SCIENTIFIC COMPUTING WORLD

Computing solutions for scientists and engineers

Spring 2023
Issue #183

Exascale arms race
UK vs Europe in HPC

Laboratory informatics
Better manufacturing with AI

Autonomous engineering
Driving innovation

SUSTAINABLE RESEARCH
CAN COMPUTING CUT ITS CARBON BILL?

www.scientific-computing.com

Contents

Spring 2023 • Issue 183

High performance computing

Modelling and simulation

Resources

Laboratory informatics

Robert Roe
Editor

Balancing impact and opportunity

As the days get longer and warmer, there is hope and excitement about technology and software that is helping to drive new opportunities for research. However, there is also concern about the sustainability of computationally intensive research, which may threaten to derail innovation if not kept in check.

Our HPC coverage starts on page 4 with a feature on the convergence of HPC and AI technology and how this might affect researchers and computing centres. The sustainability of these huge HPC and AI systems is in question. On pages 8 and 10, we have two HPC tech focus articles on cooling and memory products, respectively. On page 12, a feature explores the announcements of two new exascale supercomputers set for the UK and EU.

On page 16, we introduce a special section on lab informatics. Sustainability in biotechnology research is the article's focus on page 18, as the environmental cost and impact of its computing and data storage grow rapidly. On page 22, there is post-event coverage from SLAS highlighting the impact composable IT can have on laboratory connectivity. Following an article exploring chemistry software on page 24, there is another interview from the Paperless Lab Academy, which focuses on the use of AI in chemical manufacturing on page 26.

Battery simulation kicks off the start of our engineering coverage on page 28. This feature explores how software simulation is helping develop better battery technologies for electric and autonomous vehicles. On page 32, there is a feature highlighting how software development is helping to drive the growth of autonomous systems used in engineering.

@scwmagazine

Editorial and administrative team
Editor Robert Roe
robert.roe@europascience.com
Head of content Mark Elliott
mark.elliott@europascience.com
Group managing editor Finbarr O'Reilly
fin.oreilly@europascience.com
Marketing manager Kate Risdon
kate.risdon@europascience.com
Specialist reporters Sophia Ktori, Gemma Church, Eugenia Bahit

Advertising team
Advertising manager Lexi Taylor
lexi.taylor@europascience.com +44 (0) 1223 221039
Production manager David Houghton
david.houghton@europascience.com +44 (0) 1223 221034

Corporate team
Managing director 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 Subscriptions £180 a year for six issues to readers outside registration requirements. Single issue £20. Orders to ESL, SCW Circulation, 4 Signet Court, Swann Road, Cambridge, CB5 8LA, UK. Tel: +44 (0)1223 221 030. Fax: +44 (0)1223 213 385. ©2023 Europa Science Ltd.

US copies: Scientific Computing World (ISSN 1356-7853/USPS No 018-753) is published bi-monthly for £180 per year by Europa Science Ltd, and distributed in the USA by DSW, 75 Aberdeen Rd, Emigsville PA 17318-0437. Periodicals postage paid at Emigsville PA. Postmaster: Send address corrections to Scientific Computing World PO Box 437, Emigsville, PA 17318-0437.

Cover: Shutterstock.com

Subscribe for free at
www.scientific-computing.com/subscribe

Can AI solve its sustainability problem?

INCREASING WORKLOADS AND THE GROWTH OF HPC AND AI SYSTEMS ARE CREATING A CARBON CRISIS FOR RESEARCH, WRITES EUGENIA BAHIT

With the proliferation of artificial intelligence (AI) and machine learning (ML), the high-performance computing industry's increasing workloads are contributing to environmental damage.

It may be confusing that cutting-edge technology such as AI, which helps humanity in so many fields, can have a negative impact and put an entire industry in check. Still, the increase in AI workloads is becoming a problem that needs urgent attention.

What causes increasing AI/ML workloads?

Artificial intelligence is advancing fast, and workloads are becoming bigger due to data volume and algorithm complexity.

Developing an intelligent machine requires the ability to make inferences that involve applying logical rules to deduce a possible conclusion. Current ML models do this by storing and classifying astronomical volumes of data, which is one of the factors responsible for large workloads.

The complexity of deep learning models – used to understand the data and apply the logical rules – typically requires recursion to process millions of parameters, demanding lots of resources. Therefore, the more complex the algorithms, the greater the workload.

Optimising the ML workflows to reduce workloads is not easy. According to Dion Harris[1], Head of Product Marketing for Accelerated Computing in HPC, AI, and Quantum Computing at Nvidia, some HPC users may not have the specific skills to optimise AI/ML workflows, and AI/ML experts may not have the experience to optimise them on HPC systems. Harris sees a gap between researchers with applied AI backgrounds and those with scientific computing knowledge.

"The rate at which AI is evolving requires deep expertise in data extraction, data preparation, AI training and inference, across a growing range of AI foundational models and techniques," he says. As a result, gaining experience lags behind the pace of AI evolution. Additionally, Ying-Chih Yang[2], CTO at SiPearl – a European company working to develop a low-power exascale microprocessor – thinks the workflow optimisation is also problematic because the addition of heterogeneous computing elements, such as GPUs or AI accelerators, improves performance on workloads while increasing software complexity.

Why are increasing workloads so critical?

From a business standpoint, the shortest answer is "money", as increased workloads lead to longer processing times, higher power consumption and, consequently, higher costs. However, a short answer is insufficient to become aware of the real problem.

The increase in workloads has a snowball effect from more than one perspective. When workloads increase processing time, the hardware has to work harder, causing a significant temperature rise, which demands more cooling. This not only increases costs but also leads to a larger carbon footprint – a direct result of greenhouse gas emissions. Therefore, the increasing AI workloads should not only be considered from a technical perspective. They ought to be managed to curb the environmental damage and ecological deterioration that is impairing the globe and primarily impacting the most underprivileged populations[3].

metamorworks/Shutterstock.com

"The carbon footprint needs to be viewed systematically, including development and transport of materials, manufacturing and transport of components, system design, energy sources, and reuse"

Yang emphasises the importance of viewing greenhouse gas emissions as a whole by pointing out that "the carbon footprint needs to be viewed systematically, including development and transport of materials, manufacturing and transport of components, system design, energy sources, and reuse".

Reducing the carbon footprint of AI workloads is a shared responsibility encompassing low-power hardware production, workflow optimisation, renewable energy sources, and a holistic approach covering the three greenhouse gas emissions scopes: direct, indirect, and indirect third-party emissions.

How does AI affect the environment?

The environmental impact of AI has been studied in recent years.

In 2019, Alexandre Lacoste and a group of Canadian scientists presented a Machine Learning Emission Calculator for cloud-based platforms[3]. They found that choosing a provider based on location offered a significant difference. The same model tested on three providers resulted in a carbon footprint 50 times larger in South Africa and 45 times bigger in China than in Canada. They also mentioned that GPUs were 10 times more effective than CPUs and 4 to 8 times less effective than TPUs. In an interview with journalist Payal Dhar, Lacoste expressed that using renewable energy for training ML models was the most significant change possible[4].

After the interview with Lacoste, Dhar wrote an article summarising studies since 2018. The report, titled "The carbon impact of artificial intelligence" and published in *Nature Machine Intelligence*[4], highlighted a finding showing that training a single ML model emitted the equivalent CO_2 of 125 round-trip flights between New York and Beijing.

That same year, a study by Simon Portegies Zwart[5] found that Python, the most popular programming language in astrophysics, produces the most CO_2. The paper concludes that using GPUs in computational astrophysics can significantly reduce the carbon footprint and emphasises using more efficient languages, such as Julia or Rust.

In 2021, Loïc Lannelongue wrote the article "Carbon footprint, the (not so) hidden cost of high-performance computing,"[6] which exposed numerous findings on the environmental impact of AI. Lannelongue, a principal contributor to several studies on this matter, suggested that the lack of data could be the primary reason this issue was not considered important within the industry.

Towards the end of 2021, an editorial note from Tan Yigitcanlar, published in the MDPI *Sustainability* journal, summarised the Green AI approach. He explains how it goes beyond the transition to renewable →

"The main differentiator of LUMI is the use of 100% hydroelectric power"

→ energy sources and its importance as public policy and its reliance on effective government regulations for success[7].

In 2022, Lannelongue participated in a study analysing "The Carbon Footprint of Bioinformatics"[8], concluding that: (a) cloud-based solutions reduce overhead power consumption; (b) the relationship between carbon footprint and the number of cores is not linear; (c) power consumption depends on the allocated memory; and (d) in some cases, the use of GPUs reduces the execution time, but multiplies the carbon footprint.

Something is wrong

Studies seem to agree that cloud-based HPC is the most energy-efficient. Reading the sustainability reports of cloud-based HPC owners, everything seems to be on the right track. Most aim to reach net zero emissions by 2030 and similar goals.

However, although the HPC era began in the 1960s[14], it was not until the end of the 1990s when the first terascale supercomputer appeared and towards the end of the 2000s that the first petascale supercomputer saw the light.

Long before the current exascale era, the world already knew about the severe consequences of fossil fuel burning and the necessity of renewable energy sources. The United Nations Framework Convention on Climate Change was adopted and signed by 154 countries in 1992. The damages of greenhouse gas emissions were already in the public domain when the IBM Roadrunner astonished the world.

However, if the world already had this knowledge at the end of the 1990s, why do the sustainability plans of cloud-based AI companies set 2030 as the goal for reaching carbon neutrality or net zero? A possible answer could be found in an article[9] by Jay Boisseau – HPC & AI Technology Strategist at Dell – where he gives three reasons for energy reduction: a reliance on costly fossil fuels, the environmental impact, and government pressure.

How can AI negatively impact humanity?

In January 2023, *The Lancet* published an article titled "Climate Change, Health, and Discrimination: action towards racial justice" [10], which mentions that the Global North is responsible for 92% of historical CO_2 emissions despite representing only 14% of the global population. It makes sense, since the Global North was the first to industrialise.

However, the priorities and reasons driving the industrialised world seem more aligned with individualistic and economic interests rather than the common good. Therefore, it is worth wondering what the priorities and reasons would be for reducing the carbon footprint and power consumption if the derived costs of fossil fuel burning were not so high. Fortunately, they are. So those who suffer the most from the product of global warming – namely, the Global South, which includes low-income countries and black, brown, and indigenous populations – could potentially avoid some of the terrible effects of an economy which is intrinsically supremacist[12] and unable to act motivated solely for the common good. Unfortunately, the Global South is already experiencing the impacts of global warming[10] due to the apathy of a world economy that has been inherently racist[11] [12] for decades.

What is the HPC industry doing to face the problem?

Each organisation works for sustainability from different perspectives.

At Nvidia, Harris says it is addressing the new AI era challenges on multiple levels. They focus on helping researchers by providing training, tools, and libraries – such as TensorRT – to optimise AI/ML workflows and pre-trained models and GPU-optimised frameworks. However, they are also working on energy-efficient hardware. Harris says that towards the end of 2023, they will start distributing the Grace CPU and Grace-Hopper CPU-GPU, enabling existing data centres to maintain


Doidam 10/Shutterstock.com

"SiPearl's product roadmap targets reducing the carbon footprint for workloads"

performance while doubling energy efficiency. They also presented cuLitho, a library for NVIDIA Hopper GPU systems that optimises the inverse lithography technology processes – used to develop nanoscale semiconductors – reducing power consumption to up to six times less than current CPU systems.

Intel, AMD, and ARM, among others, are working on more efficient hardware, while organisations such as SiPearl and LUMI are going even further.

LUMI: The cleanest, most effective future experience at the present

"HPC is energy-intensive and, therefore, has a high CO_2 footprint if operated in the wrong manner," says Dr Pekka Manninen[15], Director of Science and Technology at CSC, the Finnish centre where LUMI (one of the pan-European pre-exascale supercomputers) is located.

LUMI, a high-performance supercomputer with a sustained computing power of 375 petaflops, is considered one of the most advanced centres in terms of sustainability. Its entire power consumption comes from renewable energy sources, and the waste heat generated by the centre is used by the Finnish energy company Loiste, contributing nearly 20% to the district heating system. "By operating the HPC systems with CO_2-free energy, the game changes. The main differentiator of LUMI is the use of 100% hydroelectric power," says Manninen.

Based on an AMD Instinct GPU architecture, LUMI is the fastest HPC in Europe according to the TOP500 list and the third worldwide[13]. "The tenders were evaluated in terms of their technical value, performance, supply chain considerations, and EU-added value. We valued price-performance and energy efficiency a lot, which gave the edge to AMD," says Manninen.

When it comes to sustainability, he is direct: "No HPC facility should consider any other choice than 100% CO_2-free electricity. That is the leading factor in the sustainability of an HPC installation. The rest, like providing users with information about their jobs, tinkering with energy-efficient algorithms, scheduling, etc., we have seen over the years, is just 'peanuts', and the whole equation of end-to-end sustainability is dominated by the choice of power sources."

Like Yang, Manninen thinks sustainability should be seen as a whole, building an end-to-end systematic evaluation methodology that considers the system supply chain to its decommissioning.

SiPearl: the energy-efficiency microprocessor promise

Founded in 2019 and headquartered in France, SiPearl is working to produce a low-power, high-performance microprocessor. The first generation, Rhea, aims to be the first worldwide HPC-dedicated processor designed to support any third-party accelerator and deliver energy-efficient computing at scale, explains Yang.

"SiPearl's product roadmap targets reducing the carbon footprint for workloads. When available, the processor should significantly reduce the carbon footprint by up to 30% and 50%," he says.

Rhea is based on a low-power optimised Arm Neoverse architecture, which, according to Yang, "offers an ideal choice for performance per watt and energy efficiency for scalar and vector processing."

Yang further points out that SiPearl is working with Arm to integrate the network on the chip to interconnect the compute elements. He highlights the importance of the comprehensive Arm plan to produce future processor cores for seamless software migration, facilitating consistent improvements over time.

What's next?

No one can doubt the advantages of AI when every day, the media leave us astonished with technological innovations. But the world lives two parallel realities. One is technologically perfect, filled with unimaginable advancements, where large companies invest billions in increasingly sophisticated technology. And then, there is the other reality. One in which the waste ends up, and the nations that were once colonies of the former still suffer the consequences of inequality. It is a reality where energy crises force entire populations to live in inhumane conditions.

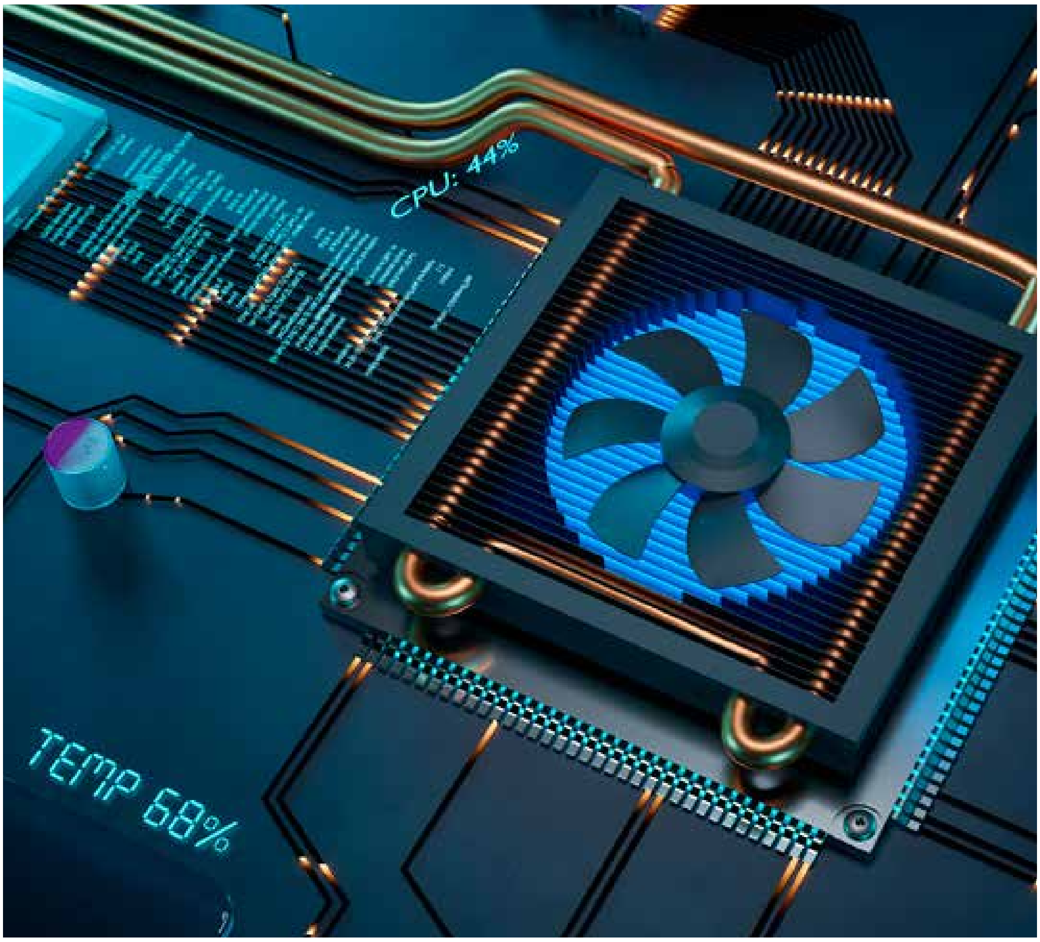
Carbon neutrality and net zero are preferable to any greenhouse gas emissions. However, they are not the ultimate solution. The workloads required to train deep learning models are growing daily, increasing power consumption. The only viable option is a joint effort to transition to renewable energy sources, build a systematic methodology for comprehensively evaluating sustainability, develop energy-efficient hardware, and optimise AI/ML workflows without neglecting the fundamental role of states, where nations commit to creating clear public policies that include minorities and underprivileged populations in decision-making.

References:

[1] www.linkedin.com/in/dionharris
[2] https://sipearl.com/en/ying-chih-yang
[3] https://arxiv.org/abs/1910.09700
[4] www.nature.com/articles/s42256-020-0219-9
[5] www.nature.com/articles/s41550-020-1208-y
[6] https://academic.oup.com/itnow/article-abstract/63/4/12/6503631
[7] www.mdpi.com/2071-1050/13/24/13508
[8] https://academic.oup.com/mbe/article/39/3/msac034/6526403
[9] www.cio.com/article/418812/whats-new-and-whats-next-in-2023-for-hpc.html
[10] https://pubmed.ncbi.nlm.nih.gov/36343651/
[11] https://bbc.in/3NRG5QA
[12] https://archive.org/details/statement-on-systemic-and-pervasive-racism-within-the-environmental-field
[13] www.top500.org/system/180048/
[14] https://digital-library.theiet.org/content/journals/10.1049/cce_19970107
[15] www.linkedin.com/in/pekkamanninen

Tech focus: Cooling

A ROUND-UP OF THE LATEST COOLING TECHNOLOGIES FOR SCIENTISTS USING HPC TO SUPPORT THEIR RESEARCH



B-th/Shutterstock.com

Cooling technologies continue to adapt to meet the changing demands and requirements of HPC users. Alongside growing demand for heat-dissipating products, other factors in choosing cooling solutions include increasing server density, infrastructure, costs, maintenance and support. This article highlights several providers delivering methods from RDHx to direct liquid cooling, passive loop cooling and immersion cooling – accommodating HPC users with varied requirements.

How to choose a HPC cooling solution

Start by determining the total power requirements for a server. This helps to estimate how much heat you will need to dissipate. For single CPU servers, RDHx or liquid cooling will most likely be sufficient; dual-socket servers with power-hungry CPUs or multiple GPUs will draw substantially more power. These require more complex technologies such as direct liquid cooling, passive loop cooling or even immersion cooling.

However, while underlying power requirements are key to choosing the right solution, there are other factors; available data centre space, infrastructure, maintenance and support – along with technologies such as chillers or evaporative cooling – all need to be considered to find the right solution.

Cooling products on the market now

Aqua Cooling supports users who need hardware to accurately cool equipment, providing comfortable temperatures for people to perform at their best. Controlling temperatures for an educational establishment or research facility can be challenging.

Through technical expertise and a range of equipment, Aqua Cooling delivers solutions designed to be reliable, energy and cost-efficient.

Aquarius' fixed cold plate warm water cooling combines efficient, dense, reliable and easy-to-use server cooling with the industry's best TCO and ROI.

Aquila's Dual Processor Intel Xeon Aquarius server systems offer performance, reliability and density. With efficiency of 95%, adopting Aquarius' liquid-cooled computing will enhance a data centre's PUE while drastically reducing operating expenses.

Asperitas offers solutions equipped with 24 server cassettes designed, optimised and certified for immersion cooling. Offering high performance and density compute for CPU and GPU workloads, all solutions can be warm water cooled and offer 100% of IT energy ready for reuse.

Calyos designs, develops and manufactures passive loop heat pipes. Its solutions provide freedom from thermal limitations, enabling the best performance from components and products.

A passive solution capable of cooling the latest high-performance processors, Calyos's technology uses a dielectric, safe, and green refrigerant inside their systems. This removes the risk of short-circuiting expensive equipment if the lines are damaged.

Calyos systems are built for the latest components and can handle heat-flux hotspots up to 250W/cm². Calyos is developing solutions to take this even further.

CoolIT Systems' Direct Liquid Cooling (DLC) uses the thermal conductivity of liquid to provide dense cooling to targeted areas. Using DLC and warm water, dependence on fans and expensive air handling systems is drastically reduced. This results in much higher rack density, overall reduced power use and significantly higher potential performance.

CoolIT Systems' CPU Coldplate is designed to accommodate lower-profile footprints, such as 1U blades and other custom chassis. These passive coldplates do not contain any high-failure-rate components such as internal pumps. Instead, coolant circulation is provided by a rack- or row-based Coolant Distribution Unit (CDU).

EcoCooling says its direct evaporative cooling system can reduce IT or data centre cooling costs by up to 90%, complies with ASHRAE 9.9 and produces a PUE (Power Usage Effectiveness) of under 1.1.

In Nordic and Arctic climates where temperatures never exceed 25°C, fresh air can be used all year round to cool a data centre. However, in warmer conditions, or when cooler inlet temperatures, are required, free cooling alone cannot meet required conditions. Supplementary cooling is required on hotter days. EcoCooling uses direct evaporative cooling in its products to offer a reliable, energy-efficient alternative to refrigeration.

Green Revolution Cooling's (GRC) patented immersion cooling technology creates favourable budget economics, which it says can reduce server energy by 10 to 20%. GRC's data centre solutions cool up to 100 kW/rack. Support for high rack densities allows for more compute per rack; the absence of hot/cold aisles allows for back-to-back racks, saving valuable space in your data centre.

Capturing nearly 100% of the heat generated from systems, the GPU-rich Ku:l Micro Data Centre solution from **Icetope** promises extreme cooling performance and high-grade heat recovery. Precision delivery of dielectric coolant cools hotspots directly. Without front-to-back air cooling and bottom-to-top immersion constraints, space usage is maximised.

Encapsulating Lenovo**'s** ThinkSystem SR670 server, each liquid-cooled chassis is sealed to protect the IT from the surrounding atmosphere. This creates a controlled environment, impervious to dust, gases and humidity. With reliable, efficient heat recovery, the Ku:l Micro Data Centre solution maintains density, offering improved efficiency.

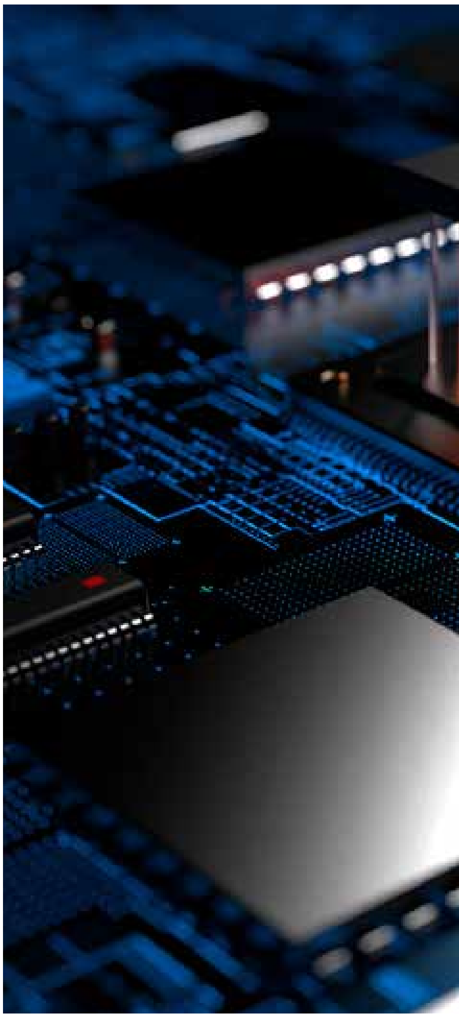
Motivair provides solutions for data centres facing ever-changing thermal challenges. The growth of artificial intelligence (AI), big data, Internet of Things (IoT) and hyper-converged infrastructure requires increasing CPU, GPU, & FPGA power densities. Motivair says its liquid cooling technology allows for an increase in processing power, using less energy and space than traditional air-cooled systems.

HPC applications have driven innovation in many industries. Companies must ensure the energy-efficient expansion of their data centre. **nVent** provides solutions that support the highest performance classes, delivering the best possible energy efficiency.

Submer is a specialist in liquid submersion cooling, a routine method for large power distribution components, such as transformers. Still rarely used for the cooling of IT hardware, this method is becoming more popular with innovative datacentres around the world.

TMGcore is a provider of data centre solutions and manufacturer of data centre hardware specialising in next-generation two-phase liquid immersion cooled technologies. Its OTTO Ready solutions are designed to provide customers with complete, holistic turnkey packages and purpose-built OTTO platforms, components and software across all industries.

Tech focus: Memory and processors



A ROUND-UP OF THE LATEST TECHNOLOGIES AVAILABLE TO HPC USERS

Memory and processor technologies are becoming more diverse and competitive in the HPC market. Buoyed by interest in artificial intelligence (AI) and machine learning (ML), demand for HPC hardware is increasing. High-performance processors and accelerators are necessary to deliver sustained application performance, but this has meant that memory and I/O bottlenecks are becoming increasingly common in HPC.

For several years, memory has increasingly been a stumbling block for HPC applications, as computational performance, or the number of floating point operations per second (FLOPs) that a system can produce outstrips the ability to feed data into the system. Known as the memory bottleneck, it has meant that innovation of memory technology is becoming crucial to driving performance for increasingly parallel systems which require vast amounts of bandwidth to stream data where it is needed consistently.

HPC systems require increasingly large memory subsystems, and this creates a problem for system designers who are trying to deliver enough data to each node to keep up with modern heterogeneous HPC systems.

Heterogeneous architectures are also becoming more commonplace. This technology focus will highlight memory, processor and accelerator technologies available to researchers and scientists using HPC hardware and help to showcase the growing diversity of hardware for supercomputing.

In recent years, the movement of data – rather than the sheer speed of computation – has become a major bottleneck to large-scale or particularly data-intensive HPC applications. This has driven demand for bandwidth increases in memory technology that can continue to support sustained application performance for HPC applications.

Memory and processor technologies on the market now

Achronix's Speedster7t FPGA family is optimised for high-bandwidth workloads. Built on TSMC's 7nm FinFET process, Speedster7t FPGAs feature a new 2D network-on-chip (2D NoC), an array of new machine learning processors (MLPs) optimised for high-bandwidth and artificial intelligence/ machine learning (AI/ML) workloads, high-bandwidth GDDR6 interfaces, 400G Ethernet and PCI Express Gen5 ports – all interconnected to deliver ASIC-level performance while retaining the full programmability of FPGAs.

Aldec has developed a portfolio of FPGA accelerator boards. Currently, Aldec provides several board configurations that can cope with the acceleration of the most demanding and sophisticated algorithms in two main categories – large-scale HPC and embedded HPC.

Derived from the family of HES prototyping boards, the FPGA accelerators were designed and optimised for large-scale HPC applications. To address the growing needs of embedded HPC, Aldec designed a special family of compact TySOM boards that utilise Xilinx Zynq devices, which integrate ARM Cortex processors and FPGA structures in one chip.

Alpha Data's ADM-PA100 is an adaptable PCIe form factor Versal ACAP data processing unit suitable for early development and rapid deployment of solutions based on a Xilinx Versal ACAP VC1902 AI Core device.

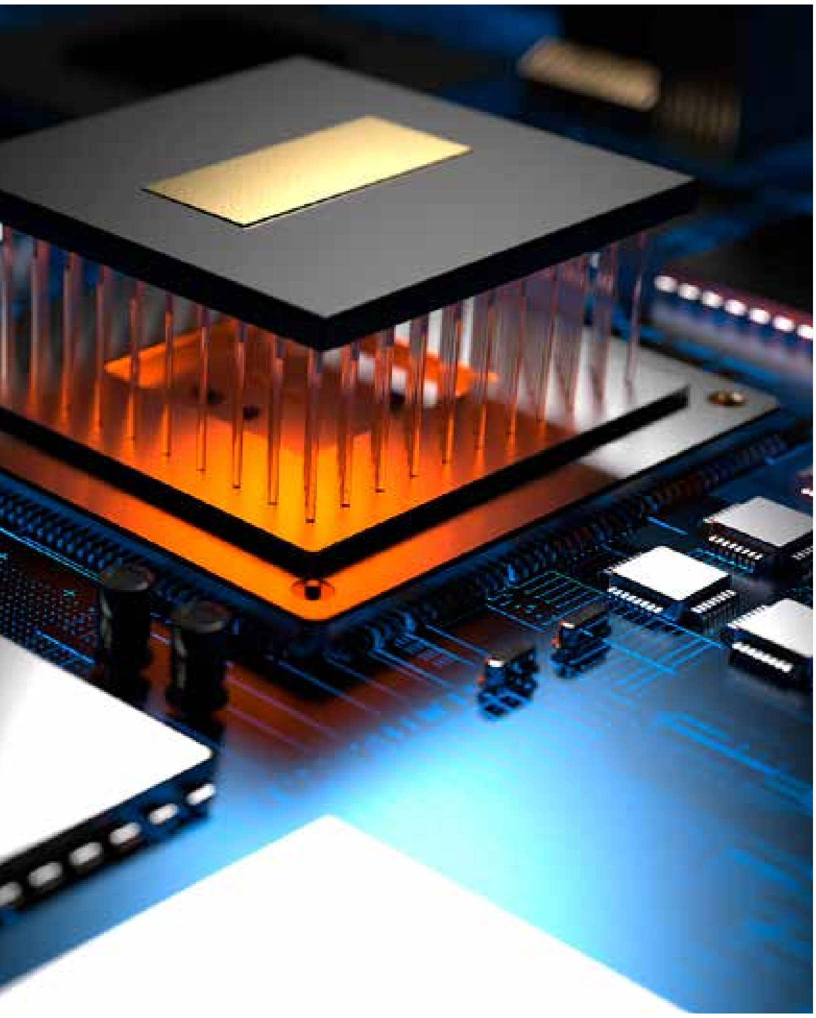
The PCIe form factor is suitable for desktop, lab, rack mount and data centre deployments in commercial temperature ranges. Additionally, the board can optionally be deployed stand-alone without any reliance on a host CPU. The FMC+ interface allows off-chip support for many standard and custom interfaces that the Versal ACAP can support through the very wide range of Alpha Data and third-party FMC IO adapters.

AMD EPYC processors and Instinct accelerators power the world's top supercomputers—including Frontier, the first system to break the exascale barrier.

AMD EPYC 9004 Series processors include up to 96 "Zen 4" microarchitecture-based cores, built on 5nm process technology. "Zen 4" enables leadership memory bandwidth and capacity with 12 DDR5 channels, as well as next-gen I/O with PCIe 5.0 and memory expansion with CXL.

Arm's Neoverse V1 includes the Scalable Vector Extension and CMN-700 mesh interconnect, delivering extreme performance while giving designers the flexibility and freedom to experiment. Neoverse V1 designs can be customised, but remain compatible with mainstream OSs, libraries, compilers, middleware and debug and profiling tools.

BittWare offers production-ready FPGA cards with Achronix, Intel and Xilinx FPGAs. The company allows researchers to develop and deploy applications quicker and with reduced cost and



evgeniy jamart/Shutterstock.com

risk by having it integrate their FPGA cards into its certified, integrated FPGA servers.

Cerebras is a computer systems company dedicated to accelerating deep learning. The largest chip ever built, the Wafer-Scale Engine (WSE) is at the heart of its deep learning system, the Cerebras CS-1.

Each core on the WSE-2 is independently programmable and optimised for the tensor-based, sparse linear algebra operations that underpin neural network training and inference for deep learning. The WSE-2 empowers teams to train and run AI models at speed and scale, without the complex distributed programming techniques required to use a GPU cluster.

Intel provides a comprehensive HPC technology portfolio that helps customers achieve results for demanding workloads. Intel HPC hardware is designed to scale from stand-alone workstations to supercomputers with thousands of nodes. Deep learning acceleration is built into the chip, so Intel processors are a great choice for solutions combining HPC and AI.

The Intel product lineup includes the Intel Xeon Platinum, Gold processors, and Intel Xe GPUs – the first Intel Xe-based discrete GPU for HPC. Intel also provides Intel oneAPI cross-architecture software development tools for HPC.

Taking full advantage of **Kalray**'s patented technology, the Kalray MPPA Coolidge processor is a scalable 80-core processor designed for intensive real-time processing. Coolidge is the third generation of Kalray's MPPA DPU (data processing unit) processors.

Coolidge is natively capable of managing multiple workloads in parallel to enable smarter, more efficient and energy-wise data-intensive applications. It offers an alternative to GPU, ASIC or FPGA, bringing value to multiple applications, from data centres, to edge computing and in embedded systems.

Kingston server SSD and memory products support the global demand to store, manage and instantly access large volumes of data in both traditional databases and big data infrastructure.

The need to store and manage larger amounts of data has increased exponentially in recent years. Data centres, cloud services, edge computing, internet of things and co-locations are just some of the business models that amass tremendous volumes of data.

Marvell offers a broad portfolio of data infrastructure semiconductor solutions spanning compute, networking, security and storage. Organisations deploy the company's products in enterprise, data centre and automotive data infrastructure market segments that require ASIC or data processing units equipped with multi-core low-power ARM processors.

Marvell's ThunderX2 family of processors enables the design of servers and appliances optimised for compute, storage, network and secure compute workloads in the cloud. Fully compliant with Armv8 architecture specifications including Arm SBSA, ThunderX2 further accelerates adoption for Arm servers in mainstream HPC deployments by providing the next level of computing performance and ecosystem readiness for commercial deployments.

MemVerge's Memory Machine virtualises DRAM and persistent memory so that data can be accessed, tiered, scaled and protected in memory. The new Memory Machine Cloud Edition introduces innovation needed for efficient Big Memory Computing in the cloud. AppCapsule snapshot technology paired with a sophisticated system and cloud orchestration service allows big memory cloud workloads to recover from spot terminations gracefully and automatically.

Micron's technology is powering a new generation of faster, more intelligent, global infrastructures that make mainstream artificial intelligence possible. Its fast, vast storage and high-performance, high-capacity memory and multi-chip packages power AI training and inference engines in the cloud or embedded in mobile and edge devices.

Micron offers a variety of HPC technologies, from GDDR6 and other memory products to SSDs and AI interference engines. Their Deep Learning Accelerator (DLA) solutions comprise a modular FPGA-based architecture with Micron's advanced memory solutions running Micron's (formerly FWDNXT) high-performance Inference Engine.

Xilinx's Alveo U55C accelerator card delivers the efficiency and scalability called for in HPC applications. The U55C delivers dense compute and HBM, with onboard 200Gbps networking enabling massive scale-out using Xilinx's groundbreaking open-standards-based clustering. Built around Xilinx's powerful Virtex XCU55 UltraScale+ FPGA, the Alveo U55C card delivers fast application acceleration.

The European exascale race heats up

THE UK AND EU HAVE BOTH ANNOUNCED THEIR FIRST EXASCALE SUPERCOMPUTERS WILL BE DELIVERED IN THE NEXT TWO YEARS, WRITES ROBERT ROE

The pursuit of exascale HPC systems has been a target of the HPC community since the first petaflop system broke into the Top500 in the June 2008 edition of the biannual list of the fastest supercomputers based on the Linpack benchmark. The Roadrunner system at the US Department of Energy's Los Alamos National Laboratory (LANL), built by IBM, was the first system to break through the petaflop ceiling with a Linpack score of 1.026 petaflop/s.

However, in 2022, the first exascale system was confirmed at US Oak Ridge National Laboratory. Frontier is currently the fastest recorded supercomputer and is driving new opportunities for research with a staggering 1.102 exaflops (equal to 1,102 petaflop/s).

Creating the technology stack for exascale has taken years of innovation beyond the simple iterative improvement of technology. In many cases, new computing architectures, networking systems, accelerators and processors are being constructed and optimised to deliver efficient exascale computing.

The drive towards exascale has often focused on delivering the highest possible raw computational power. The standard measure of exascale has generally been an exaflop or the ability to generate 10^{18} floating-point operations per second, equal to 1,000 petaflops/s. But this just scratches the surface of what is required to support an exascale system. Scientists require sustained application performance on real scientific codes.

Driving application performance at exascale requires a combination of computational power, I/O and memory bandwidth, and increases in energy efficiency that can make these future systems viable.

Exascale in the UK?

In the most recent budget from the UK Government, the Chancellor, Jeremy Hunt, announced almost £3.5bn to support the government's ambitions to make the UK a scientific and technological superpower.

The Chancellor also confirmed about £900m in investment into a new exascale supercomputer and a dedicated AI research resource. A statement from the UK Government said: "This funding will provide a significant uplift in the UK's computing capacity and will allow researchers to understand climate change, power the discovery of new drugs and maximise potential in AI – making breakthroughs that will benefit everyone in society and the economy."

Furthermore, it states that the "UK will become one of only a handful of countries in the world to host an exascale computer, attracting the best talent and ensuring researchers have access to the best infrastructure in the world".

The EU has delivered several pre-exascale supercomputers and is gearing up to deliver its own exascale systems in 2024 -2025. In recent years, the UK has lagged behind European countries such as Italy, France and Germany, all of which have delivered multi-petaflop systems in the top 20 supercomputers in the most recent Top500 list of supercomputers.

In the current iteration of the Top500, published in November 2022, the UK's highest-ranking system is ARCHER2,

Ink Drop/Timofeev Vladimir/Shutterstock.com

"This funding will provide a significant uplift in the UK's computing capacity and will allow researchers to understand climate change, power the discovery of new drugs and maximise potential in AI"

hosted by the University of Edinburgh jointly with the EPSRC, at #28 on the list. Prior to the delivery of ARCHER2, the UK's most powerful supercomputer was found at #58 on the Top500 list published in June 2021. This entry is based on the CrayXC40 and is used by the United Kingdom Meteorological Office.

The UK Government's announcement suggests that it has had a significant change of heart with regard to HPC and the value of large-scale computational research for HPC and AI.

From this new £900m investment, it is not yet clear how much will be spent on the exascale system. This new investment is roughly 10 times the cost of the UK's fastest supercomputer. The contract for ARCHER2, for example, was reported to be worth £79m in 2019. ARCHER2 was reported to be around 11 times faster than the previous system, ARCHER1. However, any new system close to exascale would be several orders of magnitude faster than anything seen in the UK.

In an interview with Edinburgh University, Professor Mark Parsons, Director of EPCC, the supercomputing centre at the university, commented on the importance of supporting UK research with its own exascale HPC systems: "The US and China already have exascale computers, Germany is fast developing its capabilities, and within a few years, all major economies will have one.

"Building and installing an exascale computer will cost around £500m, but it is essential to maintain the UK's competitive edge as a science superpower, especially after Brexit. We should be at the vanguard of innovation."

When asked for a comment regarding the budget announcement and the importance of delivering an exascale system in the UK, Parsons said: "EPCC's very pleased that the Future of Compute review recommended investment in an exascale supercomputer for the UK.

"This represents the culmination of over a decade's work and will allow the UK's computational science community to compete with their international peers", Parsons continued. "It's been a very tortuous process, and the Budget 2023 announcement is the first step along a complicated road. However, with a commitment to deliver full exascale by 2026, we're looking forward to the challenges of installing this system in Edinburgh."

The exascale computer will be housed at the existing datacentre used for Archer2. Parsons also noted that Edinburgh now →

"Building and installing an exascale computer will cost around £500m, but it is essential to maintain the UK's competitive edge as a science superpower"

Gorodenkoff/Shutterstock.com

→ has 38MW of power to its data centre – 30MW specifically allocated for the exascale system.

The EU unveils its exascale plans

In January 2023, the European High Performance Computing Joint Undertaking (EuroHPC JU) launched a call for tender to select a vendor for the acquisition, delivery, installation and maintenance of JUPITER, the first European exascale supercomputer.

This new system will be the first in Europe to surpass the exaflop threshold. This will have a major impact on European scientific excellence and help pave the way for new possibilities, with larger and more complex scientific models becoming viable for study due to the huge increase in computational power.

The system, installed on the campus of Forschungszentrum Jülich will be owned by the EuroHPC JU and operated by the Jülich Supercomputing Centre (JSC).

This new system will support the development of high-precision models of complex systems and help to solve key societal questions regarding, for example, climate change, pandemics, and sustainable energy production, while also enabling the intensive use of artificial intelligence and the analysis of large data volumes.

JUPITER will be available to serve a wide range of European users, no matter where they are located, in the scientific community, industry, and the public sector. The EuroHPC JU and Germany will jointly manage access to the computing resources of the new machine in proportion to their investment.

The EuroHPC JU will conclude a contract with the successful tenderer to acquire, deliver, install and maintain this high-end supercomputer. Therefore, the prospective contractors should have proven experience acquiring, delivering, installing and maintaining supercomputers in similar environments to those required in this procurement procedure.

The tender document delivered by the EuroJU highlights several key characteristics where this new system should outpace existing HPC systems in the EU. This includes performance/TCO; programmability and usability; versatility; system stability; power and energy efficiency; and computing density.

The estimated total value for this new system call is €273m (about £241m). The EuroHPC JU will fund 50 per cent of the total cost of the new machine, and the other 50 per cent will be funded in equal parts by the German Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the State of North Rhine-Westphalia (MKW NRW).

Dr Thomas Eickermann, Deputy Director of the JSC, says: "JUPITER will be based on the dynamic, modular supercomputing architecture, which Forschungszentrum Jülich has developed together with European and international partners in the DEEP projects funded by the European Commission and EuroHPC JU.

"It will be the first system in Europe to surpass the threshold of one exaflop. This next-generation European supercomputer represents a significant technological milestone for Europe and will have a major impact on European scientific excellence. JUPITER will support the development of high-precision models of complex systems and help to solve key challenges facing society, for example, climate change, pandemics, and sustainable energy production, while also enabling the intensive use of artificial intelligence and the analysis of large data volumes," says Eickermann.

Case study: Spain's IFISC tackles climate change with GIGABYTE NVIDIA GPU servers

Engineer studio/Shutterstock.com

Few things in the world are certain, but here are a couple of cold, hard facts: one, computing power will continue to advance as industry leaders push the envelope of Moore's Law; and two, the Earth is getting hotter and more crowded as climate change and pollution continue to affect the planet. The critical question for our generation is: how do we use the momentum of the first trend to delay or reverse the course of the second?

Revered institutes around the world are focusing on this million-dollar question. One such hallowed hall is Spain's Institute for Cross-Disciplinary Physics and Complex Systems (Instituto de Fisica Interdisciplinar y Sistemas Complejos; abbreviated as IFISC), a joint research facility founded by the University of the Balearic Islands (UIB) and the Spanish National Research Council (CSIC). Located on the island of Mallorca in the Mediterranean Sea, the IFISC is engaged in interdisciplinary research in various fields of study, including quantum technologies, photonics, environmental sciences, biosystems engineering, and sociotechnical systems.

The IFISC is devoting considerable resources to the most pressing issues of the day. Regarding the environment, the SuMaECO project studies the impact of climate change on aquatic plants in the Mediterranean Sea; and the Xylella project utilises machine learning to detect *Xylella fastidiosa* (Xf), a devastating plant pathogen that's spreading rapidly, especially during the hotter summers caused by global warming.

These projects run into a common hurdle: incredible computing power is needed to carry out high performance computing (HPC), numerical simulations, artificial intelligence (AI) development and big data management. Not just any servers can do the job. The IFISC needs servers that can satisfy the full gamut of computational needs, because the issues faced by humanity are multifaceted, and there are different approaches to solving the world's problems.

Parallel computing is vital for both of these projects. In the case of SuMaECO, highly detailed models are used to simulate the growth of Posidonia under different circumstances. In the case of Xylella, the algorithm must sift through a massive library of images before it can learn how to differentiate between healthy and stricken plants.

> "Most of us here are scientists, not computer engineers"

Through Sistemas Informáticos Europeos (SIE), a Spanish company specialising in servers, workstations, and other network and communications systems, the IFISC chose GIGABYTE's advanced server solutions as the best answer to its computational needs. The IFISC built computing clusters with three types of GIGABYTE servers: one type is the 4U 8-node G482-Z54 G-Series GPU Server.

AMD CPUs + High-density NVIDIA GPU Accelerators

One thing that's undeniable about GIGABYTE servers is their prodigious processing power. In particular, the G482-Z54 can bolster the speed of the CPUs with a dense array of NVIDIA GPU accelerators, which is a key feature in GIGABYTE's NVIDIA-based GPU Servers.

The G482-Z54 has a 4U chassis that can house up to 8 GPU cards in PCIe Gen 4.0 bandwidth, which has a maximum bandwidth of 64GB/s and is twice as fast as PCIe Gen 3.0. The G482-Z54 is a natural choice for parallel computing, high performance computing (HPC), cloud computing, and many other data-intensive applications.

For its G482-Z54, the IFISC chose dual AMD EPYC Rome 7282 processors with 16 cores and 32 threads per CPU. They supplemented the powerful processors with 6 NVIDIA RTX™ A6000 GPU cards. This set-up is exactly what the SuMaECO and Xylella projects needed. The Tensor Float 32 (TF32) precision of NVIDIA RTX™ A6000 provides up to five times the training throughput over the previous generation, to accelerate AI and data science model training without requiring any code changes. Hardware support for structural sparsity doubles the throughput for inferencing. Now, the IFISC can precisely simulate the growth of the Mediterranean's Posidonia meadows, and an AI model that detects signs of *Xylella fastidiosa* in satellite photos is in the works.

Following the successful implementation of G482-Z54, GIGABYTE would consider IFISC to deploy the next-generation successor of G482-Z54: the new G493-ZB3-AAP1 GPU supercomputing solution. With its new PCIe Gen5 bandwidth for GPU devices, the new G493-ZB3-AAP1 will further augment computational power for the IFISC research team by adopting the new NVIDIA L40 GPU. The new NVIDIA L40 is powered by the Ada Lovelace architecture and its 4th-gen Tensor cores (with hardware support for structural sparsity and optimized TF32 format), enhanced 16-bit math capabilities (BF16), and 48GB ultra-fast GDDR6 memory.

"At the end of the day, most of us here are scientists, not computer engineers," says Dr Pere Colet. "Fortunately, GIGABYTE and SIE were able to understand our specific computational requirements and offer us the best combination of GIGABYTE servers to solve our problems. We are very happy to be working with them."

To learn more about GIGABYTE products, follow the link to their website.
https://bit.ly/3JMI9ti

SPECIAL 12-PAGE REPORT

'Digital labs' free scientists to focus on the science

Robert Roe
Editor

For more than two decades, *Scientific Computing World* has been leading the way in its insightful coverage of laboratory informatics. Today is the most exciting time to be a laboratory scientist because labs are moving from traditional paper-based systems to digital transformation.

The rise of integrated systems, from web-based to cloud-based software, increasingly connects instruments and data management systems with structured and unstructured data, including raw experimental data and associated metadata.

The proliferation of computing and AI promises to unlock new avenues for research and discovery across all laboratory segments. Alongside the push for digital transformation is a look toward the future as laboratories begin to automate simple workflows. This automation will free up time for domain expert scientists to focus on their research by removing or reducing steps that previously caused bottlenecks in scientific discovery.

SCW was among the more than 12,700 people who attended this year's Pittcon, which featured more than 400 exhibitors from 80 countries

How biotech researchers cut computing carbon costs

AS BIOTECHNOLOGY RESEARCH BECOMES MORE COMPUTATIONALLY INTENSIVE, RESEARCHERS MUST FIND NEW WAYS TO INCREASE SUSTAINABILITY, WRITES SOPHIA KTORI

All pics: EMBL-EBI

Andy Cafferkey, EMBL-EBI

The European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) maintains possibly the world's most comprehensive range of molecular data resources, including databases, tools and software, which it provides to researchers globally. The organisation, based in Cambridge, UK, has been taking strides to address the significant carbon cost associated with computing data.

"People may not always think of the IT industry as consuming as much energy as it does," says Andy Cafferkey, Head of Technical Services at EMBL-EBI. "But I think that as a society, we are becoming more aware of the energy-related cost of things, such as the internet services that we all take for granted. Our response as an organisation is in line with a general awareness that the IT industry needs to be just as accountable for the environmental cost and impact of its computing and data storage, as are other industries that are more traditionally associated with a high carbon cost."

EMBL-EBI is installing a data management infrastructure that will allow for growth, but remain sustainable, and that starts at the level of the energy required for even basic data storage and computing, suggests Cafferkey. "We need to be aware of the cost of energy going through the stack, and where the greatest energy requirements lie. Our compute cluster is the largest consumer of energy, so we are constantly looking at the energy efficiencies that we can make within that cluster."

For more than a decade, EMBL-EBI has used external data centres to support its own data centre, situated at the Wellcome Genome Campus. In 2019, it teamed up with KAO Data. "Energy efficiency and sustainability were really important to us as we embarked on an open procurement process to find a new data centre," says Cafferkey. "As we were reviewing options, we were impressed to find that KAO even uses biofuel to power its back-up generators."

As part of the partnership with KAO, EMBL-EBI completed an 18-month process of migrating its most energy-intensive processes to the KAO data centre, deploying new compute and storage platforms that offer greater energy efficiency, explains Cafferkey: "We've achieved impressive savings, and importantly we have reduced power consumption by the most energy-hungry part of our infrastructure, without significantly impacting on performance. So, while over the course of that 18 months we've increased the amount of data that we house we've still achieved a reduction in the amount of energy that our data centres use."

The project touched on almost every individual at EMBL-EBI, according to Cafferkey, and involved close engagement between the IT teams and scientists. "During this 18-month migration project, we moved services over one by one as resources allowed,

Researchers at EMBL-EBI learning about data resources

to ensure that we had a constant flow of data moving, and that we were able to get everything over and retire the old equipment, while maintaining operations." Achieving this successfully involved close engagement with the scientific teams to ensure that the process of migration, "from timetable to breakdown of responsibilities", was clearly laid out.

One of the major goals for EMBL-EBI – within the next five years, believes Cafferkey – will be to move to a tiered data storage system. Data will be assigned to a tier dependent upon factors such as how that data is used, and how often it is accessed, for example. He says: "What we do with data needs to be understood both from an administrative perspective, and also from the research and the scientific perspective" so that it then becomes possible to accurately map data onto the tiered system. "Deciding to which tier any designated dataset should be assigned would need input from the users, together with implementation of a structured data management and cataloguing system." Cafferkey says EMBL-EBI is working on this process with its scientific users: "The organisation needs input from all users so that we can derive and apply rules and policies. Crucially, in this new tiered system, all the data will continue to be freely and openly available to the scientific community."

A tiered data storage structure would allow EMBL-EBI to assign data that is rarely accessed to more energy-efficient storage tiers. "This would mean that the access times would be slightly longer, but there would be significant energy and cost efficiencies to be gained from storing that data on less energy-intensive equipment. And that would also help us to address the scaling that we'll need for our next phase of growth," says Cafferkey.

At present, EMBL-EBI retains all of its →

"The IT industry needs to be as accountable for the environmental cost and impact of its computing and data storage as other industries more traditionally associated with a high carbon cost"

"Energy efficiency and sustainability were really important to us as we embarked on an open procurement process to find a new data centre"

data either at the KAO site, or on site in its own data centres. "However, a hybrid cloud approach will let us make the most of both worlds," adds Cafferkey. "We already have strategic partnerships in place with some of the major cloud providers as well as with other software companies, and we have some pilots running in the cloud. We are also working across EMBL-EBI to deliver the skills that users will require, because using cloud does require a slightly different skill set, and some of the responsibilities that previously were embedded in the IT teams will shift towards the science-oriented teams."

Alongside the development of a tiered storage system, EMBL-EBI is also generating tools to help scientists understand how their work affects energy usage. It's important for users to have what Cafferkey describes as "visibility into the carbon cost of the code that they run", so that they are aware of the implications of their activities from an energy perspective.

"We're working towards putting in place the tools that will allow users to explore the carbon footprint of the compute that they're asking to be performed on our stacks, so that they can consider the implications and potentially change their behaviour. This will drive people to look for efficiencies in their working," says Cafferkey. "It's no good just having an efficient data centre. You need to make sure the equipment is also efficient, that your scientists and users are operating efficiently, that your data is organised efficiently, because the carbon cost, if you like, is a product of all of those across the whole organisation."

The concept of sustainable data management is not new, suggests Simon Andrews, who heads the bioinformatics group at the Babraham Institute in Cambridge, UK. The group is one of eight scientific facilities at the Institute, which is focused on carrying out fundamental human health-related research in three core areas; epigenetics, signalling, and immunology, and is strategically funded by the Biotechnology and Biological Sciences Research Council (BBSRC). Andrews' team oversees how the organisation stores, manages, analyses and reuses research data derived both in house, and through access to public datasets.

"The need to keep data management sustainable has been an issue for 20 years or more," says Andrews. Storage is finite, and so for any organisation, achieving sustainability will almost inevitably mean that some data will have to be ousted, and knowing what can be removed and when requires a keen and in-depth understanding of the lifespan of that data.

Data volume and complexity affects storage cost, and the amount of energy used to run the data centre, adds chief information officer Cass Flowers, who heads the institute's IT department, overseeing the management and provision/maintenance of servers and other hardware. She suggests that her team's role is to provide "the suitcase" into which the institute's data is packed and stored. "We retain data on our own servers, and so we are continually having to review what we're storing, and trim that volume over time," says Flowers.

In practice, organisations need to consider from an energy perspective how/where their data are stored, how often that data needs to be accessed, and the energy requirements of the processes used to evaluate and analyse the data.

Any approach to decision-making needs to be founded on knowing exactly where the data is, how much of it there is already, and how often it is used. Andrews describes the institute's approach: "We have a centralised system for registering and reviewing the public datasets that we hold, and when we need to have a clean-up, we essentially send a message around to all the users, prewarning them that the data will be cleared, and providing a tick list of data they want to keep. The system works very well, and there's rarely any comeback from scientists after data has been wiped, although given that this is public data, it could just be reacquired."

Computing infrastructure supports scientists

Modern data centre facilities

For data generated in-house, the institute has some metrics in place that show when people are using what data, when it is being accessed or updated. "We rarely effect destructive deletions with our own data," says Andrews. "We can continue to store what we are legally obliged to store – for as long as we are required to – as part of our funding conditions, and with a view to future data submission to public repositories, or to publication. What we can do is then remove some derived data that we could make again, but that has been stored for convenience in the shorter term. We generally have a pretty good idea when these intermediate data files are no longer needed."

The ability to delete confidently is dependent on how well structured, and how well tracked the data is, acknowledges Flowers. "Where there are comprehensive management systems in place and a centralised repository, it's fairly easy to do." But without centralised management it can be almost impossible to know if copies of data are exact duplicates, or even where to find them.

While Flowers and her team provide the institute's data management infrastructure and supporting metrics, data management is, in effect, everyone's responsibility. So there is constant communication between Flowers' IT department, Andrews' bioinformatics team, and the scientists in the labs.

"There always has to be a compromise between what we would ideally want, and what is more realistically available or feasible," she says, and this is reached by combining the technical, budgetary, and sustainability perspectives. "Everyone understands that."

While there are some metrics in place to help keep track of data, at present, the institute relies significantly on manual archiving, and the insights provided by the scientists, or data owners, with respect to understanding how much data they have, where it is, what they need to keep, and for how long. "But there are digital tools out there that can even better monitor all of this for us," says Flowers. "We have a major storage refresh planned for a few years' time. There will be a lot of options for us to consider. There's a lot more available now in terms of technology that can give us a better visualisation of the data that we have, and that could automate processes such as archiving, and even transferring inactive data up into the cloud where it may be cheaper to store."

"The problem with cloud storage is it could become very costly for us if we transfer our active data up there"

The institute is investigating the use of cloud storage, according to Flowers and Andrews, though they acknowledged that cloud isn't always the better option for data storage. And it doesn't address the sustainability issue, they point out. "The problem with cloud storage is it could become very costly for us if we transfer our active data up there," says Flowers.

"Cloud offers a very different way of thinking about data storage. It can give more flexibility, but can be harder to control cost. When using cloud storage you don't necessarily pay a lot for the data storage itself, but for data transfer. So, for data back-ups or archives that are not being regularly accessed, the cloud option can make sense."

"And it can be useful, from a security perspective, to keep some data separate, and away from your own infrastructure," says Flowers. "On the other hand, with the size of our research data we could encounter major costs, so cloud options may not be cost-effective for us, It certainly wouldn't make sense for us to just push everything to cloud storage. If you have a known, fixed amount of computing and storage on site, then as long as you're making fairly heavy use of that, it can be quite cost-effective. Conversely, if your processing requirements change dramatically, then the flexibility that the cloud affords can be great."

How composable IT supports laboratory connectivity

GEOFF GERHARDT OF SCITARA EXPLAINS HOW COMPOSABLE IT ARCHITECTURES CAN HELP UNLOCK THE VALUE OF DATA IN A MODERN LABORATORY

What is a composable architecture?
The term composable IT is becoming increasingly popular to describe layered and connected IT systems. With modern cloud software, we can create workflows by configuring applications to have APIs and business logic that can be applied externally.

We can automate, not by hard coding but by configuring. Right now, we do this a lot, even in DevOps. We spin up Docker containers that represent an entire application. And just through configuration, we can get an application running. The idea of a composable enterprise, a term coined by Gartner, is to use these modern architectures that have exposed APIs, and then create an automation layer on top.

Modern software technologies democratise access to this approach so that employees don't have to wait for a developer to create their desired applications. With little or no coding experience, users can build a custom solution themselves using an Integration Platform as a Service (iPaaS) or integration tools to create business logic. Businesses can operate faster and don't have to wait for this bottleneck of development built on this cloud architecture.

What does a typical composable architecture consist of?
A business may choose applications such as Gmail and Slack for communications, Microsoft 365 for business document creation, AWS to store data or Snowflake as a data lake. It may choose applications such as MailChimp to send automated mailings or Airtable, an advanced spreadsheet tool, for data visualisations.

It will usually also choose an identity provider. This means a user doesn't have to log in individually nor remember each authentication. By having one identity provider across all of these applications, they can now communicate and exchange data.

Once you have all these functional applications, the identity provider and APIs enable you to use an iPAAS such as Power Automate, Zapier or MuleSoft – there are several available products – that can then be layered on top to create this automation.

Power Automate is a common one available within Microsoft, allowing users to create these sorts of scripts or automation. Users can automate their workflows by exchanging data between these different functional applications. MuleSoft is a slightly more advanced iPAAS that corporate IT could use to write business logic to create new corporate functions. But it's still not a developer. You just need IT professionals to create these automation workflows that allow data to be exchanged.

How have labs adopted this technology?
Labs are behind compared with how typical modern enterprise operations work with composable architectures. There are several modern cloud-based ELNS and LIMS systems, but even they can have difficulty connecting to the lab in the same way and getting data out of all these applications.

We can use Scitara DLX to bring that laboratory architecture into the modern world and allow users to create these applications that can exchange data with the rest of your cloud architecture. DLX is cloud-based. We have developed technology to connect prem-devices and instrumentation, simple devices such as weight balances, particle counters, pH meters and environmental devices. For instruments such as HPLC system, plate readers, or a bioreactor, we have an IoT device that can connect to those systems. Even legacy systems, where all they can do is drop a file, we have developed a file connector, allowing these on-prem systems to publish events and data into our cloud-based application.

One of our use cases is a CMO customer who wanted to modernise their upstream bioprocess development. With bioreactors that would need to run for several days to several weeks, this user wanted to be able to aggregate all the data and make decisions on the fly with the bioreactor, as well as inform customers of the progress. Typically, they were using an offline spreadsheet to try and aggregate all this data.

We helped this client to create a workflow that automatically pushed sample results to an online spreadsheet. Collaborators have a chance to see this data live, as can the analyst running the bioreactor. So it's a much cleaner process, with no transcription errors.

Like any other iPAAS, our application can connect to these general business applications via APIs. It can be connected to modern ELN or cloud-based LIMS systems, or even to customer cloud resources. We can provide the connectivity and bring these lab assets into the cloud.

Once we have everything connected, we have an automation layer that can be used to create applications to automate workflows. With the customer we spoke of earlier, we provided an editor that allows what we call "orchestrations" – an automation that starts with some sort of trigger, such as an instrument producing particular data. We also have "user-triggered" orchestrations. Any user-triggered orchestration can be assigned a QR code. A user will walk over to that QR code, scan it with a tablet, and that will trigger this automated workflow.

What has held up the adoption of composable laboratory IT?
Labs often operate in a protected LAN environment and so the systems may not be exposed to a cloud-based interface.

That is part of the reason we have this software tool that will aggregate the connections in a very safe, secure way. Connecting to DLX means cloud-based applications, such as Benchling ELN or Perkin Elmer's Signals Notebook can communicate with our cloud. We can provide this tunnel into the lab so that they can send instructions to an on-prem instrument.

Dr Geoff Grahardt is CTO of Scitara and has more than 25 years of experience leading R&D teams producing innovative technologies and products.

White Papers
now available online

VIEW FOR FREE*

AWS & TetraScience: 500 Pharma Executives Research Survey
TETRASCIENCE

Don't miss out on our AWS & TetraScience: 500 Pharma Executives Research Survey. See what pharma executives have to say about the need to replatform scientific data to the cloud.

Manual No More: Automating the Scientific Data Lifecycle
TETRASCIENCE

Some 50-80% of research scientists' and data scientists' time is spent wrestling with data before they can focus on higher value AI/ML and advanced analysis to help bring new life-saving therapeutics to market.

Progress Snapshot: Enabling the Digital Lab of the Future
ASTRIX

This report assesses the progress achieved and plans for near-term activities among laboratories in six subsectors within the broader classification of Life Sciences organisations: Biotechnology, Biomedical Research, Clinical Research Institutions, Contract Manufacturing Organisations, Contract Research Organisations, and Pharmaceutical Research and Manufacturing.

Advanced Analytics in the lab: How AI-Powered Analytics Transforms LIMS Data into Valuable, Action-Ready Insights
LABVANTAGE

When modern LIMS software overtook traditional paper-based record-keeping in the lab, operations in every industry improved. Automated data collection drove efficiencies at the bench. Cloud solutions accelerated knowledge-sharing between partners.

A Complete Guide to LIMS Selection
THERMO FISHER SCIENTIFIC

A Laboratory Information Management System or LIMS is a safe, efficient and affordable digital solution to laboratory management. Using a LIMS, you can collect data and manage your lab processes easily, giving you more time to focus on making new discoveries or ensuring quality products.

*Registration required

www.scientific-computing.com/white-papers

SCIENTIFIC COMPUTING WORLD

Tech focus: Chemistry

A ROUNDUP OF THE LATEST CHEMISTRY SOFTWARE AVAILABLE TO RESEARCHERS

With huge costs and long development timelines associated with getting new drugs on the market, a wide selection of software tools have been developed to help pharma organisations to more effectively and efficiently develop new candidate drugs.

In such a hugely competitive market, even getting a drug to market a few months early can have substantial effects on profitability. Equally, setbacks and delays in the development and research process can have a large impact on the potential cost of successful drugs or the outright success of a new compound.

Today there are several categories of software tools, from data curation to quantitative structure–activity relationship (QSAR) and even specialised ligand-based tools. There are also an increasing number of AI- and deep-learning-powered software options that offer huge potential to accelerate the discovery pipeline.

Chemistry products on the market now

ACD/Labs develops its Spectrus and Percepta platform software to support drug discovery research. The latest release of Spectrus provides functionality for machine-readable structured output, which has the potential to assist stakeholders in R&D organisations with aspirations to digitalise laboratory workflows, reduce the degree of document-driven decision-making, and mitigate the risk of manual data transposition.

These new advances to the software platform help to deliver new functionality for scientists that includes: decision support, improved data flow and hardware/software integrations; expansion of instrument control, and the expansion of browser-based technologies.

Founded in 2018, **Alvascience** is striving to boost value for cheminformatics research through high-tech software solutions using machine learning and other science tools.

Starting from data curation (alvaMolecule), the calculation of molecular descriptors and fingerprints (alvaDesc), regression and classification modelling (alvaModel), the deployment of the models (alvaRunner), to de novo molecular design (alvaBuilder), AlvaScience products can be used independently or as a suite of tools to tackle the entire QSAR workflow.

The alvaMolecule component was conceived as a molecular worksheet where molecular datasets can be visualised both as a molecule grid or as a spreadsheet; additional data provided within SMILES and MDL files are automatically imported and can be used, together with the calculated descriptors and physicochemical properties, to sort and filter the molecular dataset.

CDD Vault is a complete informatics platform that helps project teams manage, analyse, and present data for biotech companies, CROs, academic labs, research hospitals, agrochemical and consumer goods companies.

Built for drug discovery teams, CDD Vault says the web interface is intuitive for beginners and powerful for experienced users, allowing data migration without requiring any IT expertise.

CDD Vault enables users to access all their research data from any modern web browser. They can search for any experiment or explore results in SAR tables and plots. Filter, colour code, and highlight substructures to quickly screen for patterns.

CDD Vault is delivered as a hosted solution that is cost-efficient for academic laboratories, start-ups, and biopharmaceutical companies. CDD Vault comes with in-house support and can scale up as teams grow.

Certara has added deep learning capabilities to its scientific informatics software platform, D360. The addition of novel AI will enable discovery scientists to integrate multiple structured and unstructured data sources and substantially enhance predictions and analysis of content related to small molecules and biological drugs.

The rollout of D360 with advanced deep learning analytics includes automated property prediction, novel structure generation and access to unstructured data.

Predictive deep learning models trained on public and proprietary chemical structure and biological data will provide quantitative or categorical predictions that improve substance prioritisation.

Certara's Simcyp Simulator Version 22 includes new capabilities and updated features to the company's population-based modelling and simulation platform. The simulator has proven use cases across drug development, including first-in-human dosing, extrapolation to special populations, bioequivalence testing, optimising clinical study design and predicting drug-drug interactions (DDIs).

New capabilities in version 22 include an expanded compound library, enhancement of subcutaneous dosing capabilities and expanded mechanistic modelling.

Cresset software is underpinned by patented scientific methods that use 3D molecular electrostatics and shape to shed light on the properties and behaviours of chemical structures and, crucially, to understand the key interactions which underpin biological activity.

Flare is an agile ligand-based and structure-based drug design solution, while Spark is a scaffold-hopping and R-group replacement to generate innovative ideas for discovery projects.

Flare enables chemists to closely inspect the detail of their ligand-protein complexes, using a variety of methods to gain useful insights into their protein targets and ligand series. As our most feature-packed software package, both ligand-based and structure-based drug designers are supported to progress their lead optimisation.

Iktos specialises in the development of AI solutions applied to chemical research, more specifically medicinal chemistry and new drug design. Iktos' technology platform is designed to enable major productivity gains in upstream pharmaceutical R&D.

Iktos has developed Makya, a fully automated SaaS platform for de novo drug design that enables rapid identification of molecules that satisfy multiple drug-like criteria to expedite drug discovery and development. Makya is suitable for both ligand-based and structure-based virtual lead design and allows multi-parametric optimisation (MPO) of lead molecules towards potential pre-clinical drug candidates.

Intellegens' Alchemite deep learning software solves real-world data problems, accelerating innovation in materials, chemicals and manufacturing.

Alchemite provides web-based access to data analysis, model-building, optimisation, and prediction tools, with intuitive graphical visualisation of results. Scientists or engineers can get started fast, according to Alchemite – no coding skills are required. And Alchemite Analytics is lightning-quick – it takes minutes to train models that take conventional machine learning methods hours or days.

OpenEye Scientific develops Orion, a software cloud platform for molecular design and simulation. This software reduces the requirement to run and maintain expensive infrastructure hardware. With Orion, you can access OpenEye computational tools in the cloud.

Running these solutions on AWS (Amazon Web Services) means users can solve modelling questions that may have been previously impossible due to compute-time or hardware limitations.

The Orion Science Suites group together complementary functionality into solutions: Orion Small Molecule Discovery Suite for structure and ligand-based design, Orion Antibody Discovery Suite for diverse antibody library design, Orion Formulations Suite for crystal structure prediction and the Orion Gaussian Module for quantum chemistry calculations on cloud.

Optibrium's Stardrop software aims to provide a complete platform for small molecule design, optimisation and data analysis. StarDrop delivers in silico technologies within a visual interface, designed to enable a seamless flow from the latest data through predictive modelling to decision-making. This helps users to improve the speed, efficiency, and productivity of the discovery process.

Cerella is powered by Alchemite, a deep learning method developed by Optibrium's technology partner Intellegens. In collaboration with pharmaceutical and biotechnology partners, Optibrium has showcased Alchemite's benefits over conventional modelling methods in peer-reviewed studies, resulting in reductions in cost and time of discovery cycles. Cerella has collaborated with several pharma and biotech organisations to demonstrate its benefits, working with datasets from the individual project level up to global compound data repository level investigations.

Merck's Synthia retrosynthesis software is powered by algorithms that can help experts' access and make use of the vast amounts of data on chemical synthesis collated over decades of research.

Synthia provides invaluable information that can point towards the best possible route to execute – minimising cost, the number of steps, and the best chance of making the required molecule with the desired properties. This can dramatically reduce the time it takes for a chemist to think of a viable route to embark on in the lab.

The tool works by harnessing the potential of advanced algorithms powered by more than 100,000 hand-coded reaction rules – painstakingly sifting through retrosynthetic possibilities while at the same time examining what has been done, what could be done, and what starting materials are available.

Innovative manufacturing

PAPERLESS LAB ACADEMY KEYNOTE SPEAKER TONI MANZANO SHARES HIS THOUGHTS ON THE USE OF AI IN PHARMACEUTICAL MANUFACTURING

Can you talk about your experience in pharmaceutical manufacturing?
I am the co-founder of Aizon, which my partners and I created eight years ago in Silicon Valley. We have one aim – improve pharmaceutical industry processes using modern technologies, including big data and AI capabilities. We want to enable scientists to be sure that every drug is delivered at the right time, for the right patient.

In the pharmaceutical industry, there are a lot of inefficiencies. And the reason is that the pharmaceutical industry earns a lot of money every year, so there is not any kind of urgency to change the existing process. But it is very inefficient. Ultimately, these inefficiencies are passed along with the price of the drugs. The final users, in this case, patients, are paying for these.

We created Aizon to change this mindset, increase knowledge, and optimise processes in pharmaceutical manufacturing to ensure that every batch created is the best. We knew we could only do that using AI.

Is the pharmaceutical industry slow to adopt change?
I wrote an article a year ago where my colleagues and I described and calculated the gap between the pharmaceutical industry adopting innovation compared with other industries. We found that the pharmaceutical industry, on average, is 11 years behind. Sometimes you might see the pharmaceutical industry pointing to the regulator as the cause of a lack of innovation. The shocking result was that regulators are faster than the industry to adopt innovation.

We're spreading the word. We are trying to teach people at conferences, publishing articles and so on in order to democratise access to this knowledge. Nevertheless, I would like to highlight something important. In comparison with the other industrial revolutions, the difference with this one is society is driving this new change. In the previous industrial revolutions, industries drove innovation. Society adopted innovations that they introduced.

But in this fourth industrial revolution, it is society that is using 5G, it is society that is using AI everywhere. Starting with the smartphone, we are using augmented reality, virtual reality, big data, and cloud technologies. Every day, society is using and emphasising the need for these technologies. I firmly believe that industry, in this case, the pharma industry, will adopt AI because society is driving that adoption.

Today, pharma organisations are using Microsoft Teams, Google Meet, Zoom, or shared drives, and all these things are in the cloud. So they are adopting that because the rest of society is pushing.

OEMs that provide equipment such as bioreactors, filling lines, and packaging lines are already providing the equipment, with sensors connected, or the capability to connect, to the cloud. Many opportunities are provided to laboratories: the ability to control the bioreactors to detect potential anomalies before they happen; to predict anomalous behaviours in the reaction. As these kinds of opportunities increase in number, eventually the pharmaceutical industry will be convinced it has to change.

How can AI improve GXP compliance?
How to improve the GXP in general? If we can automate the control of all the processes, scientists and lab managers can be sure that everything is performed as expected.

We are improving this inefficiency, and we are improving the quality behind the GXP requirements. AI does not replace humans. It enhances human capabilities and allows humans to be dedicated to delivering higher value than just analysing things that machines can do better. Inspection is another example. AI systems now interact with filling lines or vials to detect anomalies. This is another way to improve the quality.

Lab managers consider manufacturing in pharma operations as something that is constant. This is not real. The raw material can vary depending on the provider; the processes depend a lot on the instrument and interaction, the set-up with operators and so on. Everything is changing. So why not consider the variations, the variability in the process? AI brings a lot of opportunities to improve GXP conditions.

Humans interact with a reactor based on recipes. But these recipes may not consider all the potential variabilities on raw material, sugar, the material that we are using to deliver insulin or many other raw materials that we are using to acquire this insulin or, for example, lipids, proteins and vaccines.

When humans interact with these static recipes, they try to accommodate what happens during a 15-day bio-reaction as best they can with their own knowledge.

But the process is long, there are different shifts, and there are a lot of human interactions with the process depending on each person's know-how. These inefficiencies come when we have different behaviours, attitudes, or interactions with batches, based on people's experience.

How is compliance changing over time?
In March, the US Federal Drug Administration (FDA) proposed a draft to implement AI in drug manufacturing. It aims to evaluate the technology and to try to introduce it in a regulated way. The FDA is moving ahead to promote technology that the pharma industry is still considering.

Two years ago, the FDA published the guidelines to implement AI in medical devices. The FDA is pushing in this direction, to use AI to have more knowledge, more knowledgeable actions, and improve scientific research and manufacturing.

The UN has different workshops to teach the pharmaceutical industry how to use AI in the right way, using good practices. The regulatory bodies are ready to help. The problem is they can only do something if the pharmaceutical industry adopts the technology in an industrial context.

The regulatory bodies are ready, the administration is ready, and the compliance is ready. The problem is that the industry doesn't want to move forward.

Toni Manzano is the co-founder and Chief Scientific Officer at Aizon. The company has designed a cloud platform that provides simple solutions for analysing and improving industrial processes in biotech and pharma manufacturing.

Fully charged

SIMULATION SOFTWARE IS HELPING TO ACCELERATE BATTERY DEVELOPMENT, WRITES GEMMA CHURCH

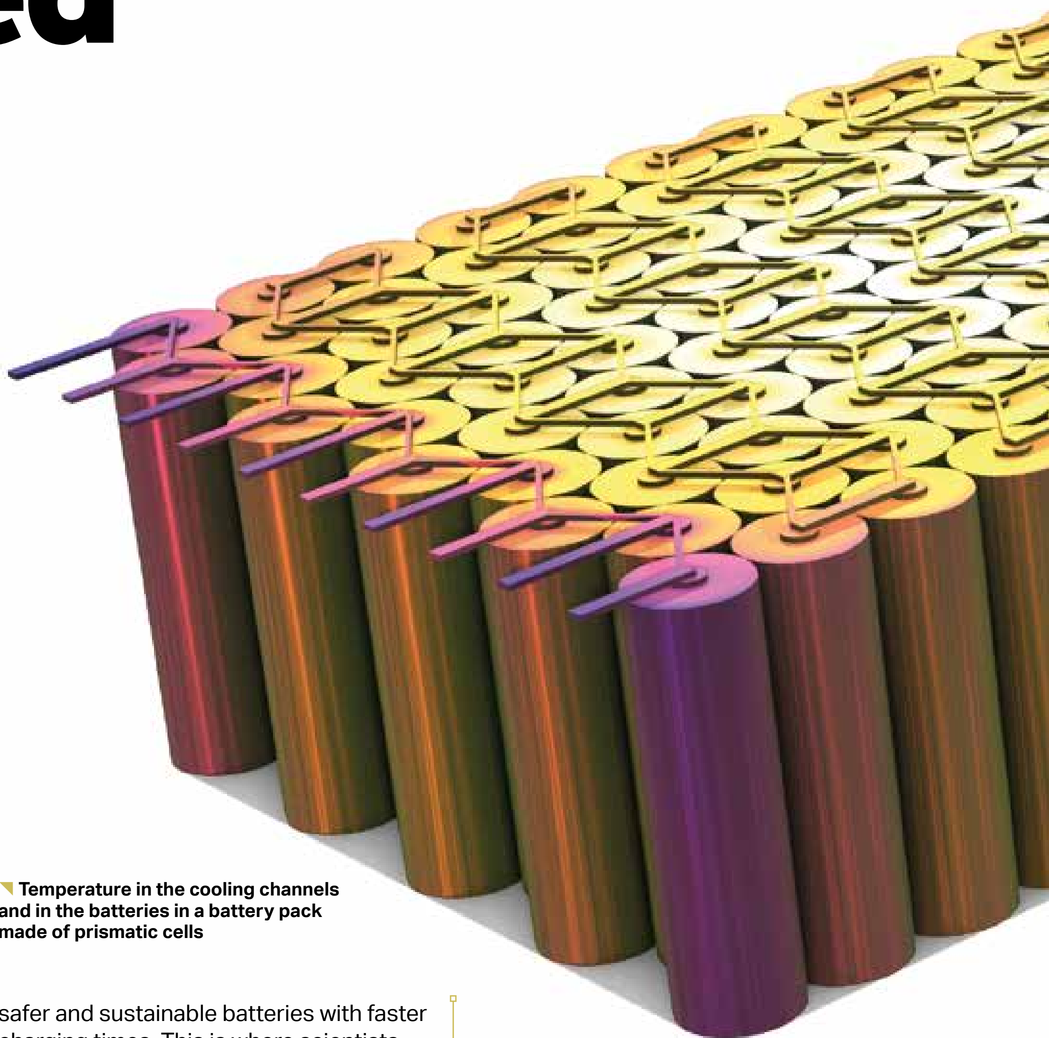
Batteries are complex devices and their electrical, chemical, thermal and structural properties must be considered during development. It's a tough, multi-physics balancing act that simulation and modelling helps to address, providing engineers and scientists with key insights and information throughout the development process and across varied length scales.

At the atomic level, for example, scientists often use simulation to explore combinations of new materials and novel technologies. Vaida Arcisauskaite, Senior Marketing Specialist at Synopsys, says: "As new materials are needed to meet evolving standards for both sustainability and performance, atomistic materials modelling with Synopsys QuantumATK helps battery designers explore the countless combinations of materials and select the most promising ones without performing experiments, building, or testing batteries for each candidate, thus shortening development time and cost."

These novel battery materials could be used in the cathodes and anodes, liquid and solid electrolytes, additives and solid electrolyte interphases (SEI) found in many batteries, for example.

Lithium-ion batteries dominate the market, powering portable devices such as smartphones and laptops as well as larger-scale batteries found in electric vehicles (EVs) and those used in renewable and grid energy applications to store generated energy.

While lithium-ion batteries are relatively cheap and rechargeable, there is a need to develop longer-lasting, more reliable, safer and sustainable batteries with faster charging times. This is where scientists are either focusing on advancing current lithium-ion technologies or on developing alternative technologies, such as solid-state batteries.



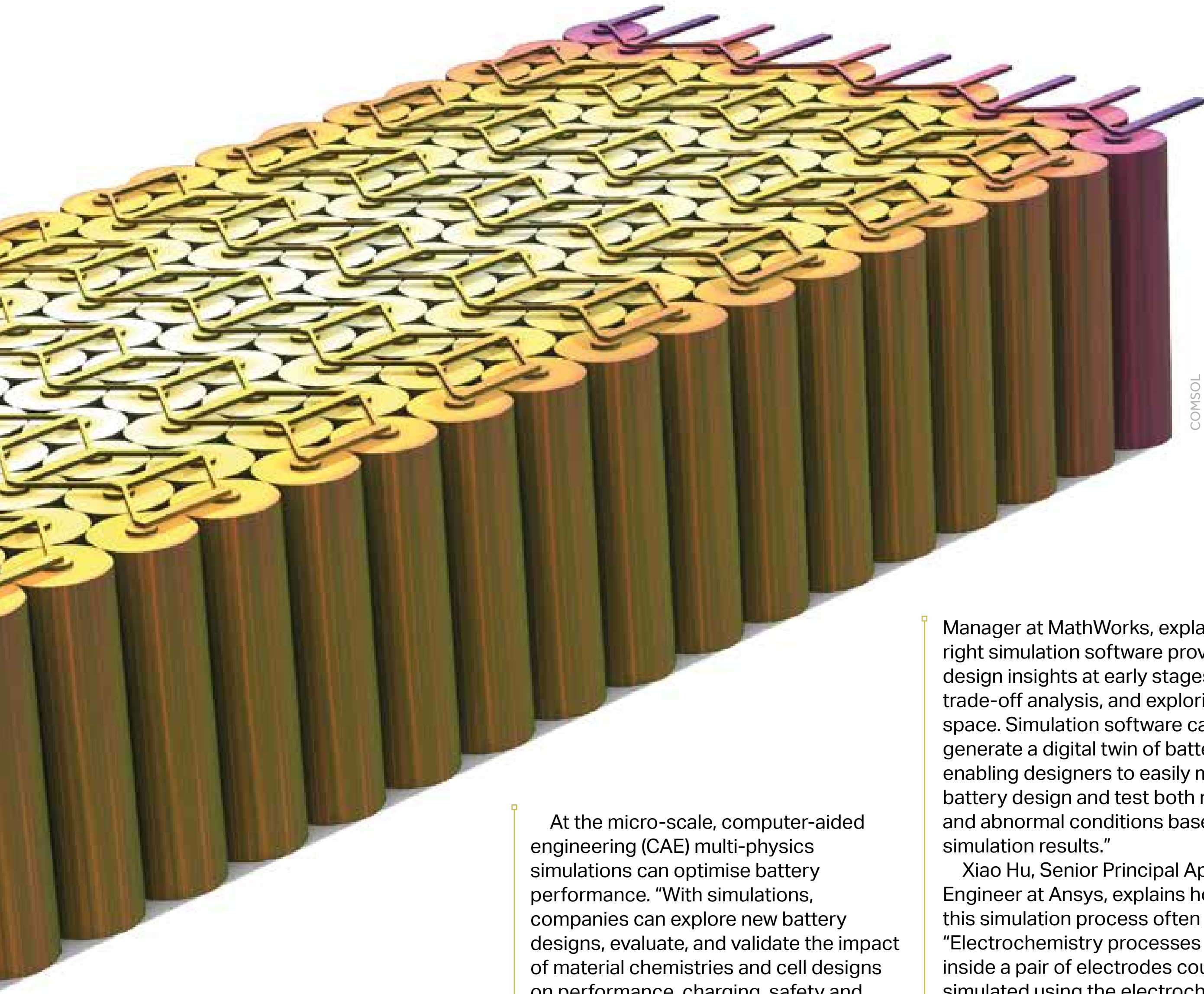
Temperature in the cooling channels and in the batteries in a battery pack made of prismatic cells

Anders Blom, Solutions Engineer, Senior Staff, at Synopsys, says: "Solid-state batteries in particular hold a lot of promise. Even though they are of similar cost to lithium-ion batteries, they are much safer (non-flammable), more reliable, have a higher capacity and faster charging time, require fewer raw materials, and allow for more charging cycles before degradation starts to happen.

"Still, there are challenges that designers need to account for, including the suppression of dendrites which can lead to safety issues, lower mechanical stability during cycling, and electrical resistance."

Lithium-air and lithium-sulphur battery technologies provide "increased capacity, lower cost, and environmental friendliness," according to Soren Smidstrup, Director of R&D at Synopsys. "However, there is still a lot of R&D needed to address challenges such as improving the stability and cycle life of lithium-sulphur batteries and prolonging the life span and improving the stability of lithium-air batteries."

Sodium and manganese ion batteries are less flammable, safer, more abundant, and easier to extract alternatives to lithium-ion batteries. Smidstrup added: "However, due to the larger size and thus lower mobility of sodium and manganese



ions, there is a need to find suitable electrolytes for efficient ion transport and improve charging rates and capacity."

This is one of the reasons why atomic-scale modelling is key to helping explore new battery materials and novel technologies. These tools allow for faster and cheaper development of new batteries at the atomic scale, and their associated chemistries. This work is vital to increase the energy density of batteries, helping scientists create faster-charging batteries that are also safer and cheaper to develop.

Puneet Sinha, Senior Director of Battery Industry at Siemens Digital Industries Software, explains: "Maximising the energy density of a battery pack by design innovations can maximise the number of cells that can be packed in a given volume without sacrificing battery thermal and structural integrity. Cell-to-pack and cell-to-chassis technologies are such examples."

At the micro-scale, computer-aided engineering (CAE) multi-physics simulations can optimise battery performance. "With simulations, companies can explore new battery designs, evaluate, and validate the impact of material chemistries and cell designs on performance, charging, safety and ageing much faster," according to Sinha. "They can develop reliable algorithms to measure state of charge, state of health for batteries and ensure safe and secure implementation of these algorithms in their battery management system."

Simulation software enables the evaluation of battery designs and battery management systems (BMS) in a system environment, which is faster, safer, and more cost-effective than building and testing physical hardware prototypes.

Danielle Chu, Product Marketing Manager at MathWorks, explains: "The right simulation software provides great design insights at early stages, that is trade-off analysis, and exploring design space. Simulation software can also generate a digital twin of battery systems, enabling designers to easily modify battery design and test both normal and abnormal conditions based on simulation results."

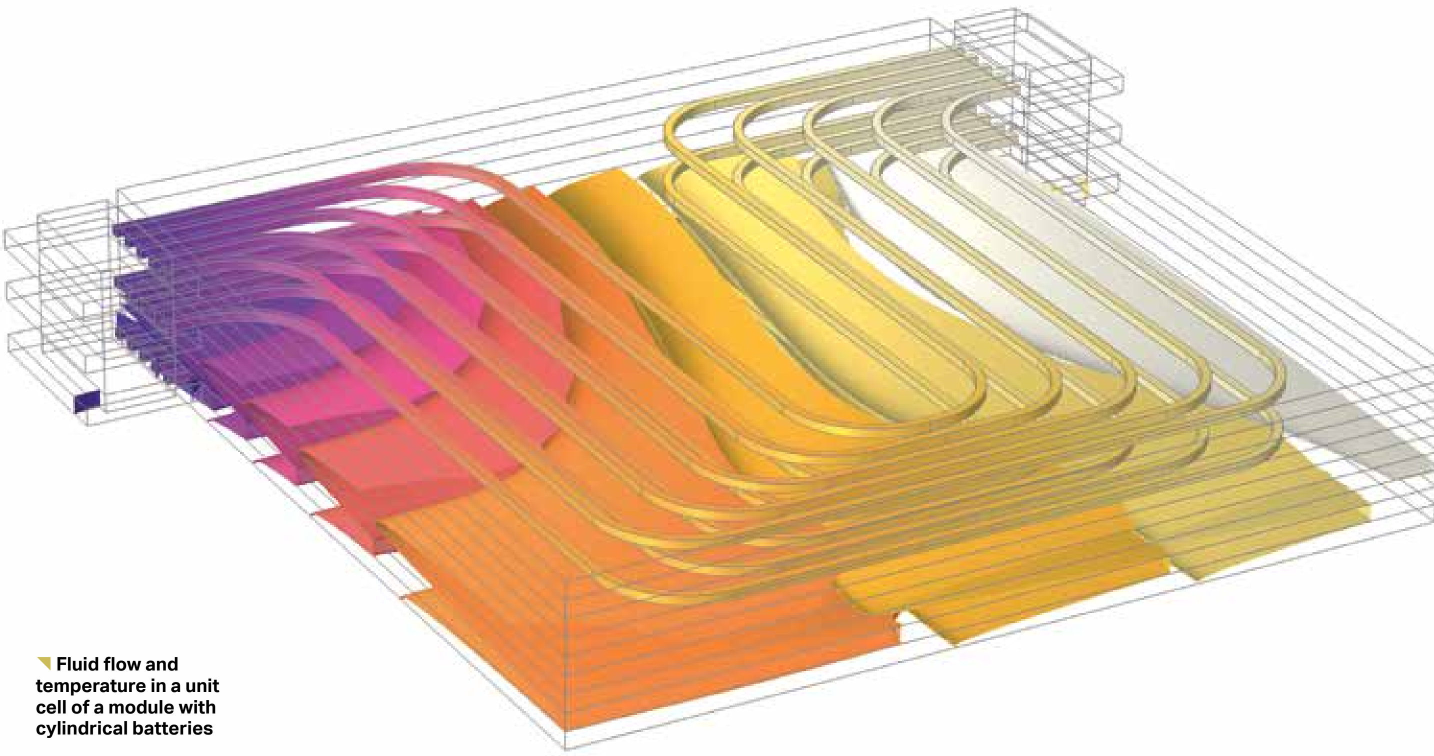
"Atomistic materials modelling with Synopsys QuantumATK helps battery designers explore the countless combinations of materials"

Xiao Hu, Senior Principal Application Engineer at Ansys, explains how this simulation process often works: "Electrochemistry processes occurring inside a pair of electrodes could be simulated using the electrochemistry model. Such a model could help cell manufacturers understand the electrochemical phenomena inside a cell including ageing, and the impact of different chemistry on cell performance."

When we get to the battery module/pack level, different cooling system designs, including air cooling, liquid cooling, and even immersion cooling, could be evaluated by simulation, according to Hu. "The impact of different thermal management designs in the case of thermal runaway could also be evaluated using simulation software."

Such work provides valuable insights that are tough to get from testing alone. says Sinha: "For instance, battery safety testing, though absolutely critical, can only tell if a cell or pack passes the safety test but doesn't necessarily offer why it failed, if it failed, and what to tweak to meet the safety requirements. Battery thermal runaway simulations provide these insights to accelerate material/design optimisation and validation."

→



Fluid flow and temperature in a unit cell of a module with cylindrical batteries

For example, electrochemical impedance spectroscopy (EIS) could be used with modelling and simulation to determine the parameters responsible for cell deterioration and ageing in a specific battery cell. Engineers at Comsol recently used simulations to determine how different battery designs and operating conditions are susceptible to ageing.

"The alternative would be to open the cell and examine the possible changes in the structure, which would be a very difficult and expensive procedure," according to Ed Fontes, VP of development at Comsol. "It is more efficient to combine experimental measurements with modelling and simulation for characterisation to determine the internal status of a battery than trying to investigate the battery post-mortem. If a battery is failing, the same methods may be used for diagnosis."

Another way modelling and simulation can be used in battery development is to determine the utilisation of the battery electrodes during charge and discharge. This provides information about how the cell should be designed and operated for a uniform utilisation and, eventually, for a longer life.

"It also provides a detailed distribution of the state-of-charge of a cell at any given point of operation, which is very difficult to achieve without modelling and simulation," Fontes adds. "Using a digital twin to track a battery or battery pack is an efficient method for continuously gathering and quantifying information about its status."

"Simulation software can also generate a digital twin of battery systems, enabling designers to easily modify battery design and test both normal and abnormal conditions"

At the macro-scale, simulations can help companies optimise their battery packing and manufacturing processes to reduce scrap rates and accelerate time-to-market.

Hu explains: "Predictive maintenance can be achieved through the Ansys framework of high-end 3D simulations and physics-based digital twin (ROM or reduced order model) technologies. Hybrid analytics, by using physics simulation and data-driven models in Ansys, can help to create lifetime modelling."

These simulations can be scaled up to cover entire production plants. Sinha adds: "With Siemens digital manufacturing solutions, companies can design plants, optimise and validate cell and pack production processes virtually before implementing on the factory floor, thereby de-risking investment and shortening scale-up time."

Driving development

Advancing electrification is driving most of today's battery developments, with the automotive industry leading the charge. Chu says: "Automotive (electric vehicles, hybrid electric vehicles, and automobiles), is a big market for lithium-ion battery innovation due to the increasing number of electric vehicles and favourable government policies.

"Energy storage and consumer electronics also have a high demand for battery usage, which tends to drive innovation. The main innovation we see is in both electrical management and thermal management. Both aspects are critically important to improve overall efficiency and safety and to extend the operational life of the battery system."

In automotive, many manufacturers are targeting the problem of recharging speeds. This is because drivers are demanding electric vehicles that recharge in times comparable to the time taken to fuel an internal combustion engine gasoline car. But drivers also want that recharge to last for hours. This is a big ask.

"The difference in magnitude of the recharge and discharge currents is a problem," Fontes explains. "The battery, heating, cooling, current collectors and feeders, and many other components have to be able to cope and operate at very different conditions during recharge and discharge. This problem remains unsolved. Therefore, a large part of battery research is dedicated to addressing it."

The aerospace industry is also exploring the use of batteries for hybrid-electric and all-electric aircraft, satellites, and spacecraft. Here, the market demands high-energy-density batteries with a compact size and low weight.

"The growth of EV and consumer electronic markets are driving the research and development of battery technologies that will benefit other industrial applications as well," says Smidstrup. "They demand batteries that offer higher capacity (thus, longer range), faster charging, and improved safety. In return, novel battery technologies, such as solid-state, could significantly improve the performance and safety factors, and become a game changer in exponentially growing the EV market globally."

"One obvious use is to create very compact and fast digital twins to make predictions during the operation of battery systems"

Innovation in wearable/implanted medical devices, such as battery-powered pacemakers, hearing aids, insulin pumps, and neurostimulators, also require the development of even smaller and more efficient batteries.

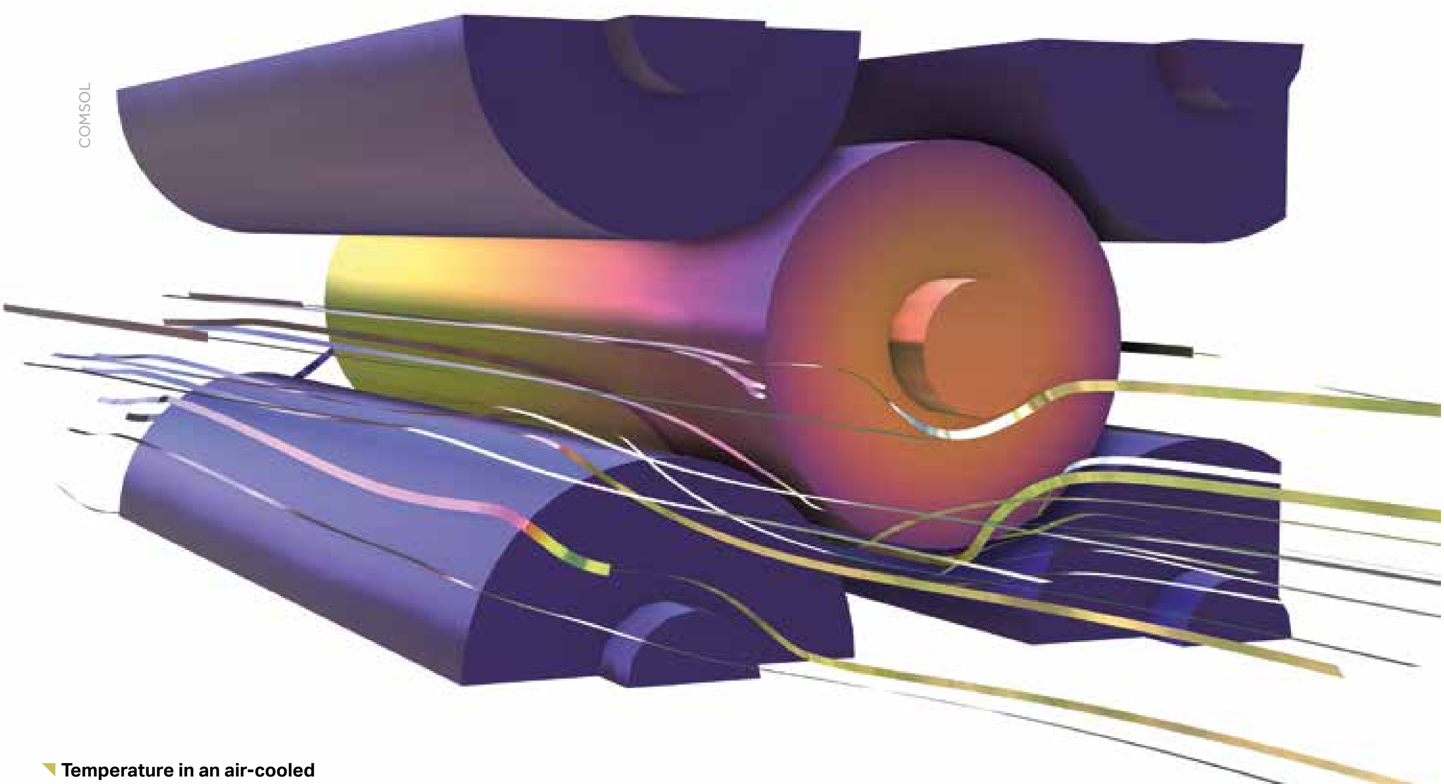
However, wind and solar power are "probably the largest future areas of application", according to Fontes: "These are so-called 'intermittent' sources of electricity. A possible future application is to use large battery systems to stabilise the grid. This would mean improving the quality of grid electricity, not necessarily storing electricity for the purpose of selling it, since doing so would require very large systems."

To develop tomorrow's batteries, simulation and modelling tools will come to increasingly rely on artificial intelligence (AI) and machine learning (ML).

But there's "a lot of buzz" around both, according to Fontes. "Almost every research institution working with batteries has programs for applying AI to batteries, mainly to predict the status of a battery based on data from operation history. We envision AI being used for different purposes.

"One obvious use is to create very compact and fast digital twins to make predictions during the operation of battery systems," he explains. Then, modelling and simulation can be used together with training data to train an AI. This, in principle, is "not very much different" from the traditional validation processes that you will find in parameter estimation and optimisation, where a subset of data is used to train the software. Then another set of data is used to validate the predictions.

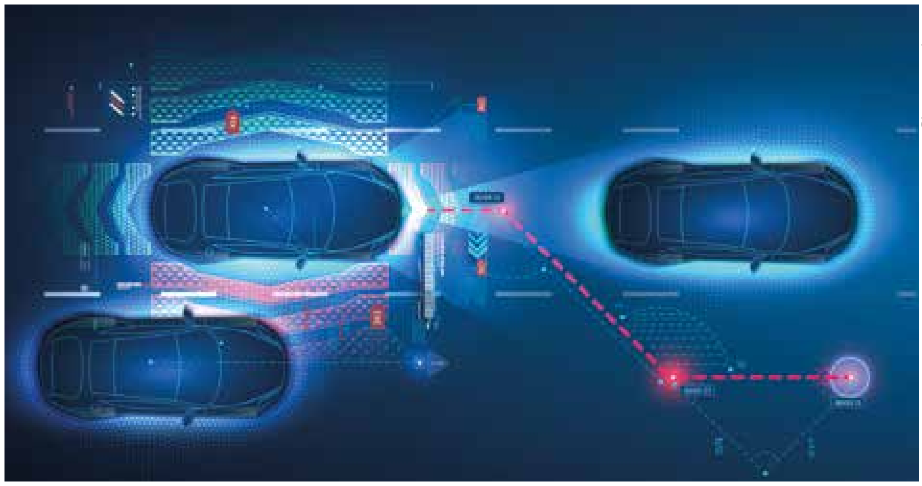
Fontes adds: "The difference is that we would invest large amounts of computing power in developing and training the AI. Once this is done, an AI-based model could be very fast and compact. It could be used for real-time decision-making on an operating battery. It could also help make decisions regarding the selection of materials and for characterisation, diagnosis, and other methods involved in battery design and operation."



Temperature in an air-cooled battery module in a battery pack for automotive applications

Autonomous systems drive innovation

AUTOMATION IS HELPING TO STREAMLINE ENGINEERING PROCESSES AND OPENING UP NEW OPPORTUNITIES FOR PRODUCT DEVELOPMENT, WRITES ROBERT ROE



Shutterstock.com

Autonomous engineering is changing the way that products are created and driving a shift in the way organisations look at internal processes and technology to support continued innovation.

From the development of new autonomous vehicles to additive manufacturing, automation is supported by digital transformation initiatives. This new wave of technological advancement is fuelling a new era of engineering that offers unprecedented opportunities to build new products and update existing processes that can better facilitate innovation.

Three main factors are seemingly driving changes in the automotive industry. First, regulators worldwide are forcing the creation of more stringent targets for greenhouse-gas emissions. The European Union (EU) has, for instance, unveiled its Fit for 55 programme, which seeks to align climate, energy, land use, transport and taxation policies to reduce net greenhouse-gas emissions by at least 55 per cent by 2030. In the US, President Joe Biden's administration wants sales of EVs to account for 50 per cent of total vehicle sales by 2030. Additionally, consumers are becoming increasingly open to more sustainable alternatives to traditional car travel. Finally, according to consultant McKinsey & Co, companies working to electrify, connect and automate driving technology have attracted more than $400bn in investments over the past decade – about $100bn since the beginning of 2020. This investment is accelerating the pace of change.

In a hugely competitive market such as the automotive industry, organisations must make changes quickly or face being left behind their competitors in a changing world of AI, automation and connected digital systems.

Autonomous vehicles

If autonomous vehicles are ever to achieve widespread adoption, engineers need to know they can navigate complex traffic situations, such as merging into heavy traffic when lanes disappear on a highway. North Carolina State University researchers have developed a technique that allows autonomous vehicle software to make the relevant calculations more quickly – improving both traffic and safety in simulated autonomous vehicle systems.

In a research paper, "Distributed Cooperative Trajectory and Lane changing Optimization of Connected Automated Vehicles: Freeway Segments with Lane Drop," published in the journal *Transportation Research Part C*, the researchers from North Carolina State University discussed how this research could improve road safety.

Ali Hajbabaie, corresponding author of the paper and an assistant professor of civil, construction and environmental engineering at NC State, says: "The [existing] programmes designed to help autonomous vehicles navigate lane changes rely on making problems computationally simple enough to resolve quickly so that the vehicle can operate in real time. However, simplifying the problem too much can actually create a new set of problems, since real-world scenarios are rarely simple.

Hajbabaie says the NS State researchers' approach allowed them to embrace the complexity of real-world problems: "Rather than focusing on simplifying the problem, we developed a cooperative distributed algorithm. This approach breaks a complex problem into smaller problems and sends those to different processors to solve separately. This process, called parallelisation, improves efficiency significantly."

"Our approach allows us to embrace the complexity of real-world problems"

The researchers have tested their approach in simulations, where the problems are shared among different cores in the same computing system. However, if autonomous vehicles ever use the approach on the road, the vehicles would network with each other and share the computing sub-problems.

In initial testing, the new technique did exceptionally well. In some scenarios, particularly when traffic volume was lower, the two approaches performed approximately equally. But in most instances, the new approach outperformed the previous model considerably. The new technique had zero incidents where →

vehicles had to come to a stop or where there were "near crash conditions". The other model's results included multiple scenarios where there were thousands of stoppages and near-crash conditions.

"For a proof-of-concept test, we're very pleased with how this technique has performed," Hajbabaie says. "There is room for improvement, but we're off to a great start. The good news is that we're developing these tools and tackling these problems now, so that we're in a good position to ensure safe autonomous systems as they are adopted more widely."

Modern automotive development

However, in the automotive industry, it is not just the software systems that must keep pace with technological advancement. The organisations themselves are also changing to improve the design and delivery of new products.

In March, Jaguar Land Rover (JLR) announced that it had partnered with Tata Technologies to accelerate its path towards digital transformation. The first phase will include UK core production facilities, with solutions deployed to other global locations.

Tata Technologies will deliver end-to-end integrated enterprise resource planning (ERP) to transform JLR's manufacturing, logistics, supply chain, finance and purchasing modules by bringing data and knowledge from multiple departments into one single source.

JLR also recently announced an update to its partnership with Nvidia to develop three new tech hubs. This project will further expand JLR's digital capability, creating nearly 100 new engineering jobs.

The hubs are being set up to develop autonomous driving systems for JLR's next generation of modern luxury vehicles and come in addition to the six existing global tech hubs JLR has invested in. The new hubs will be located in Munich, Germany; Bologna, Italy; and Madrid, Spain, with the locations chosen because of the availability of digital engineering skills in the area.

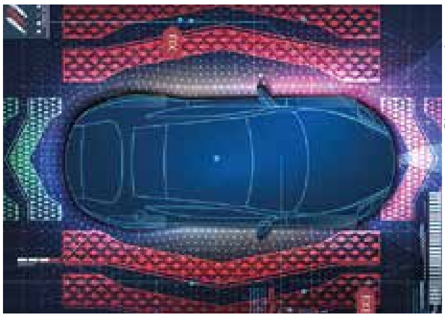
The ERP software is designed to streamline and consolidate information across core business areas. Creating a dedicated home for data and management processes across departments is fundamental to the transformation of how a business functions.

ERP systems can also aggregate, store and interpret data to provide insights instantly, helping companies respond to issues, challenges and opportunities in an agile and timely manner.

Warren Harris, MD and CEO of Tata Technologies, says: "We believe this collaboration will help JLR to build innovative and sustainable vehicles and achieve faster time to market using new-age digital technologies."

Anthony Battle, Chief Digital and Information Officer, JLR, adds: "JLR's digital transformation will play a pivotal role in the business becoming more agile as part of Reimagine and fulfilling its potential as a tech leader. As part of our Digital 2024 programme, we are transitioning to a cloud-native digital landscape. The partnership with Tata Technologies enables us to take another important step on that journey."

The expanded partnership with Nvidia focuses on developing tech hubs to further bolster skills and expertise in autonomous vehicle development in Europe. Thomas Müller, Product Engineering Director at JLR, says: "Software is essential for us to deliver a fully connected experience for our clients and creating global engineering hubs will enable seamless hybrid working across several locations.



Shutterstock.com

"Software is essential for us to deliver a fully connected experience for our clients"

"We are harnessing talent in autonomous technologies around the world to develop new autonomous technologies for our future products, which will deliver a truly modern luxury experience for our clients," Müller adds.

Under the partnership, JLR engineers are working to develop and deliver next-generation automated driving systems, and digital services. While the two activities may seem loosely connected, digitalisation or digital transformation is necessary to transform business operations. To develop AI-based maintenance systems or to further automate the creation of new components, interconnected digital systems that facilitate communication need to be in place. A process needs to be fully understood before it can be automated, but beyond that, the technology and data must support any advances in business agility or decision-making. Like autonomous vehicles, any action or decision taken by an autonomous system can only be as accurate as the data used to make that decision.

For true automation in an area such as product development, teams need to be fed information from different departments. For example, a failure in manufacturing or validation that is instantly available and analysed by digital systems can accelerate any changes that design teams need to make.

Automating manufacturing

Automation is not just reserved for the automotive industry. Increasingly it is being used across different market verticals and is now making its way into the manufacturing of components with additive manufacturing. Imperial College London and manufacturing software company AMFG received funding in February 2023 from Innovate UK to advance autonomous 3D printing.

The funding, delivered through a management Knowledge Transfer Partnership (mKTP), will go towards developing software 3D printers that can function entirely independently.

Additive manufacturing, or 3D printing, is an industry that allows users to create 3D objects from digital designs. Currently used in a variety of industries, including aerospace, medical devices, dentistry, and mechanical engineering, 3D printing has the potential to produce materials in large quantities as they become cheaper and easier to manufacture using this method.

Imperial College researchers will work with AMFG, which produces workflow automation software for manufacturing processes, to develop autonomous manufacturing. Academic lead Dr Connor Myant, from Imperial's Dyson School of Design Engineering, says: "Establishing strong collaborations with industry is central to modern academia, and together we can solve real-world problems. Teaming up with AMFG, we aim to do exactly that. Our mKTP will create exciting opportunities to build research activities across the digital manufacturing spectrum."

Central to the mKTP is the transfer of innovation and collaborative creativity skills required to transition to a culture favouring disruptive innovation. Academic co-lead Dr Céline Mougenot, from Imperial's Dyson School of Design Engineering, will support AMFG with implementing collaborative ideation workshops with cross-functional teams. The partnership aims to produce software that includes automated order management, integration with all major 3D printers, post-production management, and shipping and labelling. To do this, they will run a series of research and development initiatives that develop new solutions for each part of the workflow.

Tech Focus: Autonomous vehicle software

A ROUND-UP OF SOFTWARE USED TO DESIGN AND VERIFY AUTONOMOUS DRIVING SYSTEMS

Autonomous vehicles, capable of sensing their environment and operating without human involvement, could help to prevent road accidents and save billions in damages across the world each year. Carmakers, engineering software providers, AI chipmakers and start-ups are all developing solutions to accelerate development and make autonomous vehicles a reality.

Advanced driving assistance systems (ADAS) are helping to lay the groundwork, but automotive software increasingly covers electronics, sensors and optics and widespread connectivity to share information across other vehicles on a network.

The Society of Automotive Engineers (SAE) has clearly defined guidelines for automotive "autonomy levels" from Level 1 "Driver Only", to Level 5 "Fully Automated". Current ADAS systems are estimated to be able to prevent more than a third of all passenger vehicle crashes. These systems generally follow SAE guidelines Level 2 "Assisted Driving."

The benefits of automated driving are clear, but real-world vehicles must overcome technical hurdles to reach Level 5 autonomy. The automotive industry needs to adopt advanced software and integrate that with the design of new vehicles and components. Automotive companies must design and verify tools that can help integrate sensors, connectivity, mapping and the design and validation of optical systems and computing frameworks into vehicle designs.

Autonomous vehicle design and verification software on the market

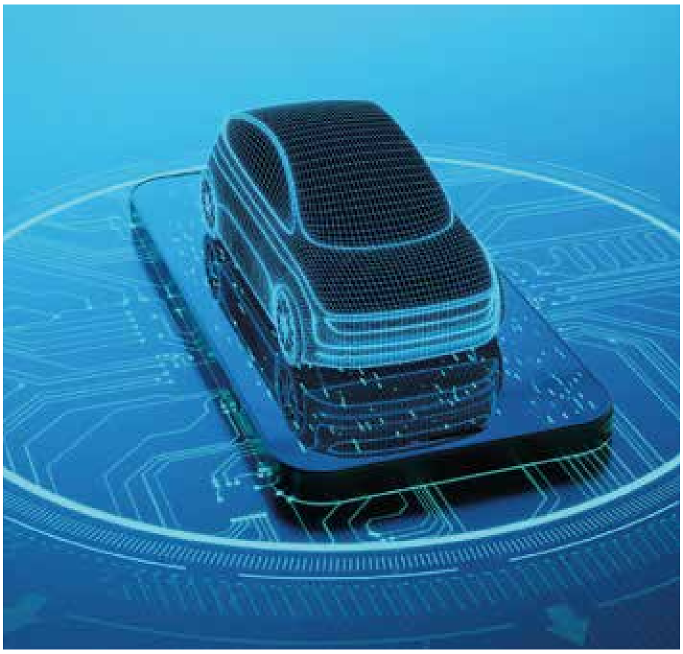
Autonomous vehicles (AVs) and ADAS bring increased complexity and a need for more testing. Exploring all the required scenarios within product development timing requires advanced simulation. **Altair** enables customers to deliver solutions that make cars and trucks safer today and on the road to driverless mobility.

Altair offers solutions in a wide array of engineering for antennas, from design to placement to communication.

Altair Feko efficiently and accurately simulates radar antenna design as well as the integration aspects, including radome and bumper effects. Feko also provides a solution for ultrasonic sensors, and the Altair Partner Alliance (APA) offers access to TracePro for lidar modelling.

Altair PollEx provides integrated printed circuit board (PCB) design viewing, analysis and verification tools for electrical, electronics, and manufacturing engineers. PollEx transfers data between popular ECAD and simulation tools and lets engineers quickly visualise and review PCB designs. Its checking tools detect issues early in the design to avoid product failures and simplify manufacture and assembly.

Ansys Speos predicts systems' illumination and



cono0430Shutterstock.com

optical performance to save on prototyping time and costs while improving a product's efficiency.

Ansys HFSS is a 3D electromagnetic (EM) simulation software for designing and simulating high-frequency electronic products such as antennas, antenna arrays, RF or microwave components, high-speed interconnects, filters, connectors, IC packages and PCBs. Engineers worldwide use Ansys HFSS software to design high-frequency, high-speed electronics found in communications systems, ADAS, satellites, and internet-of-things (IoT) products.

Ansys AVxcelerate realistic sensor testing and validation enables users to test their autonomous vehicles, ADAS and sensors faster than with physical prototypes.

The **Autoware Foundation** is a non-profit organisation supporting open-source projects enabling self-driving mobility. The Autoware Foundation creates synergies between corporate development and academic research, enabling autonomous driving technology for everyone.

The Autoware defined Open AD Kit brings the best practices of software defined vehicle (SDV) development to the Autoware ecosystem. The Open AD Kit promises to accelerate the development of optimised hardware and software solutions for autonomous driving. The kit includes a complete AD reference platform for the rapid development of cloud-to-edge solutions and a cloud-native development and verification of AD solutions with system-level parity to edge platforms.

The kit also includes a micro-services architecture orchestrating mixed critical workloads across heterogeneous compute platforms and an open ecosystem system of hardware/software components and development tools.

The open **Nvidia** DRIVE SDK gives developers all the building blocks and algorithmic stacks needed for autonomous driving. It allows developers to build and deploy various AV applications more efficiently, including perception, localisation and mapping, planning and control, driver monitoring, and natural language processing.

The foundation of the DRIVE Software stack, DRIVE OS is a safe operating system for in-vehicle accelerated computing. It includes NvMedia for sensor input processing, Nvidia CUDA libraries for efficient parallel computing implementations, Nvidia TensorRT for real-time AI inference, and other developer tools and modules to access hardware engines.

Nvidia DriveWorks provides middleware functions on top of DRIVE OS that are fundamental to autonomous vehicle development. These consist of the sensor abstraction layer (SAL) and sensor plug-ins, data recorder, vehicle I/O support, and a deep neural network (DNN) framework. It's modular, open, and designed to comply with automotive industry software standards.

The DRIVE AV software stack contains the perception, mapping, and planning layers and diverse DNNs trained on high-quality, real-world driving data. These rich perception outputs can be used for both autonomous driving and mapping. In the planning and control layer, the Nvidia Safety Force Field computational module keeps a vehicle out of harm's way to ensure that it won't contribute to or cause an unsafe situation.

Nvidia DRIVE Chauffeur is an AI-assisted driving platform based on the Nvidia DRIVE AV SDK that can handle both highway and urban traffic with increased safety. It can use the high-performance compute reference architecture and sensor set of Nvidia DRIVE Hyperion 8 to drive from address to address. For those who want to drive, the system also provides active safety features and the ability to intervene in dangerous scenarios.

Siemens Accelerated Product Development is a suite of tools and solutions designed to help automotive manufacturers to speed up the product development process. It includes a range of software tools, simulation and testing capabilities, and data management solutions.

Siemens Accelerated Product Development includes a range of software tools, such as NX for product design and engineering, Simcenter for simulation and testing, and Teamcenter for data management and collaboration.

Siemens Accelerated Product Development can help automotive manufacturers reduce time-to-market, improve quality, increase efficiency, and reduce costs. It does this by enabling the digitalisation of the product development process, enhancing collaboration between teams, and providing advanced simulation and testing capabilities.

Siemens Teamcenter enables engineering teams to use digital twins to connect and optimise designs, systems, software, simulation and visualisation processes.

Synopsys' comprehensive Safety-Aware Solution helps engineers automate the design and verification of functional safety systems-on-chips (SoCs). The Safety-Aware Solution helps to make automotive SoC designs highly predictable so that you can reach your chip's target automotive safety lLevel (ASIL).

With the rapid shift to ADAS and autonomous driving, carmakers are transitioning to a more centralised domain architecture with central compute. Synopsys believes autonomous driving compute solutions will become more centralised within SoCs to rapidly fuse multimodal sensor data and safely choose and initiate self-driving actions.

The SoC helps to ensure safe and secure operation. To design SoCs that meet these requirements, system architects integrate automotive-grade IP and ISO 26262-certified safety-aware test, design implementation, and hardware verification solutions while consulting with automotive SoC experts. Selection of design flows and IP can simplify meeting functional safety objectives and achieving power-performance-area (PPA) targets.

Suppliers' Directory

Find the suppliers you need quickly and easily

www.scientific-computing.com/suppliers

HIGH-PERFORMANCE COMPUTING

Altair Engineering, Inc
(248) 614-2400
info@altair.com
www.altair.com
■ Software

Boston Limited
+44 (0)1727 876 100
sales@boston.co.uk
www.boston.co.uk
■ Systems Integrator

Calyos SA
0032 4 75 73 54 65
newbusiness@calyos-tm.com
www.calyos-tm.com
■ Cooling

Cluster Vision
+31 20 407 7550
marketing@clustervision.com
www.clustervision.com
■ Software

Eurotech Spa
+39 0433 485 411
welcome@eurotech.com
www.eurotech.com
■ Systems Integrator

Numerical Algorithms Group (NAG)
+44 (0)1865 511 245
katie.ohare@nag.co.uk
www.nag.co.uk
■ Software

OCF
+44 (0) 114 257 2200
scw@ocf.co.uk
www.ocf.co.uk
■ Systems Integrator

LABORATORY INFORMATICS

Abbott Informatics
+44 (0)161 711 0340
AISalesUK@abbott.com
www.abbottinformatics.com
■ LIMS
■ Scientific Document Management Systems

Accelerated Technology Laboratories
+1 910 673 8165 (outside of US)
800 565 LIMS (5467) US and Canada
info@atlab.com
www.atlab.com
■ Bioinformatics
■ LIMS

LabWare

Denzell Lodge, Dunham Road
Altrincham, Cheshire,
WA14 4QE
Tel: +44 161 927 5600
infoeu@labware.com | www.labware.com

LabWare is recognised as the global leader of Laboratory Information Management Systems (LIMS) and instrument integration software products. The company's Enterprise Laboratory Platform combines the award-winning LabWare LIMS, LabWare ELN/LES and LabWare Mobile, which enable companies to optimise compliance, improve quality, increase productivity, and reduce costs. A wide range of industry template solutions are also available to provide best practice and help fast-track installations. LabWare is a full-service provider offering software, professional implementation services and validation assistance, training, and world class technical support to ensure our customers get the maximum value from their LabWare products.

ACD/Labs
+1 (416) 368-3435
info@acdlabs.com
www.acdlabs.com
■ Data management
■ Drug discovery
■ Analysis tools

Amphora Research Systems
+44 (0)845 230 0160
info@amphora-research.com
www.amphora-research.com
■ Electronic Laboratory Notebooks

Autoscribe
+44 (0) 118 984 0610
info@autoscribeinformatics.com
www.autoscribeinformatics.com
■ Data Management
■ Electronic Laboratory Notebooks
■ Scientific Document Management Systems

Biovia
+ 1 858 799 5000
www.3ds.com/products-services/biovia
■ LIMS
■ Electronic Laboratory Notebooks

Eusoft
+39 080 5426799
info@eusoft.it
www.eusoft.it
■ LIMS

IDBS
info@idbs.com
www.idbs.com
■ Electronic Laboratory Notebooks

InfoChem
+49 89 2153 7160
info@infochem.de
www.infochem.de
■ Cheminformatics

LabWare
infoEU@labware.com
www.labware.com
■ LIMS

Osthus GmbH
+49 241 943 140
office@osthus.de
www.osthus.de
■ Electronic Laboratory Notebooks
■ Cheminformatics
■ LIMS

Siemens
+1 322 536 2139
industrial-it.swe@siemens.com
www.siemens.com/simaticit-rdsuite
www.siemens.com/industrial-it/lims
■ Electronic Laboratory Notebooks
■ LIMS

Thermo Fisher Scientific
+44 (0)161 942 3000
marketing.informatics@thermofisher.com
www.thermoscientific.com/informatics
■ LIMS
■ Scientific Document Management Systems
■ Chromatography Data Systems
■ Spectroscopy Software

MODELLING AND SIMULATION

Altair Engineering, Inc
(248) 614-2400
info@altair.com
www.altair.com
■ Software