InSiDE

Innovatives Supercomputing
in Deutschland

No 24-2 · Autumn 2024

Stronger Together

Researchers at the University of Stuttgart used three GCS systems simultaneously, running one of the largest-ever simulations done on European HPC infrastructure.

Europe's First Hybrid Quantum System

LRZ is installing Q-Exa, a superconducting 20-qubit system that works with the center's flagship HPC system.

Fighting Misinformation

Researchers from JSC and the University of Bonn are developing methods to detect deep fake audio content.

GCS
Gauss Centre for Supercomputing

Mushrooms and other fungi grow through large, complex, subsurface networks of root-like structures called mycelium. Through mycelium, fungi partner with the roots of plants to create mutually beneficial mycorrhizal networks to help both organisms share resources. Much like fungi, the growth of the larger supercomputing industry comes from sharing resources and knowledge across large groups of different organizations and subject matter experts, and our resources are developed and maintained in order to nourish the greater scientific community in academia and the private sector. This issue is focused on all the ways our centers actively work on forging and expanding these connections, building a healthier HPC ecosystem during this time of change in our industry.

Welcome!

Welcome to the latest issue of InSiDE, the biannual Gauss Centre for Supercomputing magazine showcasing innovative supercomputing developments in Germany. As our industry continues to evolve – becoming more multidisciplinary, technologically diverse, and collaborative in the process – we are using this issue to highlight how our centers are forging connections between the people, technologies, and infrastructure needed to take advantage of these exciting new developments surrounding high-performance computing (HPC).

The Leibniz Supercomputing Centre (LRZ) is celebrating the installation of Q-Exa, Germany's first hybrid HPC-quantum computing system (p. 6). Researchers will benefit from the raw computing power of SuperMUC-NG and the exciting potential of accelerating certain calculations with a superconducting 20-qubit IQM system. The High-Performance Computing Center Stuttgart (HLRS) has long prided itself as being a leader in helping Germany's industry harness the power of HPC resources. Recently, the center has signed several agreements with companies designed to support artificial intelligence (AI) innovations in the private sector (p. 9). The Jülich Supercomputing Centre (JSC) is collaborating with researchers at the University of Bonn to better understand and identify so-called "deep fake" audio content being created with AI tools, helping slow the flow of misinformation online (p. 12).

Our centers are also always working on strengthening the connections we have between ourselves. Dr. Theresa Pollinger, who recently completed her PhD at the University of Stuttgart, led a team that ran the largestever simulation ever done on European infrastructure, as well as a novel experiment that used all three of the GCS centers' flagship systems simultaneously (p. 17). JSC staffers are working closely with their colleagues at other Helmholtz Association organizations to improve so-called foundation models to better leverage the power of AI in solving some of society's greatest challenges (p. 24). LRZ recently brought together a group of multidisciplinary computer and computational scientists to showcase the center's investment in the Cerebras CS-2 system, a novel architecture containing the world's largest computer chip that is well-suited to run large-scale AI applications (p. 31).

We also highlight how connections to our past shape our future. In our profile of Erika Fischer, we detail how her 60 years of dedicated service to HLRS and its predecessors continues to have a positive influence on the center's future (p. 40). The world of HPC has changed drastically in the last several years, but we embrace these changes as exciting opportunities that bring together different ideas, experiences, and expertise that will only help make us stronger.

Prof. Dieter Kranzlmüller
Prof. Thomas Lippert
Prof. Michael Resch

Prof. Dr. D. Kranzlmüller
Director, Leibniz Supercomputing Centre
Co-chair of the GCS Board of Directors

Prof. Dr. Dr. Th. Lippert
Director, Jülich Supercomputing Centre
Chairman of the GCS Board of Directors

Prof. Dr.-Ing. Dr. h.c. Dr. h.c. Hon.-Prof. M. M. Resch
Director, High-Performance Computing Center Stuttgart
Co-Chair of the GCS Board of Directors

Contents

Editorial 3

News Features

LRZ Launches Germany's First Hybrid Quantum Computer 6

HLRS to Boost AI-Based Innovation in Industry 9

Collaboration Between Jülich Supercomputing Centre and the University of Bonn Makes Progress in the Fight Against Deep Fakes 12

Science Features

Star Flyby Simulations Explain Solar System Dynamics Beyond Neptune 14

German National Supercomputers Combined to Run a Single Large-Scale Simulation 17

Next-Generation Earthquake Modelling Highlights Differences in Large and Small Earthquake Dynamics 20

Projects

Helmholtz Foundation Model Initiative Lays the Groundwork for Next-Generation AI Research 24

FFplus Holds First Open Call for HPC and AI Applications in Industry 27

CompBioMed Leaves Legacy Promoting Computational Research in Medicine and Drug Development 29

Events

Cerebras Workshop Showcases Short Paths Between AI Data Processing and Storage 31

How Simulation Could Support Crisis Response 33

JSC Organizes Premier ComputationalFluid Dynamics Conference: ParCFD 2024 35

News Briefs 36

Inside GCS

Staff Spotlight: A Life in Supercomputing, Erika Fischer 40

Education and Training 43

Center Descriptions

High-Performance Computing Center Stuttgart 44

Jülich Supercomputing Centre 46

Leibniz Supercomputing Centre 48

Imprint 50

A part of the Q-Exa project, LRZ recently combined a 20-qubit IQM quantum computer with SuperMUC-NG.

LRZ Launches Germany's First Hybrid Quantum Computer

The system combines traditional HPC architecture with a 20-qubit processor based on superconducting circuits.

A new era has begun at the Leibniz Supercomputing Centre (LRZ): in cooperation with researchers and companies, a quantum processing unit (QPU) has been successfully integrated into a high-performance computing (HPC) system. In July, LRZ presented the world's first operational hybrid system called Q-Exa, in which a 20-qubit processor based on superconducting circuits is connected to LRZ's supercomputer, SuperMUC-NG. The two systems have already begun their first jobs. These test runs show that classical computers can work together with quantum systems and can be integrated even more deeply.

Q-Exa is not installed in a physics lab, but rather in a computing center in the immediate vicinity of SuperMUC-NG and other supercomputers.

Rapid access to quantum computing

The Q-Exa system is the result of the Quantum Computer Extension for Exascale HPC (Q-Exa) project. The German Federal Ministry of Education and Research (BMBF) funded the integration of quantum processors and the development of innovative interfaces, useful control tools, and software into traditional HPC systems with around €40 million. Through innovative co-design and collaboration between science and industry, the aim was to give researchers access to a promising future technology that will enable new computing methods as quickly as possible.

Work on the first hybrid HPC and quantum computer system began in 2022. In addition to the LRZ, the European-based hardware manufacturer IQM Quantum Computers, which developed the quantum processor based on superconducting circuits, was also involved. The software specialists at HQS Quantum Simulation in Karlsruhe and the technology provider Eviden also worked on Q-Exa by providing a quantum simulator, Quaptiva, for the preliminary integration work. The integrated quantum emulators were first used to simulate and implement a digital twin of the 20-qubit IQM chip. This allowed the necessary circuits between the two computing technologies to be developed, tested, verified, and progressively harmonized. This digital twin will allow the hybrid system to be further optimized in the future and will also help the user community to better understand how the hybrid computing system works so that users can develop and test their own codes or algorithms.

Software for the first hybrid system

In addition to the hybrid Q-Exa system, specialists from LRZ have also developed the prototype of the Munich Quantum Software Stack (MQSS) with partner institutions from the Munich Quantum Valley (MQV). Its tools integrate quantum systems into the workflows of supercomputers, coordinate the data exchange between the different computer technologies, and the calculations of the quantum processor. The MQSS will be made available to researchers and users as an open source version and will be expanded in the future to include tools and programs for other quantum technologies, such as those based on ion traps or neutral atoms.

HQS Quantum Solutions has also contributed a first practical use case to Q-Exa and MQSS: the Karlsruhe start-up has developed software that can be used to rapidly calculate and simulate the composition of new materials and chemical substances on quantum-accelerated supercomputers.

With Q-Exa, the MQSS, and the first application codes, the foundations have been laid for the research and development of quantum computing and the acceleration of HPC with QPUs. In combination with supercomputers, quantum computers should become suitable for a wider range of

LRZ Director Prof. Dr. Dieter Kranzlmüller (left) celebrates the inauguration of the Q-Exa system with Dr. Jan Goetz (center), Co-CEO and Co-founder of IQM Quantum Computers, and Markus Blume (right), Bavarian State Minister for Science and the Arts.

modeling and simulation applications while incorporating QPUs can accelerate HPC and help overcome performance limits that cannot be achieved today with known processor technology or artificial intelligence (AI) methods. Initial test runs and data exchange between SuperMUC-NG and Q-Exa are promising, and the hybrid system is now being prepared for everyday use at LRZ. Selected researchers will soon get access and be able to experiment with Q-Exa.
Susanne Vieser

Q-Exa benchmarks
• 20-qubit quantum processor unit
• Quantum Volume (QV): 32
• Entangled Greenberger-Horne-Zeilinger (GHZ) state of 19 qubits without readout error mitigation
• Fully entangled 20 qubit GHZ state with readout error mitigation
• A median two-qubit (CZ) gate fidelity of 99.46 % across 30 qubit pairs, with maximum fidelity over a single pair reaching as high as 99.74 %.
• A median single qubit gate fidelity of 99.94 % across 20 qubits, with maximum fidelity over a single qubit reaching as high as 99.95 %.

“We are currently building the future of computing. Q-Exa is a key project for our activities at the LRZ Quantum Integration Center, QIC, and demonstrates the success of co-design.”
Prof. Dieter Kranzlmüller

NEWS
FEATURES

HLRS to Boost AI-Based Innovation in Industry

The High-Performance Computing Center Stuttgart recently launched several new partnerships that will support future-oriented companies in developing novel AI-based products and services.

Since its founding, a defining feature of the High-Performance Computing Center Stuttgart (HLRS) within the ecosystem of the Gauss Centre for Supercomputing (GCS) has been its focus on supporting industry. More than 70 companies – including both global brands and regional small and medium-sized enterprises (SMEs) – currently have contracts with HLRS to use its Hawk supercomputer, turning to the center to run large calculations that would not be practical on their in-house systems. In the past, these companies have primarily used traditional high-performance computing for simulation – using computational fluid dynamics to optimize vehicle aerodynamics or comsbustion, for example. With the recent explosion of interest in technologies like ChatGPT and generative artificial intelligence, however, HLRS’s place in the industrial landscape has also begun expanding in new directions.

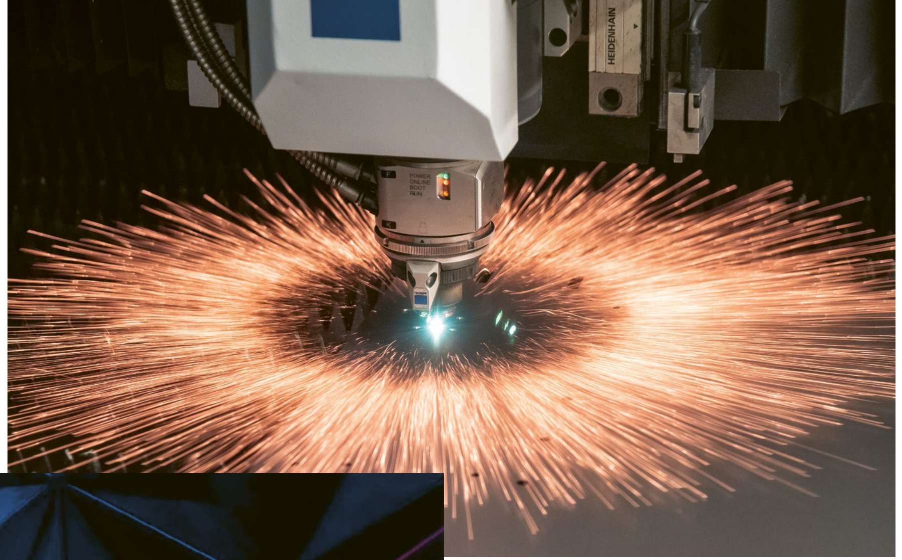
In 2024, HLRS signed several new formal partnerships with companies interested in using the center’s resources to explore the potential of AI. These initiatives have been made possible by recent improvements in the center’s computing infrastructure, including the addition of a GPU partition to its current Hawk supercomputer and the planning of its next-generation supercomputer, Hunter. The upcoming system, which is being manufactured by Hewlett Packard Enterprise and is scheduled to go into service in early 2025, is based on AMD’s Instinct MI300A APU, which combines CPUs, GPUs, and high-bandwidth memory on a single chip. This convergence of computing resources into a single architecture will not only permit increases in the speed, size, and resolution of traditional simulations, but also offer a powerful platform for deep learning, high-performance data analytics, and artificial intelligence.

According to HLRS Managing Director Dr. Bastian Koller, the new partnerships announced this year provide a taste of what the future might hold. “HLRS will always be here to support large-scale simulation, but we are also seeing increasing interest in our computing infrastructure from industries and in applications that are not just using traditional HPC,” he explained. “We have begun laying a foundation to support them by providing systems and expertise for artificial intelligence.”

A secure AI platform for German industry

In August 2024, for example, HLRS and Stuttgart-based AI consultancy Seedbox Ventures GmbH announced the beginning of a long-term cooperation agreement. Here,

HLRS recently launched a new partnership with Seedbox Ventures GmbH to bring artificial intelligence to German SMEs. (l-r): Kai Kölsch (Founder and CEO, Seedbox), Dennis Dickmann (Founder and CTO, Seedbox), Dennis Hoppe (Head, Converged Computing, HLRS), Bastian Koller (Managing Director, HLRS).



In recent collaborations with TRUMPF (engineering), MACK ONE (entertainment), and Seedbox (AI consulting and development) to expand their use of artificial intelligence, HLRS is helping improve European economic competitiveness.

NEWS
FEATURES

> **"We have begun laying a foundation to support [several companies] by providing systems and expertise for artificial intelligence."**
> Bastian Koller, HLRS Managing Director

HLRS's supercomputing systems will provide the performance Seedbox needs to develop custom AI solutions for German companies, including SMEs. In addition, Seedbox will use Hawk and HLRS's future HPC systems to train specialized, multilingual language models, making them available to the larger AI community on an open-source basis. The efforts should produce knowledge of how domain-specific AI models can be set up, operated, and scaled on local HPC environments in a cost- and energy-efficient manner.

This collaboration agreement marks the first time that HLRS has partnered with an AI startup. As Koller explains, this makes it an important step in the high-performance computing center's ongoing evolution: "With Seedbox as a strategic partner, we can not only already show what is possible with our current system but also what will be possible with Hunter and its successor, called Herder. This way, we will lose no time in the race to develop the best possible uses of AI for science and industry."

The partners also anticipate that the collaboration will contribute to improving Germany's digital independence. When developing and using AI applications based on proprietary or personal data, hosting data in a local and secure data center significantly reduces risk for Seedbox and its clients compared to the heightened vulnerabilities of uploading sensitive data to the cloud. HLRS, as a national high-performance computing center, guarantees the highest possible security standards, having recently completed a TISAX Level 3 assessment for information security and maintaining certification according to the ISO 27001 norm.

Dennis Dickmann, who has worked intensively in the field of artificial intelligence since 2008 and is responsible for technology development at Seedbox as its co-founder and CTO, sees this as a major advantage for German companies: "Such a powerful, secure and regional supercomputing infrastructure is a unique and extremely valuable resource. By cooperating with HLRS, we can guarantee the security of our clients' data and provide them with the AI technologies they need to remain innovative and competitive."

AI for manufacturing and the theme park industry

Other companies across the region have taken note of HLRS's capabilities as well. In March 2024, for example, the University of Stuttgart signed a collaboration agreement on behalf of HLRS with TRUMPF, a global leader in the production of lasers and sophisticated machine tools for industrial manufacturing. The goal of the agreement is to make HLRS's large-scale computing capacity available to TRUMPF employees. For several years, TRUMPF has offered its customers machine tools that use artificial intelligence to make their work faster and more effective. The partnership with HLRS will enable the company to expand this range of offerings with new solutions. Commenting on the agreement, TRUMPF CTO Berthold Schmidt remarked, "This cooperation strengthens our research and development efforts, enabling us to virtually simulate machine functions and train AI solutions more efficiently."

In a partnership organized by the Media Solution Center Baden-Württemberg, HLRS will also provide supercomputing resources and expertise for MACK One. Originally founded as an in-house agency at the Europa Park theme park, MACK One has evolved to become an international creator and consultant in the fields of theme park design, media-based entertainment, and media content. The new initiative, called MACK Research, is the first of its kind in Europe and will investigate innovative solutions for challenges facing the theme park industry. MACK Research will focus on interdisciplinary collaboration and the use of state-of-the-art technologies, including a focus on artificial intelligence, to build experience and develop prototypes that could later be translated into marketable products.

The modern wave of artificial intelligence is still young, but HLRS is already positioning itself to be a resource for companies interested in exploring new ideas. As the field matures, the center aims to provide powerful computing platforms as well as the technical expertise companies need to seize the opportunities AI will offer.
Christopher Williams

NEWS
FEATURES

Collaboration Between Jülich Supercomputing Centre and the University of Bonn Makes Progress in the Fight Against Deep Fakes

AI-generated content has been both a blessing and a curse for humanity—while these tools are used to streamline and support tasks, they are also being weaponized to spread disinformation. A multi-institutional team of researchers uses JSC's computing resources to uncover subtle signatures that machine-generated audio leaves behind.

A machine can imitate real human voices in a remarkably authentic way. Modern speech synthesis systems have made this possible for some time now.

Computer-generated, authentic-sounding human voices are already unlocking numerous creative applications, from video game character development to digital assistants. At the same time, these technologies also enable new digital ways to lie.

At the Jülich Supercomputing Centre (JSC) and the University of Bonn, a team of researchers is tackling one of the most complex challenges in digital security – detecting deep fakes being generated from unknown origins. This research, which also covers commercial software for publicly available generation architecture, is crucial in the fight against digital deception.

If a generator is publicly available, the researchers can use it to generate samples and use those to train detectors. Unfortunately, it is quite unlikely that scientists will always be aware of all deepfake generators in circulation. As a result, the team has focused its recent work to also train detection technology on individual generators and studies the degree to which the trained detectors generalize to uncover unknown networks it never saw during training.

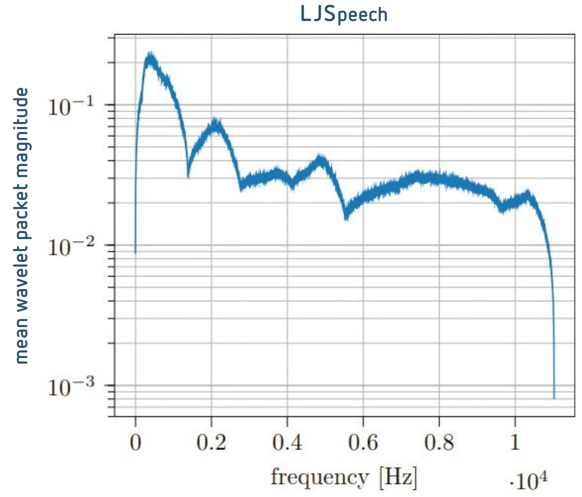
Recent studies indicate that specific types of so-called dilated convolutional neural networks are effective at detecting media from unknown audio generators (https://openreview.net/pdf?id=RGewtLtvHz).

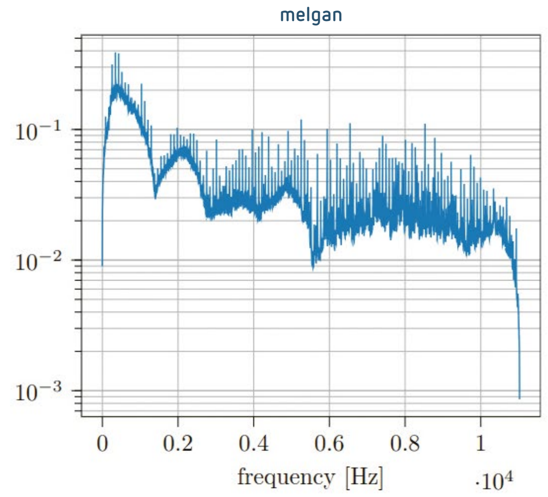
However, if a generated audio file does not noticeably differ from a recording of an actual human voice, how can scientists – or indeed anyone else – possibly identify it as a fake? Generative neural networks leave certain characteristic artifacts in synthetic media. Mathematical tools like the Fourier or Wavelet transform have been shown to effectively extract these audio "fingerprints" left behind in media generated by artificial neural networks. Both methods separate media into frequency components. Reading these frequency components is not unlike looking at the recording of an instrument such as a piano, for example: Pressing the key on the left end of the keyboard will produce a low sound, and keys on the right produce higher sounds. Similarly, Fourier and Wavelet transform assign certain numbers to the high and low parts of the spectrum.

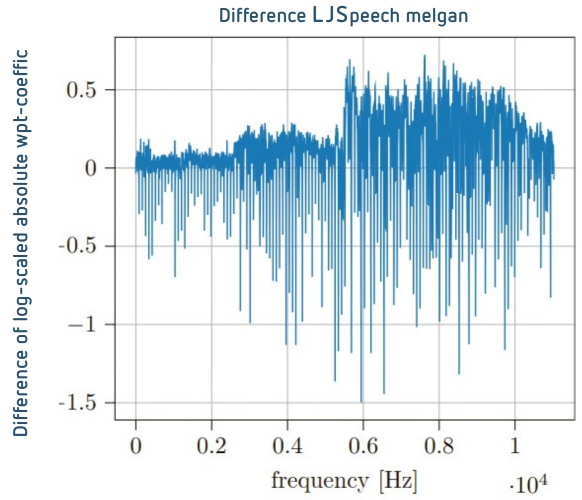
The extraction process takes a page from the world of traditional media forensics. Digital cameras, for example, are not perfectly uniform, and each one carries its own subtle signatures in images. Imperfections in the lens and the circuitry leave identical traces in every image from the same camera. If multiple photos from the same source are available, researchers can transform them into the frequency domain using the Fourier or Wavelet transforms. The fingerprint appears in the high-frequency part of the spectrum if transformed images from the same source are averaged. The exact process allows the team to visualize the fingerprints of deep fake generators.

LJ-Speech is a dataset containing samples of authentic human voice (https://keithito.com/LJ-Speech-Dataset/). Using LJ-Speech, the research team can compute a comparable representation by using MelGAN, a generative neural network, to read raw LJ-Speech-text. With this method, the researchers were able to see characteristic spikes that do not appear when they transform natural human voices.

Further, the team showed that this phenomenon is audible by transforming the high-frequency part of the spectrum back into its original form, which is also referred to as the time domain. Similar artifacts appear for most current generative neural networks. This effect is what the team's neural detectors use to identify fakes, which explains their ability to identify fakes from generators that never appeared in the training data samples. *Moritz Wolter*







Mean level 14 Haar-Wavelet-Packet decomposition of original LJSpeech (top left) recordings as well as synthetic versions generated by MelGAN (top right). The difference between both plots is shown on the right. MelGAN displays a characteristic spike-shaped fingerprint (bottom left).

The further out in our solar system we research, the more mysterious certain phenomena become. Public science facilities—both experimental facilities as well as supercomputing centers—provide world-class research tools to help us better understand our solar system and our universe more broadly.

Star Flyby Simulations Explain Solar System Dynamics Beyond Neptune

A team of astrophysicists at the Jülich Supercomputing Centre and Leiden University in the Netherlands have been using a combination of simulation and observational data to better understand our outer solar system. The team's recent findings suggest that a star may have passed close to our solar system billions of years ago that caused smaller celestial bodies to change their trajectories.

When we think of our solar system, we usually assume that it ends at the outermost known planet – Neptune. However, researchers note that there are several thousand celestial bodies that move beyond Neptune's orbit, with potentially tens of thousands of smaller objects that have diameters slightly more than 100 kilometers. According to Jülich Supercomputing Centre's (JSC's) Prof. Susanne Pfalzner, these objects do orbit our sun, just not the same way as other objects in our solar system.

"Surprisingly, many of these so-called trans-Neptunian objects (TNOs) move on eccentric orbits that are inclined relative to the common orbital plane of the planets in the solar system," she said. Up to now, these irregular, inclined orbits have remained an unsolved mystery.

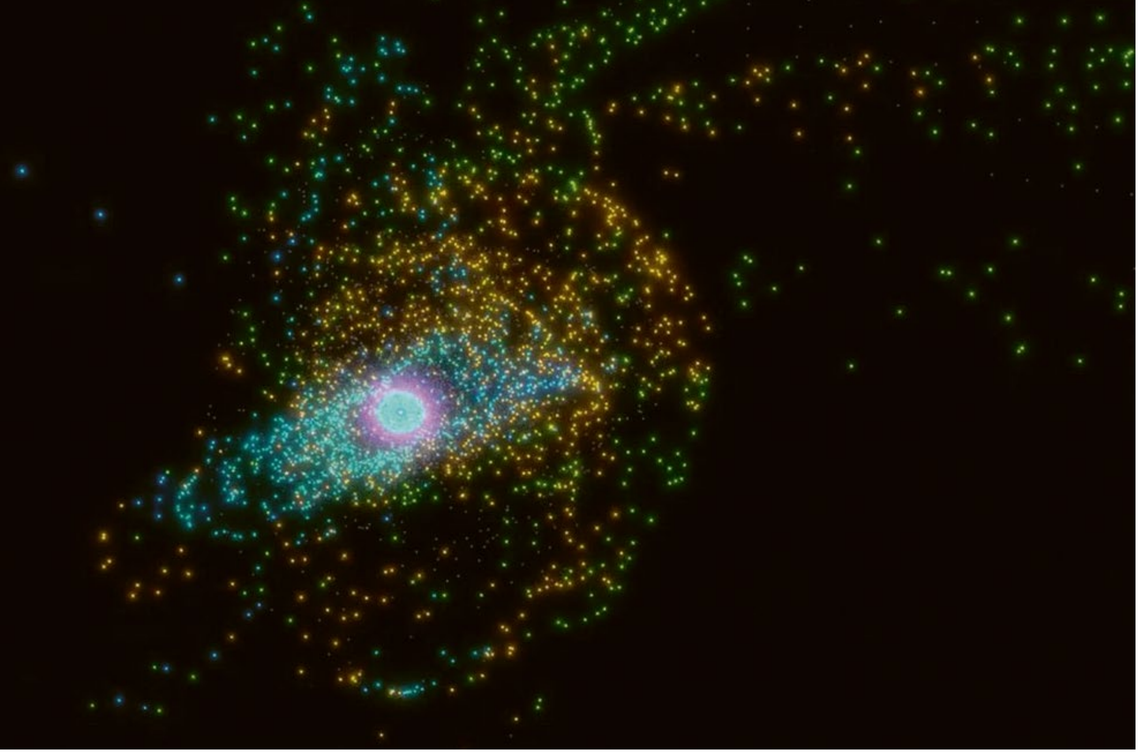
Recently, Pfalzer, PhD student Amith Govind, and Prof. Simon Portegies Zwart from the Netherlands' Leiden University employed JSC's JUWELS supercomputer to run more than 3,000 computer simulations to investigate these orbits and made a striking discovery: The team found that a close flyby of another star could explain why these trans-Neptunian celestial bodies orbit our sun in the manner that they do today. The team published its results in *Nature Astronomy* and *Astrophysical Journal Letters*.

Simulations chart the way

The most common theory for the behavior of our solar system comes from the so-called "Nice model," named after Nice, France, where the theory was developed. The model proposes that planets in the solar system started off much more compact and slowly drifted outward from the sun. The Nice model suggests that TNOs formed in between larger planets but were ejected further out into the system as the planets themselves moved away from the sun. Pfalzner noted that while this model would be possible to explain TNOs closer to the solar system, it does not account for extremely distant TNOs – such as the dwarf planet Sedna – nor would it explain why many of these objects have orbits that are nearly perpendicular to the planetary orbits in our solar system.

To better understand how TNOs came into their distant, irregular orbits, the researchers had to run a host of simulations that could both explain how a star passed by our solar system and the exact trajectories of the TNOs that were influenced by it. "We needed to run 3,000 simulations in order to try out different masses of the star, different distances of approach, and different orientations of the flyby, all of which require significant computational resources," Pfalzer said. "After we found the best match for the flyby compared to observations, we then had to investigate whether the paths of individual TNOs changes after the flyby in the long-term, which is even more computationally intensive." Pfalzner noted that having a GPU-based system like JUWELS was essential for the team to efficiently run such a large volume of precise, high-resolution simulations.

The team feels confident that simulations provided great details about the star that kicked off this TNO motion. "The best match for today's outer solar system that we found



A screenshot of a computer simulation shows the effects of another star passing by our solar system.

with our simulations is a star that was slightly lighter than our Sun – about 0.8 solar masses," Govind explained. "This star flew past our sun at a distance of about 16.5 billion kilometers. That's about 110 times the distance between Earth and the Sun."

HPC helps chart new research avenues

The team's findings not only helped explain how distant TNOs came into our Sun's orbit, but also provides a hypothesis for closer phenomena. Pfalzner, Govind, and JSC researcher Frank Wagner closely observed their simulation sets and noticed that some of the TNOs were actually hurled into our solar system after the star flyby into the region hosting the giant outer planets Jupiter, Saturn, Uranus, and Neptune.

"Some of these objects could have been captured as moons by the giant planets," says Simon Portegies Zwart. "This would explain why the outer planets of our solar system have two different types of moons." In contrast to the regular moons, which orbit close to the planet on circular orbits, the irregular moons orbit the planet at a greater distance on inclined, elongated orbits. Until now, there was no explanation for this phenomenon. "The beauty of this model lies in its simplicity," says Pfalzner. "It answers several open questions about our solar system with just one single cause."

Pfalzner also indicated that without access to world-class, public HPC infrastructure and expertise, researchers could not make these kinds of advancements. She pointed out that end-to-end support on the system, including the use of visualization resources and support in porting and scaling applications, is standard support at a facility like JSC. She also indicated that the team made its results publicly available so that other research teams could test and, hopefully, reproduce the team's results. "Considering the large computational effort involved, it should become a common aim for all researchers to make their results data available to other scientists," she said. *Regine Panknin*

The original version of this article appeared on the Forschungszentrum Jülich website.

https://go.fzj.de/jsc-stellar-flyby-effects

Related Publications:

Pfalzner, S., Govind, A. & Portegies Zwart, S. "Trajectory of the stellar flyby that shaped the outer Solar System." Nat Astron (2024). DOI: 10.1038/s41550-024-02349-x

Pfalzner, S. et al (2024) "Irregular Moons Possibly Injected from the Outer Solar System by a Stellar Flyby" Astrophysical Journal Letters (2024). DOI: 10.3847/2041-8213/ad63a6

German National Supercomputers Combined to Run a Single Large-Scale Simulation

For the first time, researchers simultaneously harnessed the capabilities of the Gauss Centre of Supercomputing's three high-performance computers, performing a massively high-dimensional calculation related to plasma physics.

Parallelization is the defining feature of high-performance computing. Researchers at the University of Stuttgart, however, have taken it to a whole new level. For the first time, they simultaneously integrated all three flagship supercomputers in the Gauss Centre for Supercomputing (GCS), creating a "superfacility" to run a single simulation.

Focusing on a high-dimensional problem modeled to the needs of plasma physics, the team used a programming approach called the sparse grid combination technique. This algorithmic method made it possible to distribute and recombine elements of the simulation across the supercomputers JUWELS (Jülich Supercomputing Centre), SuperMUC-NG (Leibniz Supercomputing Centre), and Hawk (High-Performance Computing Center Stuttgart). Leveraging the capabilities of Germany's three largest supercomputers, the enormous simulation contained a mind-boggling 35 trillion degrees of freedom. The three-system experiment built on a two-system test that used JUWELS and SuperMUC-NG. The earlier simulation was even larger, including 46 trillion degrees of freedom.

"To the best of our knowledge, this is the largest simulation ever done on European HPC infrastructure," says Dr. Theresa Pollinger, who recently completed her PhD at the University of Stuttgart Department Institute for Parallel and Distributed Systems (IPVS) and led the experiment. She will present a paper on the project at the upcoming SC24 conference, a meeting of the international high-performance computing community that takes place in Atlanta, Georgia, on November 17–22, 2024.

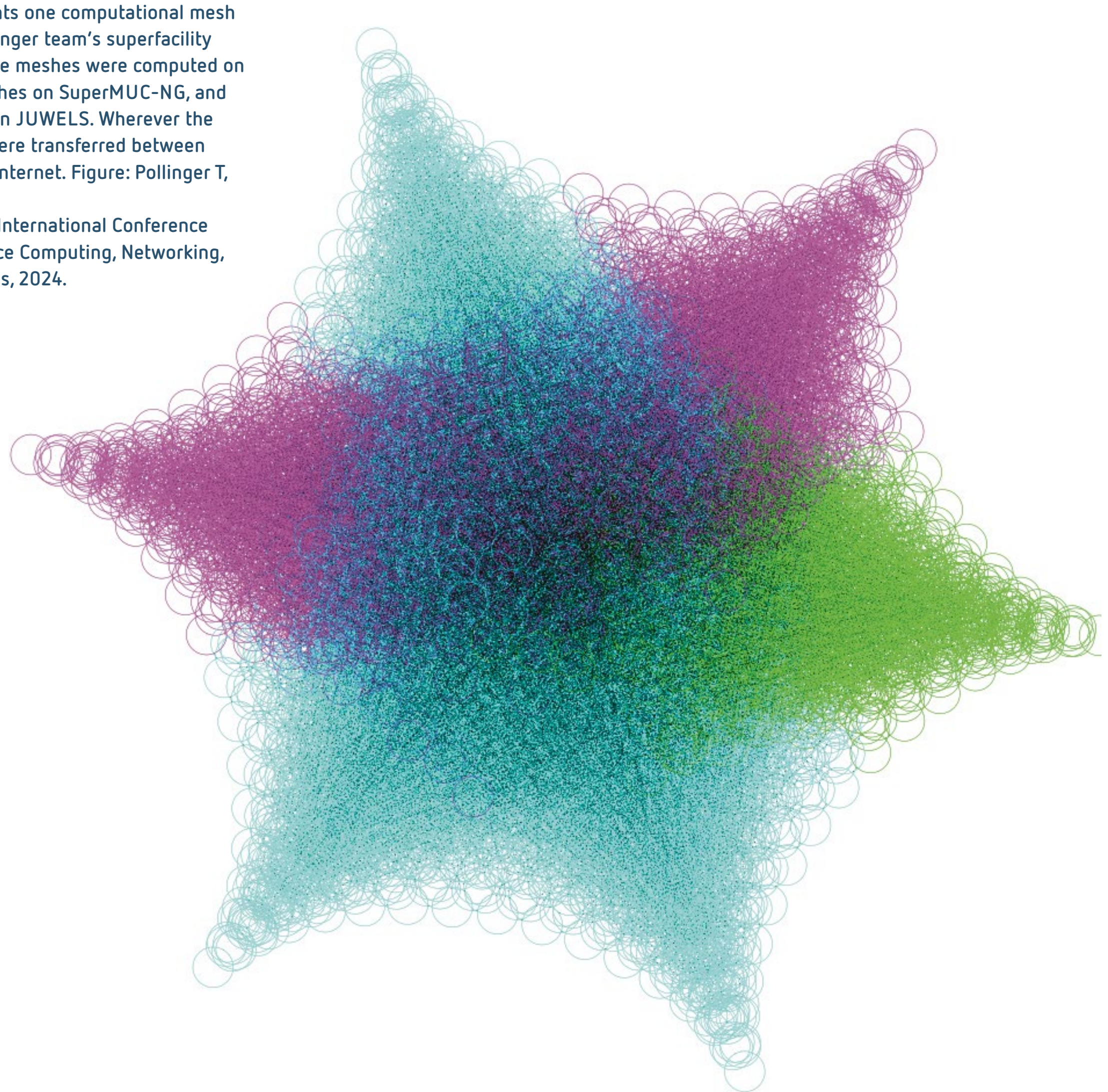
Explaining the significance of the experiment, Pollinger says, "The ability to combine the three supercomputers of the Gauss Centre for Supercomputing suggests that with enough bandwidth for data transfer and the ability to coordinate resource availability, it could be possible to use a federated system of supercomputers to run large-scale simulations for which a single system cannot offer enough memory."

Overcoming the "curse of dimensionality"

In conducting their experiment, Pollinger and her colleagues simulated a system inspired by a problem in plasma physics that is relevant for studying nuclear fusion. In the future, fusion could offer an abundant, carbon-free source of energy and so physicists and engineers have been investigating the processes involved using high-performance simulations. Because of the massive computing resources that are required to simulate such complexity, however, the calculations must typically make compromises in the resolution they can offer.

Simulating fusion in high fidelity involves mathematical models called the Vlasov equations. Here, the space containing the reaction is computationally subdivided into a grid of small virtual boxes. Within each box, an algorithm simulates changes over very short timesteps in six dimensions – 3 spatial dimensions plus 3 velocity dimensions. This approach can simulate fusion reactions at extremely high resolution. The problem, however, is that it also quickly generates enormous amounts of data, overwhelming available

Each circle represents one computational mesh included in the Pollinger team's superfacility experiment. The blue meshes were computed on Hawk, the pink meshes on SuperMUC-NG, and the green meshes on JUWELS. Wherever the colors meet, data were transferred between machines over the Internet. Figure: Pollinger T, et al. 2024.
Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2024.



data storage resources and wasting valuable computing time because of data transfer needs. Computational scientists call this the "curse of dimensionality."

Pollinger and her IPVS colleagues have been exploring how a multiscale computing approach called the sparse grid combination technique could address this problem. Instead of simulating each of the small boxes at high resolution, they fuse groups of small boxes together to form larger boxes, reducing the resolution. They repeat this approach multiple times, combining each box with different groups of its neighbors. The results of these lower resolution simulations are then combined to produce an overall result. As Pollinger explains, "All of the boxes are resolved badly, but are bad in different ways. By summing them up in a clever way, we get a good approximation that is finely resolved in all directions."

Joining algorithm with architecture

Although supercomputers have been combined to perform large simulations in the past, this has typically involved running different components of a larger algorithm on systems that are best suited to them. In the Pollinger team's experiment, however, large numbers of similar operations were executed on all three GCS machines in parallel.

The team implemented the sparse grid combination technique using a programming framework called DisCoTec. This approach distributed the computational workload by simultaneously running multiple variants of the simulation on different supercomputers (i.e., different low-resolution groupings of boxes in the computational mesh), while also following the same timesteps. This has the important advantage that simulations at the three HPC centers are loosely coupled, taking place largely independently of one another during major segments of the algorithm's overall runtime. Data exchange among them was required only occasionally, removing time consuming communication steps. Moreover, the multiscale nature of the combination technique means that only specific chunks of comparatively coarse data need to be shared and combined to provide the highly resolved final result. This dramatically simplifies memory requirements and makes it much easier to transfer data over the Internet.

"What is new is that we used this system in a highly parallelized manner as an integral part of our production environment."

Dr. Theresa Pollinger

In addition to brute computing power, the experiment relied on the high-speed, 100-Gbit-per-second research network that connects supercomputers at the three GCS centers, and on the UFTP data transfer software from the Forschungszentrum Jülich. This infrastructure was set up in 2017 under the auspices of the InHPC-DE project, supported with funding from the German Ministry of Education and Research (BMBF). "The tool already existed at GCS, but it was designed to enable users to make data backups or to move data between different systems for different processing steps," Pollinger explains. "What is new is that we used this system in a highly parallelized manner as an integral part of our production environment."

In an initial test in September 2023, the researchers used SuperMUC-NG and JEWELS to run a simulation including 46 trillion degrees of freedom, a size that would overwhelm the memory capabilities of SuperMUC-NG when running alone. They followed this up in November 2023 in an experiment that simultaneously used approximately 20 % of compute nodes on each of the three GCS systems – JUWELS, SuperMUC-NG, and Hawk. Using conventional approaches on a single supercomputer, the data requirements of such a high-dimensional simulation would have made it impossible to calculate. Using the sparse grid combination technique on the GCS superfacility, however, Pollinger and her team were able to complete this massive job.

Special considerations in system administration

Combining three of Europe's largest supercomputers in this way posed a variety of challenges. The first was organizational. During normal operation, each system executes multiple computing jobs from several users simultaneously. These jobs are coordinated using independent queueing systems at each HPC center. During tests and in the final run of the plasma physics simulations, however, it was important for Pollinger and her team to reserve large amounts of computing capacity simultaneously on all three machines.

In addition, network bandwidth on supercomputers like those in Jülich, Garching, and Stuttgart is optimized to enable users to transfer files between the HPC centers and their local computing infrastructures. In this case, nearly all of the available bandwidth was consumed by a single experiment. This also required planning between the research team and system administrators at the three centers. In the future, a unified queuing system that could automatically distribute a simulation across a federated HPC superfacility would make this approach more practical.

Staff at Hewlett Packard Enterprise (HPE), manufacturers of HLRS's Hawk supercomputer, also made important contributions to the experiment. HPE scientist Philipp Offenhäuser analyzed input/output performance on Hawk's Lustre file management system to identify bottlenecks. This included optimizing how files were written in advance to suit the I/O pattern utilized in this unusual application. Improving system performance also involved selecting domain decompositions to address potential load imbalances, and exploring compression techniques to further reduce the need for data transmission across the network.

As demand for compute, memory, and storage capacity in extreme-scale, grid-based simulations grows, this test case demonstrated a strategy for running them at a much higher resolution than would otherwise be possible.
Christopher Williams

Related publication
Pollinger T, Van Craen A, Offenhäuser P, Pflüger D. 2024. Realizing Joint Extreme-Scale Simulations on Multiple Supercomputers — Two Superfacility Case Studies.
Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis 2024.

SCIENCE
FEATURES

Next-Generation Earthquake Modelling Highlights Differences in Large and Small Earthquake Dynamics

Munich's premier universities have a long partnership working on earthquake modelling and simulation at LRZ. Most recently, the team used HPC in tandem with monitoring data to better understand the differences between large and small earthquakes.

In the last decade, more than 80,000 people globally have died from earthquakes. Advances in early warning systems and improved building codes have shrunk earthquakes' death tolls when compared to the longer historical earthquake record, but scientists, engineers, and government officials all recognize there is more to be done to improve resilience against these abrupt, violent natural disasters.

Dr. Alice-Agnes Gabriel, Professor at Ludwigs-Maximilians Universität München (LMU) and the University of California at San Diego (UCSD), dedicates her research efforts to understanding earthquakes at a fundamental level. She and her collaborators at the Technical University of Munich (TUM), led by Dr. Michael Bader, use high-performance computing (HPC) at the Leibniz Supercomputing Centre (LRZ) to model earthquake dynamics and propagation under a variety of conditions, including coupled with subsequent tsunami risks.

Recently, Gabriel and her colleagues ran a suite of simulations focused on understanding the differences in earthquake behavior and intensity based on fault size and how so-called "cascading" earthquakes can follow an initial large earthquake in unpredictable ways. "With access to larger HPC systems, we can go beyond these very large "hero" runs that model a certain historical or recent earthquake," Gabriel said. "We can run more simulations that are focused on exploring uncertainties and include more complicated physics in the process. In the past, we've focused on combing earthquakes and tsunamis in our simulations, but we've recently focused heavily on fracture mechanics, which is more of an engineering field we are trying to better understand in the geophysical context."

Understanding earthquake dynamics at scale

Earthquakes begin when two or more tectonic plates below the Earth's surface are unable to smoothly move against one another, building up pressure until rock layers eventually give way and the built-up pressure is rapidly released. Plates can grind against one another laterally (strike-slip fault earthquakes), or a portion of one plate can be pushed below the other (subduction earthquakes). Subduction earthquakes usually result in more violent shaking at the surface, but that only tells one part of the story.

Between simulations and sensors, scientists have a good understanding of how the powerful seismic waves unleashed by earthquakes propagate through Earth and at the surface that can shake and ultimately damage structures. Less understood, however, are the small-scale dynamics taking place deep below the Earth's surface and how that energy can spread to other faults.

While much of the tectonic energy released does travel as seismic waves to the surface, there is a significant portion of energy spent overcoming the subsurface friction between tectonic plates. That fracture energy can determine if earthquakes stop or jump between weak faults within the rock and result in a so-called "cascade" event. The 2023 earthquakes near the Turkish-Syrian border, for instance, cascaded out from the initial 7.8-magnitude event, triggering a 7.7 magnitude quake from a nearby fault line just 9 hours later and over 500 small and large aftershocks during the following day.

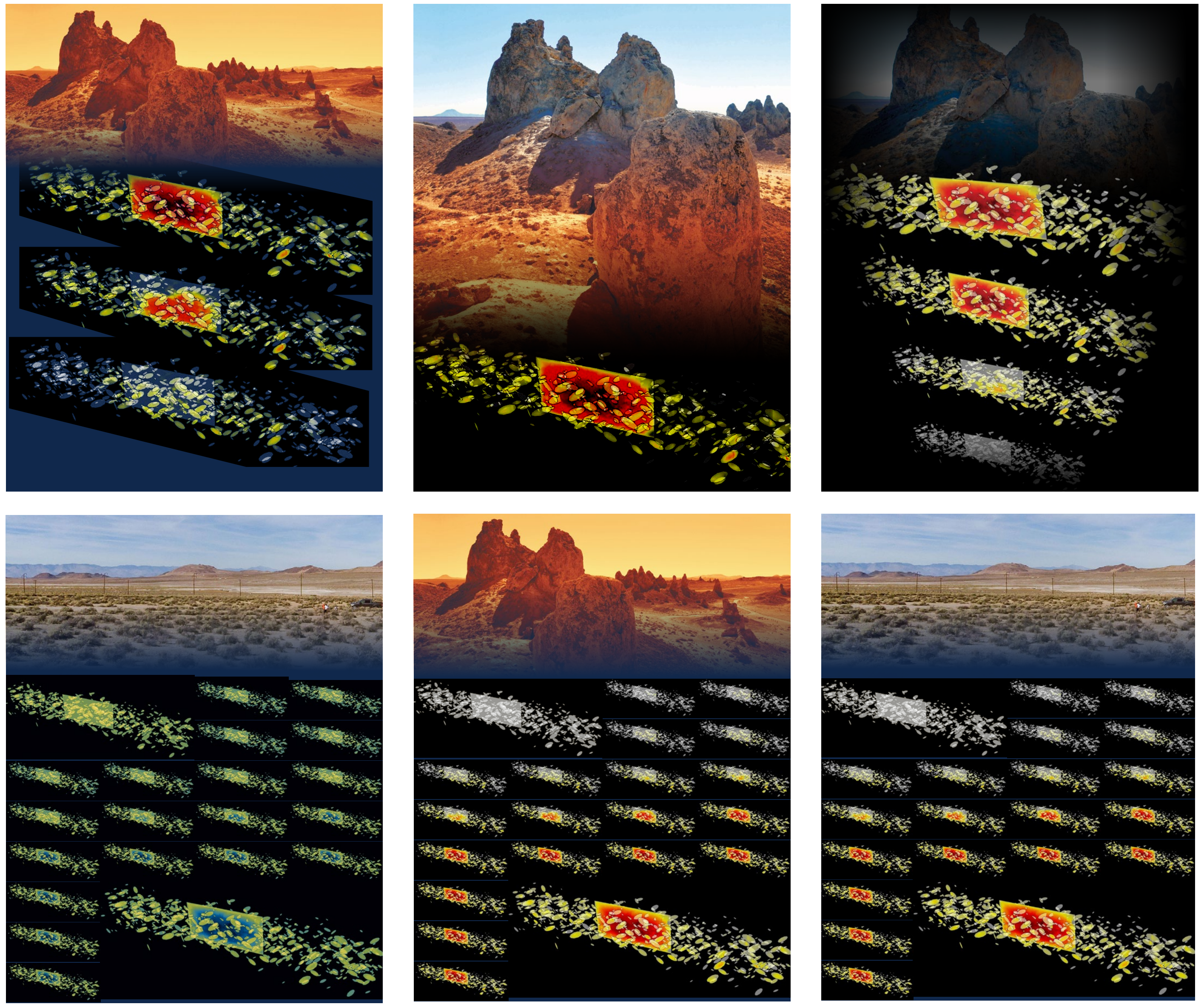
Gabriel's team wanted to update earthquake simulation approaches to better consider the distinct physics taking place in large earthquakes separately from those happening in small earthquakes and how those differences can impact an earthquake's destructiveness. "In this field, we have been assuming that large and small earthquakes are fundamentally the same, so in studying these phenomena, we hypothesize with that assumption in mind," she said. "By revisiting data that we've collected and with the help of modelling, we are starting to collect evidence that maybe they do not actually play by the same rules."

Gabriel added that researchers have access to far more data about small earthquakes because they are more common, and that even today, collecting earthquake data directly tens of kilometers deep under the Earth's surface is impossible. To that end, the team compiled the data available to it, developed mathematical models that could more accurately express the physics happening at different magnitudes and scales of earthquakes, then wrote equations to update the physics behavior at different scales of a supercomputer simulation. With the help of the SuperMUC-NG supercomputer at LRZ, the team discovered a linear scaling relationship between the fracture energy released and the size of a fault that can improve the accuracy of future simulations and fundamental understanding of earthquakes, ultimately helping better inform city planners, engineers, and other stakeholders of the risks of cascading earthquakes and their effects.

Public HPC supports scientific state-of-the-art

Gabriel and her collaborators have a strong appetite for HPC resources, using systems in Europe, the Middle East, and North America in their research. Her long-running relationship with LRZ staff and the enthusiastic, multi-

Despite making significant gains in early warning technologies and improved building codes in recent decades, earthquakes have still killed more than 80,000 people in the last decade.



Supercomputing simulation of a 3D multi-scale and multi-fault earthquake cascade across more than 700 fractures and triggering a main fault, revealing that small and large earthquakes don't play by the same rules. Earthquakes vary in size, with the largest and most complex also being the most dangerous. In their recent paper, Gabriel and her collaborators use modeling to determine whether information cleaned from more frequent, smaller events can be scaled up to better understand larger events.

"We are scientists who have very specific requirements, and while some of our needs are not immediately, commercially fruitful, we are doing work that is important for society."
Dr. Alice-Agnes Gabriel, LMU

disciplinary group at both LMU and TUM has kept her coming back to use LRZ's resources for over a decade. "It is just a long-term, steady, stable support system. Even my students that I'm onboarding in my role at UCSD, they are also using the Munich supercomputer because our workflows are so well integrated on LRZ's systems," she said.

Gabriel pointed out that working with public facilities provides researchers with added value that is not available when using commercial cloud computing providers. For instance, she noted that projects designed to help create and maintain "data lakes," or databases where topically relevant data can be stored and organized in a way that is accessible to other researchers, are only possible through large, publicly funded projects like the Europan-Union-funded Geo-Inquire project. In her other role at UCSD, she participates in efforts to develop and deploy a so-called "science gateway," at the San Diego Supercomputing Center (SDSC) that helps researchers more easily move and access relevant data and software being hosted at public research facilities.

"It's this level of support where I think public HPC facilities really shine," she said. "Private cloud computing companies are generally not going to host and share large amounts of your data, let alone work on something like a science gateway that is getting developed with other scientists in the community. We are scientists who have very specific requirements, and while some of our needs are not immediately, commercially fruitful, we are doing work that is important for society, has an interdisciplinary nature, and benefits from having data managed following open access and FAIR (findable, accessible, interoperable, reusable) principles."

Gabriel and her collaborators gained early access to LRZ's upcoming flagship supercomputer, SuperMUC-NG Phase 2, set to fully deploy in 2025. She indicated that the machine is well-suited to helping the team further its development of machine-learning (ML) approaches that can help the team more efficiently explore suites of hybrid HPC-ML simulations rather than single large, expensive simulations of a particular earthquake. *Eric Gedenk*

The Helmholtz Foundation Model Initiative is developing next-generation computer models to tackle large problems facing humanity, such as climate modelling, improving photovoltaics, and soil fertility.

Helmholtz Foundation Model Initiative Lays the Groundwork for Next-Generation AI Research

Three-year initiative focuses on expanding the use of foundation models in a broad range of research disciplines, focusing on environmental topics. JSC supplies both world-class supercomputing infrastructure and expertise.

The rapid developments in the field of artificial intelligence (AI), along with ever-increasing amounts of available data, opens new possibilities for unlocking the full potential of data in tackling previously unsolvable problems. Researchers can reach these new computational heights by using a new generation of AI models: the so-called foundation models (FMs). Currently, FMs are known in the context of natural language processing (ChatGPT, for example) and multi-modal applications, such as image generators. However, FMs are not limited to language or image applications and can be extended to many fields.

As a member of the Helmholtz Association, Forschungszentrum Jülich – specifically the Jülich Supercomputing Centre (JSC) – is participating in an ambitious three-year project that aims to integrate FMs into a wide range of scientific workflows. The Helmholtz Foundation Model Initiative (HFMI) started at the beginning of May to bring these powerful tools to public scientists.

In addition to building fully functional models that benefit scientific research, project participants are also focused on considering the complex, evolving nature of AI ethical and legal standards. Four pilot projects have been selected, involving scientists from twelve Helmholtz centers. By design, HFMI is a close collaboration between AI experts and domain scientists within individual pilot projects. Helmholtz ensures that not only do the models make a meaningful contribution to research, but also that the results withstand empirical scrutiny. Over three years, the projects will receive funding of €11 million with an additional €12 million invested in expanding the necessary infrastructure.

JSC is involved in three of the four pilot projects and the overarching synergy unit:

3D-ABC – Calculation and Visualization of the Global Carbon Budget of Vegetation and Soils

To mitigate the consequences of global climate change, decision makers need in-depth knowledge of the global carbon budget, including sources and sinks of CO_2 such as peatlands, forests, and permafrost soils. Until now, researchers have struggled to quantify how changes in land areas, vegetation, or soils affect the carbon cycle due to the complexity in merging data sources into uniform, usable sets. 3D-ABC aims at exploiting multimodal data from various sources, such as satellites, drones, and local CO_2 capture stations. This approach will allow key parameters of the global carbon cycle in vegetation and soils to be captured, quantified, and characterized with high spatial resolution.

HClimRep – Capturing Interactions Between Atmosphere, Ocean, and Sea Ice in a Novel Climate Model

What if researchers could make climate predictions more accurately and efficiently? Could humanity better combat the causes of climate change and mitigate its effects? Could researchers make the impacts of global warming clearly visible to everyone? HClimRep aims to answer these questions. By building one of the first FMs for climate research that combines data from the atmosphere, the ocean, and sea ice, the HClimRep team is developing a novel tool for fast and accurate climate predictions, trained on Europe's first exascale computer.

SOL-AI – Development and Optimization of Photovoltaic Materials

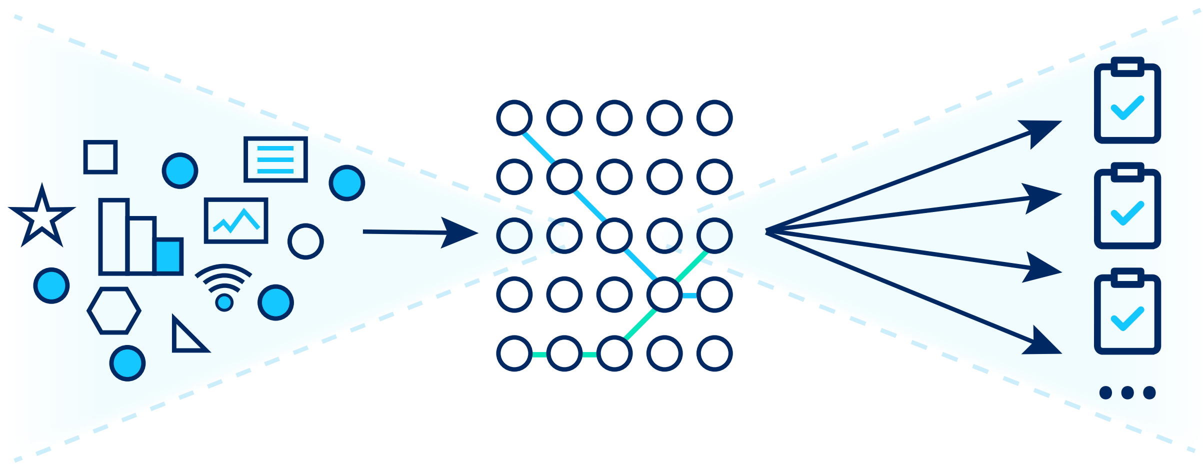
Photovoltaics is a key technology for the Germany's – and indeed the world's – energy transition. For it to be adopted on a sufficient scale globally, innovative solar cell concepts need to be implemented much faster. Despite tremendous scientific investment in this field, the sheer volume of data limits how quickly scientists and engineers can implementat the latest findings. SOL-AI is creating an FM to reform materials informatics from both scientific and industrial points of view. It will integrate diverse experimental data and research results on photovoltaic materials to drive innovations from accelerating component development, optimization, and discovering new solar materials.

The Synergy Unit

Project partners in the Synergy Unit research interdisciplinary questions, promote the exchange of knowledge between individual projects, and carry out overarching activities. For this team, a special emphasis is placed on methodological questions that relate to the understanding of how FMs function. Understanding – including explaining and reproducing results – is crucial for building trust in FMs. Therefore, HFMI places a strong emphasis on quality assurance measures, providing added value for science, and commitment to the principle of open science, by making the final results available to society. *Ehsan Zandi*

PROJECT	Helmholtz Foundation Model Initiative
FUNDING AGENCY	Helmholtz Association
FUNDING AMOUNT	€11 millon with €12 million in infrastructure support
RUNTIME	2024 to 2027
PARTNERS	12 Helmholtz centers

How Are Foundation Models Trained?



FFplus Holds First Open Call for HPC and AI Applications in Industry

In a EuroHPC JU-funded project coordinated by HLRS, companies submitted proposals for Business Experiments and Innovation Studies aimed at taking their products, services, and operations to the next level.

High-performance computing and artificial intelligence have emerged as game-changing technologies with the potential to boost industrial competitiveness and provide myriad economic and societal benefits. For many small and medium-sized enterprises (SMEs), however, technical and financial factors can mean that making the most of these technologies remains a challenge. In May 2024 the EuroHPC Joint Undertaking (EuroHPC JU) launched a new research and innovation project aimed at addressing this need. Called Fortissimo Plus (FFplus), the project will hold six open calls over the next four years for SMEs and start-ups in two categories: Business Experiments that use comptational methods running on high-performance computing systems, and Innovation Studies that harness the power of artificial intelligence, particularly generative AI, to develop new software and services. Applicants that receive awards will receive financial and technical support linked to the use of state-of-the-art EuroHPC resources.

FFplus is being coordinated by the High-Performance Computing Center Stuttgart (HLRS), in cooperation with a consortium including past Fortissimo partners Teratec, scapos AG, CINECA, CESGA, and Arctur, as well as new member CYFRONET.

"FFplus is an important link in Europe's strategy for HPC and AI, connecting SMEs and start-ups to expertise and powerful computational resources that offer enormous innovation potential," says Dr.-Ing. Bastian Koller, FFplus Project Coordinator and HLRS Managing Director. "The project will give such companies the framework they need to be successful and to create value in the digital economy."

An enthusiastic response to the first FFplus open call

FFplus is funded by the Digital Europe Programme, a European Union funding initiative focused on bringing digital technology to businesses, citizens, and public administrations. The total budget for the four-year FFplus program is €30 million, which will enable FFplus to offer SMEs access to world-class tools and expertise in modelling and simulation, data analytics, machine learning, and AI.

FFplus launched its first open call for SMEs and start-ups in the summer of 2024. A total of 126 business experiment proposals, involving 223 organisations from 30 countries associated with the EU Digital Europe programme, were received. Additionally, 62 proposals for innovation studies were submitted from 24 countries associated with the EU Digital Europe programme, involving 83 SMEs and 36 other organisations. These proposals addressed challenges across various industrial sectors. FFplus partners anticipate selecting approximately 33 business experiments and innovation studies, which will receive funding for execution.

An important additional component of FFplus will be to document the results and impact of these Business Experiments and Innovation Studies. In this way, FFplus aims to gather success stories that will inspire other startups and SMEs to investigate how HPC and AI could be used to improve their products, services, operations, and ultimately competitiveness.

In FF4EuroHPC, the predecessor project of FFplus, Sofia, Bulgaria-based engineering company SoftSim Consult developed a comprehensive urban physics modeling infrastructure that utilizes computational fluid dynamics running on a high-performance computer.

Building on the Fortissimo legacy

FFplus is the latest chapter in the Fortissimo story. Spanning a decade-long history in which they established a strong reputation across Europe, the Fortissimo, Fortissimo 2, and FF4EuroHPC projects successfully executed more than 130 experiments involving more than 300 partners. With a total budget of €42.8 million, these efforts produced 120 success stories from more than 20 EU countries, demonstrating how SMEs can effectively use HPC, high-performance data analytics, and AI to develop new products and services.

Dr. Guy Lonsdale, CEO at scapos AG, has been an integral member of past Fortissimo projects and has been leading the first FFplus open call. Reflecting on how the use of HPC in industry has evolved in recent years, he explains, "Over the last decade, we have seen high-performance computing evolve beyond numerical methods to include data analytics, machine learning, and AI. This is reflected in a key aspect in FFplus: our focus on generative AI technology development. In the Innovation Studies – a new type of Fortissimo activity – SMEs or start-ups with existing AI business models and business prospects will use large-scale EU computational resources to develop and customize generative AI models, such as foundation and large language models."

At press time, a panel of external experts has been evaluating the proposals submitted to the first FFplus open call. The project team anticipates that the results will be announced by the end of 2024, with Innovation Studies set to begin in December 2024 and the Business Experiments launching in January 2025. The second FFplus open call is expected to start in the second quarter of 2025.
Christopher Williams

PROJECT	FFplus
FUNDING AGENCY	Digital Europe Programme
FUNDING AMOUNT	€30 million
RUNTIME	May 2024 to April 2028
PARTNERS	HLRS, scapos AG, Teratec, CINECA, CESGA, Arctur, CYFRONET

CompBioMed Leaves Legacy Promoting Computational Research in Medicine and Drug Development

Over the last decade, the Centre of Excellence CompBioMed worked on the Virtual Human, a digital twin of a human, and helped create a suite of technologies to support computational medical research. However, there are still questions about the usage of personal health data and how patients can store and organize it without risks.

The list of results is long: 310 scientific articles, two books, two films and about a dozen software tools and databases are the productive legacy of CompBioMed. Funded by the European Union's Horizon 2020 program, 52 universities, companies, and research institutes – including the Leibniz Supercomputing Centre (LRZ) – have been working for a decade toward a new era of medicine that can fully leverage the power of computing in diagnosis, therapy, and research.

"One of the evaluators of the CompBioMed project commented that there's a lot to be gained for a very large number of people and for the future from the software and methods that have been developed," said Dr. Peter Coveney. The Professor of Physical Chemistry and Computer Science is Director of the European Centre of Excellence CompBioMed and teaches at University College London (UCL), Yale University, and the University of Amsterdam. "It is good that the use of mechanistic modelling and other computational methods, including predictive methods, is starting to gain ground in biology and medicine, because that's how we're going to make more progress in the long term."

Building high-fidelity human models

At the heart of both phases of CompBioMed was the Virtual Human, a digital twin of a human being that will help doctors treat patients and speed up the development of drugs and new treatments. "Digital twins of humans will be part of a virtual future of medicine. Building accurate, high-fidelity models of people is a huge task, and CompBioMed contributed to this in several ways," says Coveney. "But we're not talking about a single representation of a whole human being soon. It's more about components of the human body that can be accurately simulated on supercomputers."

The simulations calculated on the supercomputers at LRZ and elsewhere have laid the groundwork to help train doctors in digitally assisted treatment methods. If they are fed with individual patient data, doctors can use them to test the effect of drugs. After all, simulations are already rapidly accelerating drug development and trials on humans. However, issues of security remain. "Medical data needs to be secure enough to help with patient therapies and treatments, but not pose a privacy risk, so data security is one of the biggest challenges for digital health," says Coveney. "On the other hand, it's good to get data – artificial intelligence comes up in all these conversations about the digital twin and medical data. AI needs access to data to run statistical models and methods. But that has led to a lot of ethical and moral questions."

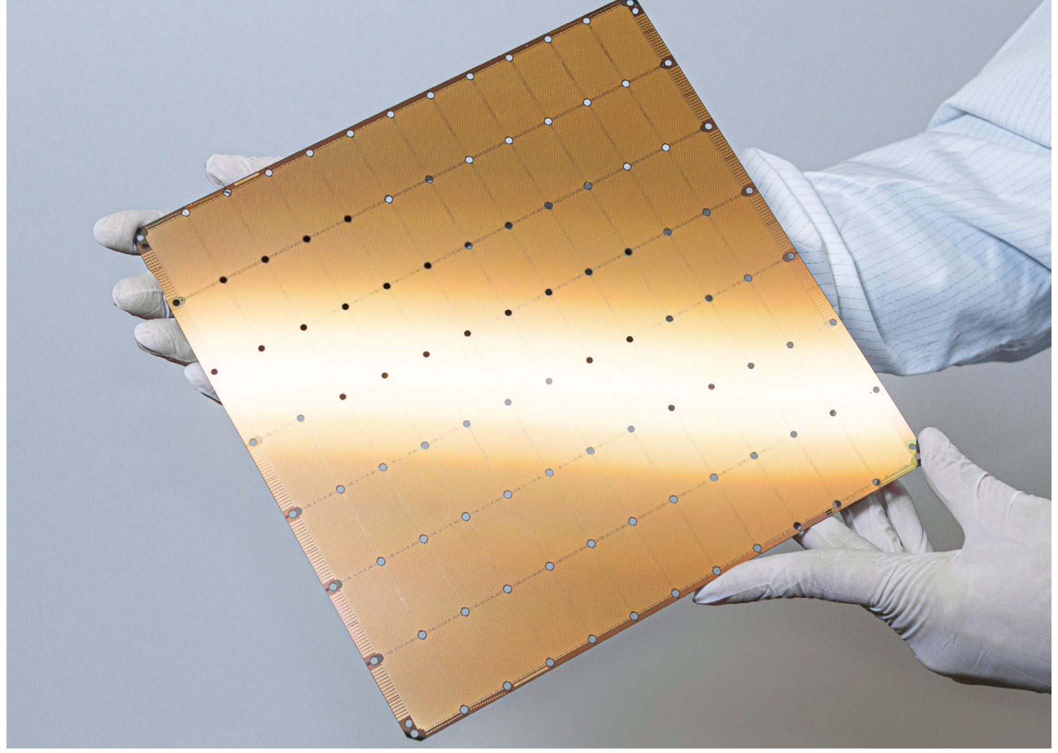
Coveney suggests, patients should be able to securely store and organize their health data in the future like their bank accounts and financial resources, so they could have active choice in when and how they could make it available for digital examinations or studies.

Interacting with simulations like a game

Visualization of simulation data is also an important part of CompBioMed's Virtual Human. For example, the LRZ has visualized blood flow in the forearm and brain and developed workloads and a toolset that can be used to map similar processes in other organs. "Visualization is really important for the work we do," says Coveney. "I'm always pushing for a scenario where you have a reliable simulation that you can interact with in the same way that you interact with video games. This could allow clinicians to work with what-if scenarios before performing invasive surgery." A third phase of the project is in doubt, but CompBioMed has played a major role in moving efficient, safe computationally enhanced medicine forward for a decade, and future projects can expand on what it has built.
Susanne Vieser

Facts & Figures: CompBioMed
18 Core Partners; 52 associated partners
Funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 675451 (phase 1) and grant agreement No. 823712 (phase 2).
310 scientific papers, 2 books, 2 films
139 events, 26 workshops,
Education, trainings, information, software development, reached round about 190 mio. People
Project website: https://www.compbiomed.eu/
Software hub: https://www.compbiomed.eu/compbiomed-software-hub/
Media: https://www.youtube.com/@computationalbiomedicine2363

LRZ was an active partner in the decade-long CompBioMed project, helping researchers create realistic simulation conditions of blood flow. CompBioMed finishes with a productive legacy: 310 scientific articles, two books, two films, and a dozen software tools.



The Cerebras Systems Wafer Scale Engine 2 is the largest computer chip today, and is well-suited for a variety of AI research tasks.

Cerebras Workshop Showcases Short Paths Between AI Data Processing and Storage

A workshop at the Leibniz Supercomputing Centre demonstrated the potential of the Cerebras CS-2 system for science: In addition to training large models, it can also accelerate high-performance computing applications.

New technologies often require explanation, but the world's largest computer chip – Cerebras' Wafer Scale Engine 2 (WSE-2) – also intrigues and excites. "The architecture of the CS-2 system is unique," says Michael Gerndt, Professor for Computer Science at the Technical University of Munich (TUM) "I want to use it in my lectures on computer architecture, and it can also be used for or high-performance computing (HPC) applications, which is interesting for us," he said. The researcher was one of about 30 scientists in July that took part in a workshop at the Leibniz Supercomputing Centre (LRZ) to learn about the possibilities of the Cerebras System's CS-2, a computer specialized in artificial intelligence applications. "The aim of the workshop was to start community building with researchers interested in using the system," said Dr. Michael Hoffmann from the LRZ's Big Data & Artificial Intelligence (BDAI) team. "We were also able to learn a few strategies for using it ourselves."

The CS-2 system, which uses Cerebras' WSE-2 chips, is well-suited to training large language models (LLMs), because each large WSE-2 chip is doing more of the work, ultimately spending less time sharing data between many nodes: "Large AI models have challenges when researchers are primarily using graphics processing units, because developers have to break the models down to distribute them across many GPUs," explains Gokul Ramakrishnan, technical lead of Cerebras Systems, during the workshop. "Often, models that scale for use on larger systems must be rewritten for use on a cluster." This is not necessary with the CS-2, as its 46-square-centimeter chip integrates some 2.6 trillion transistors, as well as memory capacities of up to 40 gigabytes of SRAM. While many classical AI clusters load data block by block onto processors and write back the results to memory, the Cerebras chip allows data to flow from processing units to on-chip memory at a speed of 20 petabytes per second. This accelerates the data transfer and deep or machine learning.

"The architecture of the CS-2 system is unique. … I want to use it in my lectures on computer architecture, and it can also be used for HPC applications, which is also interesting for us." Michael Gerndt

For LRZ's long-time users focused primarily on traditional modeling and simulation, the workshop also highlighted that CS-2 is also suitable for HPC: "Projects that require high memory bandwidth and benefit from having working data stored close to the processing cores can be effectively executed on the CS-2," says LRZ researcher Jophin John. "The first HPC projects using the technology are currently underway at LRZ."

For the development of models and programs, the CS-2 system uses a PyTorch-based software stack, as well as a selection of widely used AI models and transformers on their "Model Zoo," a large Github repository. The workshop primarily focused on the use of the system and included practical exercises. Participants learned how to work with models, prepare training data and process it on the CS-2. "When pre-training LLMs, it's crucial to first experiment with smaller versions of the data set. This allows users to train in an energy-efficient manner," said Hoffmann. "If users work with the models from the Model Zoo, the CS-2 is already quite user-friendly, and if implementing custom models becomes challenging, we are always excited to support users in tailored solutions." "Computing resources are always in high demand in research," said Dr. Niki Kilbertus, Professor at the Technical University of Munich's School of Computation "The promise of being able to train large AI models from scratch with the CS-2 attracted me to participate. In particular, projects where we cannot easily reuse or fine-tune existing large language models, such as models for microbiome DNA data, could benefit greatly from Cerebras." As the summer workshop showed, LRZ's investments in a diverse range of new architectures and computing paradigms expands the ways that scientists and engineers can advance research. *Susanne Vieser*

Cerebras Systems' Wafer Scale Engine-2
850,000 cores optimized for sparse linear algebra
46,225 mm² silicon
2.6 trillion transistors
40 gigabytes of on-chip memory (SRAM)
20 PByte/s memory bandwidth
220 Pbit/s fabric bandwidth
7 nm process technology

The Model Zoo for the CS-2
The software stack for the CS-2 is based on PyTorch. Cerebras Systems has also implemented common AI models and transformers such as: Bert, Bloom, Codegen, Dpr, Falcon, Flan-ul2, GPT (2,3,4, j, neox), Llama, LlaVa, Mistral, Mpt, Octocoder, Roberta, Santacoder, SqlCoder, T5, Transformer, Ul2, Wizardcoder.
You can find it all on their "Model Zoo" at: https://github.com/Cerebras/modelzoo

EVENTS

How Simulation Could Support Crisis Response

In the project CIRCE, HLRS has been investigating how supercomputers could help public administrations across Germany prepare for and manage emergency situations like natural disasters or pandemics. A recent symposium in Berlin presented the preliminary findings.

In 2024, floods in southern Germany, Thuringia, and Saarland provided sobering reminders that natural disasters and other crisis situations can arise quickly and have devastating consequences. For public administration at the local, regional, state, and national levels, managing such events can pose enormous challenges, particularly considering the difficult-to-predict future consequences of climate change.

The High-Performance Computing Center Stuttgart (HLRS) has been investigating how simulation using supercomputers could assist public officials in addressing such challenges. With funding from the German Federal Ministry of Education and Research, and the Baden-Württemberg Ministry of Science, Research and Art, HLRS has been conducting a study to plan a Computational Immediate Response Center for Emergencies (CIRCE). Such a center would offer public agencies rapid access to high-performance computing (HPC) resources for large-scale simulation and data analysis during crisis situations.

This summer, the CIRCE team held a full-day symposium in Berlin that presented its preliminary findings to representatives of public administration and experts in disaster response. In addition to offering general insights into the requirements for establishing a computational emergency response center, the talks highlighted potential applications of simulation, artificial intelligence, and data visualization for crisis management. Data scientists at German federal governmental institutes also enriched the meeting by describing how data is currently used in government planning and decisionmaking, and offered insights into the practical challenges of navigating Germany's complex, federalized system of government.

Potential applications of simulation, and challenges

In a qualitative, interview-based survey of representatives from federal, state, and local agencies, as well as crisis response organizations, CIRCE found that many recognize opportunities for using simulation to address challenges they face, including migration, pandemics, floods, wildfires, chemical or nuclear accidents, and terrorist attacks. The dialogue has also identified technical and administrative considerations that will need to be addressed to ensure that supercomputing resources and expertise are available when they are urgently needed.

These outreach efforts also led to a new collaboration between HLRS and the Duisburg fire department,

"We can't simply expect to be able to manage crises on an ad hoc basis, but need to be active even before crises arise." Dr. Julianne Braun, Federal Ministry of the Interior and Community (BMI)

"Just looking at what HLRS can do is huge, but such things are currently not available for many public administrations. From our perspective it is not so important to strive to make better tools, but to attempt to implement what already exists."
Christian von Spiczak-Brzezinski, Duisburg Fire Department

focusing on development of a simulation tool for predicting flooding from a potential dike breach on the Rhine. HLRS's Dr. Ralf Schneider has been leading the center's contributions to the cooperation, and explained the lessons that have emerged thus far. He pointed out, for example, that despite the growing digital resources of many public agencies – including geographic data, building information, demographic data, and environmental quality measurements – the Duisburg project revealed that customized data preparation can still be required; in this case, information concerning the geometries of bridge underpasses and the topography of the Rhine riverbed.

In addition, Schneider explained, creating and testing such simulations is not something HPC centers can do alone, but requires technical expertise and local knowledge. The Duisburg project has involved experts in flood control and fluid dynamics, and onsite inspections of the area were necessary to confirm that the data input into the simulation adequately represent physical reality on the ground. As the simulation nears completion, Schneider also anticipates that experts will need to validate the results to be sure that they are accurate. Only in this way will users of the simulation feel confident that the information it provides is reliable.

Improved planning and collaboration are needed

Such experiences demonstrate why advance preparation for emergency situations is essential. "We can't simply expect to be able to manage crises on an ad hoc basis, but need to be active even before crises arise," explained Dr. Juliane Braun, Chief Data Scientist of the Federal Ministry of the Interior and Community (BMI). This means not only developing the right simulation software, but also ensuring that monitoring procedures, communication systems, and guidelines for decision making are in place before a crisis computing application goes into service. Moreover, sustainable collaboration between HPC centers and public agencies will require long-term contingency plans for maintenance or upgrades in supercomputing infrastructure. Such preparations will guarantee that data gathering and preparation, calculation of the simulation, and evaluation of simulation results happen quickly, automatically, and without interruption.

With Germany's 294 regional districts and an additional 107 self-organized cities, speakers argued that it will not be practical or efficient for each community to develop its own specialized solutions. Instead, increased collaboration among public agencies is needed to design applications that will be useful across the country. "We need to develop a stronger network so that together we can determine how best to arrive at solutions," Braun said.

Christian von Spiczak-Brzezinski has been representing the Duisburg fire department in its collaboration with HLRS, and observed that public organizations need a better understanding of what simulation tools are available and how to implement them. Better coordination could also have financial benefits. "If every community contacts CIRCE or HLRS directly and develops a custom solution for its own needs, it could end up costing, 3, 4, or 5 times more than if we develop a single concept that is made available to all communities," Spiczak-Brzezinski argued. Collaboration could also make such simulation tools more consistent and interoperable.

To realize this potential, however, high-performance computing centers like HLRS will need partners at the local, state, and federal levels who can articulate challenges and commit to developing and implementing solutions. This will require not only that public administrations invest more time in thinking about potential scenarios that could affect their communities, but also improving digital competencies of their staffs.

"From where I sit, I see the beginnings of a change in thinking," said Dr. Bastian Koller, Managing Director at HLRS. "As more people see what we are doing and find it valuable, my hope is that interest will snowball."
Christopher Williams

ParCFD 2024

JSC Organizes Premier Computational Fluid Dynamics Conference: ParCFD 2024

Annual event brings together multidisciplinary interests for three days of lectures, discussions, mini-symposia, and sessions on various novel topics.

On September 2–4, the 35th Parallel Computational Fluid Dynamics (ParCFD) Conference took place at the University Club in Bonn, Germany. The event brought together around 110 participants from around the world.

ParCFD is an annual conference series dedicated to the discussion of recent developments and applications of parallel computing in the field of computational fluid dynamics (CFD) and related disciplines. Since the inaugural conference in 1989, many new developments and technologies have emerged in the field, including current generations of world-class high-performance computing (HPC) systems that serve as large-scale research investments research for governments, academia, and the private sector worldwide. This year, the conference was organized by the team from the Jülich Supercomputing Centre (JSC), led by the conference chair Andreas Lintermann and the co-chairs Jens Henrik Göbbert and Sohel Herff. This year's sponsors were Eviden and ParTec AG.

During the three-day event, attendees were involved in a full program of keynotes, special sessions, and mini-symposia. The five keynotes addressed a wide range of topics related to cutting-edge CFD research. Among other focuses, participants discussed the adoption of CFD algorithms to novel and heterogeneous hardware technologies – including GPUs – to increase execution performance and scalability. The CFD community is always looking for more efficient simulation technologies and methods that accurately solve equations related to fluid physics. The conference demonstrated that CFD applications that take into account multiscale modeling and multiphysics demand real exascale performance. The EuroHPC Joint Undertaking (JU) also presented the current state of play regarding the procurement of different HPC and quantum computing systems as well as the various calls for projects supporting EuroHPC JU activities.

In total 84 talks were given across 4 parallel sessions. The nine mini-symposia concentrated on different CFD topics, such as quantum computing for CFD applications (a novel topic to the community), advances in parallel simulations of reacting flows, or the convergence of artificial intelligence and HPC for CFD. Operational topics, such as tool support for developing highly parallel CFD applications also came up.

The other topic sessions were reserved for in-depth discussions on numerical methods, academic flows, scalable solvers, and CFD in aerospace.

In the first of two special sessions, JSC staff presented the first European exascale supercomputer JUPTER, which will be hosted at JSC. In the second, staff unveiled a the Partnership for Advanced Computing in Europe (PRACE) as the new European Association for Users and HPC Centers. All sessions were well attended and the lively discussions showed significant interest and engagement from the CFD community. Most of the submissions will be published in the conference proceedings.
Maximillian Tandi

News Briefs

JSC

nekCRF Plugin Supports Extreme-Scale Simulations for Green Energy

As a result of the Center of Excellence in Combustion (CoEC), funded by the European Union's Horizon 2020 research and innovation program (grant agreement No. 952181), an international team of researchers (K2/Microdat, ETH Zürich, AU of Thessaloniki, U of Illinois UC, Jülich Supercomputing Centre) published the GPU-accelerated chemistry plugin nekCRF. nekCRF, together with nekRS, a leading exascale-capable CFD code, enables the efficient solution of highly accurate finite-rate chemistry simulations on tens of thousands of GPUs. Compared to its CPU-optimized predecessor, the time-to-solution is reduced by more than an order of magnitude. nekCRF thus has the potential to significantly accelerate important developments for the green energy transition. It is currently deployed simulations utilizing the full JUWELS.

Hero Simulations Leverage 3,360 GPUs to Decipher Rayleigh-Bénard Convection

Rayleigh-Bénard convection (RBC) is a type of natural convection that has been studied extensively to understand the fundamental mechanisms of heat transfer. A key parameter of RBC is the Rayleigh number, which quantifies the strength of the buoyant force that drives the convective flow. As the Rayleigh number increases, the level of turbulence increases, making numerical simulations increasingly expensive. An international team of researchers from TU Ilmenau, Jülich Supercomputing Centre, Occidental College and New York University used the CFD code nekRS with 3,360 GPUs of JUWELS Booster to decipher the physics behind the emerging boundary layers of RBC at Rayleigh numbers of up to 10^{12} in a periodically extended, open domain. A significant result is that the velocity field is characterized by strong fluctuations in all three components around a non-existing or weak mean flow. This is in contrast with convection in small, confined domains, which typically exhibit a pronounced mean flow with small fluctuations about the mean. These simulations were among the largest, or perhaps the largest, ever carried out on the JUWELS Booster, and the results have been accepted for publication in the *Journal of Fluid Mechanics*.

A preprint is available on arXiv: https://arxiv.org/html/2403.12877v2

JuWinHPC is Gaining Momentum

Founded about two years ago, the "Jülich Women in HPC" (JuWinHPC) chapter of the international initiative "Women in HPC" (WHPC) has become active and received notable attention during the past months. JuWinHPC's main goal is to strengthen the local community of women involved in HPC and to promote equality, diversity, and inclusion. The team organizes monthly networking meetings that are open to everyone. Each meeting covers various topics, such as a discussion about the gender dimension in AI on International Women's Day. Furthermore, the JuWinHPC team, together with other WHPC chapters, successfully organized "birds of a feather" (BoF) sessions called "Super(computing)heroes" at ISC24 and SC24 to showcase the amazing work of women in HPC and to provide role models for young community members. The network gained attention through invited talks at international events and its blog. It also won the internal Forschungszentrum Jülich "JuEngage" award for their engagement on campus. Lastly, JuWinHPC collaborates with other WHPC chapters, like the recently founded NHR group, and supports others considering whether to start their own local initiatives.

For more information, please visit: https://www.fz-juelich.de/en/ias/jsc/juwinhpc

Study Conducted by JSC and LAION Researchers Reveals Dramatic LLM Reasoning Breakdown

Even the best AI language learning models (LLMs) fail dramatically when it comes to simple logical questions. This is the conclusion of researchers from the Jülich Supercomputing Centre (JSC), the School of Electrical and Electronic Engineering at the University of Bristol and the LAION AI laboratory. In their paper, "Alice in Wonderland: Simple Tasks Showing Complete Reasoning Breakdown in State-Of-the-Art Large Language Models," (preview available at https://arxiv.org/abs/2406.02061), the scientists attest to a "dramatic breakdown of function and reasoning capabilities" in the tested state-of-the-art LLMs and suggest that although language models have the latent ability to perform basic reasoning, they cannot access it robustly and consistently. The authors of the study – Marianna Nezhurina, Lucia Cipolina-Kun, Mehdi Cherti and Jenia Jitsev – call on, "the scientific and technological community to stimulate urgent re-assessment of the claimed capabilities of current generation of LLMs." They also call for the development of standardized benchmarks to uncover weaknesses in language models related to basic reasoning capabilities, as current tests have apparently failed to reveal this serious failure.

To read the full article, please visit: https://go.fzj.de/jsc-study-llms-reasoning-breakdown

HLRS

Hunter Installation Begins at HLRS

Installation and testing have begun at the High-Performance Computing Center Stuttgart for its next flagship supercomputer, called Hunter. The system was delivered on October 1 and is scheduled to go into service early in 2025. Manufactured by Hewlett Packard Enterprise, Hunter is based on the HPE Cray EX4000 platform, and is designed to enable large-scale, state-of-the-art applications of simulation, artificial intelligence, and data analytics, including hybrid computing approaches that combine diverse methods into powerful workflows. Hunter marks a milestone for HLRS, as it moves away from the center's previous CPU-oriented approach to make greater use of GPUs. At its core is the AMD Instinct™ MI300A accelerated processing unit (APU), which combines CPUs, GPUs, and high-bandwidth memory in a single package. With an expected peak performance of approximately 39 petaFLOPS, Hunter will be 50 % faster than its predecessor, Hawk, while its utilization of more energy-efficient GPUs will slash energy requirements at peak performance by approximately 80 %. Hunter is conceived as a transitional system that will support HLRS's user community in preparing for an upcoming exascale system, called Herder, which is currently planned for arrival in 2027.

HPC Training Program to Offer New Certification Opportunities

The HPC training program at HLRS is one of Europe's most experienced and comprehensive programs for continuing professional education in high-performance computing and related technologies. In a new initiative launched this fall, the program also now offers the chance for trainees to achieve professional certification in key high-performance computing occupations. HLRS offers professional certification tracks in simulation, machine learning and artificial intelligence, parallel programming, and HPC administration. Each track offers a flexible curriculum, enabling trainees to select from a menu of courses and to customize their training experience in a way that aligns with their professional interests and career goals. All participants in the certification program must complete at least three courses and pass an exam. Once the requirements are fulfilled, a certificate documents successful completion of the program. Ultimately, professional certification at HLRS offers trainees the chance to develop in-depth knowledge and skills, and to stand out in today's rapidly evolving tech landscape.

Information about HLRS's professional certification program is now available at www.hlrs.de/training/professional-certification.

HLRS Signs Memorandum of Understanding with TalTech

The High-Performance Computing Center Stuttgart and the School of Information Technologies at the Tallinn University of Technology (TalTech) have launched a new formal research collaboration. Signed in Tallinn, Estonia on September 12, a five-year memorandum of agreement pledges to build on the organizations' complementary strengths in high-performance computing, applications of simulation for automotive research and industry, global systems science, applications for climate research, and professional training. The memorandum of agreement will facilitate collaborative activities including staff exchange programs, joint research projects, professional services, and continuing education programs, as well as technology transfer and collaboration with industry. HLRS and TalTech have also identified several preliminary scientific areas for collaboration, including the development of numerical algorithms for engineering applications, the application of emerging technologies for parallel programming and networking computing to models in engineering applications, and the use of immersive visualization technologies in numerical simulation for engineering challenges.

HPC-AI Convergence at AISys 2024

The convergence of high-performance computing (HPC) and artificial intelligence (AI) has begun to unlock exciting new opportunities for science, industrial R&D, and many other fields. Realizing the potential of this convergence, however, will mean ensuring that HPC architectures satisfy the needs of the AI community, and that solutions are available to integrate AI applications and HPC workflows. This means promoting close collaboration and knowledge transfer between experts in HPC and AI. A special track at the First International Conference on AI-Based Systems and Services (AISyS 2024) organized by Prof. Michael Resch and Dennis Hoppe of HLRS offered a chance to explore synergies between HPC and AI. Held in Venice, Italy, the special track explored both how HPC can leverage AI to enhance system performance and efficiency, and how AI can harness the immense computational capabilities of HPC to tackle increasingly complex modelling and simulation challenges. Talks discussed how AI could be used to address global challenges such as sustainable city planning and wildfire prevention, how AI surrogate models could improve traditional applications of HPC for simulation, and cybersecurity considerations that arise when integrating AI into traditional HPC systems.

HLRS Wins Datacenter Strategy Award for "Transformation"

The High-Performance Computing Center Stuttgart was named winner of the Datacenter Strategy Award in the category "Transformation" at the 2024 Datacenter Strategy Summit in Bad Homburg. The prize recognizes HLRS for its sustainable computing center strategy and its visionary new construction project, HLRS III. HLRS III will house the center's upcoming exascale system, called Herder, when it arrives in 2027. It will combine sustainable materials, photovoltaic systems, and dynamic power management, and importantly, will also be accompanied by construction of a facility for waste heat reuse. Heat generated by the new supercomputer will be captured and distributed across the Vaihingen campus of the University of Stuttgart, contributing to its decarbonization. HLRS III will also implement dynamic power management, which will enable additional efficiency gains. HLRS already uses this approach in the operation of Hawk, limiting the system's maximum performance to ensure the safe functioning of critical infrastructure components such as the uninterruptible power supply (UPS). By combining dynamic power management with waste heat reuse, HLRS III will enable a more efficient, dynamic distribution of unused energy, enabling HLRS and the University of Stuttgart to reach a new milestone on the way to sustainability.

HLRS Completes TISAX Level 3 Assessment for Information Security

Since 2021 the University of Stuttgart has been registered as a participant in the Trusted Information Security Assessment Exchange (TISAX) on behalf of the High-Performance Computing Center Stuttgart (HLRS). Governed by the ENX Association on behalf of the German Association of the Automotive Industry (VDA), TISAX prescribes a standardized set of strict requirements for handling sensitive data that are intended to protect data confidentiality, data integrity, and data availability. The TISAX framework closely follows the international standard ISO 27001 for information security management systems, but also prescribes additional requirements for the implementation of controls that must be followed. Recently, the audit provider TÜV Nord CERT GmbH conducted a new evaluation of HLRS's information security management system at TISAX Level 3, the most comprehensive auditing approach within the information security framework. It is designed to ensure adherence to information security standards that are appropriate for handling data with "very high protection needs." The results of HLRS's TISAX assessment have been published in the ENX portal.

LRZ

LRZ's Invites Participation in its "HPC Lounge"

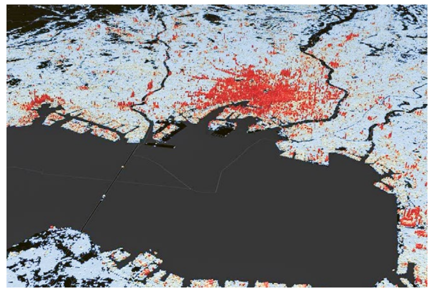
Do you want to apply for computing time on LRZ supercomputers? Are you stuck in coding or would like to optimise your algorithms? Do you want to know which software or AI model you can use on SuperMUC-NG or learn more about the suite of LRZ's resources more broadly? Then join the LRZ HPC Lounge to regularly discuss your open questions.

For more information about the event, as well as a schedule with dates and times, please visit: https://doku.lrz.de/hpc-lounge-813636379.html

Plasma: Beautiful Chaos

With the help of SuperMUC-NG, a German-Australian team researched the composition of the plasma that flows through our universe. The group simulated gas densities and how electrical radiation is distributed in the plasma clouds, resulting in stunning visualizations. The work and what it means for our planet and the universe as a whole was recently detailed in a feature article published in *New Scientist magazine*.

For the full article, please visit: https://www.newscientist.com/article/2433764-stunning-image-reveals-the-intricate-structure-of-supersonic-plasma



Big Data from Satellites

More power, more memory, faster connections: The terrabyte high performance data analytics platform at the LRZ has been further expanded. The tailor-made system for analysing earth observation data has been in heavy use by researchers from the German Space Agency (DLR, Deutsches Zentrum für Luft- und Raumfahrt). The researchers recently published a report on how the new technology is advancing the exploration of Earth.

To read the full report, please visit: https://www.dlr.de/en/media/publications/magazines/all-digital-magazines/dlrmagazine-175/big-data-in-earth-observation

Fresh Air for Supercomputers

The ventilation systems in the computer rooms at the LRZ are currently being modernised. The center's building management team is working with a supplier of fans and ventilation technology to convert and upgrade the technology so that it will both provide fresh air for the supercomputers and improve energy efficiency by modernizing the building infrastructure at the same time.

For more information about the center's recently sustainability efforts surrounding ventilation, please visit: https://www.lrz.de/presse/ereignisse/2024-08-08-Economical-Fans/

"It has always been important to think years into the future." Erika Fischer

Erika Fischer moved to Stuttgart in 1961 and has made important contributions to the growth of supercomputing in the city ever since. She is still involved in planning HLRS's next major expansion, HLRS III.

INSIDE
GCS

Staff Spotlight: A Life in Supercomputing

Erika Fischer, HLRS

When looking back at the history of high-performance computing, few people alive today have experienced as much as Erika Fischer. After fleeing East Germany in 1961 just in time to escape the construction of the Berlin Wall, Fischer arrived in Stuttgart as a refugee, sharing a three-bedroom apartment with two other families who had also left the GDR. Having previously worked at the Bauhaus University Weimar, she found a position working under Prof. John Argyris at the Institute of Aerospace Statics and Dynamics (ISD) of the University of Stuttgart. She soon joined the ISD's technical department, where she received extensive training on computer maintenance. At that time it probably never occurred to anyone that she would spend more than 60 years making important contributions to the growth of supercomputing in Stuttgart.

Fischer started her career in the predigital era, when computers were made of moving parts, and algorithm inputs and outputs consisted of holes punched in paper. She soon became responsible for making sure that the systems remained operational, including performing emergency repairs. "This work was very important because computers at that time were not nearly as stable as they are today," she recalls. "From the beginning it was always vital to me to offer our users the resources they need, and to ensure that our systems ran as consistently as possible."

At that time, this was more than a full-time job. Fischer often found herself studying circuit diagrams at her kitchen table or being summoned to the computer if a part malfunctioned. The fact that her husband was often on the road as a computer technical support specialist also meant that it was sometimes a difficult juggling act for the young mother. "When my son was still very small and there was a problem with the computer I would bring him along in a bag and set him down beside the console while I worked," she says, remembering those times affectionately. "When he was a little older and could use a telephone, he knew that he could call me if he woke up at night."

As Fischer's responsibilities grew she became department manager, which encompassed a larger range of tasks including user management, personnel management, resource management, and oversight of the infrastructure. As supercomputers steadily became more powerful and complex, she also needed to keep at least one eye focused on the future. This meant participating in the procurement of new IT systems and overseeing the infrastructure improvements that would be necessary to support them.

"It has always been important to think years into the future," Fischer explains. "You need to anticipate how performance will increase, how much physical space the systems will take up, and what other resources will be needed for their operation. And at some point, you realize that your current facility is no longer capable of meeting those needs."

As scientific computing advanced in Stuttgart, the ISD computing systems were incorporated into the regional computing center, which then evolved into the University of Stuttgart Computing Center (RUS). With the growing workload that came with a rapidly expanding user community, Fischer eventually shifted away from operations management to concentrate on the planning of new facilities.

In 1996 the High-Performance Computing Center Stuttgart (HLRS) was founded as a new entity, and Fischer recognized that it would require more space, higher power capacity, and as much flexibility as possible to address the unique needs of HPC users. In the years to come she wrote position papers and worked hard to persuade the relevant committees to approve construction of a new building. Without her advocacy, HLRS's current, permanent home at Nobelstraße 19 might never have been built.

As Fischer approached retirement age in 2002, planning for the new facility was underway and she was deeply involved. Because continuity and expertise were still needed,

> **"It's possible for an individual to achieve great things when you have reasonable goals and the determination not to give up in the face of difficulties."** Erika Fischer

she continued working at the center on a contractual basis to ensure that the project came to a successful completion. Even in this more limited role, however, there was no shortage of work to do. Planning a facility as complex as a high-performance computing center requires attention to countless details – from power supply, to security, to fire prevention, to the planning of comfortable workspaces, among many other considerations – and Fischer remained a key contact for planners, engineers, and construction personnel.

Even after HLRS's new home opened in 2004, Fischer continued to oversee the numerous service contracts necessary to keep the center operating smoothly. When it became time to expand HLRS further, she oversaw the planning and construction of the building housing the center's state-of-the-art training facility, which was inaugurated in 2017.

Despite her age Erika Fischer shows little sign of slowing down. Working together with HLRS's Dr. Ralf Schneider, she continues to play a central role in the planning of HLRS's next major expansion – construction of a new building called HLRS III, which will begin in early 2025. The facility will be a home to the center's upcoming exascale supercomputer, called Herder, when it arrives in 2027. A long way from holes punched in paper, indeed.

Reflecting on her life in high-performance computing, Fischer mentions three things of which she is particularly proud: "It has been very important to me that I have been in a position to follow the development of information technology from its very beginnings. The University of Stuttgart was one of the pioneering universities, and it is wonderful to have seen what happened here. I also learned that it's possible for an individual to achieve great things when you have reasonable goals and the determination not to give up in the face of difficulties. And finally, it has brought me joy to see that the things I have been a part of have great value to others, including scientists and others who need their computing results quickly and reliably."

Although Fischer's "retirement" has not been typical, it is clear that her dedication to HLRS continues to be a source of great personal motivation. "For practically my whole life, working at the University of Stuttgart has been extremely interesting," she says. "I have had the pleasure of working with many outstanding colleagues and system users. I really can't complain, and it is a wonderful thing when a person can say that." *Christopher Williams*

Education and Training

GCS and its member centers provide first-class training opportunities for the national and European HPC communities. More than 50 highly qualified scientists work as trainers for over 100 courses (or 280 course days), which positions GCS as one of the leading training institutions in all of Europe.

Over the past three years, we have seen a significant shift in the way training events are conducted. Before the Corona pandemic, virtually all courses and workshops took place on site at one of the supercomputer centers, typically spanning multiple full days in a row. As face-to-face meetings were suspended in early 2020, all our training offers were switched to online events, employing a combination of video conferencing software, chat programs, and other collaborative tools.

We currently see the training landscape changing, with a mixture of on-site courses, online workshops, and hybrid events becoming the norm. As a side effect, training collaboration between the different GCS centers has further increased, with more events being co-organized by multiple locations, and instructors from different centers supporting each other's training efforts.

Our training events are always growing to adapt to technological changes, and our course offerings reflect an increased emphasis on GPU programming, AI methods like deep learning, and other emerging applications.

For a full list of training courses offered, please visit:

GCS	HLRS	JSC	LRZ
https://www.gauss-centre.eu/trainingsworkshops/	https://www.hlrs.de/training/	https://go.fzj.de/jsc-courses	https://www.lrz.de/education

The Hawk supercomputer at HLRS.

High-Performance Computing Center Stuttgart

The High-Performance Computing Center Stuttgart (HLRS) was established in 1996 as Germany's first national high-performance computing center. A research institution affiliated with both GCS and the University of Stuttgart, HLRS provides infrastructure and services for HPC, data analytics, visualization, and artificial intelligence to academic users and industry across many scientific disciplines, with an emphasis on computational engineering and applied science.

Supercomputing for industry

Through a public-private joint venture called hww (Höchstleistungsrechner für Wissenschaft und Wirtschaft), HLRS ensures that industry always has access to state-of-the-art HPC technologies. HLRS also helped to found SICOS BW GmbH, which assists small and medium-sized enterprises in accessing HPC technologies and resources. Additionally, HLRS cofounded the Supercomputing-Akademie, a training program that addresses the unique needs of industrial HPC users.

Guiding the future of supercomputing

HLRS scientists participate in dozens of funded research projects, working closely with academic and industrial partners to address key problems facing the future of computing. Projects develop new technologies and address global challenges where supercomputing can provide practical solutions. With the support of the EuroHPC Joint Undertaking, HLRS is also currently coordinating efforts to build and integrate HPC competencies across Europe. The center is certified for environmental management under the EU's Eco-Management and Audit Scheme (EMAS) and for information security under the ISO 270001 standard.

High-Performance Computing Center Stuttgart

CONTACT

High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart

Prof. Dr.-Ing. Dr. h.c. Dr. h.c. Hon.-Prof. Michael M. Resch
Nobelstraße 19, 70569 Stuttgart, Germany
Phone +49 - 711 - 685 - 8 72 69
resch@hlrs.de
www.hlrs.de

CENTER
DESCRIPTIONS

Compute servers currently operated by HLRS

System	Size	Peak Performance (Tflop/s)	Purpose	User Community
HPE APOLLO 9000 "HAWK"	5,632 nodes 720,896 cores 1.44 PB memory	26,000 TF	Capability Computing	European and German Universities, Research Institutes, and Industry
HAWK GPU EXTENSION	24 nodes 192 NVIDIA A100 GPUs	120,000 TF AI performance	Machine Learning, Artificial Intelligence applications	German and European (PRACE) research organizations and industry
NEC CLUSTER (VULCAN, VULCAN 2)	653 nodes 17952 cores 117 TB memory	974 TF	Capacity Computing	German Universities, Research Institutes, and Industry
NEC SX-AURORA TSUBASA	64 nodes 512 cores 3072 GB memory	137.6 TF	Vector Computing	German Universities, Research Institutes, and Industry
CRAY CS-STORM	8 nodes 64 GPUs 2,048 GB memory	499.2 TF	Machine Learning Deep Learning	German Universities, Research Institutes, and Industry

Modular Supercomputer JUWELS at the Jülich Supercomputing Centre.

Jülich Supercomputing Centre Forschungszentrum Jülich

The Jülich Supercomputing Centre (JSC) at Forschungszentrum Jülich is committed to enabling scientists and engineers to explore some of the most complex grand challenges facing science and society. Our research is performed through collaborative infrastructures, exploiting extreme-scale supercomputing, AI at scale, quantum computing, and federated data services.

Provision of supercomputer resources:

JSC provides access to supercomputing resources of the highest performance for research projects coming from academia, research organizations, and industry. Users gain access for projects across the science and engineering spectrum in the fields of modeling and computer science.

Core tasks of JSC are:

- Supercomputer-oriented research and development in selected fields of physics and other natural sciences by research groups and in technology, e.g. by doing co-design together with leading HPC companies.
- Implementation of strategic support infrastructures including community-oriented simulation and data laboratories and cross-sectional teams, e.g. on mathematical methods and algorithms and parallel performance tools, enabling the effective usage of the supercomputer resources.
- Cutting-edge quantum computing research and access through the Jülich UNified Infrastructure for Quantum computing (JUNIQ).
- Higher education for master and doctoral students in close cooperation with neighbouring universities.

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

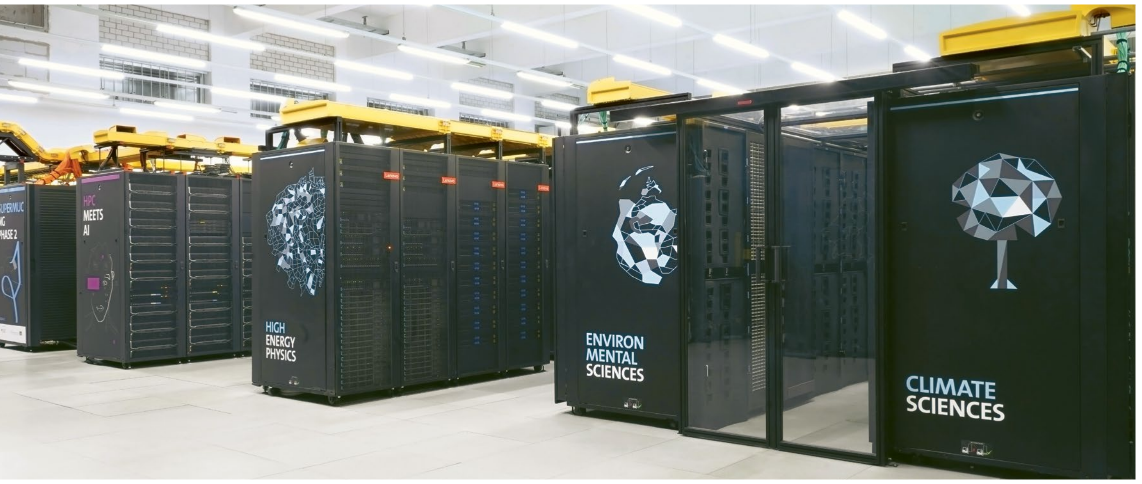
CONTACT

Jülich Supercomputing Centre (JSC)
Forschungszentrum Jülich

Prof. Dr. Dr. Thomas Lippert
Wilhelm-Johnen-Straße, 52425 Jülich, Germany
Phone +49 - 24 61- 61- 64 02
th.lippert@fz-juelich.de
www.fz-juelich.de/jsc

Compute servers currently operated by JSC

System	Size	Peak Performance (Tflop/s)	Purpose	User Community
Modular Supercomputer "JUWELS"	Cluster (Atos): 10 cells, 2,567 nodes 122,768 cores Intel Skylake 224 NVIDIA V100 GPUs 275 TByte memory	12,266	Capability Computing	European and German Universities and Research Institutes
	Booster (Atos): 39 racks, 936 nodes 44,928 cores AMD EPYC Rome 3,744 NVIDIA A100 GPUs 629 TByte memory	75,020		
Supercomputer "JURECA"	Data-Centric Cluster (Atos): 768 nodes, 98,304 cores AMD EPYC Rome 768 NVIDIA A100 GPUs 443 TByte memory	18,515	Capacity and Capability-Computing	European and German Universities, Research Institutes, and Industry
JUPITER Development System "JEDI"	1 rack, 48 nodes 13.824 cores 192 NVIDIA GH200 Superchips 40,5 TByte memory	10,264	Capability Computing	European and German Universities and Research Institutes
Atos Cluster "JUSUF"	205 nodes, 26,240 cores AMD EPYC Rome 61 NVIDIA V100 GPUs 52 TByte memory	1,372	Capacity and Cloud Computing	European and German Universities and Research Institutes through Human Brain Project
Modular Supercomputer "DEEP-EST" (Prototype)	Cluster: 50 nodes, 1,200 cores Intel Xeon Gold 6146 9.6 TByte memory + 25.6 TByte NVM	45	Capacity Computing (low-/medium-scalable code parts)	Partners of the "DEEP" and "SEA" EU-project series and interested users through Early Access Programme
	Booster: 75 nodes, 600 cores Intel Xeon Silver 4215 75 NVIDIA V100 GPUs 6 TByte memory 38 TB NVM	549	Capacity and Capability Computing (high-scalable code parts)	
	Data Analytics Module: 16 nodes, 768 cores Intel Xeon Platinum 8260 16 NVIDIA V100 GPUs 6.7 TB memory + [statt 7.1] 43 TB NVM7.1 TByte memory + 32 TByte NVM	170	Capacity and Capability Computing (data analytics codes)	
D-Wave Quantum Annealer "JUPSI"	More than 5,000 qubits	No classical performance measure applicable	Quantum Computing	German Universities and Research Institutes (10%) Industry Applications and D-Wave customers (90%)



The SuperMUC-NG supercomputer at LRZ.

Leibniz Supercomputing Centre

For more than six decades, the Leibniz Supercomputing Centre (Leibniz-Rechenzentrum, LRZ) has been at the forefront of its field as a world-class high-performance computing center dedicated to providing an optimal IT infrastructure to its clients throughout the scientific community – from students to postdocs to renowned scientists – and in a broad spectrum of disciplines – from astrophysics and engineering to life sciences and digital humanities.

Leadership in HPC and HPDA

Located on the research campus in Garching near Munich, the LRZ is a leadership-class HPC and HPDA facility delivering top-tier supercomputing resources and services on the national and European levels. Top-notch specialists for HPC code portability and scalability support the broad user base at LRZ and ensure that the systems are running their operations in the most energy efficient way possible.

Quantum and Future Computing at LRZ

LRZ is leading the way forward in the field of future computing, focusing on emerging technologies like quantum computing and integrating AI on large-scale HPC systems. A robust education program that touches on HPC, machine learning, artificial intelligence, and big data complements LRZ's offerings.

IT backbone for Bavarian science

In addition to its role as a national supercomputing center, LRZ is also the IT service provider for all Munich universities as well as research organizations throughout Bavaria.

lrz Leibniz Supercomputing Centre
of the Bavarian Academy of Sciences and Humanities

CONTACT

Leibniz Supercomputing Centre (LRZ)

Prof. Dr. Dieter Kranzlmüller
Boltzmannstraße 1, 85748 Garching near Munich, Germany
Phone +49 - 89 - 358 - 31- 80 00
leitung@lrz.de
www.lrz.de

Compute servers currently operated by LRZ

System	Size	Peak Performance (Tflop/s)	Purpose	User Community
SUPERMUC-NG PHASE 1 INTEL/LENOVO THINKSYSTEM	6,336 direct hot-water cooled compute nodes, 304,128 cores, Intel Xeon Platinum 8174, 608 TByte of memory, Omni-Path 100G interconnect	26,300	Capability Computing	German universities and research institutes(Tier-1)
	144 direct hot-water cooled compute nodes, 6,912 cores, Intel Xeon Platinum 8174, 111 TByte of memory, Omni-Path 100G interconnect	600	Capability Computing	
SUPERMUC-NG COMPUTE CLOUD	64 air-cooled nodes, 5,120 cores, Intel Xeon Gold 6148, 64 Nvidia Tesla V100	644 (CPUs + GPUs) 7,168 AI Performance*	Cloud Computing	German Universities and Research Institutes(Tier-1)
SUPERMUC-NG PHASE 2	240 direct hot-water cooled compute nodes, 26.880 Intel Xeon Platinum 8480+ compute cores (Sapphire Rapids), 122,88 TByte of memory, 960 accelerators (Intel Ponte Vecchio), NVIDIA HDR Infiniband interconnect	27,960	Capability Computing & Machine Learning, AI applications	German Universities and Research Institutes(Tier-1)
COOLMUC-2 LENOVO NEXTSCALE	812 direct hot-water cooled compute nodes, 22,736 cores Intel Haswell EP, 51.968 TByte of memory, NVIDIA FDR 14 Infiniband interconnect	1400	Capability Computing	Bavarian Universities (Tier-2)
COOLMUC-3 MEGWARE SLIDE SX	148 direct hot-water cooled compute nodes, 9,472 cores, Intel Knights Landing, 15 TByte of memory, Omnipath interconnect	459	Capability Computing	Bavarian Universities (Tier-2)
LRZ AI SYSTEMS	17 nodes (NVIDIA GPU-based), HDR Infinband 88 NVIDIA GPUs, 3,328 GB HBMemory 1,424 CPUs, 5,824 GB DDR4 Memory	2,302 66,690 AI Performance*	Machine Learning, AI applications	Bavarian Universities
CEREBRAS CS-2	1 node with 850,000 compute cores, 40GB SRAM, 20 PB/s memory bandwidth and 220Pb/s interconnect	3,570,000 AI Performance* (estimate based on arXiv:2204.03775)	Purpose-built Deep Learning System	Select users
LRZ QUANTUM COMPUTING RESOURCES	Eviden Qaptiva 1	N/A	Quantum simulation	Bavarian Universities
	Eviden Qaptiva 2	N/A	Quantum simulation	Select users
	QMWare	N/A	Quantum simulation	Bavarian Universities
	IQM 5-qubit system	N/A	Quantum computation	Select users
	IQM 20-qubit system 1	N/A	Quantum computation	Select users
	IQM 20-qubit system 2	N/A	Quantum computation	German, Bavarian and European as pre.Euro-Q-Exa program
	AQT 20-qubit ion-trap system	N/A	Quantum computation	MQV partners

*AI Performance refers to GPU peak performance for FP16 operations. For Nvidia GPUs, it is specific to different architectures. P100 architecture: CUDA core performance. V100 architecture: Mixed precision Tensor Core performance. A100: Structured sparsity Tensor Core performance.

Would you like to keep up with the newest research and news happening at the GCS centers?

Please send your postal address to pr@hlrs.de and you will recieve InSiDE magazine twice a year.

IMPRINT

InSiDE is published two times a year by the Gauss Centre for Supercomputing e.V., Alexanderplatz 1, 10178 Berlin

Editor-in-Chief: Michael Resch, HLRS Höchstleistungsrechenzentrum Stuttgart, Nobelstaße 19, 70569 Stuttgart, resch@hlrs.de

Executive Editor: Eric Gedenk, GCS, eric.gedenk@lrz.de

Art Direction and Design:
GROOTHUIS. Gesellschaft der Ideen und Passionen mbH für Kommunikation und Medien, Marketing und Gestaltung: Ulrike Jänecke; www.groothuis.de

Print:
Gutenberg Beuys Feindruckerei GmbH

This magazine has been printed climate-neutral on paper that has been certified by FSC®.

RECYCLED
Paper made from recycled material
FSC® C009051

Print product with financial climate contribution
ClimatePartner.com/10951-2404-1002

Contributing authors: Eric Gedenk, GCS, eric.gedenk@lrz.de; Christopher Williams, HLRS, williams@hlrs.de; Susanne Vieser, LRZ, susanne.vieser@lrz.de; Moritz Wolter, University of Bonn, moritz.wolter@uni-bonn.de; Regine Panknin, FZ Jülich, r.panknin@fz-juelich.de; Ehsan Zandi, JSC, e.zandi@fz-juelich.de; Maximillian Tandi, JSC, m.tandi@fz-juelich.de

Photo credits:
Cover page: istock/Lakes4life; cover page 2: unsplash.com / Michael Krahn; pg. 6/pg. 8, Veronika Hohenegger, LRZ; pg. 9, HLRS; pg. 10 TRUMPF Gruppe, MACK ONE, Seedbox; pg. 3, Gasenzer, K. and M. Wolter (2024). DOI: https://doi.org/10.48550/arXiv.2305.13033; pg. 14 HLRS/generated with Firefly by Groothuis; pg. 16, Forschungszentrum Jülich; pg. 18, Pollinger T, et al. 2024. Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2024; pg. 22, Jennifer Matthews, UCSD; pg. 24 istock / Stasz D Sirotkin; pg. 26, Helmholtz Foundation Model Initiative; pg. 28, SoftSim Consult; pg. 30, CompBioMed project; pg. 31, Cerebras Systems; pg. 33/34: Miguel Ángel Cano Santizo; pg. 36, Samuel et al (2024) https://doi.org/10.1017/jfm.2024.853, JSC; pg. 37, HLRS; pg. 38, HLRS; pg. 39, LRZ, Veronika Hohenegger, DLR; pg. 40, Christopher Williams; pg. 43, HLRS

InSiDE magazine (German: Innovatives Supercomputing in Deutschland) is the bi-annual publication of the Gauss Centre for Supercomputing, showcasing recent highlights and scientific accomplishments from users at Germany's three national supercomputing centers. GCS was founded in 2007 as a partnership between the High-Performance Computing Center Stuttgart, Jülich Supercomputing Centre, and the Leibniz Supercomputing Centre. It is jointly funded by the German Ministry of Education and Science (Bundesministerium für Bildung und Forschung – BMBF) and the corresponding ministries of the three states of Baden-Württemberg, North Rhine-Westphalia, and Bavaria.

GCS
Gauss Centre for Supercomputing

www.gauss-centre.eu

Funding for GCS HPC resources is provided by:

Federal Ministry of Education and Research

Bavarian State Ministry of Science and the Arts

Baden-Württemberg
MINISTRY OF SCIENCE, RESEARCH AND ARTS

Ministry of Culture and Science of the State of North Rhine-Westphalia