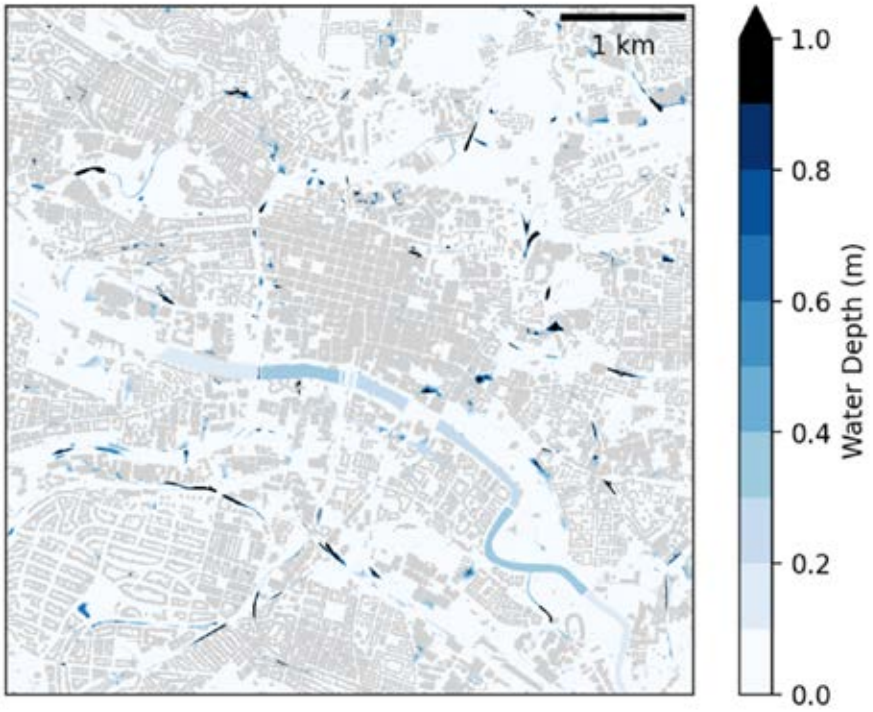
1 SCRC: **https://www.gla.ac.uk/research/az/scrc/**

2 RAMP: **https://royalsociety.org/topics-policy/Health and wellbeing/ramp/**

OpenCLIM: Assessing the Impact of Climate Change

Climate change is now top of the world's agenda. The Glasgow COP26 meeting in November 2021 brought into focus the urgency of understanding and mitigating the effects of climate change on the world around us. Climate scientists have developed detailed computational models to project the likely effects on our climate of continuing to emit carbon. In the UK, the Meteorological Office produced a detailed set of climate scenarios going out to 2100. These predict how weather patterns are likely to change if the global temperatures rise by an average of 2°C or 4°C, with potentially more extreme weather such droughts, storms and heatwaves.

The Open Climate Impacts Modelling Framework (OpenCLIM) project (2020-2023) takes this to the next step. If we are to understand the risks of climate change, we need to study how these new weather patterns will affect our lives. For example, changes in rainfall may affect what areas might flood after storms, or differences in temperatures might alter what crops we can grow. Further, this information needs to be combined with predictions of where we might build new houses and businesses to understand who might be affected.



Glasgow flood

OpenCLIM is a project funded by the Natural Environment Research Council within the UK Climate Resilience Programme. It is led by the University of East Anglia, with partners Newcastle University, Bristol University, and the UK Centre for Ecology and Hydrology, together with the Data and Analysis Facility for National Infrastructure (DAFNI) within SCD. It brings together research in different computational models to provide a common framework to explore the likely impacts of climate change over the rest of this century. It is concentrating on four risk areas: the threat to urban areas of flooding from extreme rainfall, or rivers bursting their banks; the effect on people and farm animals of extreme heatwaves, which cause health problems and even death; the effect of changing temperature and rainfall on the yields of common crops like wheat or potatoes; and the effect on the diversity of our wildlife, which may thrive or decline in different parts of the country, with the possibility of new invasive species taking hold. These scenarios aim to contribute to the next national climate change risk assessment, CCRA4, that the government will undertake beginning in 2023.

The OpenCLIM framework brings together models in these areas using the DAFNI computing platform. DAFNI offers a common environment where the models prepared by different groups using different approaches can be shared, and executed using the UK Climate Prediction (UKCP18) data. The DAFNI platform allows the models of different features to be brought together in new 'workflows', so for example, growth in urban areas can be combined with flood models.

Further, OpenCLIM is also exploring how we might adapt to help soften the impact of climate change. For example, we might grow more trees to absorb flood water, add green spaces to cities to soak up water and cool the city down, or grow different crops adapted to a hotter, drier environment. In OpenCLIM, the model scenarios can be changed to see how effective these strategies might be; DAFNI offers the flexibility to easily change the workflow to explore these options.

Authors: Dr Brian Matthews & Marion Samler, DAFNI Group

Global Optimization of Parameters for ISIS Neutron and Muon Source

Neutron spectroscopy is used to study the magnetic structure and behaviour of materials, but the resulting data often require fitting to a model to extract physically relevant parameters and information. The SCD Computational Mathematics Group has been investigating novel numerical algorithms that will better address this challenge. They provide the expertise and numerical tools that enable users of ISIS instruments to analyse their data and publish their findings much faster. Importantly, they also enable the scientists to focus their efforts on the experiments rather than having to deal with the complexity of the models.

The ISIS Neutron and Muon source, as the name implies, produces beams of neutrons and muons that allow scientists to study the composition, structure and dynamics of materials, including magnetic rare-earth metals used in many electronic devices. The magnetic behaviour of these materials is determined by their crystalline environments (i.e. the arrangement of neighbouring atoms to the rare earth atoms, which vary depending on the compound). Researchers use ISIS instruments to investigate these and other materials in a range of experiments where magnetic behaviour needs to be understood – such as for novel quantum materials or to make stronger permanent magnets for electric motors or hard disks.

In particular, neutron spectroscopy can be used to study transitions between electronic states in the rare earth atom, to understand the crystalline electric field that surrounds it. This field is the origin of the *magneto-crystalline anisotropy*, a physical property of magnets related to how much energy it takes to magnetise (or de-magnetise) the material. Larger anisotropy energies are desirable in permanent magnets as it allows them to be used at higher temperatures or in smaller quantities. However, obtaining these physically important anisotropy parameters from the measured neutron spectra requires fitting to a complex mathematical model.

Such theoretical models are applicable to a large class of magnetic compounds and depend on a number of material-specific parameters that are generally unknown. Thus, the task of the experimentalist is to determine the model parameters that best match the observed data arising from the experiment, data that is often noisy and incomplete.

This is where the technique of numerical optimization comes in: here optimization concerns itself with finding the model parameters that best match the experimental data. For many such problems, it is sufficient to use local optimization; that is, mathematical techniques that search for parameters that are locally optimal. Here locally optimal means that there are no nearby parameters that achieve a better fit to the data. This is a very well understood problem, and numerous excellent local optimization algorithms have been developed and implemented over the years, several of which were developed by our Computational Mathematics Group in years past. However, there are situations where being locally optimal is not enough, situations where we desire parameters that are not merely better than nearby ones, but achieve the best possible fit across all parameter values, that is parameters that are globally optimal. This is precisely the situation that arises at ISIS when modelling the neutron spectra of magnetic compounds. Let us consider an illustrative example of this using real experimental data from ISIS.

Figure 1 shows the spectrum (in green) arising from an attempt to *locally optimize* the parameters in a model of the magnetic structure of $NdOs_2Al_{10}$ as compared to the noisy experimental data (denoted as black stars in the figure).

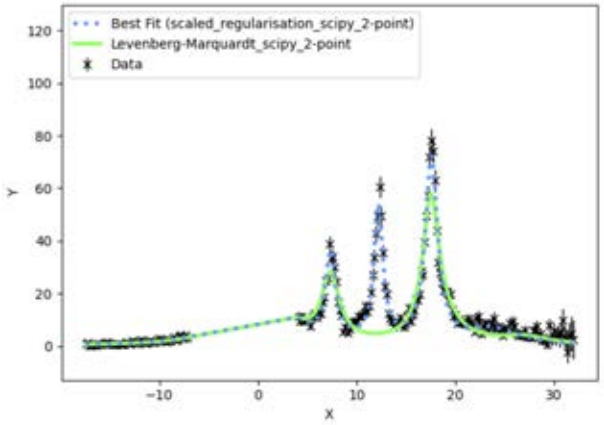


Figure 1: Spectrum of the magnetic structure of $NdOs_2Al_{10}$ under local optimization

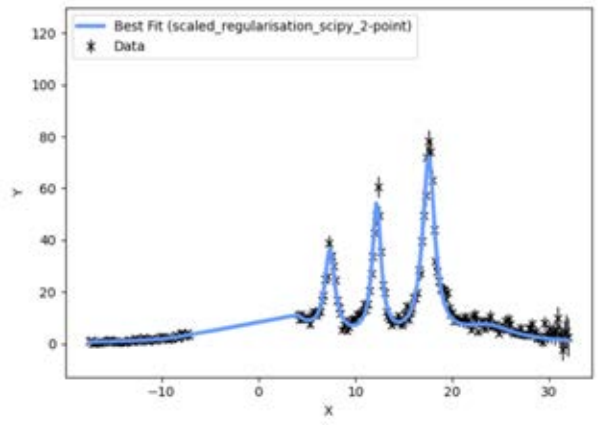


Figure 2: Spectrum of the magnetic structure of $NdOs_2Al_{10}$ under global optimization

Clearly there is a problem here, as the model parameters returned by the local optimization algorithm lead to a proposed spectrum (in green) that entirely misses the middle of the three obvious peaks in the data (denoted as black stars). Here the starting parameters for the local optimization, that is the initial guess for the best parameters, were only slightly different from the best possible initial guess. In fact, they were simply rounded up from the best initial guess parameters. This example clearly demonstrates the fragility of the current model fitting process for such compounds at ISIS. Not only this, but at present expert knowledge is evidently required to determine sufficiently accurate estimates of initial parameters for the optimization, demanding a large amount of ISIS expert time be devoted to supporting facility users in this task, time that could be better spent elsewhere.

Contrast this with Figure 2 showing the spectrum (in blue) arising from an attempt to *globally optimize* the parameters in the same model of the magnetic structure of $NdOs_2Al_{10}$ as in Figure 1.

Here we see that the global optimization algorithm that we have developed in the Computational Mathematics Group at SCD[1] is able to correctly determine the best model parameters no matter how well the initial parameter guess is specified. The benefits of this approach are self-evident: any facility user at ISIS wishing to study the magnetic structure of metal compounds will be able to quickly and easily find the model parameters that best match their experimental data without requiring expert insight into the modelling process and the complex mathematical theory that underpins it. Indeed, the success of this Ada Lovelace Centre project has led to it being extended to another year and we are currently in the process of directly integrating our newly developed global optimization algorithm into the fitting software used at ISIS, where it will become the default option for such problems.

Author: Jaroslav Fowkes, Computational Mathematics Group

References

1 O'Flynn, M., Fowkes, J., & Gould, N. (2022). Global optimization of crystal field parameter fitting in Mantid. RAL Technical Reports, RAL-TR-2022-002.

Intelligently Deciphering Small Angle Scattering Data

Machine Learning techniques and frameworks are being used to help with experiments carried out at some of the UK's large-scale facilities. Scientists from SCD's Scientific Machine Learning Group (SciML) are collaborating with researchers at the Diamond Light Source and the ISIS Neutron and Muon Source to help automate processes and speed up analysis of data coming from the sample materials being investigated.

One such project focuses on identifying the size of particles in solution or the size of holes in a medium – all on the nanoscale. The particles in solution could be from titanium dioxide in a water solution that forms the paint you put on your walls, micro-plastic contaminants in water, or even lipid vesicles (tiny sacs filled with liquid) used to deliver mRNA vaccines. The holes in a medium could be from catalytic materials (such as those in a car), water filtration membranes or dye-sensitized solar cells.

Small-angle scattering (SAS) is an experimental technique that allows scientists to observe features by analysing the scattering pattern from an incident X-ray (SAXS), neutron (SANS) or light beam (SALS) source. The key outcome sought through SAS analysis is obtaining statistically significant information about the shape, size, orientation, and contrast of objects or voids in the sample, usually on the nanometer to micrometer scale. Typically, X-ray and neutron scattering (SAXS and SANS) can resolve objects on a length scale of 1 nm to 100s of nm.

One central task in SAS data analysis is model inversion. That is, to determine the parameters of a selected theoretical model that can best explain the experimental data. These model parameters are "polydisperse" in general, meaning that each parameter is a density distribution across a permissible range instead of a fixed number. Such polydispersity introduces a huge search space to this inverse problem. This model search is particularly acute for models with multiple parameters, where model inversion can become the most challenging part of SAS data analysis. Existing methods are limited in their capability, such as the lack of generality (i.e., they can only address the inversion problem for a fixed model), prescribed functional form of the parameter distributions or compromised accuracy.

The SciML Group collaboratively worked with researchers at ISIS and Diamond to develop an intelligent solution to this inversion problem. The overarching idea was to develop a novel method for polydisperse SAS inversion with high accuracy and efficiency.

Among various approaches developed and evaluated, the route to solution was the careful manipulation of mathematics. Regardless of the complexity of the underlying model or scattering physics, the forward modelling process in SAS can always be abstracted as something called a multilinear map. This multilinear map can be characterised by a unique high-dimensional tensor, which can be considered as higher dimensional matrices, and often known as Green's tensor. This theoretical generalisation was the critical point to the novel solution, leading to a unified formulation and solution of SAS inversion. With this generalisation, the overall inverse problem was formulated as a constrained nonlinear programming problem solved with high accuracy and efficiency with several theoretical and computational enhancements.

We illustrate the architecture of the novel approach in Figure 1. This is a perfect example where mathematics and machine learning (particularly deep learning) can offer a remarkable benefit when carefully combined. The final software is built on a deep-learning library to take advantage of modern accelerators, such as GPUs, purely for performance. This is also a good example of using cutting-edge deep-learning libraries for general-purpose scientific computing.

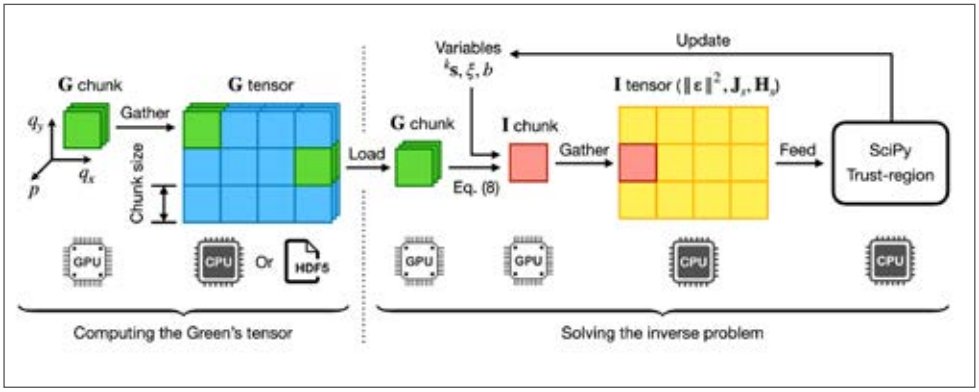


Figure 1. Architecture of GPU-accelerated chunk computation for large-scale multi-parameter problems. Left: given a SAS model and its parameter space, the Green's tensor G is computed in chunks on GPU. Right: the scattering intensity I is also computed in chunks on GPU and then assembled on CPU; this assembled I is fed to a trust-region method for nonlinear programming.

The approach has been simplified as a Python package, or library, and named FFSAS (Free Form Small Angle Scattering) so that scientists can use this numerical approach without knowing the intimate details or relevant mathematics or deep learning aspects. Here the free form signifies generalised functional form for polydisperse materials, unlike some existing approaches, where these forms are often fixed.

Compared to existing algorithms where numerous parameters are required to be calculated in advance and iterative fitting of the data is always required, FFSAS presents a fundamental change in approach to the analysis of small angle scattering data. It helps scientists to deduce if their experiment is working or failing, so makes the best use of their time at the facility. It also highlights relevant data within the vast sea of results, making it easier for scientists to identify where they should focus attention when publishing their findings.

The FFSAS library is freely available for public use (released under a BSD-3 license) at **https://github.com/stfc-sciml/ffsas/**. The library also comes with a detailed User Guide and many examples. We show an example output of this software in Figure 2 below.

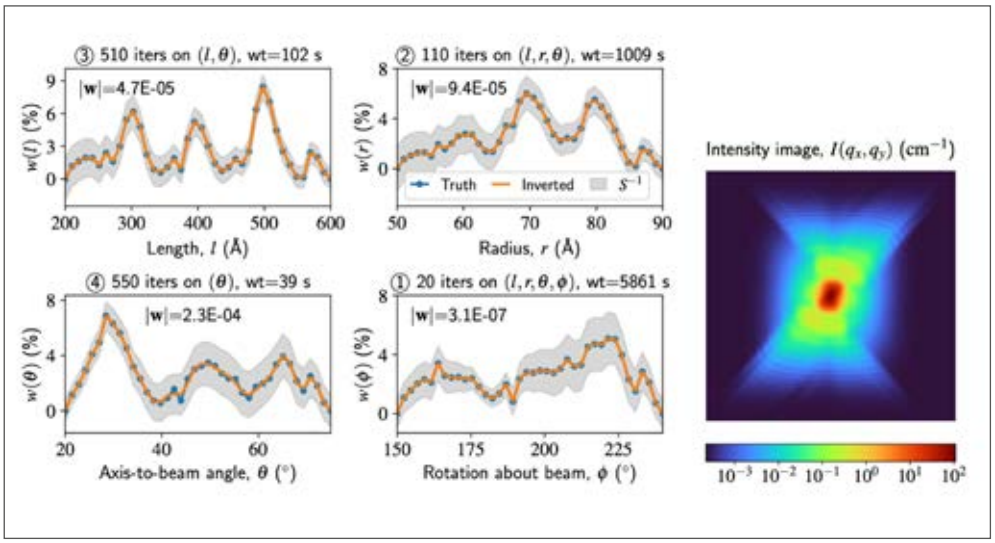


Figure 2. A synthetic test on size and orientation distribution inversion of polydisperse cylinders. The parameter distributions (truth and inverted) are shown on the left. The intensity image is shown on the right (truth and fitted look identical).

The example is a synthetic example, where the structural properties of a material are simulated to have polydisperse cylinders, and FFSAS is used to invert the space. This is a large-scale, multi-parameter problem that could not be solved by any previous approaches. The library is also under consideration for integration into the SasView project, which is a GUI based software package, for wider outreach as it is widely used throughout the SAS community today.

Authors: Kuangdai Leng and Jeyan Thiyagalingam, Scientific Machine Learning 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		
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/ Director 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

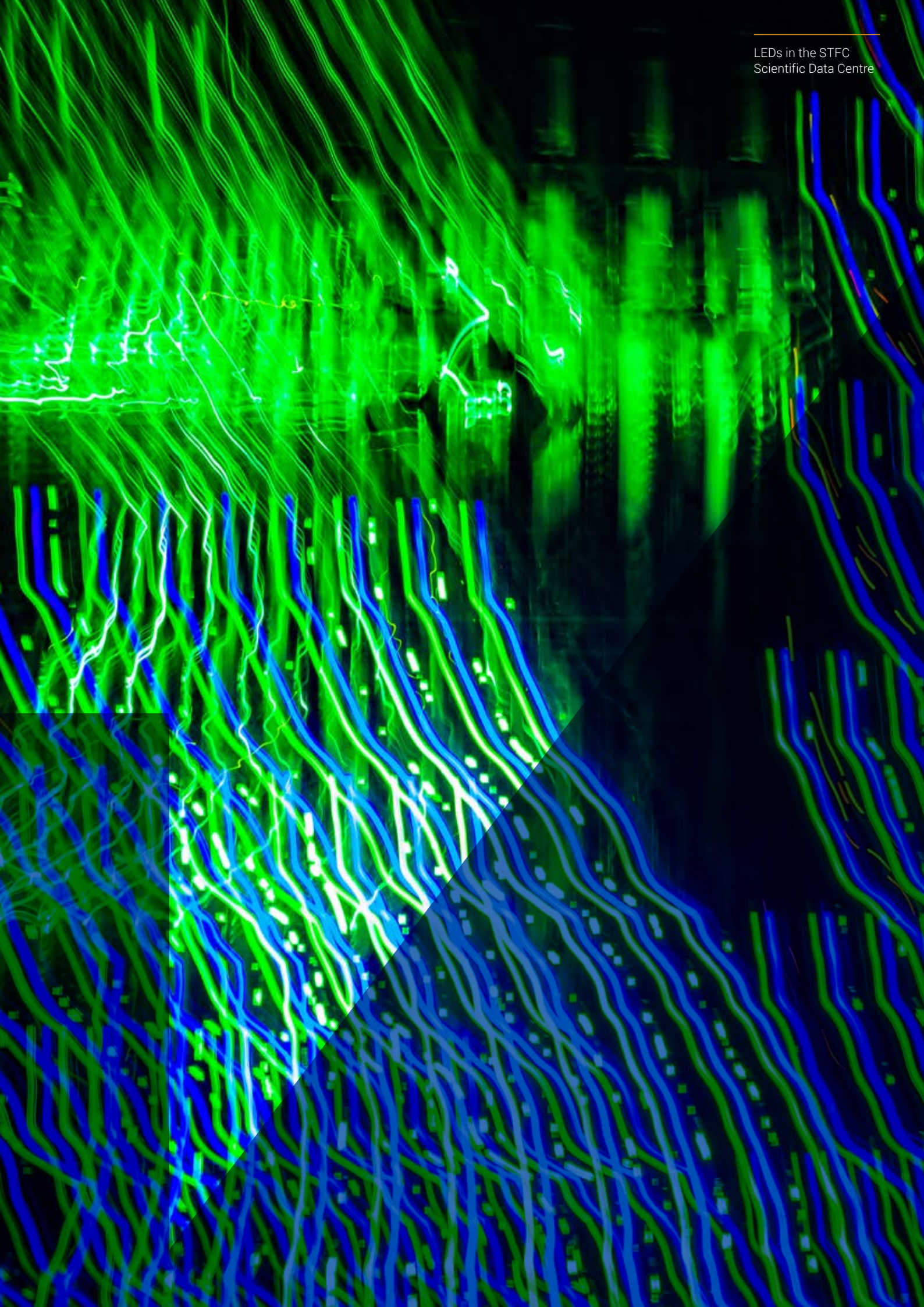
Administrative Support Team

Mrs Karen McIntyre	karen.mcintyre@stfc.ac.uk
Ms Georgia Lomas	georgia.lomas@stfc.ac.uk
Mrs Rocio Garavito Ramirez	rocio.garavito-ramirez@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



Science and Technology Facilities Council

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

7152.08.22