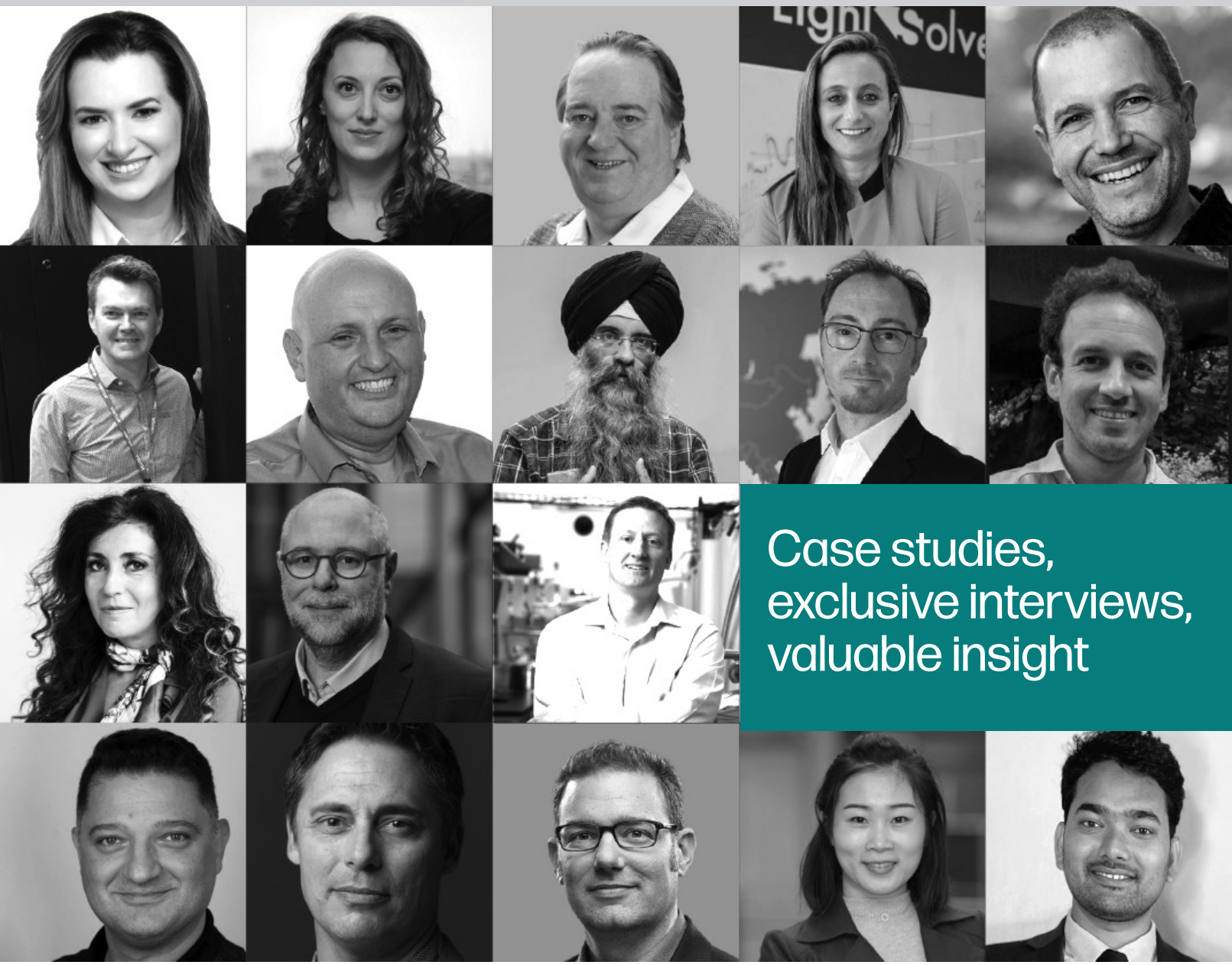
breakthroughs 2025

Unleashing scientific discovery with the power of computing

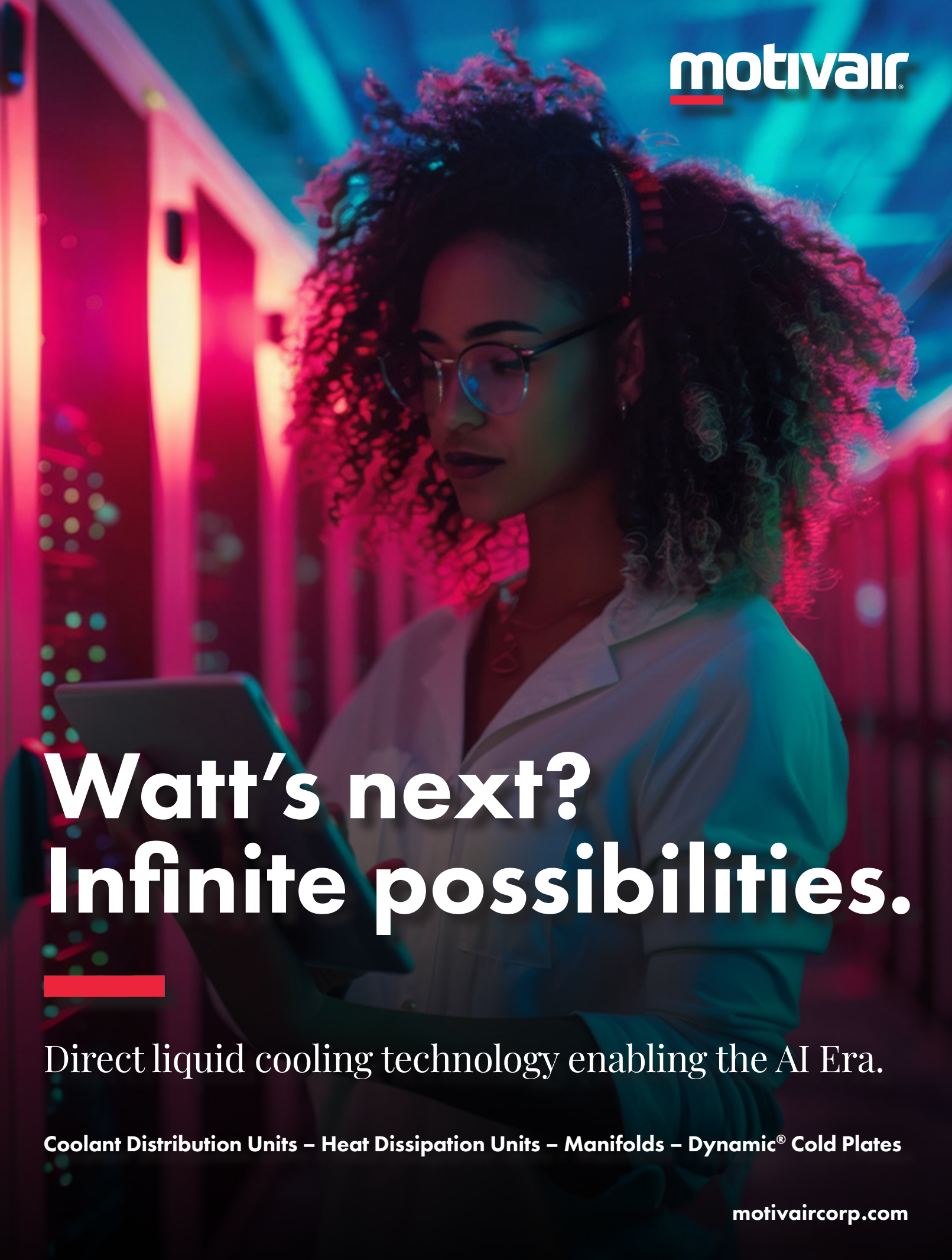


Case studies, exclusive interviews, valuable insight

Including

Digital laboratories: Accelerating ALS research with AI
Engineering: Pioneering automotive simulation for students
AI: Advancing material science using AI
Quantum: Solving real-world challenges
HPC: Powering the future with ARM

Brought to you by SCIENTIFIC COMPUTING WORLD


motivair
Watt's next?
Infinite possibilities.
Direct liquid cooling technology enabling the AI Era.
Coolant Distribution Units – Heat Dissipation Units – Manifolds – Dynamic® Cold Plates
motivaircorp.com

Unlocking the Future: Breakthroughs in Scientific Computing

In today's rapidly evolving technological landscape, scientific computing is at the forefront of innovation, driving discoveries across various disciplines. From life sciences to engineering, advancements in high-performance computing (HPC), laboratory informatics, quantum computing, and artificial intelligence (AI), they are transforming how research is conducted and accelerating the pace of discovery. The implications are profound, reshaping everything from drug discovery and climate modelling to the design of new materials and the democratisation of computing power.

HPC is essential in tackling some of the world's most complex problems. Whether simulating molecular interactions for pharmaceutical development, optimising the aerodynamics of a next-generation aircraft, or predicting weather patterns with unprecedented accuracy, HPC provides the computational muscle needed to perform calculations on a previously unimaginable scale.

Laboratory informatics plays a crucial role in harnessing the full potential of these computational advances, particularly in the life sciences. As the volume of data generated in laboratories grows exponentially, there is a pressing need for sophisticated tools to manage, analyse, and interpret this information.

Quantum computing, although still in its infancy, promises to be revolutionise scientific computing. By leveraging the principles of quantum mechanics, quantum computers have the potential to solve problems exponentially faster than classical computers, particularly in areas such as cryptography, material science, and complex system simulations.

AI, with its ability to process and learn from vast amounts of data, is already revolutionising scientific computing. In life sciences, AI is being used to accelerate drug discovery, model protein folding, and personalised medicine.

With this special issue, we delve into the latest breakthroughs in these areas, examining how they reshape industries, drive innovation, and pave the way for future discoveries. From the laboratories of pharmaceutical giants to the experimental qubits of quantum labs, we explore the cutting-edge work being done to push the boundaries of what's possible in scientific computing. Join us as we take a deep dive into the technologies and innovations that are not just changing the world but redefining what's possible.

Editorial and delivery team
Editor: Robert Roe
robert.roe@europascience.com
Head of content: Finbarr O'Reilly
fin.oreilly@europascience.com
Chief operating officer: Mark Elliott
mark.elliott@europascience.com

Commercial team
Partnerships director: James LeMasurier
sales@europascience.com +44 (0) 7792 053919
Senior graphic designer: Zoe Wade
zoe.wade@europascience.com

Corporate
Chief executive officer: Warren Clark

Scientific Computing World is published by Europa Science Ltd, 4 Signet Court, Cambridge, CB5 8LA ● ISSN 1744-8026
Tel: +44 (0) 1223 211170 ● Fax: +44 (0) 1223 213385
Web: www.scientific-computing.com/

Subscriptions: Free registration is available to qualifying individuals. Register online at www.scientific-computing.com

Registered address: Europa Science Ltd, 4 Signet Court, Cambridge, CB5 8LA, UK. Tel: +44 (0)1223 221 030. Fax: +44 (0)1223 213 385.

US copies of any printed issues: Scientific Computing World (ISSN 1356-7853/USPS No 018-753) is published by Europa Science Ltd, and distributed in the USA by DSW, 75 Aberdeen Rd, Emigsville PA 17318-0437. Periodicals postage paid at Emigsville PA.
Send address corrections to Scientific Computing World PO Box 437, Emigsville, PA 17318-0437.

THANK YOU TO OUR PARTNERS

Breakthroughs would not have been possible without the support of its sponsors. *Scientific Computing World* is proud to partner with these inspiring organisations and to help showcase their work in this project

ACD/Labs

ACD/LABS

More than 4,600 labs worldwide choose ACD/Labs software because it is designed with scientific tasks and workflows in mind. Serving the scientific community since 1994 via offices in Asia, Europe and North America, it employs a team of more than 200, including PhD-level scientists, and specialists in informatics and chemical disciplines. It has an international client base in the chemical, pharmaceutical, food and beverage, environmental, agrochemical, flavours, fragrances, government, and educational sectors. Its product mix includes enterprise applications, processing software and databases, expert systems for analysis, predictions and simulation, as well as content libraries. ACD/Labs software is designed with scientific tasks and workflows in mind, so it can communicate science; streamline and automate complex workflows; understand molecular behaviour; support decision-making in HTE & CMC; extract answers from analytical data; share knowledge and enhance collaboration.
www.acdlabs.com

CDD VAULT® Complexity Simplified

CDD (COLLABORATIVE DRUG DISCOVERY)

CDD's mission is to provide an unparalleled software experience for humanitarian and commercial collaborative research and discovery, and it aims to achieve this by focusing on its people, values, goals, innovation, and technology. CDD was established in 2004 when founder and CEO Barry Bunin was an entrepreneur in residence at Eli Lilly & Co. Its flagship product, CDD Vault, was designed to address the problem of inefficient data management in the field of biological and chemical research. Over the years, it has been able to significantly expand the capabilities of its platform, as well as its user base, and is now helping hundreds of organisations and thousands of scientists globally to collaborate more effectively every day. It is a hosted, out-of-the-box solution that aims to streamline software to simplify data complexities for users. It provides a RESTful API with clear documentation to enable users to integrate with other applications in their workflows, with exemplary support systems.
www.collaborativedrug.com

GIGABYTE™

Gigabyte

Gigabyte, along with its subsidiary Giga Computing, is a leader in enterprise computing, renowned for its engineering, innovation, and hardware expertise. With strong industry insights and eco-partnerships, Gigabyte offers a comprehensive range of solutions for both traditional and emerging workloads, including high-performance computing (HPC), artificial intelligence (AI), data analytics, 5G/edge computing, and cloud infrastructure. Whether powering AI models or enabling edge computing, Gigabyte's solutions empower businesses to tackle complex challenges, drive performance, and stay ahead in the rapidly evolving digital landscape.
www.gigabyte.com

MERCK

MERCK

Merck is made up of a vibrant team of specialists in electronics, healthcare and life science. Across those three areas, there are common values of a love of science and a passion for technology. Its teams of research scientists are working tirelessly to develop groundbreaking medicines and technologies to treat everything from cancer and multiple sclerosis to infertility. Innovation is at the core of the value that it delivers – it is a global innovator, creating new technologies and markets for its customers through cross-industry collaborations and partnerships. It has come up with groundbreaking liquid crystal displays and windows, effect pigments for coatings and colour cosmetics, as well as high-tech materials for integrated circuits. Its main focus is in the electronics market, with its materials and solutions changing the way we access information.
www.merckgroup.com

motivair

MOTIVAIR

Motivair's wide range of products are focused on providing customised cooling solutions for the critical infrastructure and technology behind the world's leading businesses. Data and analytics are allowing those businesses to increase production speeds, reduce costs and improve quality at rates never seen before, but this increase in productivity is creating thermal challenges that Motivair is already addressing inside scores of factories, power plants, research laboratories, data centres, super computers and more. The company focuses on three core areas: direct-to-chip cooling; data centre and IT cooling; and thermal management.
www.motivaircorp.com

Merging medical data, modelling and HPC to personalise disease treatment

Duke University's **Amanda Randles** highlights the challenges of modelling complex biological systems, emphasising the importance of parallel computing and large-scale simulations.

Amanda Randles' research in biomedical simulation and high-performance computing (HPC) focuses on developing new computational tools that her laboratory uses to provide insight into the localisation and development of a series of human diseases ranging from atherosclerosis to cancer.

Randles' early work included creating accurate 3D simulations of how blood flows through the circulatory system. More recently, she and her team developed biomedical simulations that yield direct and concrete impacts on patient care, including simulations of 700,000 heartbeats, the interaction of millions of cells, and cancer cells moving through the body.

In 2024, the Association for Computing Machinery named Amanda Randles the recipient of the ACM Prize in Computing for groundbreaking contributions to computational health through innovative algorithms, tools, and HPC methods for diagnosing and treating disease.

Your work spans the intersection of medical science, molecular biology, and HPC. How did you initially become interested in combining these fields to address medical challenges?

I started out working in bioinformatics and liked the idea of using computing to solve biological problems. I worked at IBM after undergrad before going to graduate school, where I got to work on the Blue Gene supercomputer. I was exposed to high-performance computing and what you could do with these systems. I returned to graduate school because I wanted to use the supercomputer rather than just build it. I had experience with the supercomputing side, and I wanted to use it to address

→

Amanda Randles

→ different medical problems. I went back for my physics PhD, and now we're taking physics-based approaches to target biomedical questions.

When I joined grad school, we wrote the code from the ground up with the intention to run on Blue Gene. And it was written to be parallel from the beginning versus going back and retrofitting it. HPC lets you work with larger spatial or temporal domains than you would have been able to do without large-scale computing.

Throwing in biology just makes everything complicated. You can't take into account absolutely everything. And it's trying to make those decisions: what must we have in the model? What's necessary for the model? It gets complicated because, a lot of times, what we're trying to predict is disease. We're trying to predict where something went wrong. And so you're trying to predict the edge cases. This type of research introduces interesting problems – you're trying to capture all the potential paths that could occur, which is where supercomputing is helpful. You can try all these different options and then isolate some of the edge cases versus just trying to figure out the average of what's actually likely to happen.

HPC lets you make it more personalised. For example, in cardiovascular projects, the geometry of someone's arteries varies between patients, and the geometry affects the flow patterns, which is critical. So it's essential to have that 3D anatomy tailored to that patient, which makes it more computationally intense.

"What we're trying to predict is disease... You're trying to capture all the potential paths that could occur, which is where supercomputing is helpful"

Could you elaborate on the importance of these long-duration simulations for patient care and medical diagnostics?
[Randles and her team developed biomedical simulations, including simulations of 700,000 heartbeats - the previous state-of-the-art was 30 heartbeats]
That's one of the things I'm most excited about right now. A large part of my lab is looking at this network. Everything researchers have done before is trying to diagnose the patient by looking at one single heartbeat, and that's very useful. Still, when we couple it with wearables, we can get a realistic view of how blood flow changes over months or weeks, and that gives an accurate view of your patient's state. Instead of saying in this one representative heartbeat during exercise, you happen to hit this value of whatever metric we're looking at, we can say over six weeks, you spent 10% or 20% of your time in this at-risk state.

We know that exercise changes your susceptibility to disease, but having that connection between how much the exercise state or even just basic activity influences your blood flow is important.

Pathways are getting turned on. It's not how much time you spend in a single heartbeat in this one at-risk state. But how much of the next six weeks did you spend in these at-risk states? That will help us figure out if a patient is likely to have the blockage come back. Are they likely to develop further heart disease?

At a certain point, it switches from being a model to a digital twin, or at least closer to a twin. With wearable sensors, you have this constant coupling with the feedback loop of information from the patient continuously driving the flow model. It's not just this decoupled entity; it's not a virtual representation that's just based on the information you got from the health record, but it is continuously informed by the data we're getting from the patient, which is a massive shift in approach.

HaveaniceadayPhoto/shutterstock

What have been some of the biggest challenges in developing these precise simulations to simulate biophysical processes?

When we started working on this, in 2008-2009, it would take the world's biggest supercomputer to try to run just one heartbeat and capture the entire coronary tree. If you want to put this in the clinic, you can't use the world's largest supercomputer for every simulation. It took about six hours in 2010 on one of the biggest supercomputers at Argonne, just to get one heartbeat. When we're trying to get to millions of heartbeats or hundreds of thousands of heartbeats, that is incredibly difficult. A lot of what we did initially was how do you make that more tractable? So we looked at implementing new computational algorithms, different ways to save space on the supercomputer, and ways to save memory and speed up the calculations. We were doing this just when GPUs had started coming out. How do we start making use of the new hardware?

Now we've got bulk fluid simulations where you could run a coronary tree simulation between five minutes and under an hour. You can at least get to a flow simulation where you can make a diagnostic in less than an hour and do that in the cloud with a reasonable number of nodes. That took a lot of computational and algorithmic advances. Then, we tried to figure out whether we needed the entire tree. What kind of boundary conditions do you need? Where can we make assumptions? When can we get away with using reduced order models, and then how do we add in all of the individual cells and do that in a computationally efficient way?

What advice do you have for young professionals who aspire to make interdisciplinary contributions in fields such as computational health?

One thing is to be open to trying risky ideas. I've been fortunate – the National Institutes of Health (NIH) has been supporting crazy out-there ideas, which is great and helpful; not being afraid to try those ideas. However, collaborations have also been critical. So, for us, every project has a doctor or an experimentalist who either gives us data or provides some sort of feedback.

From a computer scientist and physicist side of things, there are 5 million questions you could be asking, but not all of them are translatable. They might not be able to be implemented in the clinic. They may be feasible computationally but, when you get the doctor's perspective, they're like, 'That's great, but I would never actually be able to do that, and here's why'. It's essential to get their perspective. Ensure you have those interdisciplinary collaborations and ongoing brainstorming sessions with the domain experts. Having them all in the same room really is critical.

We have surgical fellows who come to our lab meeting every week and are part of the lab and can weigh in and help give that expertise alongside the HPC expert grad student. And then we have teams from the Lawrence Livermore Lab that visit us, and it's not just on Zoom. They spend a week or two with us a year to work with us on the code and provide hands-on help. Building interdisciplinary collaborations with people and bringing those teams together allows you to ask these questions. **B**

Amanda Randles is the Alfred Winborne and Victoria Stover Mordecai Associate Professor of Biomedical Sciences at Duke University's Pratt School of Engineering

Accelerating ALS analysis with AI

Dr Ilary Allodi

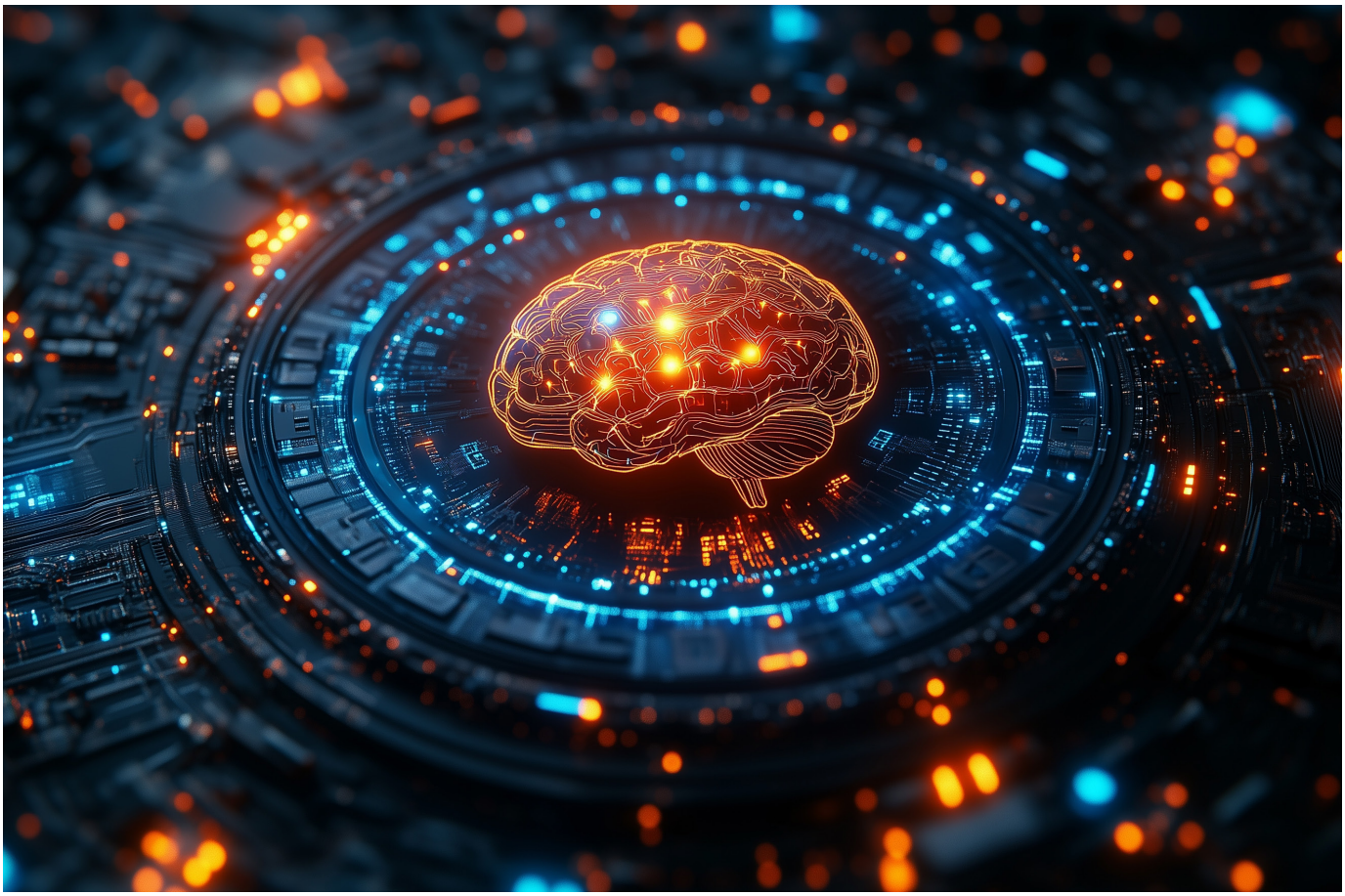
Dr Ilary Allodi from the University of St Andrews explains how she and her team used AI to deal with huge and varied datasets while exploring the progression of ALS motor neuron disease

Research from the University of St Andrews published a paper detailing how they harnessed the power of AI to generate new insight into the progression of Amyotrophic lateral sclerosis (ALS). *Scientific Computing World* spoke to lead researcher Dr Ilary Allodi, lecturer in Systems Neuroscience at the School of Psychology and Neuroscience, about how this tool was opening up new avenues to understand the mechanism of this disease faster than was previously possible.

The research, led by researchers from the School of Psychology and Neuroscience at the University of St Andrews in collaboration with the Department of Neuroscience at the University of Copenhagen, shows that specific cell circuits that control movement are affected early in the disease, while others are affected later during disease progression.

ALS is a fatal disorder in which motor neurons – the cells which control movement – progressively die. The disease has no cure, and life expectancy after diagnosis is generally only between two and five years.

Using techniques that enable researchers to simultaneously study multiple cell types in spinal cord tissue, combined with a novel AI-based analysis method, the researchers identified the specific networks of cells affected early in the disease before



AnggalihPrasetyashutterstock

‘These data sets are so big that, basically, manual annotation and manual quantification are just not feasible’

motor neurons die. These are subgroups of inhibitory interneurons – a type of cell found in the spinal cord known to activate motor neurons. In healthy individuals, these cell circuits are required to perform movements such as walking and running. There are specific cells, called inhibitory or excitatory interneurons, which control different aspects of movement by activating motor neurons. These inhibitory and excitatory cells are highly heterogeneous and intermingled within the spinal cord and are difficult to investigate simultaneously.

Researchers developed an AI-based method to facilitate data quantification. The work uses cutting-edge methodology to identify the cell types that are contributing to disease. These inhibitory and excitatory cells are highly heterogeneous and are intermingled within the spinal cord and are often difficult to identify and investigate simultaneously. The computational approach that the researchers developed allows them to overcome these limitations while shedding light on potential new targets for treatment.

What is your background and how did you come to implement AI in ALS research?

My background is in neuroscience, especially in system neuroscience which focuses on how the different neurons and the different systems within the brain and the spinal cord are connected to each other. There is a lot of work to understand the connectivity and the function and how that allows human beings, and other species, for example, to move around, sense stimuli, and interact with the environment.

I have a very deep interest and longstanding interest in neurodegenerative disorders. My background is trying to combine these two elements with a particular focus on studies of neuronal neurodegenerative disorders, especially ALS and frontotemporal dementia, using this system neuroscience approach. Looking more at the connectivity and how the disease affects the system and not just a single cell, not the single neurons. It’s a holistic and more systemic approach to researching disease. We started applying AI, which is a very useful tool – this is one case – but we also use it for other types of analysis such as behavioural analysis and other types of data analysis because we generally produce very large datasets.

Why did you choose to implement AI in this type of research? What challenges did it solve?

These data sets are so big that, basically, manual annotation and manual quantification are just not feasible; it will take years. And, also, there are issues with experimental bias. So you want to have a tool that will allow you to speed up your annotation and whatever data set you’re handling,while at the same time reducing the amount of experimental bias that you’re introducing in your qualification.

I started my PhD in 2008. So I’ve been doing this for quite some time. And there was a lot of manual annotation. I remember being on slides, counting neurons, and thinking there must be a better way of doing this. The methods improved over time, but you need to find the balance between having your analysis be fast enough and accurate. There have been issues in the past with software and different tools not being →

PopTika/shutterstock

→ accurate enough. That was one of the main issues, I think, with some of the tools in the earlier days and why scientists often did a lot of manual annotations.

In our recent study, we quantified 90,000 cells. We work with very complex tissue, specifically the spinal cord and brain. If you look at a section of the spinal cord, it is this beautiful butterfly shape with thousands of neurons and other cells. All these cells have different identities and functions. So you're trying to pull apart the elements of the cells, in this very complex issue – everything is highly intermingled. So, it is really quite difficult to find a good tool to isolate your cells of interest.

We started applying AI because we had this very big data set and didn't know how to approach it. We tried to apply different types of software, but it was not accurate enough. Quite often, what is done here – in these types of research studies – is that the shape of the cell is computed from the size of the nucleus. But this doesn't work well for neurons, because neurons have very peculiar shapes. If you look at a spinal cord section, motor neurons are huge cells. So you have some very big cells on one side of the spinal cord that we call the ventral horn but then, on the other side of the spinal cord, you also have very round and very small neurons.

You can't try to pull apart the cells just by using a basic algorithm based on the shape of the nucleus; when you try to infer a probabilistic shape of the cell, it just doesn't work. We used a Deep Ensemble model based on multiple convolutional neural networks; I think it's 16 that we trained from annotated data. There was another issue that we had quite a limited amount of data. We wanted to be very specific, look at multiple populations, but also the sub-populations within these inhibitory and excitatory cells, and try to pinpoint what was going on. The cells are very heterogeneous, and depending on this heterogeneity, they also have different functions. So it was very important for us to identify the neurons first.


PeopleImages.com - Yuri A/shutterstock.com

How will the use of AI and computational methods further impact this research in the future?

We did the machine learning in collaboration with the Department of Computer Science at the University of Copenhagen. But you really need someone who likes coding and bioinformatics because it's a very computational approach to solving this research challenge. You have to spend a couple of weeks trying to make sense of the data and get it ready to run on the compute cluster.

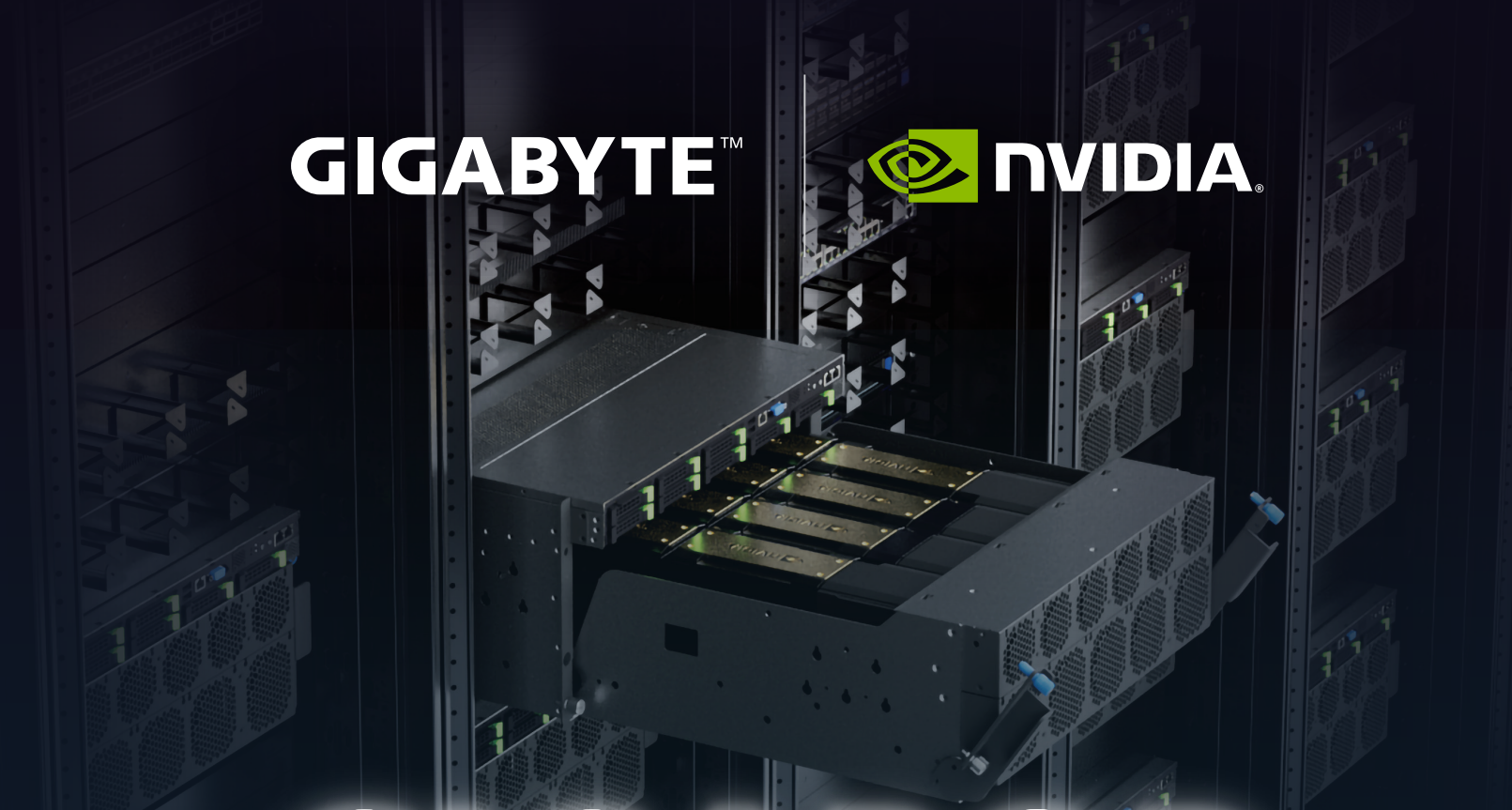
Not only in neuroscience, but I think a lot of fields have, for a very long time, been very wet lab-focused.

Nowadays, because of the techniques that have been developed throughout the years, we produce very big datasets. would say that we can look at things in a very different way, which is fantastic. But if you look at how research was conducted just 10 years ago and what I'm doing right now, you get a very different picture.

You're still going to need wet lab research; experimentation is not going away. But I think, especially for the new generation coming into the field, they need to understand that if they want to produce certain types of datasets and answer certain types of questions, they might spend a lot of time coding and implementing computational analysis. I'm not saying that we will replace the wet lab by any means, because we still need that as we need model organisms. So it's still very much needed. But I think we can reduce the amount of time that we spend doing that. **B**

Dr Ilary Allodi is a lecturer in Systems Neuroscience at the School of Psychology and Neuroscience at the University of St Andrews

GIGABYTE™ | NVIDIA®



GIGAPOD

A Scalable, Turnkey AI Supercomputing

GIGABYTE enterprise products not only excel at reliability, availability, and serviceability. They also shine in flexibility, whether it be the choice of GPU, rack dimensions, or cooling method and more. GIGABYTE is familiar with every imaginable type of IT infrastructure, hardware, and scale of data center.

Large Language Models

Science & Engineering

Generative AI

Discover More

Cloud computing and the ‘democratisation’ of protein structure data

Jason Cole tells us how a collaboration with Intel and AWS helped the Cambridge Crystallographic Data Centre to democratise access to protein structures

Jason Cole

The Cambridge Crystallographic Data Centre (CCDC) is a non-profit organisation that specialises in structural chemistry data. The centre’s primary activity is the compilation and maintenance of the Cambridge Structural Database (CSD) along with chemistry data, software, and knowledge for materials and life sciences research and development. CCDC has developed the CSD, a resource containing fully curated organic and metal-organic structures widely utilised by researchers since it was first created in 1965.

In a recent collaboration with Intel and Amazon Web Services (AWS), the CCDC created a curated data set of protein structures from the Protein Data Bank (PDB) with predicted hydrogen positions. The PDB is a database for the three-dimensional structural data of large biological molecules such as proteins and nucleic acids. Biologists and biochemists obtain and deposit these structural data using experimental methodologies such as X-ray crystallography, NMR spectroscopy, and, increasingly, cryo-electron microscopy.

Historically, collaborations with the pharmaceutical industry have enabled the development of reliable methods for interpreting interactions within protein binding sites using proprietary information not publicly available. Repeating these studies with PDB structures presented a challenge due to the absence of hydrogen positions in water networks within the proteins. Reliable predictions require databases of augmented protein structures where hydrogen positions are assigned.

Generating this information computationally is intensive, considering multiple possible models. Overcoming this computational challenge was possible for the CCDC by collaborating with Intel and AWS. The CCDC generated a comprehensive snapshot of protein cavities in the PDB, identifying potential binding sites for small molecules with accurately predicted hydrogen positions for all components. *Scientific Computing World* spoke to Jason Cole, Senior Research Fellow at the CCDC, about the centre, the collaboration programme and the benefits for researchers.

How did you come to work at CDCC and what is your role?

I’ve worked at CCDC for nearly 30 years, which is pretty unusual in the modern environment; to stay in one place. I’ve been here since I finished my PhD and, in some ways, I was associated with the organisation even before I finished my PhD. At the end of my PhD, I was asked if I

'In a protein structure... you have to infer and guess where the hydrogen atoms are. Some people have sophisticated software that allows them to do it, but not everyone has [access]'

would like a job at CCDC. I first started as a software developer at CCDC because of my work with crystallography, my background in chemistry, and my experience writing software.

Then I worked on a project called Gold (Genetic Optimisation for Ligand Docking), which is a software for protein-ligand docking. At that time, CCDC was commercialising that software.

[Eventually, after running teams of software engineers and scientists] I was Deputy Director. I did that for about seven years but never aspired to be the organisation's director. That's a very different role and not really suited to me. So, after quite a few years, I decided to step away from that. The CCDC still wanted me in the organisation as a sort of mentor to help others. So, I decided to take on a role as a Senior Research Fellow. Today, I work within the organisation in a research-oriented capacity. I look after other researchers and contribute to software development and science in general.

I'd always regarded myself as a scientist, first and foremost, and then I learned how to write some Fortran, which was the coding language we were using at the time. While at CCDC, I've learned other languages such as C and C++, Python and Tcl/Tk and any others I've had to code in along the way. It's true that once you've grasped one or two of them, it's much easier to pick up others – just like languages!

I never thought of myself as becoming a software engineer but, during my PhD, because I ended up doing quite a lot of software engineering in it, I got a bug for it, and I wanted a job where I could use science, but also use software engineering and write code to serve a scientific purpose.

What was the main motivation behind the CCDC's collaboration with AWS and Intel to create this curated data set of protein structures?

We were keen to try and develop a set of structures that, in essence, were curated a bit more in an automated fashion than you get out of the PDB. From the Protein Data Bank, you get the structures that the crystallographer produced at the time. And, invariably, anyone using one individual one of those has to go away and do a little bit more work on it.

In this particular case, they have to develop or map out these hydrogen positions. Because when you examine a crystal structure with X-rays, you can't see the hydrogen atoms. You can see the hydrogen atoms with a small molecule structure but, in a protein structure, it's impossible to see the hydrogen atoms in most cases. You can only really see the heavy atoms. That means you have to infer and guess where the hydrogen atoms are. And everyone does that. Some people have sophisticated software that allows them to do it, but not everyone has [access]. CCDC wanted to provide a set of data that was a bit more curated. It would be more modelling-ready for someone to use in a large-scale research project.

PDB does a fabulous job at curation, but they don't model additional things around structures. They provide the record of publication, a representation of the structure that was put out by the person who did it. That is their purpose.

The motivation was to provide a big data set, for everyone, [of] pre-generated data that people could use in their modelling studies. Everyone has access to the structures. It's democratising access to structures that have had a bit more curation work done on them.

The other motivation, from a purely CCDC point of view, is that we wanted better software to do this and didn't feel our software at the time was up to the level we needed. It allowed us to do a research project while improving the workflows we had to do that particular job.

Could you try to explain to us the significance of hydrogen positions i n protein structures?

If I am trying to look at docking on a particular protein, for example, I have to add the hydrogen position. There are probably docking programs that could try to do without them, but they're essential for most modelling applications. If you want to do molecular dynamics, you'd want hydrogen positions added to your structure before you even start modelling.

DFT (density functional theory) and many other methods require a complete model for reliable results. There might be a few things such as meso-scale modelling where hydrogen positions do not matter as much. Researchers use unified atom types but, in essence, you want to have the atomic positions of hydrogen atoms present for most people's job.

What was the primary motivation behind the CCDC's collaboration with AWS and Intel to create this curated data set of protein structures?

I think AWS and Intel were interested because it was a significant volume of computational time you would need to complete a project such as that. So that was where their interest was.

In fairness, we probably didn't use as much computational power for the project as AWS and Intel would have expected. Because they're thinking in the millions of CPU hours, whereas it only took about 300,000 CPU hours – not as much as they would expect – possibly an order of magnitude less than expected. But that's still a really significant investment – thousands of pounds' worth of CPU time. If you scale that across the whole world and everyone repeating that piece of work, you can get a very big number very quickly in terms of cost, which is a crucial point driving this work.

Should new researchers coming into the field learn coding skills?

These days, I would strongly encourage it... for any student getting into modelling. Not necessarily medicinal chemists, someone doing synthesis or making chemicals on the bench, not all of them need to code, although it does help. But with a PhD student, yes, I'd strongly encourage them to try and learn coding.

It isn't for everybody, of course. Not every student you have really takes to it, and you could find a pathway for that student to still succeed. But for the students we have, I'd strongly encourage it.

If the student is going to do computational modelling, it becomes a necessity these days to be able to write in one of the higher-level languages, such as Python or Julia, or something along those lines. They don't necessarily have to know C++ or Fortran.

There are high-level things you can use, such as KNIME. That's not coding; it is more workflow development, but that's another thing that chemists can use. It's a higher-level way of programming where you're pulling together modules and connecting them in a workflow. It's all very visual. So that, I think, is another skill that some will take to that rather than getting into the coding directly. **B**

Jason Cole is a Senior Research Fellow at the Cambridge Crystallographic Data Centre (CCDC), a not-for-profit organisation based in the UK.
You can see a selection of its work at www.ccdc.cam.ac.uk

Laser-based computing unlocks options for NP-hard optimisation

Can lasers deliver the next leap in performance for scientific computing? **Dr Ruti Ben-Shlomi** discusses the potential of this new technology

LightSolver has created a laser-based computing paradigm that aims to surpass the performance of HPC and quantum systems through its laser-based computing system LPU100. The LPU100's laser array represents 100 continuous variables and can tackle problems with up to 120,100 combinations, enabling organisations to accelerate their processes and make mission-critical decisions faster than previously possible.

Dr Ruti Ben-Shlomi, CEO and co-founder of LightSolver, highlights several important points that make this new computing system exciting for researchers. "What's so interesting about this computer is that it operates at room temperature," she says. "It's as small as a desktop PC, and is built from off-the-shelf components; it's not quantum, but mimics some quantum effects, such as tunnelling.

'You can find NP-hard problems everywhere... from synchronisation and job scheduling to optimally routing vehicles or even folding a protein'

"We found a way to tackle hard optimisation problems (NP-hard)," Dr Ben-Shlomi continued. "We know how to imprint these problems on the phases of the lasers, and the lasers interfere simultaneously with all-to-all connectivity in such a way as to fulfil most or all of the conditions. Our processing method doesn't use any electronic components, but only lasers, which enables us to overcome the performance barriers of classical computers. For complex problems with many constraints, the time-to-solution (TTS) for classical computers is very large or even unfeasible, but our lasers can solve some of these tasks very easily. Dr Ben-Shlomi stressed that NP-hard problems are much more ubiquitous than it may seem. Many common challenges in HPC and AI could be translated into optimisation problems. This means that the list of use cases for this technology continues to grow as organisations begin to really understand the capabilities of the LPU and translate their computational challenges into optimisation problems.

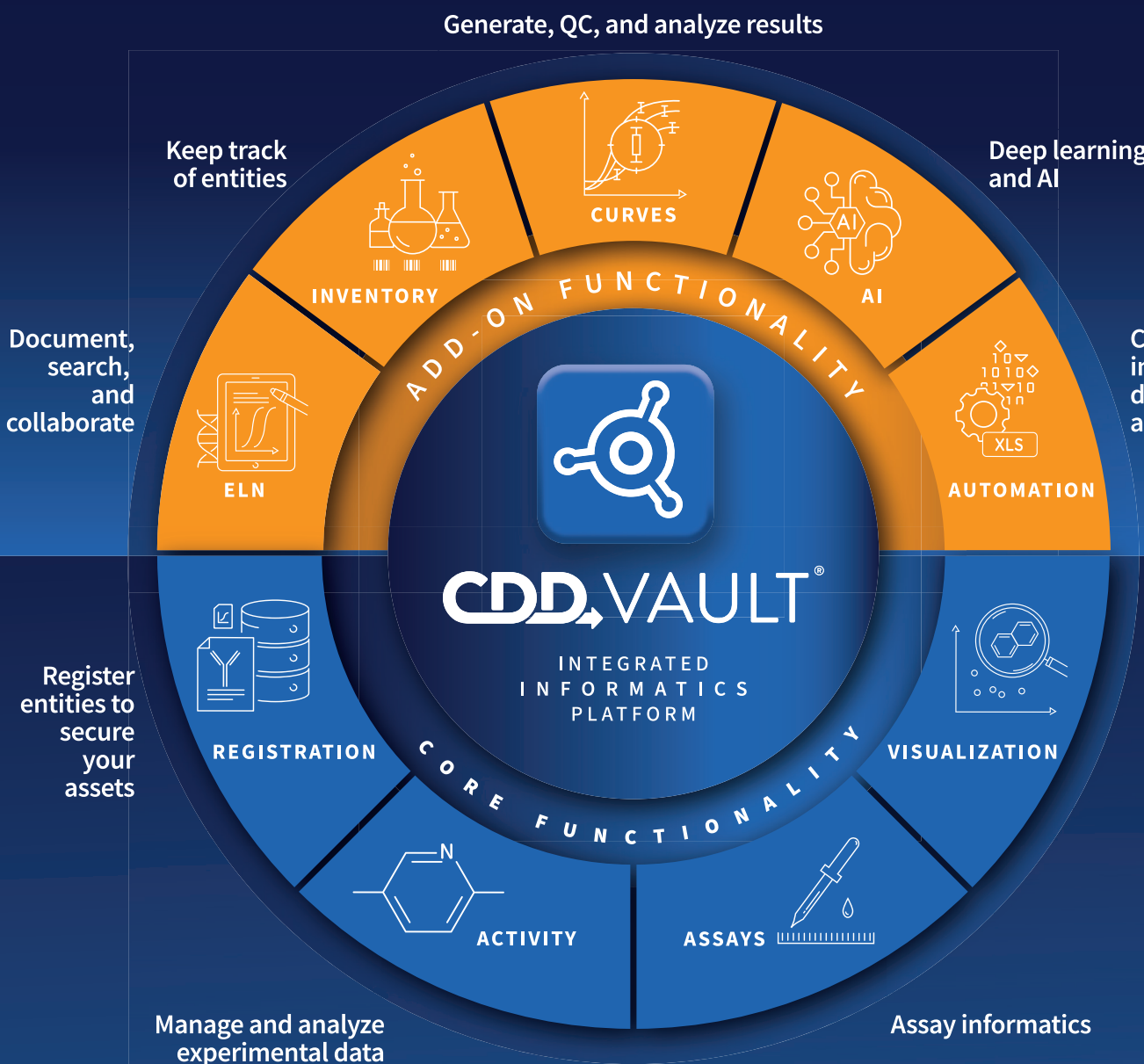
Dr Ruti Ben-Shlomi

"I'd say you can find NP-hard problems everywhere in the industry today. I can give you specific problems that we're working on and that we know we're very good at, such as image processing, credit scoring in finance, and all kinds of problems in machine learning and AI such as feature selection. These problems include everything from synchronisation and job scheduling to optimally routing vehicles or even folding a protein," Dr Ben-Shlomi told *Scientific Computing World*. "Every Fortune 500 company I've spoken with has multiple →

CDD VAULT® Complexity Simplified

Scientific data management

with new add-on functionality to make discoveries faster



cddvault.com

Request your **FREE** trial

optimisation problems that they can't solve. Sometimes, they don't even know they are looking at optimisation problems. For example, if I'm talking with a quantitative analyst from a financial institute or an algorithm developer, they might say, 'I'm using Monte Carlo or gradient descent. Can you help me provide a better solution?' That's probably because they're dealing with an optimisation problem and they don't realise it.

"Another reason the use cases continue to grow is that the company has developed a new method that gives an edge in solving differential equations because we map differential equations into linear equations," said Dr Ben-Shlomi. "We're really good at solving linear equations in that form, so, suddenly, we have opened a huge market of opportunities hat did not obviously seem to be 'optimisation problems', because differential equations are not typically thought of in that way."

Quantum sparse coding

Sparse coding refers to a method for representing data in a more concise and informative manner. It involves finding a sparse set of basis vectors or features that best describe the underlying structure of the data. In this context, 'sparse' means that only a small subset of features or basis vectors are actively used to represent each data point, leading to a more compact and interpretable representation and reduced noise. Sparse coding has applications in various domains, including image processing, computer vision, natural language processing, and neuroscience. Sparse coding also plays a key role in feature selection in machine learning where it is used to generate a predictive rule on which features manifest the strongest effect, given a large pool of features.

Sparse coding problems are NP-hard and typically solved through approximation methods such as Lasso and OMP (Orthogonal Matching Pursuit) because they are computationally hard and very complex to solve via traditional methods.

LightSolver has developed a quantum-inspired algorithm for sparse coding that implements the exact problem. They translate the problem into quadratic unconstrained binary optimisation (QUBO) matrices, achieving more accurate results than classical methods, particularly for complex cases with a large number of features and a small number of measurement.

Referencing LightSolver's quantum sparse coding paper published earlier this year in the *Quantum Machine Intelligence* journal, Dr Ben-Shlomi noted: "Because it scales exponentially, you need to deal with a lot of data sets to solve such a problem. So, researchers will use approximations, which is called L1 regularisation. In a recent research paper, we showed that we can use an exact method, not an approximation. We know that it's NP-hard, but we have a solver that is very good with these problems."

In the research paper that Dr Ben-Shlomi co-authored, the researchers explain how this method was established. "We formulate the most general sparse coding problem as a QUBO task, which can be efficiently minimised using quantum-inspired methods. To derive a QUBO model that is also efficient in terms of the number of spins (space complexity), we separate our analyses into three different scenarios. These are defined by the number of bits required to express the underlying sparse vector: binary, 2-bit, and a general fixed-point representation. We conduct numerical experiments with simulated data on LightSolver's digital platform to verify the correctness of our QUBO formulation and to demonstrate its advantage over baseline methods."

The researchers found significant improvements in challenging regimens with under-determined matrices where the noise level and the cardinality are relatively large. "While our QUBO formulation can handle the most general fixed-point representation of the underlying sparse vector, we believe it would be of great interest to extend the proposed framework to a floating-point representation," Dr Ben-Shlomi explained.

The researchers also noted that this method could be particularly useful in biological sciences: "There is a growing interest in using quantum technologies in biological sciences and the methods presented in this paper may provide performance improvements in applications where sparse linear regression plays a role. For example, in genome-wide association studies (GWAS), scientists are interested in accurately identifying which of the thousands of single-nucleotide polymorphisms (SNPs) are associated with high cholesterol or any other phenotype of interest. The ability to find a sparse, interpretable set of 'important' SNPs can be used to improve medical treatment as well as to expand our understanding of the human genome."

Dr Ben-Shlomi added: "We can deliver better results than the state-of-the-art machine learning tools that are available today to deal with these problems, such as OMP and Lasso. We also showed that we can deliver accurate results much faster and that we can use less data and deliver the exact same results." She also stressed

'What's so interesting about this computer is that it operates at room temperature. It's as small as a desktop PC, and is built from off-the-shelf components; it's not quantum, but mimics some quantum effects, such as tunnelling'

that these types of problems can be found in many HPC market verticals and are also present in AI. "When you're looking at the HPC market, you can see it is divided into the verticals of defence, weather forecast, biological, and other verticals. All of these tasks involve challenging optimisation problems. You need big models and complex simulations. Any kind of combinatorial problem that has a lot of constraints can be an optimisation problem. Even in AI, you have them. Gradient descent is one example."

Benchmark results

LightSolver has delivered several papers and results benchmarks highlighting how the LPU performed against state-of-the-art computing technologies such as AI and ML.

For example, in a head-to-head challenge, LightSolver achieved 2X-1,000X faster Time-to-Solution. LightSolver develops laser-based computing systems that are particularly effective for optimisation problems, specifically NP-hard.

Taking on the '3-Regular 3-XORSAT Challenge', LightSolver broke the exponential barrier, thereby beating classical and quantum computers and extending the maximal problem size to more than 16,000 variables.

LightSolver was also recently benchmarked against Gurobi for solving an Equal Demand Capacitated Vehicle Routing Problem. The work demonstrated how LightSolver optimised field service technician scheduling for a US telecommunications provider by using a QUBO formulation and running it on the LightSolver platform, benchmarking its performance against the Gurobi solver. **B**

Dr Ruti Ben-Shlomi is the CEO and co-founder of LightSolver, with CTO Dr Chene Tradonsky from the world-renowned Weizmann Institute. The company was founded in 2000 and is based in Tel Aviv, Israel. You can see its work at www.lightsolver.com

Via Nova Therapeutics boosts research with CDD Vault for secure data

How Via Nova Therapeutics uses CDD Vault to streamline global collaboration and securely manage antiviral research data

Via Nova Therapeutics deploys state-of-the-art drug discovery modalities to identify novel, first-in-class antivirals for the individual infections it studies. These modalities include phenotypic screening in infected cultured cells, biochemical screening in vitro on virus-encoded enzymatic targets and structure-based drug design, as well as combinations of the above approaches. Where appropriate, the firm also targets host factors essential for viral infection.

The problem

The application of these approaches has generated specific antiviral programmes targeting influenza, rhinovirus, BK virus, and adenovirus. Initially, the company uploaded spreadsheets into a shared repository, but soon outgrew this solution and required a secure and robust central repository to support collaboration across the globe. Via Nova Therapeutics deployed Collaborative Drug Discovery's CDD Vault – the hosted drug discovery informatics platform that securely manages both internal and external biological and chemical data. "Prior to CDD Vault, if someone wanted to dive into the data, we would waste time trying to find the exact data sets we needed," says Benjamin R Taft, PhD, Executive Director, Chemistry, at Via Nova Therapeutics. "CDD Vault really helps us get organised and have all of our data in one place where we consider it published and 'final', searchable and sortable."

The benefits

Via Nova Therapeutics has found a number of benefits since adopting CDD Vault:

- Gaining 'one true secure data repository'
- Supporting a globally dispersed workforce
- Facilitating collaboration
- Creating the flexibility to upload rich data
- An integrated visualisation tool
- Pricing that represents a 'huge value'
- A working arrangement that 'feels like a partnership because CDD is a great company to work with'

After trying to store and track data within its previous shared repository, Via Nova Therapeutics has a deep appreciation for the ease of use and precision of CDD Vault. "We've gained something mission critical from a business standpoint – having one true secure data repository," says Taft. "All the new data we generate as a company gets uploaded into our CDD Vault. From there, we can sort it, query the most important data really effectively and efficiently to find and then visualise it – with plots and graphics, if needed. And we can do this from wherever we have connectivity."

motorolka/Shutterstock.com

Via Nova Therapeutics uses advanced drug discovery techniques to identify novel, first-in-class antivirals for specific infections

Supporting a global workforce

With its globally dispersed workforce, Via Nova Therapeutics values the unifying force of its CDD Vault. "We're rarely in the same building, but we're all working together using CDD Vault as our one true source of data," says Taft. "Our people can access data from anywhere in the world, which is critical for our work."

The company also appreciates that it can provide granular access to data to CROs and contractors. "We have contractors and part-time employees that are supporting some of our projects," says Taft. "We can give them access to the Vault, with layers of security so that they can only access the data required for their work, allowing them to register structures, upload data, as well as download their allowed data into other tools as needed."

Facilitating collaboration

Via Nova Therapeutics values the ways in which CDD Vault facilitates collaboration across its operations. "We have files come in almost every morning from our contract labs in Asia," Taft says. "I can review the data from my home, send over a note to a colleague in a different county, who will QC the data and upload it into the Vault. I can then pull the data into a query, visualise it, create slides and show visualisations of the data to our executive team – it's all pretty seamless. The collaboration that CDD Vault enables is awesome."

The flexibility to upload rich data

Via Nova Therapeutics is impressed with the wealth of supplementary information it can store within CDD Vault. "We're doing exclusively small molecules but, for some of our more advanced programmes, we have lots of additional data associated with each molecule. One of our favourite features is how flexible the CDD Vault is. You can pretty much upload any data you want."

Taft was also impressed by the ease with which the extra data was uploaded into CD Vault. "We recently got some really complex PK data, which is analysing three different tissue types across different time points," Taft says. "It was a huge web of complicated data. My colleague was able to do it within half an hour. This included plots of the PK time course curves, dose-response curves, in vitro, in vivo, and DMPK – it's all in there. CDD Vault allows us to put everything we want in there very easily."

Integrated visualisation tool

The visualisation tool built into CDD Vault has proven popular with scientists at Via Nova Therapeutics. "We like the fact that visualisation is built in," Taft says. "If I'm querying data, visualisation is just one click away. Something as simple as a scatter plot can sometimes really help you look at the data. The visualisation tool is convenient, fast, and it allows you to do a lot of the visualisation that you need on a day-to-day basis in a very convenient way." The company also likes the fact that CDD Vault integrates seamlessly with third-party visualisation tools. "If I need to do something more complex, I can easily export the data into our StarDrop™ software – it's pretty seamless. It's nice it has that function to interface with other software because you can't expect one piece of software to do everything."

CDD pricing represents a 'huge value'

Via Nova Therapeutics wants to dedicate its resources toward research – not infrastructure. "We want to stay lean as we move towards our next value inflection point in our company," Taft says. "Coming from a large pharmaceutical background, you have a good idea of the kind of immense infrastructure those organisations have for data storage. We're able to achieve our centralised data repository needs with CDD Vault – without having to invest in infrastructure. And we are getting this at a great price point. We see CDD Vault as a huge value." **B**

Supporting biological research with bioinformatics skills development

The shift from traditional to computational biology has highlighted the increased complexity of data due to high-throughput imaging and single-cell analysis, says **Simon Andrews**

Simon Andrews

The Babraham Bioinformatics facility supports Babraham researcher's computational biology requirements. The team is composed of experienced bioinformaticians and statisticians that provide advice or practical assistance on how to apply computational methods to extract biological information. As biological research shifts from a traditional wet-lab focus to more computational research, the facility aims to support researchers by offering services, including assistance in the processing and analysis of biological data, and developing custom applications or pipelines to meet specific user requirements. The Bioinformatics facility also develops and maintains a range of publicly available software used by bioinformaticians and offers a variety of modular bioinformatics and statistics training courses.

The Bioinformatics group has a wide range of experience in high-throughput sequencing, including primary processing and transcriptomics and epigenetics techniques. The research centre works on various research projects focusing on life-science research with a long list of publications. For example, in 2023, the researchers released a paper entitled *Intra- and Interchromosomal Contact Mapping Reveals the Igh locus has Extensive Conformational Heterogeneity and Interacts with B-lineage Genes*. This study explored chromosomal alterations and uncovered patterns that show how chromosome dynamics underpins antibody diversity. This project used Seqmonk, a software package developed by the Babraham Institute, to visualise and analyse high throughput mapped sequence data. We meet its Head of the bioInformatics facility, Simon Andrews.

What is the role of the Babraham bioinformatics facility?

Our job is to provide the resources, guidance and infrastructure that people need to get the computational part of their research done. That means providing the hardware and software infrastructure alongside the teaching – whether that be about statistics, data handling or data management.

We provide he bits of infrastructure that plug everything together to get the data into the right place and to have the right software available and knowledge to apply that correctly to people's data. We're an

enabling facility made up of scientists. The people who are within my group were biologists originally, but now work exclusively computationally.

Why did you build a team of biologists rather than computational experts?

We've tried it both ways, in that we've taken computational people and tried to give them enough biology knowledge, and we've taken biological people and given them the computational training. For the type of work we have traditionally done, it's generally worked better to take the biologists who understand the kind of nuances of the science that they're doing and how messy the data is. We find that the biologists know how messy the process of getting those numbers has been, and that often has quite a big impact on how you interpret the data.

We have a sneaky way of getting people. One of the things that we do is a lot of training courses. So we do lots of programming courses and then application-specific things. We've had people who have come on those courses, have done well on them, or have an aptitude for it. Some people just get it, and you can see they will pick it up. They ask all the right sort of questions and things like that. Taking people like that has been a productive way forward. You don't necessarily need to get people with the skills; get people with the right mindset.

This is important because you're learning new stuff every day when you're doing this. Even if you've got the perfect skills now – in six months, half of those skills will be out of date.

How is biological research changing and how does this affect the computational requirements?

Experiments are routinely getting larger, but the complexity of the data has also increased. That's both through a change in the place the data has come from – so, for example, we are now generating a lot more high throughput imaging data than we used to. I'd say that imaging used to be a fairly low throughput thing, but now, with automated acquisition platforms, you can generate a lot of data from that. Just processing anything out of imaging data is generally more computationally intensive than sequencing data.

But even on the other genomics-type experiments, we're moving a lot to single-cell data now. So instead of getting one set of values from a sample, you're getting quantitations for every cell in that sample, which vastly increases the dimensionality of the data and multiomics. We are not just measuring one thing, such as the RNA content of the cell. You may now measure two, three or even four different types of biological information from each cell. So it's the complexity of data, and as that complexity increases, the methods you need to make any sense of that also increase complexity. That has a knock-on effect on the computational requirements.

It's the same kind of biological systems we were looking at before, but previously the resolution was limited by your ability to collect data.

If you have a complex tissue sample – such as an organelle or something like that – previously, to get the amount of material you would need to measure, say, the RNA content, you'd have to mush the whole thing up.

That would mix all of the signals from all of the different structures that were in there together. Then you had to hope that what you were looking at was a big enough effect that, even though you've done all this sort of averaging and mixing of things together, something interesting still came out. No one really wanted to do that in the first place, but you're constrained by the limitations of the machines that can collect the data.

There are some examples where you can go down to single cell level, and all the cells behave the same. But there are plenty of circumstances, and when we've got a lot of groups that work in very early stage development, so you're moving from individual gamete cells, up to the initial stages of development.

In this example, almost every cell in the first stage does something different. Then there are different cell subpopulations.

The significant advantage of bioinformatics is that the computation can replace something that you would previously have done in the lab. If you have a collection of immune cells, such as a blood sample where you take your white blood cells and separate those cells into different subpopulations using a flow cytometer. If you knew how to identify those populations, you could physically separate them, take your cells, make a sample, and measure them separately. Today, we can profile the whole collection of cells at a single cell level and divide that at will computationally.

Now, if I have this collection of cells, I suddenly think maybe that's two separate sets of cells. I don't need to go back and run another experiment in the lab. I can take the same data, split it a different way computationally, and then analyse those against each other.

This has meant the speed of turnaround and the flexibility of analysis is way better than it would be if you've physically carried out the experiment.

How does the size of data sets impact computational biology research?

I'd say that one of the things that's certainly true now is that when you collect a sufficiently large data set, even when you published it and you are finished with that data set, you haven't got everything that you could from that data.

There are more interesting stories in there. So this is why we currently have such a big push to have open data and for people to publish their data at the ends of studies so that we can now pull in data that other people have generated and ask our own questions.

Do these changes impact the skills that bioinformaticians need?

These days, you pretty much can't get away without having informatics skills... if you consider yourself a wet lab scientist, you've got to have enough skills to do at least basic exploration of your data – even if you've given it to someone else to analyse, you've got to know enough to interpret what they give you back.

A lot of the training courses that we do, even if they're never going to do it, just gives people enough context to understand how the results that they're presented with were generated so that they can then make a sensible evaluation of how much they trust them or believe them or are prepared to invest their future time in them. For example, where we do our statistics training, a lot of that is not necessarily about the practicalities of running statistical tests. It's understanding the concepts, what you're measuring, and how you interpret or understand that data.

We run those courses in two different environments: we have a user-friendly desktop program. We tend to use a program called GraphPad Prism, which is a commercial program, but it's super friendly. But we also do the same thing in R, so you can control your statistics there. And more and more scientists are picking up some degree of coding skill, mostly informatics.

It's mostly either R or Python that covers the majority of our work. So many of the methods and frameworks to handle the types of data that people generate are written initially in those languages and. if you want to be able to access them, that's where you need to be. **B**

Simon Andrews is the head of the bioinformatics facility at the Babraham Institute. You can see some of its work at www.babraham.ac.uk

‘We genuinely didn’t know whether it would work, but… all the hyperscalers have adopted Arm’

Professor Simon McIntosh-Smith highlights Arm’s growing role in HPC

In 2023, GW4, a consortium of four research-intensive universities in South West England and Wales, secured £10m in UK Research and Innovation (UKRI) funding to build Isambard 3, an Arm Neoverse and Nvidia Grace supercomputer, the latest generation of G4W’s general-purpose HPC supercomputing systems based on Arm CPUs.

In parallel, the UK government provided £225m to create the UK’s fastest supercomputer at the University of Bristol – known as Isambard AI. This supercomputer is expected to be 10 times more powerful than the UK’s current fastest supercomputer and among the most powerful in the world.

Isambard 3 will utilise the Arm Neoverse-based NVIDIA Grace CPU Superchip to provide a production system of at least 55,000 cores. The new system – one of the first in the world based on NVIDIA Grace – will have more than six times the computational performance and six times the energy efficiency of Isambard 2. While Isambard 3 is a continuation of G4W’s validation of the Arm architecture for general-purpose HPC, being predominantly CPU-based, Isambard AI uses Nvidia’s Grace Hopper architecture (NVIDIA GH200 Grace Hopper Superchip) chip to provide performance for AI and HPC codes that rely on highly parallel and mixed precision computational performance to drive AI workloads.

The funding injection was part of a £300m package to create a new national Artificial Intelligence Research Resource (AIRR) for the UK, announced at the government’s AI Safety Summit at Bletchley Park.

The new Bristol facility will be used by a wide range of organisations from across the UK to harness the power of AI, which is already the main driver of emerging technologies such as training large language models (LLMs), big data and robotics. The new supercomputing facility will also accelerate automated drug discovery and climate research. Isambard 3 and Isambard AI are being installed in a new modular data centre at the National Composites Centre, near Bristol.

Scientific Computing World spoke to Simon McIntosh-Smith, Professor in High Performance Computing at Bristol University, about the new supercomputer and what it means for researchers.

Why did you choose to focus on Arm as a technology for HPC?

The main idea behind all the Isambard systems was to try Arm and see whether it was a genuine alternative to x86 for general-purpose HPC.

We started in 2016 with Isambard 1, which first went online in May 2018. That was the first production Arm-based supercomputer in the world. And it’s still running today. In fact, we’re only turning it off next month, so it’s run all that time. Amazingly, it was upgraded in 2020. We doubled its size, and it’s based on a Cray/HPE system that utilises Cavium’s Thunder x2 processors. We genuinely didn’t

Gorodenkoff/shutterstock

Is Your Data AI-Ready?

Insights depend on rich data

Don’t reduce data to text and images

- Natively standardize and normalize analytical data from >150 instrument vendor formats
- Automate data marshaling and processing
- Assemble data with experimental context and relevant metadata
- Provide FAIR, on-demand data to scientists
- Deliver AI-ready data for data science

Choose the Spectrus® Difference

30 YEARS

ACD/Labs

Learn more: www.acdlabs.com/spectrus

info@acdlabs.com ☎ 1-800-304-3988 (Toll free, USA & Canada)

+44 (0) 1344 668030 (UK)

know whether it would work, but it has worked brilliantly. In the meantime, all the hyperscalers have adopted Arm to build their servers. You've got the Amazon Graviton ARM-based stuff, we now know Google's doing something similar (Google Axion Processors), which they've announced, and Microsoft's doing something similar (Arm Neoverse Genesis CSS), so they're all going that way. Nvidia has adopted it for their CPUs as well. None of that was certain when we started, but it all worked out well.

We've gone from 32-core CPUs in Isambard 1 to 72-core CPUs in Isambard 3. The clock speed's gone up. The performance of Isambard 3 is roughly six times what Isambard 2 is, so we multiply the performance by a factor of six. Nearly half of that improvement has come from the system being twice as big, twice as many cores, but it's about the same number of nodes. So, node-for-node performance has increased by about a factor of six in the six years between the systems.

The storage has gotten much faster, and the network speed has increased from about 55 gigabits to 200 gigabits. However, the power use has only gone up a bit, so the energy efficiency has improved by about a factor of six. That's quite a significant improvement in six years.

Professor Simon McIntosh-Smith

How has the Isambard project developed over the years?

I'd say that the Arm-based systems were surprisingly good for HPC, even early on, because lots of work had already been done on various European projects that got a lot of groundwork done in advance of the production systems such as Isambard.

Early on, it was better than we expected, and for most HPC codes, which were open source, you could just recompile and run them – they would just work. In the meantime, more third-party software has been ported over to support Arm. Support for Arm as a first-class peer alongside x86 has increased, so you don't have to build everything from source anymore. For a lot of codes, you can find pre-built binaries, just like you can on x86.

Maybe the biggest change is that some third-party ISV codes also support Arm. So you're getting more and more Arm support in not just code you build yourself, but codes that you buy from someone else.

The other thing that helped with that was Apple shifting over to Arm. All the Macbooks have been ARM-based for four years, and that's helped because people run a lot of software on their Macs. That means all that software also has to work on Arm. That is probably the biggest thing that's happened in the past five or six years. Nvidia's adoption of Arm will also have a significant impact, but that's mostly to come because Grace Hopper systems are only just starting to come out, so the impact of those will be over the coming years.

People tend to be reluctant to port software. It takes a lot of work, so they want the market proven before they go to the trouble. So they don't tend to do it in advance.

Why did you choose to build these systems in a modular data centre?

The keyword in the name is modular. I would probably describe it as being like a big Lego set. As long as you have space and power, you can keep plugging more of these racks in when needed. As and when you've got funding or as and when you've got the power – it's not an all-or-nothing decision.

Of course, you can change things over time, and each module in the data centre can be fitted out quite differently. For example, you could have one module data centre full of air-cooled equipment and another one full of direct liquid-cooled equipment from a different supplier and sit them alongside, giving you a lot of flexibility.

This is the first time we've done this in Bristol, but we love it. We'll do everything this way in the future. We won't go back now. This is it. This is the way forward for us!

Thinking about the main benefits of building modular data centres over traditional 'bricks and mortar'. What would you say that they are?

It's a lot more flexible. If you build one big brick-and-mortar building, then it's hard to change what you do with that. Another benefit is cost. This is a very cost-effective way to build data centres. It is much cheaper to do this than to construct buildings. Another is speed. Those are all the things you worry about when doing this sort of stuff. How quickly can you do it? What's it going to cost, and how flexible are they? You're going to get them right, or it's something you're going to have to do an expensive retrofit at some point. You don't have those problems with modular data centres.

People probably aren't too worried about this yet, but they should get more and more worried about it. The embedded carbon in a modular data centre is a lot lower because they're just steel boxes the size that you need them, as opposed to a giant brick-and-mortar building, which is probably a lot bigger than you need, but you only get to build it once, so you want it big enough. The estimates are that a modular data centre contains at least 40% lower embedded carbon in the total solution than a brick-and-mortar building, and that's pretty substantial. **B**

Simon McIntosh-Smith is a Professor in High Performance Computing at the University of Bristol

Digitalisation helps streamline analytical lab data at Pfizer

Baljit Bains, Marketing Communications Specialist at ACD/Labs, reveals how Pfizer's PSSM team tackled inefficient data processing by adopting Spectrus software, standardising data management, enhancing collaboration, and enabling AI integration

Inefficient and non-standardised data processing, analysis, and reporting

Navigating the complexities of laboratories equipped with various instruments and diverse software packages is a common challenge in R&D organisations. Like many others, the team of process chemists in Pfizer's Pharmaceutical Science's Small Molecule (PSSM) team faced this problem. They had a variety of instruments and processed their LC/UV/MS data using multiple software packages. This resulted in scientists having to learn multiple applications with different interfaces. Equally problematic was data inconsistency, access, and lack of context (metadata, structures, etc.), the data was not findable, accessible, interoperable, and reusable (FAIR).

Implementing digitalisation for more efficient data integration and management

Most pharmaceutical organisations, including Pfizer, have already taken steps to optimise their workflows by adopting computerised tools to convert their manual processes and operations into digital formats (including digitisation of documents, automation of workflows, etc.). This not only minimises laborious, error-prone manual completion of tasks but also enhances data accessibility and sharing.

The PSSM team at Pfizer required consistent and standardised data that was easily accessible and shareable. To effectively meet these requirements, as well as the data management and efficiency goals of the team, a workflow using ACD/Labs Spectrus platform software was introduced. The Spectrus Platform allows for scattered data to be collected and integrated from all major LC/MS instrument sources into a single normalised format and enables the creation of workflows for semi- or fully-automated processing, analysis and customised data reporting.

Knowledge management capabilities ensure that this normalised data is stored with its chemical context for use by scientists and data science projects. Leveraging Spectrus software empowers users to harmonise and manage data more effectively, enables accessibility to data, and allows informed data-driven decisions to be made faster.

Harmonise data for easier retrieval

Spectrus provides a vendor-neutral and standardised solution to capture, unify, store, and access FAIR analytical data. It allows multiple file formats and data types to be harmonised into a single standardised data format and all analytical analysis and processing, no matter the technique the PSSM team use (NMR, LC/MS, and PXRD), can be done in a single interface, eliminating data silos and allowing data continuity.

Make informed data-driven decisions faster with better data management

With the new workflow, the PSSM scientists can store their homogenised data in centralised, easily accessible, chemically intelligent databases. These strategically curated databases of live analytical and chemical information can be easily searched by various parameters (structure, spectra, text-based queries) and ensure that scientists have access to the right data at the right time – empowering them to make insightful, confident, and informed decisions.

Gorodenkoff/Shutterstock.com

Navigating the complexities of laboratories can be a common challenge in R&D organisations

Easier collaboration

Good communication and collaboration across teams, facilities, and partners is essential for an efficient and effective workflow. Spectrus software digitalizes analytical and chemical data – ensuring it is consistent, reproducible, and readily accessible in real time. With it, the PSSM team can create customised reports and add data to their ELN, enabling them to quickly share their data with the Analytical Research and Development (ARD) department, with whom they work closely.

Automation for improved efficiency

Workflow automation reduces repetitive work, significantly reducing the scientist's relative time burden, freeing up valuable chemist time, and increasing the team's productivity and efficiency. Previously, they manually processed each sample, which would take an expert analyst about 5–10 minutes per sample. Using the tools available in the Spectrus-enabled workflow, this processing could be done in 1–2 minutes, reducing data processing time by more than 50% and saving scientists hours per week.

Enabling integration into artificial intelligence (AI) and machine learning (ML) initiatives

Spectrus prepares data for machine use by homogenising and storing it with its chemical context in curated databases. Export of this data in machine-readable format (i.e., JSON, XML) helps further prepare it for AI and ML applications. AI can very quickly analyse masses of data to identify anomalies and create simulation programs to accelerate the prediction of experiments.

Returns from FAIR, digitalised data

Expert data analysis and management tools, such as Spectrus, help support the digital transformation journey by delivering FAIR data to R&D organisations. Implementation of the Spectrus-enabled workflow provided transformative results for the PSSM team at Pfizer, boosting efficiency and accuracy of data collection, making collaboration easier, and allowing scientists to preserve and leverage their knowledge. Automated normalisation and assembly of data also means it will be available to machine learning and AI projects in the future. **B**

For more information, please visit: www.acdlabs.com

Putting energy efficiency at the heart of HPC strategy

The Wellcome Sanger Institute's **Nathan Cunningham** discusses the importance of sustainability in creating a research infrastructure by balancing HPC usage with energy needs

As data usage increases globally, the institute aims to reduce its energy consumption while maintaining computing power. With global research organisations generating and consuming ever-increasing volumes of data, the Wellcome Sanger Institute has acknowledged the need to maximise compute capacity and speed while, in parallel, minimising or offsetting energy use.

Achieving this means ensuring each of the thousand-plus researchers within the organisation understand and take on accountability for the energy implications of creating, storing, analysing and sharing the data that their research both creates and consumes. Opened in 1994, the existing Wellcome Genome Campus at Hinxton, Cambridgeshire, was constructed around the newly established Sanger Centre – now the Wellcome Sanger Institute – and, today, is also home to the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) and the BioData Innovation Centre.

A major site expansion was announced in May 2023, including an electrical grid connected to a substation that will support renewable generation on site. The ultimate aim is for the development to be net-zero carbon. Nathan Cunningham leads the design, delivery and management of Sanger's digital infrastructure and is a central figure in strategising and implementing the organisation's drive to create a scalable but sustainable supercomputing infrastructure. The stated concept marries the need to maximise capacity and speed with the drive to improve energy efficiency and achieve a zero carbon footprint.

"Within the past year, we began transferring the Sanger critical disaster recovery services to Crown Hosting Data Centres, a joint venture between the UK Cabinet Office and Ark Data Centres dedicated to serving the public sector, so it is ideally set up to support Sanger," he told *Scientific Computing World*. "In essence, every data user, from the bench scientist upwards, must be aware of the energy cost of their research and compute requirements. Energy sustainability is built into the Crown Arc ethos; for example, with the use of backup generators that harness biomass-recovered energy, which is as good as it gets.

"Ultimately, the Wellcome Sanger Institute aims to achieve a PUE (power usage effectiveness) of 1.05. In its simplest terms, the goal is to minimise energy loss and waste – I'd say that heating and cooling are perhaps prime culprits – to improve PUE as much as possible. And, yes, achieving that will mean investing in state-of-the-art, liquid-cooled computing systems that can run at 100% capacity."

What is Sanger's strategy for reducing energy usage?

So, you get 1MW of computing power out for every 1.05MW of energy put in. In reality, Crown Arc has achieved a PUE of about 1.15, so we still put in 15% more energy than you get out. But that's about the best you can buy commercially.

But in parallel with making our computing infrastructure as energy and operationally efficient as possible, we also have to consider the energy cost associated with each parcel of data that we generate, store, analyse and move into and out of that computing infrastructure.

Data is not 'free' from an energy perspective. I estimate that I spend about £4.5m a year on computing power and am effectively responsible for perhaps 25% or more of the energy budget for the current site. We are all individually accountable for that energy use.

Today, I believe it's incumbent on every individual, from the point of experimental design up to paper publication, to consider how they can achieve their objective in the most energy-efficient way.

We need to build that into the culture of science and to do that, we really need to educate people and provide support at the most basic level.

'It's important to optimise the HPC methodologies... poor code can represent a significant bottleneck and energy drain'

Cunningham believes we need a culture change such that energy use and accountability are key considerations when designing an experimental or data analysis workflow: "We need to educate at every level," he told us. "For a start, we can signpost individuals to the most appropriate existing online resources and develop new resources so that users can understand the implications of their activities from an energy perspective and put best practice into design and execution."

Can HPC be a sustainable technology?

It's important to optimise the high-performance computing methodologies that we use to help further reduce energy consumption. Education and the right tools for the front-end users are important. This is a recurring and enduring theme.

So, even at the level of basic coding, we'll find that poor code can represent a significant bottleneck and energy drain. Poor code offers no sustainability. Sanger encourages partnerships between the scientists, the research software engineers and other data specialists to make sure that they ask the right questions, and know where to find or how to derive the best data to answer those questions. It's about developing an effective 'pipeline'.

What we try to do at the Sanger is make sure that the back-end support is there for the scientists so that they can make informed decisions, input the right code, and generate – most efficiently – the data that can be utilised and answer those questions. I'd suggest it's a bit like thinking you own all the music in Spotify. You don't. What you have is access to what you want to hear through your curation. It's a subset; you can curate from the corpus of all relevant data to generate the dataset you will use to answer questions.

How does the growing volume of data processed and analysed create challenges for sustainability?

In fact, for each pipeline, we work a lot on optimisation of that long tail back of data creation and flow right back to the instruments themselves. Ultimately, then, we want to be able to say that a published paper resulting from that research and data pipeline – in terms of compute required – has been derived as efficiently as possible because all the knowledge and data used in the creation of those results, analyses and conclusions, has been considered from the perspective of need and energy. However, we have to do this to maintain the effectiveness of that research and the data that results.

Nathan Cunningham

It's not about how much data you have – in fact, you may not need a lot – it's about how you get to the endpoint using your data most efficiently. Training programs are critical when building an ecosystem of energy-related best practices. We can start with easily accessible platforms such as LinkedIn for professional development and O'Reilly. GitHub is also really big for us. We can then invest in follow-on or intermediate programs and use worked examples or tutorials to give the researchers the insight and knowledge to go back to their labs and work effectively to develop and utilise their algorithms, improving efficiency. As we said earlier, if you give people signposts and you give them the tools, then very few would object to being shown how to work more efficiently. That will then be able to drive a cultural change.

How do you decide where and how to store data most efficiently?

Many cloud environments don't support the level of high-performance computing that we require, and the public cloud is still very expensive. It's why the MET Office and the military have their own data centres. This ownership gives the Sanger as an organisation the ability to scale and prioritise how it stores data in an energy-efficient way, according to how it will be accessed, analysed, and transferred into and out of the organisation.

Sanger led the Covid response in the UK precisely because of this ability to scale. We carried out perhaps 20% of all the sequencing for the government. Software research engineers were integral to that drive to most effectively devise how to answer the questions using the data available, normalise that data and disseminate that data. That's also important for 'greening' that data, so the data used is all relevant and can be disseminated across organisations, picked up and utilised quickly. In effect, this also minimises data – and hence energy – wastage. **B**

Nathan Cunningham is the Associate Director of Platform Solutions at the Wellcome Sanger Institute

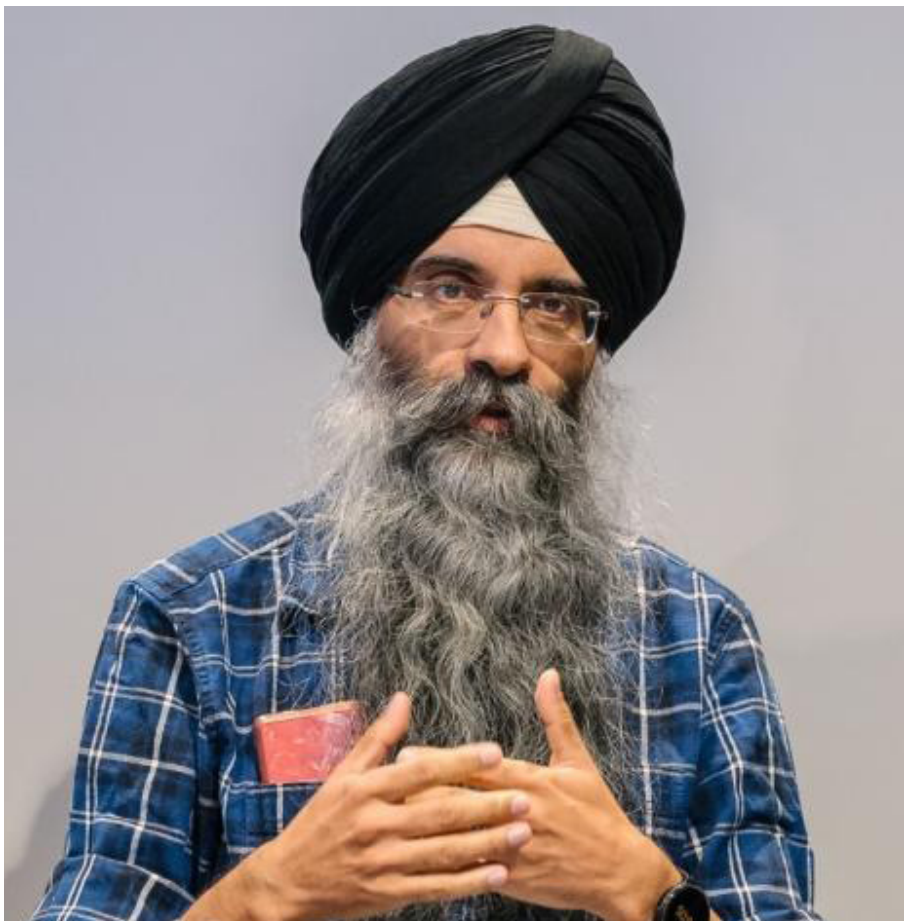
How causal AI can help scientists to build better cancer treatments

Raminderpal Singh explains how causal AI can lead to better patient interventions

Incubate.bio's founder, Raminderpal Singh, discusses the use of causal AI to help scientists better understand the mechanism for treating cancer and how this knowledge leads to better patient interventions.

There are several cancer drugs in development that target DNA damage response (DDR), following the early successes of drugs such as olaparib. Pinpointing appropriate biomarkers to identify patients who will respond to treatment and identifying combination therapies that improve efficacy without increasing toxicity presents an opportunity to optimise the development of new drugs that can better serve patients with a reduced chance of severe side effects. While identifying biomarkers could aid in overcoming these challenges, the existing methods primarily generate lists of potential genes and proteins that display changes in cancer patients without exposing the underlying and often critical mechanisms that underpin resistance.

In the preprint paper *Signaling pathway evaluation of leading ATRi, PARPi and CDK7i cancer compounds targeting the DNA Damage Response using Causal Inference*, researchers used the Adaptable Large-Scale Causal Analysis (ALaSCA) software platform, which applies Pearlian Causal Inference (PCI) techniques to specifically transcriptomic, proteomic and phenotypic multi-omics data. ALaSCA quantifies the causal contributions of different biological pathways to an outcome, such as responsiveness to treatment. The researchers applied ALaSCA to transcriptomic data for several compounds related to three known inhibitor types targeting DDR proteins: an ATR, a CDK7, and several PARP inhibitors. The major research study aimed to use →



Raminderpal Singh

causal methods to evaluate biological signalling pathways and identify resistance mechanisms in breast, ovarian, and non-small cell lung cancer (NSCLC).

Can you tell our readers how your AI platform can benefit scientists working in drug discovery?

Forget the notion of AI and words like that. Let's just think about the biologist trying to solve a problem. They have a question. And their choice is to run some experiments. They can run the experiments in the lab. They can outsource it, but it is a slow, expensive process. They know that, and their management knows that. They're limited in how much they can do. There's a bandwidth problem, but there's also a solution space problem.

What computers do, in general, is allow you to go global and deep-dive almost for free, and at speed. We use cloud systems, such as ALaSCA, which is on the cloud. It is parallel. If you push a button, it'll run massively parallel systems, and you can run across different systems. Suppose you've got a scientist who has a hypothesis about an experiment, and they want to take that hypothesis to a lab. Before stepping into the lab, a scientist can model that hypothesis into a network diagram. And let's run it 1,000 times in different directions for far less cost and faster on a computer.

The scientist may run the lab experiment anyway, but they can first explore in silico. We enable this at scale, cost-effectively and very fast. This provides the scientist with lots of additional evidence very quickly. But, crucially, this approach also looks at spaces they wouldn't have thought of. Think of it as a suite of informative experiments, not just an application of AI.

I have personal experience with when IBM Research took Watson Genomics to market. I worked with several hospitals and saw the resistance to public literature and publications because of these issues. At Incubate Bio, I am cautious about not hiding from these problems and finding ways to get through.

How can AI developers encourage the adoption of AI technologies in biopharma?

The problem in life sciences, which is a bit different from other use cases, is there's a huge amount of resistance just to begin with. There was a huge amount of resistance to AI. AI came along, over-promised, and didn't deliver drug discovery, and then people became even more apathetic or cynical. And, now, what's happening is you arrive with another topic, which is that I can implement LLMs; it's a hard wall to climb.

The other problem is that big companies can spend a billion dollars or 100 million dollars on something. What about normal biotechs? It's a strange curve; you've got a few big and AI ones. But if you add those two buckets together, you may have a couple of hundred, but you still have a couple of thousand biotechs behind them. What are they doing? For them, this is weird; they might hire a bioinformatician to do some data processing, genome processing, or multi-omics processing. But that's about as far as they will go.

How do they profit from it or impact their biology without just throwing money at the problem or hiring loads of new staff? They don't. The confusion creates a fog; they listen to it, they know there's something there because he has all the buzz coming from the big companies, but they end up saying, "I don't know how to do this". So if you can't take baby steps, why are you going to leap?

Can you explain the DDR paper and how AI helps to shed light on the underlying mechanism?

That study combines data engineering, machine learning, causal AI, and an LLM, in this case, AskMaddy. It shows us that to use AI effectively, sometimes you need to stitch these things together. People come from different backgrounds. One person wants to talk about Gen AI, but the scientist doesn't see it that way. You need to stitch all these different-sounding tools together. These are a series of pathways, and the

'The scientist may run the lab experiment anyway, but they can first explore in silico. We enable this at scale, cost-effectively and very fast'

Gorodenkoff/shutterstock

reason they're essential to cancer is not all cancers but several cancers, including ovarian, lung, breast, pancreatic, colorectal and leukaemia. The point about DDR is the response to changes in specific pathways. The damage will change over time, almost like they evolve. That's what the paper is about: predicting causality within pathways.

Well, you look for that, and you're trying to find out because there will be hitting something, so there'll be hitting a specific protein with that, and you'll be working out where is it inhibiting, in which pathway is it creating the inhibition, which one is is not really that important, which pathway is not being used so much, and so on.

By understanding that, you can understand the dynamics in this protein pathway and, therefore, the dynamics in the biological mechanism. If I can understand that mechanism, then I can understand a little bit about what is or is not going well, or I can add something to it to make it even better. So, we're finding that there are opportunities to add an additional intervention to make the treatment more effective for patients. Once you can predict and feel about the mechanism of your primary intervention, you can add a second intervention, which goes after the weaknesses you predicted based on the mechanism of your primary intervention.

'There was a huge amount of resistance to AI. AI came along, over-promised, and didn't deliver drug discovery, and then people became even more apathetic or cynical...'

If I'm putting a drug in, and it seems to work fairly well, why is it [working] fairly well? Because it's missing things. If it's missing things, can I add another drug to hit those things? People get that. That's not a problem. But it's subtle. It's art; you're painting with brushes, and you're not looking at the absolute number but the relative effects of things. This is a human, and these are hidden systems in the body.

So, the paper took three studies, and the researchers went after them individually to find out 'What are we learning about the mechanism? What can one do about that?' Once you understand the mechanism, can another intervention increase efficacy? Of course, when looking at pathways, you must also be conscious of toxicity.

When people started this type of research 20-30 years ago, they started to use drugs at the top of these pathways because they were more effective. If I drug at the top of a pathway, it will hit all branches on the way down. So it's more effective. That is true, but it will also present more side effects. So, they started pushing at the bottom of the pathway – the logic being that if I put it at the bottom, I've got less effects because I'm now in the branches and missing the stuff further up. The point is there is always a balance of efficacy over toxicity; you find where you're trying to hit, and that's also part of the painting with brushes. What do I want to do? What kind of interventions make sense in this specific case? B

Raminderpal Singh is the CEO and founder of Incubate Bio. You can find out more about the company's work at www.incubate.bio

Reference:
https://www.biorxiv.org/content/10.1101/2024.02.15.580418v

COMPUTING INSIGHT UK 2024

Science and Technology Facilities Council

5-6 DECEMBER 2024
Manchester Central, UK
www.ukri.org/CIUK

Early Bird Fees From Just £100

Registration Now Open
www.ukri.org/CIUK

The theme for the conference this year is **"Catalysing Research"** with sub-themes including future technology trends, collaborative research, ethics in HPC, energy consumption and AI focus.

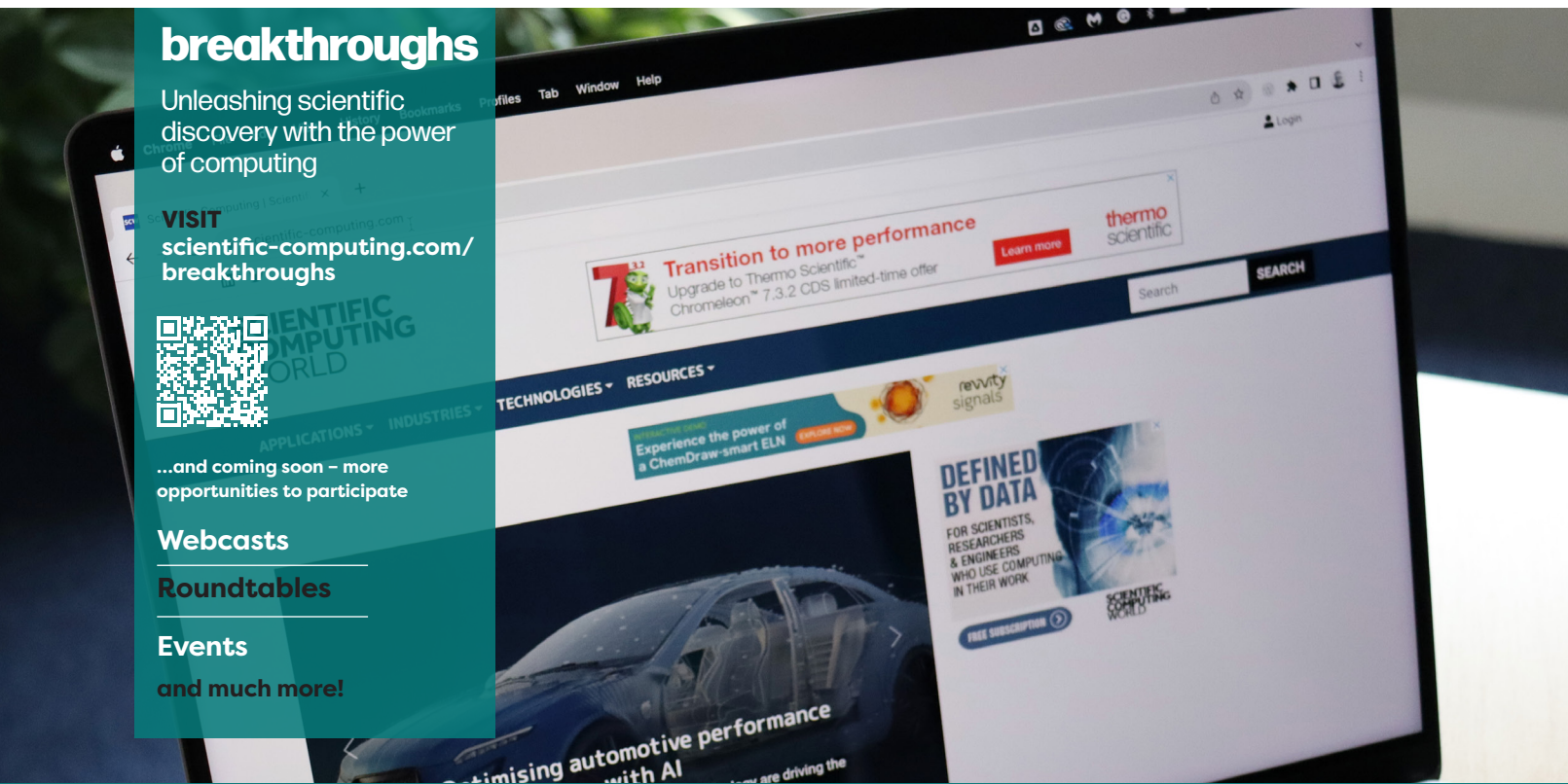
CIUK 2024 will include an exhibition of the latest hardware and software releases plus a two day programme of presentations and a series of parallel breakout sessions. There will be a student poster competition plus the fifth instalment of the CIUK Student Cluster Challenge. We will also present our annual Jacky Pallas Memorial Award to an early career researcher who has completed a high impact project.

Early bird fees available until 1 November.

CIUK 2024 - the UK's premier conference for HPC and associated science!

You can get involved!

Submit an application to our student poster competition or make a nomination for the CIUK Jacky Pallas Memorial Award. If you would like to get more involved in CIUK 2024 all details can be found on our website... www.ukri.org/CIUK.



Help translate technical benefits into tangible results for engineers, scientists and researchers, and help leaders equip their teams with the best available computing tools.

Follow Scientific Computing World on LinkedIn or register at www.scientific-computing.com to receive our emails to find out more

AUDEx: Pioneering hands-on automotive simulation for future engineers

Professor Dirk Engel discusses the implementation of continuous integration and continuous delivery processes in software development, particularly for autonomous driving and advanced driving-assisted systems

AUDEx, or Automotive Development in 1:x, uses realistic remote-control vehicles and a combination of simulation and sensing tools commonly used in industry to enable students to undertake modern automotive development tasks first-hand. The project aims to prepare students for work in industry by giving them access to the tools and training needed to work on modern automotive development.

The vehicles are fitted with cameras, microcontrollers, sensors, and actuators to help students develop and implement complex control algorithms. In addition, the vehicles can be controlled via a motion system, thus making it possible to assess how autonomous development impacts the driving experience for drivers and passengers.

As the project progressed, organisers saw the need to implement continuous integration and delivery (CI/CD) processes in software development. This is becoming important as software development becomes critical for automotive development – particularly for autonomous driving and advanced driving-assisted systems.

Scientific Computing World caught up with Professor Dirk Engel to discuss the AUDEx project and how it has evolved over the past 12 months.

Can you tell us how has the AUDEx project changed in the past 12 months?

I'd say the biggest and most important thing is that we had a look towards the Continuous Integration (CI) and Continuous Delivery (CD) process, which is a big deal right now, concerning software development and also the software development for autonomous driving, advanced driving-assisted systems. This is because these vehicles are increasingly related to software development.

We started by having a look at the available tools. What is the industry doing? What is affordable for universities? Accessibility is another key concern for us. Once we made a decision, we started to implement this whole process using the tools. We now have four students working on their final thesis, working on this topic, and they were able to install a process that suits our needs pretty well with all those tools that had to be handled from Git, Jenkins and Docker to install this process that was probably the largest change to the project.

Continuous integration (CI) is a software development practice that merges code changes into a central repository and then automatically builds and tests the software. CI focuses on preparing code for release, whereas CD involves actual code release. Git is a distributed version control system that tracks versions of files. It is used to control source code by developers working collaboratively. Jenkins is an open-source automation server that enables developers to build, test, and deploy their software reliably. Docker is a set of platform-as-a-service products that use OS-level virtualisation to deliver software in packages called containers. These tools allowed the students and the project managers to work collaboratively to build and test software before it is deployed to production and used to control the scaled vehicles and driving simulator.

Professor Dirk Engel

What was the existing process for managing software development?

Previously, some students used Git to manage their software. We didn't use Jenkins or Docker because we hadn't implemented →

a process like that before. There was no need to work with those tools to create pipelines to check for updates in the whole project. So, we decided to have some students try to do some of our software management on a very small basis.

Now that we have developed these internal processes, it will be mandatory for all students to get at least some experience with them. Because, if they have to bring their code to Git, they need to have a solid knowledge base of what is going on in this process and how they will have to tackle this process later on in industry.

Approximately 90% of our students will write their final thesis within the industry. So they're looking for an industry partner. They go there and will do their final thesis there. I have already supervised some final thesis and, sometimes, these organisations use more complex processes than the process we installed. However, MathWorks asked us if we could build something like a training model based on AUDEx because more and more customers are asking them 'What is CI? What is CD? What is continuous testing (CT)? How do you think I can use this in my process?'.

From my point of view, when I supervise those theses, I think these software development processes are standard, but probably it is the beginning of becoming a standard. Organisations will take a student and get them to explore and implement the methods, and the organisation will build up a picture of how to do things within their company. It's on the way to becoming the standard in the automotive industry.

Another key point is that the students can help bring these processes to smaller companies. Not all of our students go to the big OEMs. If they go to tier one or tier two companies, the students can bring some ideas to those companies because they are still working in this old-fashioned way.

metamorworks/shutterstock

'If you think about processor-in-the-loop (PiL), you can build something for the scale vehicles using the NVIDIA hardware and then scale that to a real car with real hardware'

How would you say that the hardware changed during the AUDExproject?

The second most significant change is that we started to install newer, more up-to-date hardware. The challenge is how to bring Nvidia hardware into those scaled models. The Nvidia Jetson Nano is the tool that we selected. We have already seen some ways and projects on how to use this hardware and scale it to the real hardware.

If you think about processor-in-the-loop (PiL), you can build something for the scale vehicles using the NVIDIA hardware and then scale that to a real car with real hardware. The hardware is comparable. You can start on the small-scale cars, and then use the general setup of the hardware, and of the processor, and you can scale it up so you don't have to reinvent everything. Because you already tested it on a small scale, like PiL testing, and then you bring it up to the one-to-one scale.

The Jetson Nano module is a small computer designed for embedded computing applications that provides the performance and power efficiency to take on modern workloads, including AI. It can be used to process data from several high-resolution sensors simultaneously and run multiple neural networks in parallel. This platform makes it possible to bring new capabilities to millions of small, power-efficient AI systems. I would say that it opens new worlds of embedded IoT applications, including entry-level Network Video Recorders (NVRs), home robots, intelligent gateways and scaled remote or autonomous vehicles.

PiL refers to testing and validating embedded software on the processor. The algorithms and functions are usually developed on a computer within a development environment. The resulting code is generally compiled with a special 'target' compiler for the processor, which is then used, in this context, in the vehicle's ECU. PiL tests are used to check whether the compiled code also works on the target processor.

Furthermore, we have fully implemented the driving simulator. We have this motion platform with vertical and lateral actuators. Initially, we had a problem as we could not implement all the hardware parts because they had different Software Development Kits (SDKs).

We created a solution based on the Unity engine as a middleware. We brought in Unity and now implement the steering wheel, pedals and actuators in real time. We can also integrate with a software tool like Carmaker from IPG, a software capable of real-time driving simulation. We use this for our mixed reality projects. We have images or video from the camera in a virtual environment. Our body-in-white colleagues create fully 3D virtual cars that we implement in the Unity engine.

Using Unity, you can have an overlay. If you use VR goggles, you can look around in your model vehicle or the general environment." B

Dirk Engel is a Professor at the Hamburg University of Applied Sciences

Advancing materials science research with AI

MaterialsZone is an R&D solution for advanced materials, chemicals, and other industries that helps scientists use AI to engage with their data, merging artificial intelligence with human expertise to unlock insights from complex datasets

In 2021, the AI cloud-based materials discovery platform MaterialsZone raised $6 million in Series A funding to develop its online materials discovery platform. MaterialsZone funnels R&D and manufacturing data into an interoperable and structured database, enabling users to achieve meaningful AI/ML insights, reduce R&D research times, and efficiently collaborate. The domain-agnostic platform can serve customers from different verticals, such as energy storage, renewables, and nano-materials, supporting simple and complex materials architectures.

Combining data within an organisation and using AI tools enables additional capabilities, such as multi-dimensional analysis, to make data-driven decisions and building models to predict experimental results.

Scientific Computing World caught up with Ori Yudilevich, the CPO of MaterialsZone, to discuss how the data-driven platform is helping scientists and how GenAI tools could further bolster the capabilities of this platform in the future.

Thanks for talking to us. Can you tell our audience about MaterialsZone and how the platform works? It seems like an interesting project gathering a lot of attention.

MaterialZone is a platform that addresses the material-based products industry. We work with companies that produce products, and these products involve some kind of research in material science.

A scientist or team of scientists develop some kind of new materials, mix different materials, or improve the processing of materials. It ranges from 3D printing to battery development, concrete, pharmaceuticals, and cosmetics. So these industries are very broad, but they have one thing in common: they design and create their products. They have an R&D team that does this research, and we're here to help these R&D departments, or R&D processes within companies, to accelerate their processes of discovering materials.

How does this improve existing processes?

Anyone who has been involved with AI in machine learning knows that it's about the data, and you can't get insights from your data without good quality data feeding the model. We see that as a significant challenge in materials science. I believe these industries are often working in ways that are relatively backwards. Many of them are still documenting their R&D process in Excel spreadsheets. They're scattered all over the place, so they don't learn from iteration to iteration or from product to product.

A researcher might do some experiments, then throw the spreadsheet somewhere, or that researcher moves on, and the next researcher comes in and doesn't know that this was done. MaterialsZone identified this problem many years ago, and we decided to build a data platform where scientists in these companies can aggregate all of the data related to R&D in one place, one single source of truth, or a unified database where all the data related to R&D is accessible, findable, and usable by tools that let you leverage it. And that's where AI comes in.

This data is not only about experiments. This data is data from the whole process within a company. So it's also procurement. You want to know what materials are available. You want to know the pricing because the prices of raw materials might influence your choice. You want to know regulatory information, and you want to know even information coming from sales sometimes because some input from sales might affect

DCstudio/shutterstock

Sigma-Aldrich®
Lab & Production Materials

MERCK

Discovery at your fingertips

Accelerate synthesis design with SYNTHIA™ Retrosynthesis Software

SYNTHIA™ Retrosynthesis Software allows to quickly discover novel pathways for new and published target molecules, as well as entire libraries of targets. The SYNTHIA™ software platform develops synthetic routes using expert-coded rules created based on proven chemical transformations.

One of the key advancements of SYNTHIA™ is its easy customization of search parameters to eliminate or promote reactions, reagents or classes of molecules. With a catalog of over 8 million commercially available starting materials and building blocks, SYNTHIA™ provides direct access to real-time local pricing and availability through our web shop.

SIMPLIFY ROUTE DESIGN

Quickly and efficiently scan hundreds of pathways to help you identify the best option according to your needs.

SAVE TIME AND REDUCE COSTS

Explore the most cost-effective routes to your target molecules with state of the art visualization and filtering options.

MEET PROJECT DEMANDS

Easily customize search parameters to eliminate or promote reactions, reagents or classes of molecules.

GENERATE NEW IDEAS AND IP

Explore unique and innovative syntheses that may be unknown for building your desired molecule.

QUICKLY BUILD A SHOPPING LIST

Easily generate a list of commercially available starting materials for your synthesis.

Pioneer Your Pathway at SynthiaOnline.com

ISO 27001 CERTIFIED by schellman

The secure cloud infrastructure is ISO 27001-certified and scalable for cost-efficient, seamless access.

MK_AD14028EN 58818 09/2024

→ the direction that you'll take when starting to develop a product.

MaterialsZone is a data platform. It's a platform where we aggregate all the data. We call it the Materials Zone Knowledge Center. On top of this knowledge centre, we have three pillars. We have the collaboration pillar, called the Collaboration Hub, which allows people to collaborate once all the data is there. You can upload your experiments or experiment internationally because some organisations will do some tests in one place and others in another. You can share insights with disparate teams to ensure data is not siloed. Then, we have the Insight module. We call that the Visual Analyzer. Once you have all the data, you want to get insights from it. You want to learn. You want to look for patterns in your data and correlations between different properties.

How is AI used in this platform?

Well, then, we reach AI. We want to use AI and machine learning to accelerate this process. Now, the main thing is that AI doesn't replace us because there's a lot of knowledge experts in their field have accumulated throughout their careers that the AI doesn't immediately have. What AI does better than human researchers is it's capable of looking at very large parameter spaces. A human can see that A affects B, but it's hard for them to see that some combination of A, B, C and D affect E and F in some way. For machine learning, this is very natural, very simple.

AI is designed to do that. It's not limited by brain power. MaterialsZone helps companies accelerate their development by reducing the number of experiments they want to do. Because the AI tool can look for patterns, multi-variable patterns, and help you pinpoint what experiments are worthwhile pursuing.

The platform can save up to 60, 70, or even 80% of the experiments to reach the product and go to market. That's reducing time, and it's also reducing resources because every experiment can be very costly. Typically, you do several experiments to reach a product, but that might take a few months because of expensive devices and materials.

What are the main challenges facing companies using materials science?

One problem we see repeatedly is the amount of knowledge and data out there, that is there for you to use, is immense and very hard to get to it. One of the big problems we see people coming back with is we need help in knowing which materials or which processes we should even try to attempt.

It's not like you have this simple recipe and just try a few things. You have a world of thousands or hundreds of thousands of possible raw materials, and after that, all the processing you can do and the use of different technologies create a complex challenge for the creation of new products. That information is scattered in suppliers' databases, academic literature, or patents. Having access to this data and being able to access it and link it with your workflows is a significant challenge.

Ori Yudilevich

Using AI means that, instead of being stuck trying out 10 materials I have purchased from certain suppliers, I can consider other raw material that's available from a supplier, maybe far away – or even nearby – that I don't know of. This material could make my product cheaper, make my product more sustainable, or make my product more fit for specific use cases. For example, if it's for food, you have different regulations limiting the type of materials you can use.

Can AI tools be innovative or help scientists to innovate in materials science?

That's a direction we're exploring right now. AI can be innovative, old-school AI and machine learning can be innovative to a certain degree. The limitation is that you programmed it. You've told it what its playground is. What the features should look like. You can make it much more extensive than what you would try if you were a human doing the same work. Because it has a lot of CPU power and can scan much larger spaces.

In that sense, it can be innovative, but it's not going to be innovative in the sense of discovering by chance because it's spilt something in the lab or these other amazing stories that we know from the history of science. It won't be as innovative as that, not at this stage of machine learning. However, with the Gen AI tools, you see more of this, and that's something we're exploring.

You can combine the existing data platform trained on your data with Gen AI. It's not that Gen AI is very creative, but it's trained on a much larger data set in such a way that we're amazed by it. This is because it seemingly knows everything. It has all this knowledge of humanity. So, AI is not very innovative by itself, but humanity is innovative.

Usually, as researchers, we don't have access to all of this knowledge, and we're working now on a tool that is a chatbot that the user can access from our platform, and it combines the power of large language models (LLMs), together with the data and the capabilities that MaterialZone has. Connecting the two enhances scientists' access to literature and available information beyond the scope of a company's internal data.

We see these types of applications in every domain these days. How do we now specialise this general tool to a specific domain and make it very good at a specific job, such as materials science? We want AI to guide the user and to provide them with the correct recommendations. In that sense, it is innovative and, with time, we'll see more and more innovation generated by AI. **B**

Ori Yudilevich is the CPO of MaterialsZone

Computer-assisted synthetic route optimisation using SYNTHIA retrosynthesis software

How SYNTHIA retrosynthesis software is revolutionising organic chemistry by optimising synthetic routes, improving efficiency, and reducing costs

Various challenges and considerations exist when designing an efficient process for synthesising a target compound depending upon the scope of the project and industry.

SYNTHIA retrosynthesis software is a powerful tool for organic chemists that employs expert coded chemistry rules to predict feasible routes starting from commercially available raw materials. Using SYNTHIA to aid in the route design process allows a chemist to quickly select pathways that meet key criteria of the project, including using commercial starting materials within a set price range or avoiding certain hazardous reagents or reaction types.

In addition to the expert-coded rules, SYNTHIA utilises the published literature to incorporate previously reported synthetic steps alongside the computer-generated steps. Because the algorithms are designed to optimise potential routes based on price, step count, and atom economy, it often leads to more efficient and cost-effective routes than a chemist might design on their own or than those which have already been reported in literature.

To demonstrate the practical utility of incorporating SYNTHIA retrosynthesis software into the route planning process, this case study highlights the improved synthesis of a known compound that was designed with the help of SYNTHIA to reduce step count and increase overall yield. An evaluation regarding the Green Chemistry aspects of proposed routes was also performed using the DOZN tool.

Case study – lithium chromoionophore

Dyes have been used throughout human history to colour textiles. Early dyes consisted of natural products derived from insects, bark, leaves, berries, and fungi. The synthetic dye industry was born in 1856 when an 18-year-old, William Henry Perkin, serendipitously discovered mauveine during a failed attempt at synthesising quinine.[1] Since the commercialisation of the first aniline dye, originally named 'aniline purple' for its rich purple hue, the dye industry has expanded beyond colouring textiles to include applications in photography and optical information recording media.

One such purple-blue coloured molecule with potential applications in optical information recording media is the indoaniline dye shown in Figure 1. This molecule also carries a mono aza-crown ether moiety, making it a chromoionophore that can selectively complex ions to cause a change in absorption and emission properties. Such ion-sensors have potential applications in trace metal detection in biological systems, as well as for molecular data processing.[2]

The synthesis of 1 was originally reported

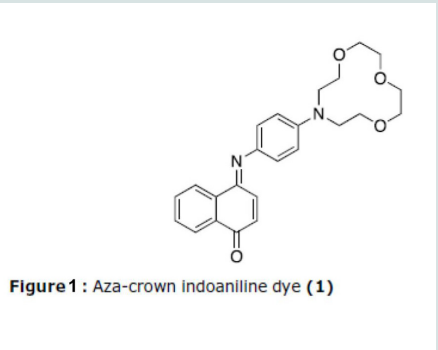


Figure 1 : Aza-crown indoaniline dye **(1)**

in 2000 by S.-H. Kim et al. (Scheme 1). The route starts with a series of functional group conversions from tetraethylene glycol 3, followed by a low-yielding intramolecular cyclization with aniline to generate the aza-crown ether 6. This is functionalized with HNO2 to generate the nitroso intermediate 7, which is subsequently reduced to amine 8 with Zn/HCl. The final step consists of a condensation with alpha-naphthol in basic solution at room temperature with air oxidation. This six-step process produced 1 with a 5% overall yield.

When tasked with preparing this molecule, the goal was to find a shorter route to improve the overall yield and reduce the synthesis time. SYNTHIA identified a commercially available aza-crown ether that could be directly coupled with p-fluoronitrobenzene in the first step, eliminating the need for the first two functional group transformations and the low yielding cyclization. This reduced the overall cycle time for the steps to core structure 7 from almost nine days to one day, after which nitro intermediate 9 was isolated in 93% yield after purification via silica-gel chromatography.

'Using SYNTHIA to identify a key first step from commercially available starting materials allowed us to develop a three-step synthesis to 1 with a 13% overall yield'

In the second step, SYNTHIA recommended reduction of the nitro group, which was performed via catalytic hydrogenation with activated Pd on carbon. After filtration of the reaction mixture through celite and evaporation of the solvent, crude 8 was obtained as a pale yellow liquid and used immediately without further purification. Although not proposed by SYNTHIA, the last step was adapted from the original publication. The condensation of 8 with alpha-naphthol in the presence of hydrogen peroxide (H2O2) produced 1 in 20% yield as a purple-blue solid after purification via neutral alumina chromatography.

Using SYNTHIA to identify a key first step from commercially available starting materials allowed us to develop a 3-step synthesis to 1 with a 13% overall yield. This represented a 260% overall yield improvement and a 50% step reduction over the original synthesis. Additionally, the labour costs were reduced by 60%, which led to an overall savings of 49% compared to the original route, assuming a labour rate of $150/hr.

For more information, please visit www.synthiaonline.com

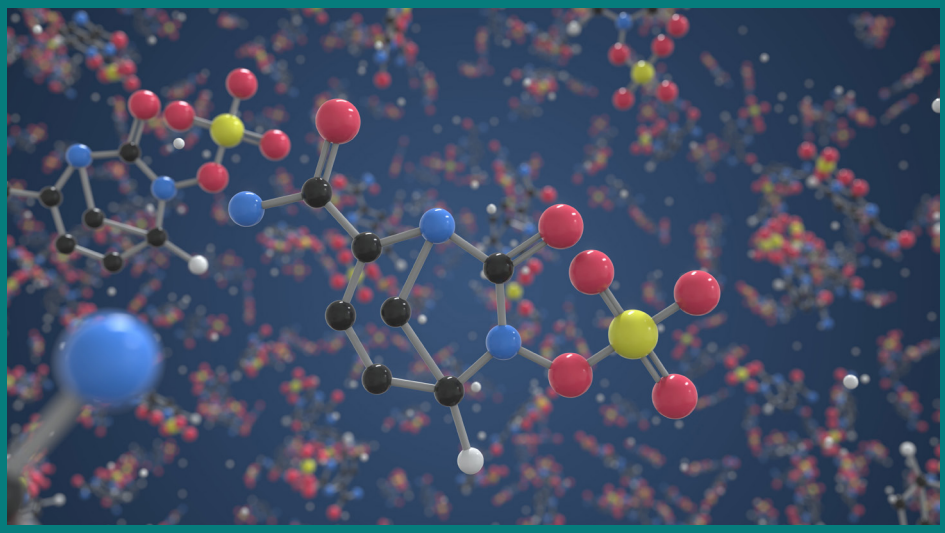
References

[1] *Perkin, W.H. The origin of the coal-tar colour industry, and the contributions of Hofmann and his pupils. J. Chem Soc. 1896, 69, 596-637. https://doi.org/10.1039/CT8966900596*

[2] *Kim, S-H; Kim, J-W; Han, S-K; Koh, K-N; Park, S-W; Heo, N-H. Synthesis and X-ray structural characterization of indoaniline dye carrying a mono azacrown moiety. Dyes and Pigments, 2000, 46 (1), 43-48. https://doi.org/10.1016/S0143-7208(00)00027-9*

Revolutionising drug discovery

How Standigm and Synthia software accelerate pharmaceutical innovation



With the increasing use of artificial intelligence (AI), particularly in the identification of drug targets, there is a growing interest in digital solutions that empower chemists to innovate and synthesise new molecules.

Standigm uses Synthia Retrosynthesis Software from Merck KGaA, Darmstadt, Germany, to accelerate its drug design and synthesis processes driving advancements in pharmaceutical development. Standigm has developed a leading platform, in partnership with Merck and its Synthia Retrosynthesis Software, to streamline the development of pharmaceuticals.

"There are many challenges in molecule pathway development, such as thousands of possible reactions, issues with reproducibility and speed of designing reasonable pathways," according to Ewa Gajewska, Head of Product Management, Synthia.

"Within the Synthia Retrosynthesis Software suite, there are a number of tools that can help in drug discovery, including Synthia WebApp," says Gajewska. "This can be used for path planning for single targets, automatic retrosynthetic analysis for entire target libraries (through our Shared Path Library), and target library generation (through our Diversity Library).

"In addition, Synthia API can help with prediction of synthetic accessibility scores, with results integrated into a customer's molecule discovery solutions.

"Synthia is built on several pillars, including expert-coded rules (programmed and verified by PhD-level scientists), algorithms responsible for chemical logic as well as searching and strategising algorithms, which helps identify pathways.

"There are also limitless customisations available to users, according to the demands of the particular project, such as avoiding certain reagents or toxic molecules."

Gajewska explains that the results can be presented in three different ways: as building blocks, as bond disconnections, or as pathways.

"The Shared Path Library allows users to prioritise routes with shared synthetic steps. Users can upload a library of target molecules they want to synthesise and Synthia will identify the synthetic pathways that have the highest number of common intermediates and starting materials, prioritising the pathways with late-stage functionalisation," she says.

"The Diversity Library allows for the generation of a target library based on a synthetic pathway. The user selects the starting materials that they want to diversify.

"Then, Synthia searches for target structure variations based on its huge database of commercially available starting materials and the entire power of its reaction rules and chemical logic algorithms. This ensures that generated molecules will be chemically correct and synthesisable."

The Synthia API comprises two main components: the Synthetic Accessibility Score (SAS) API and Full Retrosynthesis API.

"The SAS predicts how easy it will be to make a molecule based on the molecular structure, with calculation times as fast as 25ms per molecule or up to 100,000 molecules per hour," says Gajewska.

"The Full Retrosynthesis API allows users to connect results from Synthia WebApp with users' own molecular design software.

"Synthia helps users in multiple ways: it reduces the risk of overlooking a valuable pathway; it shortens the time of finding the best synthetic route; it finds pathways best suited to a particular user's needs; it can optimise synthesis globally for multiple targets; it can generate novel target structures; it can quickly filter molecule sets based on synthesisability; and it can connect with other tools using the API."

Standigm, a Synthia customer, uses AI to accelerate the drug discovery from target identification to market-ready therapeutics, operating both an AI Research Centre and a Drug Discovery Centre. You Young Song, who is Standigm's principal AI scientist, says: "We use two major AI platforms that we have developed: ASK, for novel target identification; and BEST, for novel compound design.

"Our in-house drug discovery team uses ASK and their own knowledge to identify targets, before handing over to our AI team, who use the STELLA functionality within BEST to ensure the right molecules are identified for multi-parameter optimisation. We use the Synthia SAS API to filter out any molecules that are difficult or impossible to synthesise. The SAS API applies a score to the candidates within minutes, allowing for rapid decision-making. At this stage, we reduce the candidate pool to around 1 to 2,000 molecules.

"After generation, we use Synthia Full Retro API to further filter and screen the molecules for synthesisability. This enables us to further reduce candidates to around 200, before they are reviewed by our team of medicinal and computational chemists in the Drug Discovery Centre. Using a combination of their expertise and the retrosynthesis path information contained within [the] Synthia web app, they decide which molecules are most suitable for synthesis. After the Drug Discovery Centre experiments with the design candidates, feedback is provided to the AI Research Centre, who use this to refine and design the next generation of molecules. By using Synthia, we have increased the number of molecules our computational chemists review by 40%, as well as improving the ratio of synthesisable molecules (to hard to synthesise ones) by 16%. It has also improved collaboration between the AI Research Centre and the Drug Discovery Centre." **B**

To find out more, please visit www.synthiaonline.com

Celebrating the people that drive our industry forward – the innovators, the boundary-pushers, the disruptors, the out-of-the-box thinkers and the R&D 'rock stars'.

Brought to you by Electro Optics

The PHOTONICS 100 2025

The industry's most innovative people

In association with ADMESY

www.electrooptics.com/thephotonics100

Making quantum computing more accessible and applicable to real-world challenges

Algorithmiq develops quantum algorithms that integrate classical computers with advanced quantum computers. Its quantum solutions are designed to push the boundaries of scientific discovery at an atomistic-scale resolution to solve life science problems that are currently deemed impossible...

One of the biggest challenges facing scientists and developers working with current quantum computing systems is noise – unwanted disturbances that affect the quality of the quantum computation. Noise mitigation strategies are crucial for improving the utility of near-term quantum devices. While these algorithms can course-correct for a certain period, they're prone to diminishing returns, particularly as the problem size and number of qubits increase.

Algorithmiq's Tensor Network Error Mitigation (TEM) method is a hybrid quantum-classical algorithm designed for performing noise mitigation entirely at the classical post-processing stage. TEM is also designed to integrate with error correction techniques to help extend the scale and accuracy of quantum simulations, the combination of which will become increasingly relevant, as quantum processors increasingly become more and more sophisticated.

Algorithmiq's error mitigation algorithm, TEM, is now commercially available on IBM's Qiskit Functions Catalog. Using the Qiskit Functions Catalog, developers can release Qiskit Functions that unlock capabilities for enterprise developers and quantum computational scientists.

Scientific Computing World interviewed Algorithmiq CEO Sabrina Maniscalco to discuss the use of quantum computing in life sciences and how IBM is opening up new opportunities for developers of quantum technology and making it easier for users to access quantum computing.

How does the recent expansion of the IBM Quantum Data Center impact global quantum computing capabilities?

IBM's expansion of its Quantum Data Center in Poughkeepsie marks a significant leap forward for global quantum computing capabilities. The centre now houses the world's largest concentration of utility-scale quantum computers, offering up to 16 times better performance, and is 25 times faster than IBM's 2022 quantum systems.

This expansion brings tremendous potential to the entire scientific community and enhances Algorithmiq's ability to refine our algorithms. It also intensifies the competitive landscape, which history has shown is one of the most significant drivers of innovation.

As IBM continues to push the boundaries of quantum hardware, it is creating new opportunities for algorithm discovery and practical applications; for example, in healthcare and life sciences, as well as in materials science. This progress is accelerating the global quantum ecosystem's growth, attracting top talent and bringing us closer to achieving quantum advantage in real-world scenarios.

What role does the IBM Quantum Cloud service play in making quantum computing more accessible to global clients and industries?

While there's been a great deal of promise and progress in quantum computing, we've only really begun to scratch the surface of unlocking the true potential for scientific discovery. IBM's Quantum Cloud service is a significant moving of the needle, and is offering clients and industries access to leading-edge quantum hardware. This cloud-based approach eliminates the need for organisations to invest in quantum infrastructure, which is significantly lowering the barrier to entry.

Additionally, when combined with IBM's Qiskit software stack, the Quantum Cloud service simplifies quantum programming. This accessibility is crucial for nurturing a global quantum software and services ecosystem. It's enabling startups and established companies alike to contribute to the field, accelerating the pace of discovery and innovation.

'While there's been a great deal of promise and progress in quantum computing, we've only begun to scratch the surface of unlocking the true potential for scientific discovery'

What is the significance of operating at utility-scale in quantum computing, and how does it distinguish quantum systems from classical computing?

Operating at utility-scale represents a crucial milestone in quantum computing. It means our quantum computers can now simulate certain systems as accurately as best-in-class classical methods, even outperforming brute-force classical simulations. This capability opens up a tremendous potential for computational exploration and algorithm development that was previously out of reach. The differentiator here lies in the →

quantum computer's ability to harness quantum mechanical phenomena like superposition and entanglement. At utility-scale, these quantum properties allow us to tackle more and more complex problems.

For Algorithmiq and the broader industry, utility-scale quantum computing signifies a transition from theoretical potential to practical applicability. Whilst it's not yet full quantum advantage, it enables us to explore use cases, on today's computers, which could bring new computational territories, potentially leading to breakthroughs in fields such as molecular simulation for chemistry and materials.

Algorithmiq has developed tools within the Qiskit Functions Catalog. How are these collaborations helping to address real-world challenges using quantum computing?

Algorithmiq's integration into the Qiskit Functions Catalog represents a significant step towards making quantum computing more accessible and applicable to real-world challenges. Our TEM has achieved levels of error mitigation never before witnessed without the need for additional quantum circuits. This optimised error mitigation with minimal use of the quantum hardware enables access to utility-scale quantum experiments.

Error mitigation is crucial in quantum computing, especially as we push towards utility-scale applications. Quantum systems are inherently noisy, and managing this noise is key to achieving reliable results. TEM helps address this challenge, enabling users to run more complex quantum algorithms more accurately.

Could you explain how Qiskit simplifies the process of programming quantum computers?

Qiskit has emerged as a leading tool in quantum computing, and it is known for its performance and user-friendly approach,

Sabrina Maniscalco

making quantum programming more accessible and intuitive.

For newcomers to quantum computing, Qiskit provides a gentle learning curve. I'd say its Python-based interface allows developers familiar with classical programming to transition more easily into the quantum realm. Much like the tools developers use daily, Qiskit offers a host of documentation and tutorials and it enables new users to grasp quantum concepts and start coding quickly.

For those already familiar with working in and around quantum computing, Qiskit also really allows for low-level control of quantum circuits while also offering a range of high-level abstractions for complex quantum operations. This versatility enables seasoned developers to really fine-tune their algorithms for optimal performance on IBM's quantum hardware.

How can quantum computers be used to solve challenges in life science or healthcare?

By simulating molecular interactions at the quantum level, quantum computers can accelerate drug discovery and development, enhance our understanding of diseases such as Alzheimer's, and pave the way for more personalised medicine. At Algorithmiq, we are working on integrating multiple approaches like AI and network medicine to reshape the future of healthcare and life sciences by developing quantum algorithms that unlock the power of quantum computers. Instead of the slow trial-and-error methods used in labs today, advanced computer models will rapidly analyse billions of options to discover the best medicines.

While many of these applications are still in the early stages, recent advancements in quantum hardware, such as IBM's utility-scale systems, bring us closer to practical implementation. **B**

Sabrina Maniscalco is CEO and co-founder of Algorithmiq, a professor of quantum information and logic at the University of Helsinki, and an Adjunct Professor at Aalto University, Finland

Fault-tolerant quantum computing 'will solve exciting science problems'

Quantinuum's **Dr Harry Buhrman** and **Russell Stutz** join us to discuss the technological innovation behind the company's development of fault-tolerant quantum computing systems

Quantinuum recently published an accelerated roadmap to universal, fault-tolerant quantum computing, which the company aims to achieve by 2030.

The updated roadmap is designed to accelerate the path to commercial quantum computing systems. It also unveiled the company's plans for its fifth-generation quantum computer, Apollo. Quantinuum says that this new system will be a fully fault-tolerant and universal quantum computer capable of executing circuits with millions of gates.

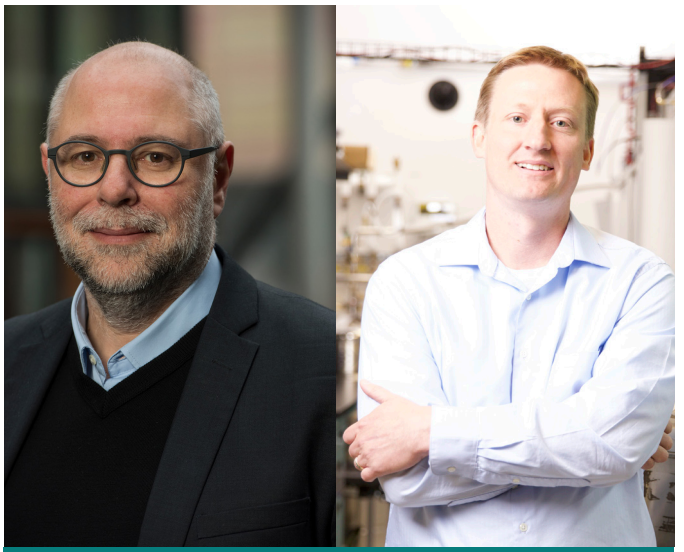
The roadmap is built on the foundations of its fully scalable quantum charge-coupled device (QCCD) architecture, including a universal gate set and high-fidelity physical qubits uniquely capable of supporting reliable logical qubits. For the past four years, Quantinuum has provided data and peer-reviewed papers to show the science and engineering work behind its methodical advances.

In a recently published pre-print research paper: *Scalable Multispecies Ion Transport in a Grid Based Surface-Electrode Trap*, the company outlined the scientific advancements that led to this accelerated technology development. *Scientific Computing World* spoke to Dr Harry Buhrman, Quantinuum's Chief Scientist for Algorithms and Innovation, and Russell Stutz, its DIrector of Commercial Hardware.

What specific advancements have enabled the acceleration of your hardware roadmap towards a commercial quantum advantage?

Harry Buhrman: Our new roadmap is an acceleration of what we were previously planning. The benefits of this are obvious: Apollo brings the commercial tipping point sooner than we previously thought possible. This acceleration is made possible by a set of recent breakthroughs.

First, we solved the 'wiring problem' – we demonstrated that trap chip control is scalable using our novel centre-to-left-right (C2LR) protocol and broadcasting shared control signals to multiple electrodes. This demonstration of qubit rearrangement in a 2D geometry marks the most advanced ion trap built, containing approximately 40 junctions. This trap was deployed to three different testbeds in two different cities and operated with two different collections of dual-ion-species, and all three cases were a success. These demonstrations showed that the footprint of the most complex parts of the trap control stay constant as the number of qubits scales up. This gives us the confidence that Sol, with approximately 100 junctions, will be a success.



Dr Harry Buhrman and Russell Stutz

Second, we continue to reduce our two-qubit physical gate errors. Today, H1 and H2 have two-qubit gate errors less than 1x10-3 across all pairs of qubits. This is the best in the industry and is a key ingredient in our record >2 million quantum volume. Our systems are the most benchmarked in the industry, and we stand by our data – making it all publicly available. Recently, we observed an 8x10-4 two-qubit gate error in our Helios development test stand in 137Ba+, and we've seen even better error rates in other testbeds. We are well on the path to meeting the 5x10-4 spec in Helios next year.

Russell Stutz: By leveraging our all-to-all connectivity and low error rates, we expect to enjoy significant efficiency gains in terms of fault tolerance, including single-shot error correction (which saves time)

and high-rate and high-distance quantum error correction (QEC) codes (which mean more logical qubits, with stronger error correction capabilities, can be made from a smaller number of physical qubits).

Studies of several efficient QEC codes already suggest that we can enjoy logical error rates that are much lower than our target of 10-6 – we may even reach 10-10, which enables exploration of even more complex problems of both industrial and scientific interest.

Error-correcting code exploration is only the beginning – we anticipate discoveries of even more efficient codes. As new codes are developed, Apollo can accommodate them, thanks to our flexible high-fidelity architecture. The bottom line is that Apollo promises fault-tolerant quantum advantage sooner, with fewer resources.

How does achieving logical error rates better than 10-6 help researchers and what are the implications of possibly reaching 10-10 in future systems?

Stutz: Apollo will achieve error rates of 10-6 and lower via quantum error correction (QEC). Reaching these error rates means that we can run circuits deep enough to solve problems of both industrial and scientific interest. This is enabled by starting with physical error rates at the 10-4 level, well below the 'threshold' where quantum error correction starts working. Our lower physical error rate makes it easier to reach a logical error rate of 10-6 and beyond. Reducing the logical error rate to 10-10 in future systems means you could run quantum circuits with order 1010 gates and still receive the correct answer with a reasonable probability. This opens up many more applications for quantum computers.

How do features such as all-to-all connectivity, mid-circuit measurement, and real-time classical co-compute improve

'Apollo will achieve error rates of 10-6 and lower. Reaching these error rates means that we can run circuits deep enough to solve problems of both industrial and scientific interest'

Apollo's performance and flexibility?

Stutz: Many error correction codes require connectivity beyond the basic nearest neighbour connectivity offered by other quantum computing architectures. Our ability to move our qubits around in space gives us all-to-all connectivity because we can bring any two qubits into contact with each other. With our all-to-all connectivity, we can explore many more QEC codes than those with limited connectivity.

Mid-circuit measurements enable you to perform quantum error correction during a circuit – so, without it, you can't perform fault-tolerant quantum computing. The ability to make real-time corrections is also a requirement for fault-tolerant quantum computing; thus, there is a need for real-time classical co-compute to determine the corrections to perform.

Beyond the ability to perform quantum error correction, these features also enable applications run on the physical layer (i.e. do not require logical qubits and QEC). Almost all applications require gating qubits that are not nearest neighbours, so in fixed qubit geometries logical 'swaps' are needed. Those drive up the number of entangling gates you must perform, reducing the probability of getting the correct answer. Our algorithm teams have devised clever ways to take advantage of these added features (mid-circuit measurement and real-time classical co-compute) for applications that are being run on the physical layer.

What role do quantum error correction (QEC) codes play in advancing towards fully fault-tolerant quantum computing, and what is 'single-shot error correction'?

Stutz: All fault tolerance is based on QEC, but not all QEC is fully fault-tolerant. We 'tolerate' faults, or errors, by detecting and correcting them. I would say QEC is a necessary but insufficient condition for creating a fault-tolerant computer.

Quantum error correction is essential for bringing the error rates lower than their physical limits. QEC is what allows for error rates of 10^-5 and lower. Low error rates mean that you can run deeper circuits – which means solving more complex problems. [Single-shot error correction:] For a fixed physical error rate, you can scale the logical (aka 'corrected') error rate down by increasing the 'size' of the code, i.e. by using more physical qubits per logical qubit. The most studied QEC code, the surface code, requires increased numbers of measurements (recall, mid-circuit measurement discussion above) as you increase the code size. Many quantum computing modalities favour the surface code since it only requires nearest-neighbour interactions. By relaxing this requirement and allowing codes with higher qubit connectivity, you can find QEC codes that only require a single round of measurements per QEC round, independent of the code size.

Our all-to-all connectivity enables us to run these efficient single-shot codes.

Because of our connectivity, low crosstalk mid-circuit measurement, qubit reuse, ability to do real-time corrections enabled by long qubit coherence times, and low physical error rates, our architecture allows for virtually any error correcting code to be run. This flexibility will be crucial as we scale up towards fault-tolerant quantum computing.

What scientific applications do you foresee Apollo driving forward, particularly in fields such as materials science, chemistry, and high-energy physics?

Stutz: Apollo will give us the ability to simulate complex quantum mechanical systems beyond the reach of classical computers, which will have important impacts on fundamental science. No one could have predicted the internet when classical computers were first developed. We don't know everything that quantum computers will be used for, but we are excited about the future potential of solving important problems.

How did you overcome the 'wiring problem' for scalable quantum systems?

Stutz: Full details can be found in this pre-print research paper. In summary, we were able to perform low heating 'junction transport' (meaning having qubits cross an intersection). We were able to do this across a grid of junctions. We co-wired different sections of the trap such that all corresponding electrodes in the different sections used a single analog voltage signal. We then devised a clever control system that minimised the signals required to arbitrarily rearrange the qubits (or ions) on the grid.

Both of these techniques combined allowed us to drastically reduce the number of physical wires in the system. This means we can increase the number of qubits and grow the size of the quantum computer while keeping the number of unique analogue electrical signals constant, or near constant. **B**

Dr Harry Buhrman is the Chief Scientist for Algorithms and Innovation at Quantinuum. Russell Stutz is the Director, Commercial Hardware, Quantinuum

Removing the biopharma bottlenecks from the drug discovery roadmap

Dr Mutlu Dogruel introduces a series of potential solutions to implement AI in the drug discovery landscape

How can AI transform drug discovery workflows?

The use of artificial intelligence (AI) is creating a paradigm shift in the drug discovery and optimisation processes, allowing for the analysis of large-scale datasets, the design of new molecules, and the prediction of drug properties more efficiently. This can unlock new opportunities for pharmaceutical and agricultural companies to develop innovative solutions for unmet clinical and agronomic needs.

AI-driven drug discovery is increasingly commonplace, with companies actively exploring hundreds of AI-driven candidates in the discovery and early clinical stages. However, no AI-discovered or designed drugs have yet reached the market. Gartner predicts that, by 2025, more than 30% of drugs and materials will be systematically discovered using generative AI (GenAI) techniques. Considering that no AI-discovered or designed drugs are currently on the market, this forecast represents a significant shift within a short timeframe. The biopharmaceutical industry is already making considerable progress toward this goal, with several biotech companies advancing AI-guided drug candidates to clinical trials.

Foundation models (FMs), including large language models (LLMs) and image-based models have the potential to revolutionise fields such as drug discovery and materials science. These models can process vast datasets, recognise patterns and generate valuable insights – capabilities essential in many scientific domains. To harness AI's full potential in drug discovery and deliver life-saving medicines to patients, the biopharmaceutical industry must overcome the challenges creating 'AI bottlenecks' in the discovery process.

LLMs are not limited to human languages – they can also be trained to understand and generate biological and chemical languages, such as amino acids, genes, ligands and proteins. This capability enables their use in various bioinformatics and cheminformatics applications, including predicting DNA regulatory elements and optimising drug properties.

How should AI be used to benefit an organisation?

Despite claims of improved speed and accuracy, AI-based methods may not always outperform traditional, well-established screening and docking techniques. In these cases, AI can still play a crucial role by orchestrating the execution of the right methods in the correct order.

Beyond revolutionising drug discovery, AI can also optimise internal processes. Integrating AI internally streamlines operations. Automating cumbersome or repetitive tasks with AI frees up time for employees to focus on high-value activities requiring creativity, expertise, and collaboration. Improvements in internal operational efficiency can fuel innovation in drug discovery, and integrating AI into every aspect of an organisation can drive future success. In traditional computational chemistry, quantum mechanics (QM), also known as 'ab initio', is the gold standard for atomic-level simulations, providing highly accurate system analysis based on fundamental physical laws. However, ab initio simulations are notoriously computationally expensive. Deep learning models are currently being developed to address this challenge using datasets of hundreds of thousands of molecules with pre-calculated properties. These models can generate QM-quality atomic charges for a molecule in just a fraction of a second. Technological breakthroughs such as this enable precise yet rapid calculations in classical (non-quantum) atomic simulations, which are essential for conformer generation, molecular dynamics, and free energy perturbation tasks.

AI offers the tools to examine complex diseases comprehensively. Unlike traditional approaches that often focus on one specific area (such as a key protein), AI can consider the broader disease pathway and leverage knowledge from the vast scientific literature. This capability allows AI to identify potential treatment strategies that consider multiple factors simultaneously, leading to the →

Dr Mutlu Dogruel

discovery of drug candidates that might not have been previously considered. I believe that AI can help develop drugs targeting complex and hard-to-treat diseases by providing researchers with valuable insights.

Leveraging AI's capabilities allows drug developers to enhance efficiency, accuracy, and patient-centricity in drug discovery, ultimately leading to improved treatments and healthcare outcomes. I would say the primary strategic benefit of AI lies in its ability to look at the big picture and focus on multiple aspects simultaneously. In the context of drug discovery, this means analysing various biological pathways simultaneously and integrating results with supporting datasets in a meta-analysis. However, we should recognise fully integrating AI into drug discovery remains a complex undertaking with many challenges.

Where do you think an organisation should aim to implement AI in its business?

Building on these principles, GraphRAG, developed by Microsoft Research, combines the strengths of RAG and knowledge graphs. By integrating knowledge graphs into the retrieval-augmented generation process, hybrid solutions such as GraphRAG enhance the AI model's ability to generate responses that are not only contextually accurate but also semantically rich.

Agentic frameworks can further improve handling AI hallucinations and complex problem-solving. By breaking down tasks and distributing them among multiple autonomous agents, these frameworks can enhance the accuracy and relevance of generated responses. Each agent can focus on specific aspects of a problem or validation process, ensuring a more thorough and adaptive approach. Cross-verification among agents helps to identify and correct hallucinations, while contextual anchoring maintains the coherence and accuracy of the outputs.

Additionally, adjusting parameters such as temperature and top-p sampling of each LLM further refines the generation process. Lowering the temperature reduces randomness, making responses more focused and coherent, while top-p sampling ensures that only the most probable and contextually relevant information is considered. Together with agentic frameworks and majority voting, these strategies enable LLMs to deliver more accurate and meaningful responses, improving their effectiveness in complex applications and reducing the likelihood of erroneous outputs.

Employing multiple LLMs with a majority voting approach can enhance output reliability by comparing responses across different models and selecting the most consistent or frequent answers. However, this method has its caveats: to avoid consensus bias, it is essential to use truly independent LLMs. They may exhibit similar biases or errors if the models are too similar (e.g., trained on the same datasets, architectures, or even using the same underlying framework). Additionally, I would say that majority voting is best suited for short, factual answers rather than complex reasoning tasks. While this technique helps filter out unreliable outputs and mitigates the impact of hallucinations, it should be supplemented with other validation methods to avoid developing a false sense of confidence in the results.

Finally, implementing post-processing algorithms to assess and filter AI outputs can further mitigate the effects of hallucinations. These algorithms can evaluate the coherence and relevance of generated content before it is used in practical applications.

Above all, implementing a scalable roadmap allows companies to successfully navigate the complexities of AI integration and fully benefit from the transformative power of these technologies. Now, let's consider the importance of having a robust AI strategy in place and the benefits it can bring to drug discovery and wider organisational goals.

How can an organisation build a roadmap to create a robust AI strategy?

Successfully implementing AI requires a well-thought-out strategy that takes into account drug discovery objectives and wider organisational goals. Researchers can take several steps to prepare for potential challenges and solve complex drug discovery problems with a robust AI strategy.

To maximise benefits, AI integration must span the entire drug discovery life cycle. AI can be leveraged to identify new drug targets, design and optimise drug molecules using a streamlined 'design, make, test, analyse' (DMTA) cycle and predict drug-target interactions. By adopting AI across all drug discovery stages, productivity is enhanced and researchers gain actionable DMTA insights quickly. I believe that this can help reduce the time and cost of drug discovery and development, improving the success rate of clinical trials.

A structured and methodical approach, as part of a broader organisational strategy, is crucial to ensuring seamless and effective AI integration in drug discovery. This involves meticulous planning, execution, and management of AI solutions to align with the organisation's overarching objectives.

I would say that a systematic approach enables thorough implementation across the entire organisation and ensures consistency in applying AI, maximising its benefits in driving drug discovery forward. To achieve this, ML and LLM operations must align with the overall digital transformation goals and the organisation's broader commercial objectives to create efficient and scalable AI solutions. Executing this AI roadmap can be challenging for some companies due to the complexity of integrating AI into existing workflows, the substantial resource and skill requirements, and the need for effective change management, but doing so can unlock AI's benefits and transform drug discovery and development.

'To harness AI's full potential, the biopharmaceutical industry must overcome the challenges creating 'AI bottlenecks' in the discovery process'

Incorporating AI throughout the drug discovery process may require a strategic and innovative partner who can leverage AI's full potential to enhance outcomes.

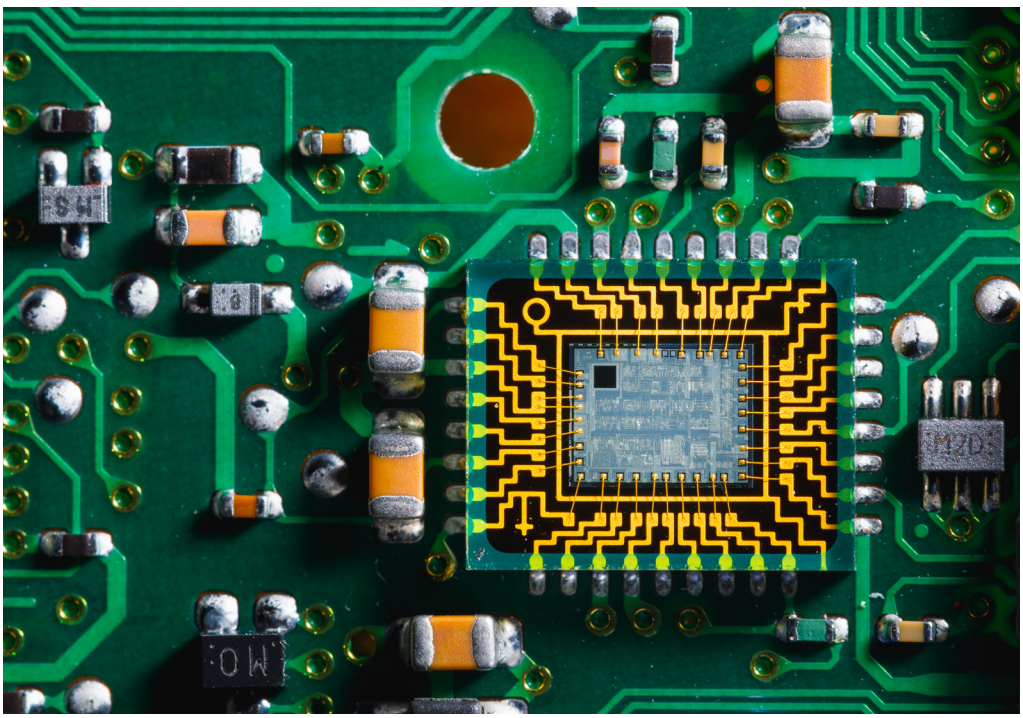
Developing the right strategic alliances can provide access to cutting-edge AI insights and help implement modular and scalable AI solutions.

Given the rapid evolution of AI and the need for continuous improvement, opting for a modular approach is essential to simplify AI integration and future-proof drug discovery. Employing modular architectures combined with cloud-based solutions and integrating ML and LLM operations provides the flexibility necessary to consistently advance AI-driven systems, allowing them to be seamlessly adapted and continuously improved. Embracing scalable AI empowers researchers to unlock new possibilities in drug discovery and look forward to a future where medicines are delivered to patients at an accelerated pace. **B**

Dr Mutlu Dogruel is the Vice President for AI Solutions at drug discovery specialist Cresset. You can find out more about its work by visiting www.cresset-group.com

Silicon photonics illuminates a path for future AI models

AI systems are straining traditional computing and memory infrastructure. **David Lazovsky** discusses how photonics could deliver a transformative leap



As artificial intelligence (AI) models grow at an unprecedented pace, traditional data centre infrastructure is strained to its limits. The explosive growth in AI model size is driving the need for a step function increase in memory bandwidth and capacity at low latencies. Celestial AI is a company that has been collaborating with some of the largest hyperscalers to develop a deep understanding of compute, memory and network system infrastructure chokepoints.

First, can you tell us about yourself and why your company was created?

Celestial AI is more than four years old. I founded the company in April of 2020 – we have just passed 120 employees and we're growing at about a 50% headcount this year. And I don't see that slowing down. My background is now approaching 30 years, with 28-plus years of semiconductor experience and experience in optoelectronic devices over the past 10 years roughly.

The market that we're addressing is AI, and the magnitude of growth is just absolutely unprecedented. It's like nothing that the world has ever seen. And not only is the size of this market immense, but the rate of growth is staggering.

Accelerated computing is not just about computing; it's not just about moving from CPUs to GPUs. There's a fundamental change to the requirements for artificial intelligence workloads that I'll describe, which impacts the processing infrastructure, the interconnectivity of processors, and the memory systems that support AI workloads.

What's interesting about this market is not just how big it is, and how rapidly it's growing, but how concentrated it is. Currently, four companies represent 70% of this market, these four hyper scalers. All of them are building their own accelerators. Some of them are further along, weaning off of Nvidia, or have never been reliant on Nvidia, such as Google. They started building their TPU in 2016. Amazon also did a good job; they acquired the Annapurna team in 2015 and have mitigated reliance on Nvidia; I think the analysts projected that both Amazon and Google invested about $1.5 billion buying NVIDIA GPUs last year, whereas Microsoft and Meta invested about $4.5 billion each buying NVIDIA products last year.

How does the growth of AI change the requirements for computing and memory infrastructure?

The AI memory wall is by far the most significant challenge facing hyperscalers and all companies that are reliant on advanced AI workloads. These models have been growing exponentially. Generative AI has got a lot of visibility right now. And it should because it has been the catalyst for people gaining an understanding of AI. What's been running in the background for companies such as Amazon, Pinterest, Netflix, and Meta's family of apps, including Instagram and Facebook, are AI recommendation models. There's a multi-trillion dollar market that 'recommenders' are serving globally. For companies such as Meta, that's a company that did about 140 billion in revenue last year – 99% of that was advertising. And the vast majority for those advertising dollars is this one algorithm – deep learning recommendation model (DLRM).

What's coming is a new wave of multimodal models, that is a combination of the modes of training, which will not just be text, or still imagery, but learn the way that humans learn, which is by using a variety

of sensory input, including, basically, streaming video. The limitation today is infrastructure that prevents the ability to train. As infrastructure becomes more advanced, those capabilities will evolve. The problem is that there is no amount of memory that you can possibly put in a single package to keep up with the evolution of these models. The models are currently growing two orders of magnitude faster than what you can possibly put in the package and it's not slowing down...

This growing demand for AI has resulted in a complete shift in what is limiting overall system performance for these workloads. It used to be that, just a few years ago, you could fit every model in existence comfortably inside of the memory capacity of a single processor. Today, you need thousands or, in some cases, tens of thousands of processors that are interconnected together to house a single model, even for inference. And then when you're training it, you need 10 times that.

What that's done is that it shifted the bottleneck for overall system performance, from compute to interconnectivity, and that interconnectivity is memory bandwidth; chip-to-chip interconnectivity bandwidth. Today, I would say that is where the battlefield lies in what drives all AI performance and efficiency.

Currently, the gold bandwidth standard for memory bandwidth is High Bandwidth Memory (HBM). HBM3E provides up to eight terabytes per second. The most advanced optical interconnectivity technology delivers two terabits per second of optical and interconnectivity bandwidth. And that two will go to four over roughly the next 12 to 18 months; that means they will go from 25% of HBM3 bandwidth to 50% of HBM3 bandwidth, which in either case doesn't really matter because it's not enough.

We have developed a technology called a photonic fabric, which provides not just HBM3 but HBM4 level bandwidth and provides the scalability of bandwidth that meets requirements for artificial intelligence workloads, generative AI and what will come beyond it.

'Accelerated computing is not just about computing; it's not just about moving from CPUs to GPUs'

For the first time in history, we have the ability to optically interconnect high bandwidth memory, which is not possible with existing technology. The result of that is we're creating an interconnectivity infrastructure that is complementary to Ethernet, which is currently used to connect racks of servers in data centres.

We have what is a technology, which is called optical compute interconnectivity, that provides the full bandwidth of memory to support a compute-to-compute data transfer. But it's not just about bandwidth, there are four requirements. It is bandwidth, cost – really, it's dollars per gigabit per second, latency because these applications translate directly to revenue for both recommenders and for generative AI. And the fourth metric that matters, and it will matter increasingly, is, of course, energy efficiency.

In our case, our photonic fabric interconnectivity is 1/10 The power of NV link, even for the current generation that's running in Hopper based systems such as the H100 H200 hundred, a remote direct memory access (RDMA) – that power consumption is roughly 62 pico joules per bit. Our RDMA is closer to six pico joules per bit so we're an order of magnitude lower cost from an energy standpoint than what is available state-of-the-art moving information over copper and Nvidia is definitely state-of-the-art, it has the highest bandwidth and uses the best interconnectivity technology in existence for processor-to-processor interconnectivity. That's the beauty of using light, right? You're moving information at the speed of light with zero resting mass photons that are far more efficient and moving things electronically.

David Lazovsky

How did your company overcome this memory bandwidth problem?

We had visibility that this need was coming because we were working with a hyperscaler two years ago. This was a company that has got the world's most advanced AI workload, and it told us that it needed the ability to disaggregate memory. So, we started with that requirement, which is HBM, level bandwidth. And we worked backwards from there. The rest of the industry was working on evolving on this roadmap of evolving Ethernet. So we're years ahead of this in development terms... and we've built a war chest of intellectual property (IP), protecting the use of thermally stable modulators in an optoelectronic system and package to provide a sustainable competitive advantage from an IP standpoint.

Does this technology expand out potentially into the telecoms market and other users inside data centres?

Absolutely, anyone who's buying GPUs and the GPU manufacturers – it's the entire ecosystem. That's like saying, 'Who's the market for PCIe?' I would say that we're an optical version of PCIe, but we're a technology that allows companies to leapfrog NV link, which is openly available to the entire market. We are working with design services partners – for example, the Broadcoms of the world – to enable the technology; we are offering the technology up to the processor manufacturers, including AMD, Nvidia and others. And we're working directly with the hyperscalers to build their own silicon and the systems integration companies, including the larger scale systems integrators. **B**

David Lazovsky is the founder and CEO at Celestial AI

How generative AI can help researchers beyond the lab

Generative AI can speed up life sciences researchers' work in areas such as writing of clinical trial protocols or support documents, says **Bryan Hill**

On average, it takes about seven years to develop a new drug and bring it to market. For ambitious life science businesses, generative AI's ability to generate insights and content in a fraction of the time of a human means wiping months, or even years, off that average. When it comes to clinical development, saving time translates to saving lives, or at least improving them, through the faster availability of medical treatments.

It can also translate into a significant revenue opportunity. Some industry sources estimate that bringing new treatments to market ahead of schedule can be worth between £500,000 and £6.5 million a day. However, due to uncertain regulatory landscapes, coupled with the rapidly evolving nature of the technology, some companies are taking a wait-and-see approach to adopting generative AI, delaying investments until the course forward is clearer. While this may seem prudent, organisations may regret it in the long run as they miss out on the opportunities generative AI holds, such as drug discovery and speed to market, and as competitors take the lead.

For life sciences companies looking to seize a competitive edge and supercharge their speed to market, there are a few key areas of the clinical development lifecycle they should focus on first.

Simplifying the drug discovery research process

Research and development (R&D) is often the most time-consuming part of the drug development process, but AI can accelerate this process by up to 50% as the technology has a multiplier effect wherever it is applied.

Life sciences can implement generative AI at the very beginning of the R&D cycle, to aid in searching and synthesising available literature on a specific potential drug. Instead of beginning with a manual keyword search and sifting through hundreds of articles across various sources, teams could prompt a generative AI-enabled tool to rapidly search, gather and distil relevant articles – or even suggest unanticipated information pathways to explore. Generative AI also has the potential to change how researchers find existing literature. Usually, researchers simply type keywords into the search box. But with a generative AI tool, they could state their goal into the prompt, providing context and intent, for the technology to find reference materials to support that specific task, saving significant time while broadening the research horizon.

Automating the writing of clinical trial protocols

Compiling a clinical trial protocol document is a lengthy process that can take anywhere from a few months to more than a year. Generative AI technology's capabilities can automate a substantial proportion of the protocol writing process, bringing it down to days or even mere hours.

Generative AI can be trained on thousands of existing protocols in industry databases and each company's own research data in order to identify the patterns relevant to investigational products, certain conditions, specific patient populations, or other factors. As the generative AI tool identifies relevant patterns, it can combine all the insights to design a baseline study, with a defined narrative that determines eligibility, drafts exclusionary criteria, and provides other necessary details. It can generate a number of draft options that would later be evaluated and refined by a human.

'Generative AI technology's capabilities can automate a substantial proportion of the protocol writing process, bringing it down to days or even mere hours'

Expediting launch processes in secondary markets

Once a new therapy has been approved to launch in one market, many companies will be looking to expand the launch into others. This process takes a tremendous amount of time and resources, from strategy development and market research to agency engagement, content creation and material development.

Much like in the research and protocol writing processes, a lot of the steps in this part of the process could be automated with generative AI.

For instance, when the drug is close to gaining approval, generative AI could support commercial teams' research and compile strategy documents for secondary markets, taking into account specific regulations the therapy will need to adhere to in the new country. Similarly, generative AI can be used to adapt existing content – including website copy, brochures and other promotional materials – to the language and culture of the secondary market. This could shave up to a year off the go-to-market timeline in new countries and massively reduce marketing and design costs.

Taking the first steps

Bryan Hill

Introducing generative AI into a business should be done one step at a time. It starts with fostering a culture of AI literacy, where every employee understands how the technology can be used to reshape and empower their role. It is also important to build a solid ecosystem of partners, which includes relationships with academic institutions, data providers, and specialty generative AI vendors that will support the business' knowledge growth and internal capabilities. Once generative AI is introduced, it is a good idea to establish a body within the business to supervise how the organisation uses the technology and manages the upskilling and development of employees engaging with the tech. This body should also establish best practices and develop frameworks that guide the deployment of generative AI across the business.

Introducing generative AI into a pharmaceutical business is no mean feat and is understandably very daunting. It is, however, essential for companies that want to stay ahead of their competitors and the market to invest in generative AI. Likewise, it is crucial to ensure employees are provided training on how to best optimise the technology and create a body that supervises how the technology is being deployed across the business to avoid any misuse. I believe that, as companies continue to experiment with generative AI across its various use cases, they will begin to lay the foundation needed to harness the full potential of this truly transformative technology, discovering, testing and, of course, bringing their drugs to market sooner.

This improves patient outcomes due to safer, more effective and affordable drug development and increases revenue opportunities in what is a highly competitive market, thereby driving value and improving patient outcomes at the same time. **B**

Bryan Hill is Chief Technology Officer for Cognizant Life Sciences

Developing open-source software to advance the development of energy-efficient microelectronics

Microelectronics, materials science, and computer modelling experts team up to fast-track the development of ultra-low-power microelectronics with open-source software

A multidisciplinary team of researchers at the US Lawrence Berkeley National Laboratory and UC Berkeley have developed an atomistic understanding of the origins of negative capacitance, enabling them to enhance and customise this phenomenon for specific device applications.

The advance was made possible by FerroX, an open-source, 3D simulation framework that the team custom-designed to study negative capacitance. The work was reported in the journal *Advanced Electronic Materials* and represents a significant milestone of a multiyear project, Co-Design of Ultra-Low-Voltage Beyond CMOS Microelectronics, which was funded by the USA's Department of Energy in 2021 to design new microchips that could perform better and require less energy than conventional silicon chips.

The open-source modelling framework will help researchers develop ultra-low-power microelectronics much faster and cheaper than current approaches that rely on costly investments in material growth and device fabrication.

Berkeley Lab's co-design approach to microelectronics research strengthens the link between material properties at the atomic level and the specific requirements of devices, aligning research objectives across all stages of device development. This approach, which leverages the lab's renowned interdisciplinary team science, aims to speed up the transition from R&D to commercialisation.

Prabhat Kumar and Zhi (Jackie) Yao

Scientific Computing World caught up with Zhi (Jackie) Yao and Prabhat Kumar to discuss the development of FerroX and how it will impact research into next-generation transistors.

How did you get involved with the 'Co-Design of Ultra-Low-Voltage Beyond CMOS Microelectronics' project?

Yao: I started as a postdoc in Berkeley Lab and, during my PhD, I was doing some experimental and modelling work using functional materials for new devices. Back then, I wasn't focusing on computing devices. I was set on microwave and radiation devices such as antennas. That was the starting point of my career and the reason I researched ferroelectric and ferromagnetic materials.

In the spring of 2021, the Department of Energy in the US started work on the co-design of microelectronics. By Co-design – what it meant is that the modelling work that I do will be directly connected to the material scientists and circuit scientists. We work together to ensure that our research fits the larger scope.

Kumar: I did my PhD in applied math from Stony Brook University, where I worked mostly at Brookhaven National Lab in a very different area. I was doing high-energy physics modelling for plasma-based particle accelerators. After that, I was a postdoc in the aeronautics department at Stanford, working on modelling plasma-based electric propulsion devices.

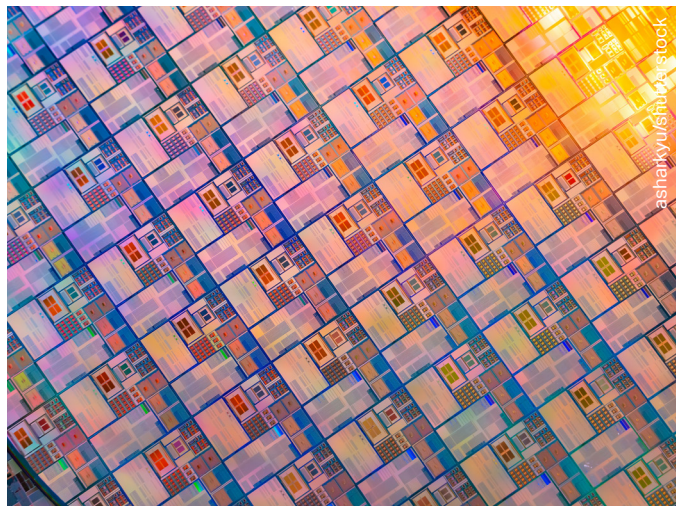
In this large collaboration, as Jackie mentioned, we have experts from material science to device physics, circuit design, system and architecture design, and we worked very closely with them, particularly on this paper. We worked in a tight partnership with Professor Sayeef Salahuddin's group at UC Berkeley to understand the physics of ferroelectric materials and how they would be used to design energy-efficient microelectronics.

Why was the co-design approach critical to the development of FerroX?

Yao: Co-design is a strategy to do the research. This should have been done for a long time because no research topics are completely isolated from each other. Think about a computer – you buy a computer, you have the display that is, maybe, photonics and optics, and then you have the keyboard – maybe that comes with sensors, and you have the computing unit, the CPU and GPU – that's a different topic – but you then have element devices within each of these components.

You have millions of transistors in your computing unit and then use transistors. You have semiconductor materials, and dielectric materials, so you have different material properties that require inputs from material scientists. Without co-design, no research of this scale would be achievable; and no commercial technology innovation could really be achieved.

Kumar: One additional benefit of Co-design that benefits us researchers is that we come from different backgrounds and, when we work together in a Co-design fashion, we learn from each other, and everyone has different expertise. And to be successful in your career, you need people from different backgrounds to guide you and help you through different things. In addition to innovating very interesting scientific products, as researchers, we grow and learn a lot when we work in this Co-design fashion.



Could you elaborate on the importance of phase composition in the thin films of ferroelectric materials and how this affects negative capacitance?

Yao: Phase computing is one of the standard methods to predict ferroid materials' behaviour. There have been phase-field simulations, and researchers have done this for decades. But there is still some unknown physics, which we're exploring.

The importance of making this open-source software came to our attention by noticing that this phase field community has been limited to a few groups of researchers. If you look globally, 20 to 30 researchers are focused on this, and they have their own in-house built code, which is not open source. When you read their papers and want to reproduce or do some design with their code or model, you just cannot do it because their source code is unavailable. Having this code open source, and accessible to a broader community – including the students – will largely benefit the field, advancing the area of phase field calculations and motivating more discoveries in this bit of research.

Why is the negative capacitance important to this research and how does FerroX help researchers explore this phenomena?

Kumar: When it comes to utilising it for more energy-efficient transistors, researchers are interested in studying ferroelectric materials in a transistor structure. This is a heterostructure with dielectric, pyroelectric and semiconductor layers. So, how would ferroelectric materials interact with this structure? Then, there are multiple optimisation problems: how would you choose the properties of the ferroelectric layer or the adjacent layers, such as the semiconductor and dielectric layer, so that your negative capacitance phenomena are enhanced, and what would be the effect of that on the overall energy efficiency of the transistor? There are two critical challenges here: understanding negative capacitance phenomena, and optimising it for use in transistors. FerroX allows us to do both.

FerroX is a general-purpose modelling tool for ferroelectric materials, and we have been

'There are two critical challenges here: understanding negative capacitance phenomena and optimising it for use in transistors'

using it for other applications as well. However, in the context of negative capacitance, we can study standalone ferroelectric materials or in a transistor structure to optimise it.

Furthermore, having this code run on laptops to large-scale supercomputers, allows researchers to do fundamental studies on simple problems they can run on their laptops. Then, if they want to go into more detail and run simulations at experimental conditions on large-scale devices, they can do that on large-scale supercomputers, which are accelerated by GPUs up to exascale machines.

Yao: We really wanted to explore the origin of multi-domain contribution in negative capacitance, improve the performance of negative capacitance in a rational way, and further explore what we see in the experiments and optimise the design of the material, and that required software such as FerroX. So, we set out to create software to achieve this so that we could explore the fundamental physics of ferroelectric domain evolution; and how the material is actually involved with the environment. **B**

Zhi (Jackie) Yao is a research scientist in the Computing Sciences Area at Lawrence Berkeley National Laboratory

Prabhat Kumar is a postdoctoral researcher in the Center for Computational Sciences and Engineering (CCSE) in the Computing Sciences at Lawrence Berkeley National Laboratory

ISC High Performance
The HPC Event.

CONNECTING THE DOTS ↘

ISC 2025
JOIN AS A
CONTRIBUTOR

JUNE 10–13, 2025
HAMBURG | GERMANY
isc-hpc.com

SUBMIT YOUR PROPOSAL NOW!

- PAPER
- TUTORIAL
- BOF
- POSTER
- WORKSHOP

FIRST CALLS CLOSE ON OCTOBER 21, 2024.

THE EVENT FOR HIGH PERFORMANCE COMPUTING, DATA ANALYTICS & QUANTUM COMPUTING

ISC 2025 offers an interactive community platform for scientists and engineers to present their work and interact with over 3,500 attendees.

As a contributor, you can guide the discussions in a way that helps transform high-performance computing and gain recognition for your project efforts.

isc-hpc.com/submissions