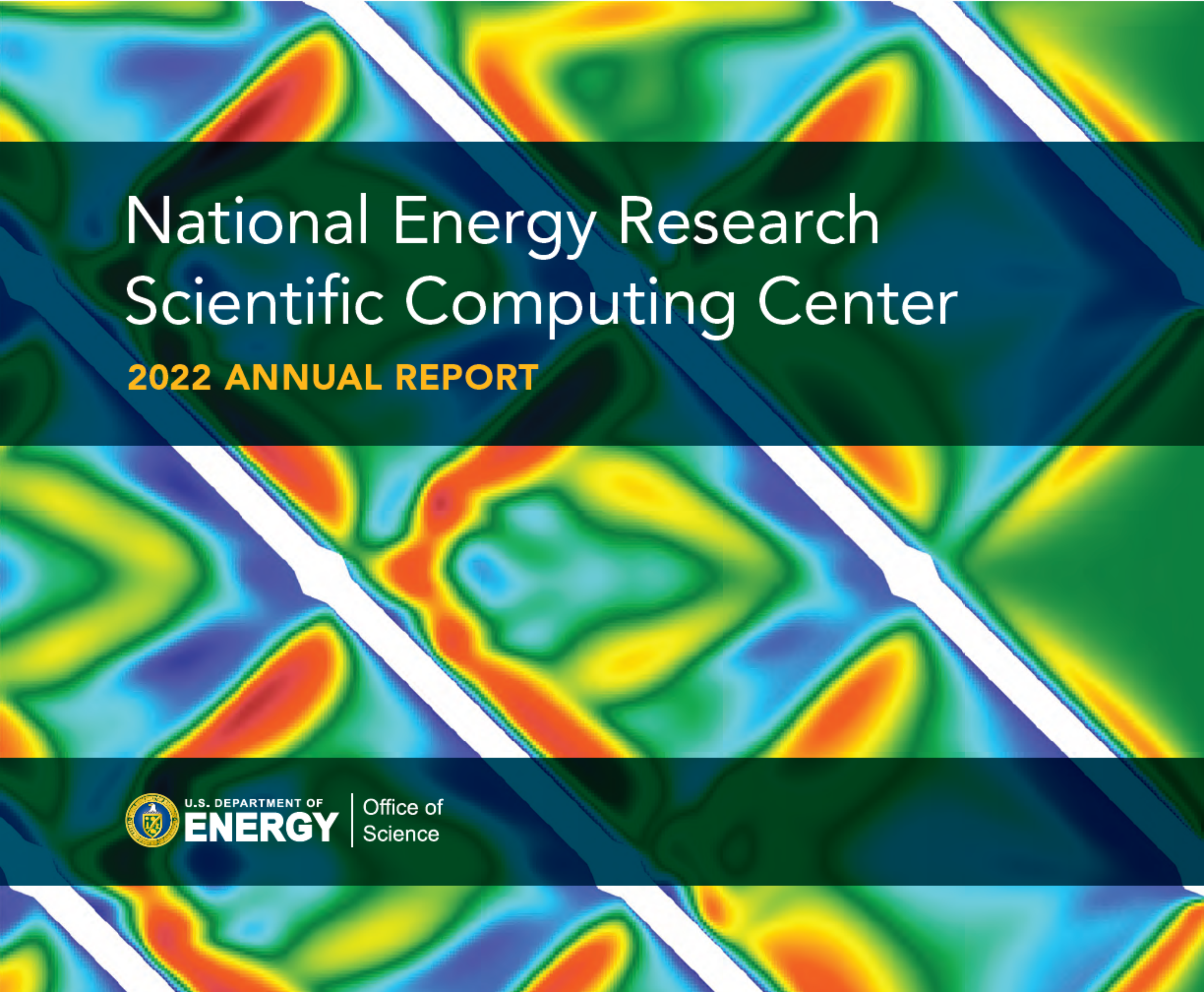
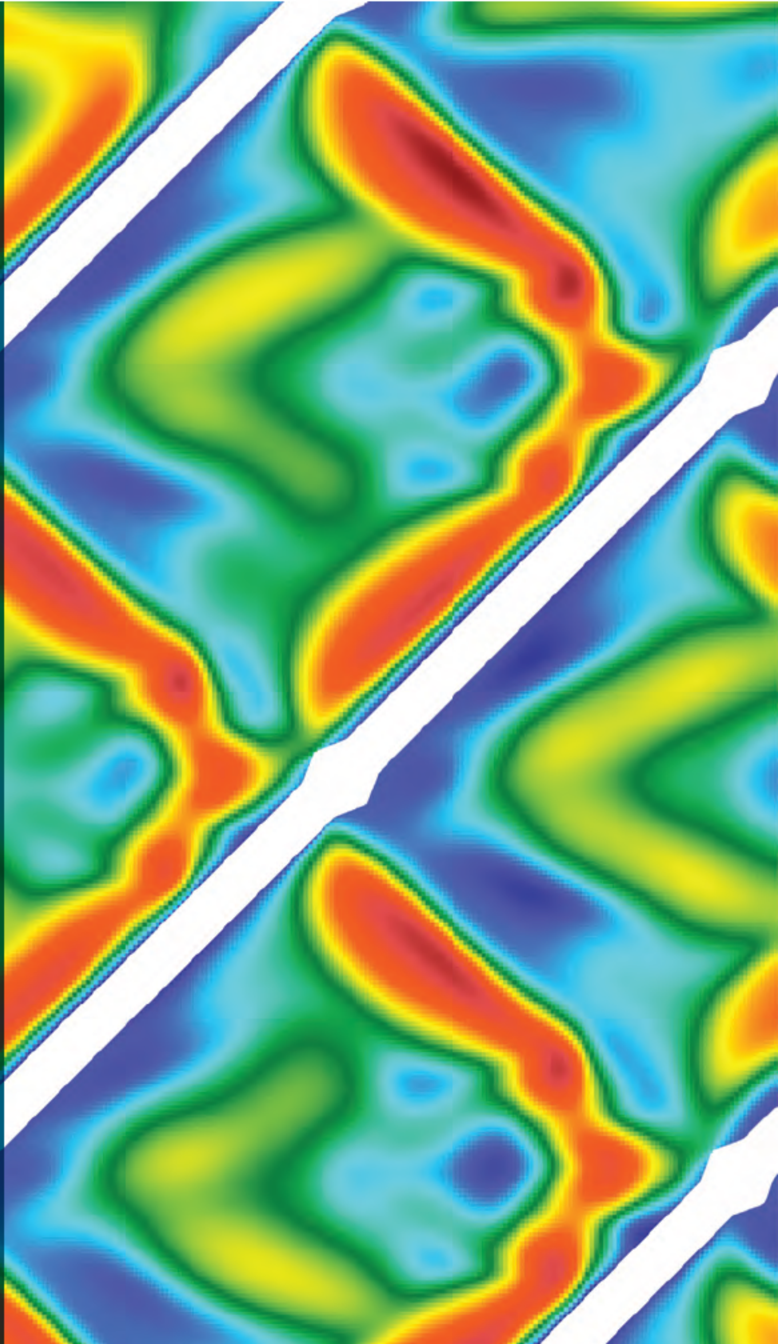


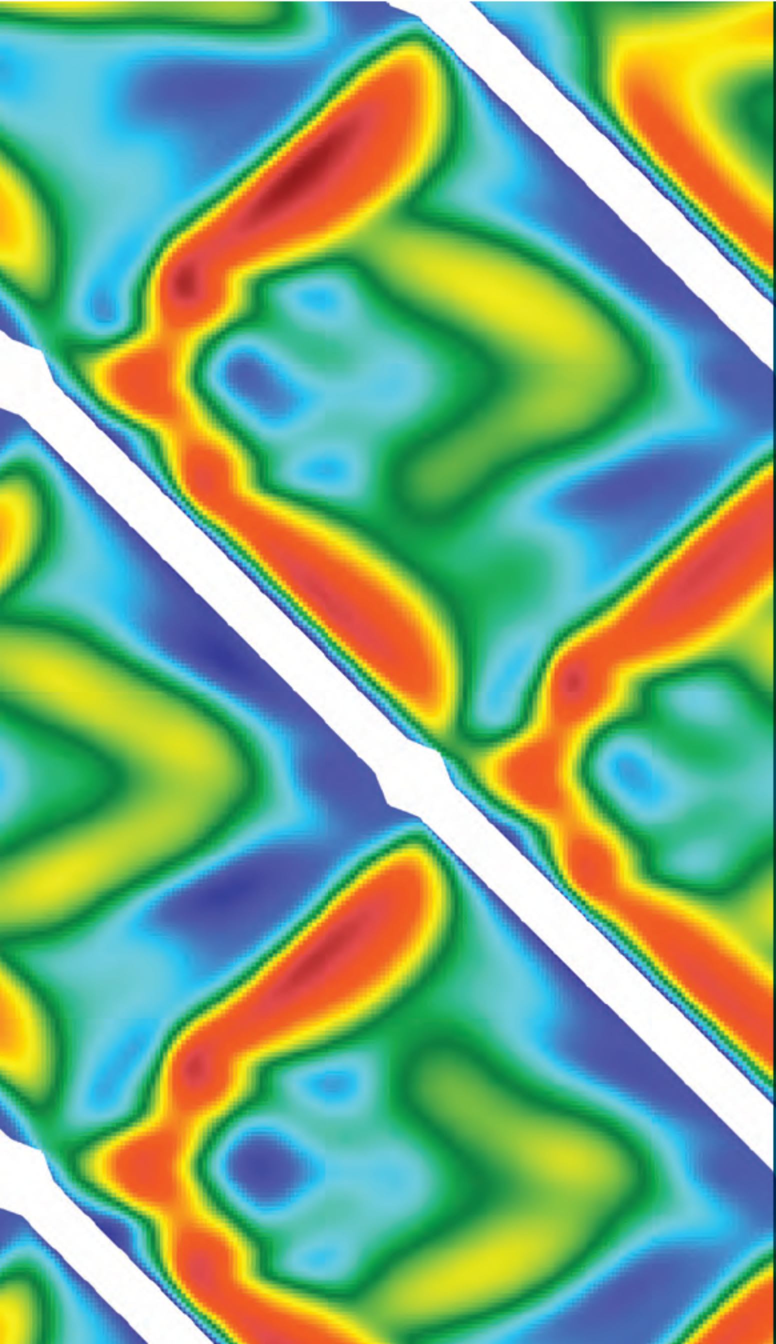
National Energy Research Scientific Computing Center

2022 ANNUAL REPORT

U.S. DEPARTMENT OF ENERGY | Office of Science

Table of Contents





The National Energy Research Scientific Computing Center is the mission high-performance computing facility for the U.S. Department of Energy's Office of Science (SC). NERSC is managed by Lawrence Berkeley National Laboratory and funded by the U.S. Department of Energy's Office of Science Advanced Scientific Computing Research Office. NERSC's mission is to accelerate scientific discovery at the U.S. Department of Energy Office of Science through HPC and extreme data analysis.

Director’s Note

The year 2022 was a busy and successful one at NERSC. The center grew its flagship supercomputer, its user base, its project load, and its reach through academic publications. System upgrades and organizational initiatives improved the center’s capacity to provide world-class service to users and connect with other facilities for seamless team science. And thoughtful action continued to enhance the strength and cohesion of NERSC staff and our user community.

Here are some highlights from the year:

Expanding *Perlmutter*

In 2022, NERSC deployed Phase 2 of its flagship supercomputer, *Perlmutter*. Phase 2 added 12 CPU cabinets housing more than 3,000 nodes, each composed of two CPUs. It also upgraded the interconnect network to HPE’s Slingshot 11, which is much faster than Slingshot 10 and will allow NERSC users to run large problems at scale and analyze very large data sets. The deployment of Phase 2 brought *Perlmutter* up to full size and capability for science.

As *Perlmutter* Phase 2 came into production, the number of researchers using the system for science expanded with it, yielding a number of exciting early results.

Additional System Upgrades

In addition to growing *Perlmutter*, NERSC pursued other upgrades in 2022 to provide better service to current users and prepare for future workloads.

NERSC’s network connection to ESnet, Berkeley Lab’s dedicated network for science, was greatly increased to 1 TB/s in aggregate. This change significantly increased users’ ability to stream large data sets from experimental and observational facilities and will help NERSC support DOE Office of Science’s Integrated Research Infrastructure.

Additionally, Spin continues to be a key tool for developing science gateways, APIs, and databases, and was used by more than 300 NERSC users working on more than 90 projects. In 2022, staff replaced the original cluster orchestrator and installed new large-memory nodes for an enhanced management and security experience.

Working Toward an Integrated Research Infrastructure

Increasingly, some areas of science research require seamless connections between observational or experimental facilities and HPC facilities, allowing for near-real-time data analysis. Since 2019, NERSC has led the Berkeley Lab Superfacility

Project, formalizing the work of supporting this model and enabling better communication across teams working on related topics. The final Superfacility Project Report was published in June 2022.

Later in 2022, members of the NERSC superfacility team also participated in the DOE Advanced Scientific Computing Research (ASCR) program's Integrated Research Infrastructure Architecture Blueprint Activity (IRI ABA), a project aiming to lay the groundwork for a coordinated, integrative research ecosystem. The IRI ABA process broadened the superfacility conversation at the DOE level and harnessed the experience of the DOE Office of Science (SC) community.

Building the Quantum Future

Quantum information science (QIS) and quantum computing (QC) are expected to one day provide a powerful approach to solving complex computational problems. With the NERSC user base in mind, NERSC has been developing staff expertise and engaging scientists, QIS researchers, and quantum computing companies.

In 2022, the QIS@Perlmutter program provided GPU hours on *Perlmutter* to more than 16 groups working in quantum simulation, error correction, chemistry, materials science, condensed matter physics, and optimization. Additionally, NERSC hosted the first annual Quantum for Science Day virtual conference, collaborated on quantum algorithms with the Advanced Quantum Testbed at Berkeley Lab, and worked toward strategic collaborations with industrial partners in the quantum space.

Artificial Intelligence and Machine Learning

HPC centers are preparing for a shift toward new artificial intelligence (AI)-enhanced workflows, and NERSC continues to co-lead the MLCommons HPC working group, taking steps to extend the industry-standard MLPerf HPC benchmarks for the HPC and scientific AI workloads.

Additionally, in 2022 planning continued for the NERSC-10 system to follow *Perlmutter*. It's clear that future systems will be required to leverage new technologies and support emerging needs in AI and experimental/observational science to accelerate workflows and enable new modes of scientific discovery through the integration of experiment, data analysis, and
simulation. Preparation for NERSC-10 continues with these requirements in mind.

Investing in the NERSC Community

NERSC's most important resource is its people, and 2022 saw the continuation of efforts to invest in the NERSC community and make the center an even better place to work. As part of this initiative, the User Engagement Group began developing a Community of Practice, defined as a group of people who "share a concern or a passion for something they do and learn how to do it better as they interact regularly." The NERSC Community of Practice is intended to expand user skills and effectiveness and foster cross-disciplinary discussion and collaboration.

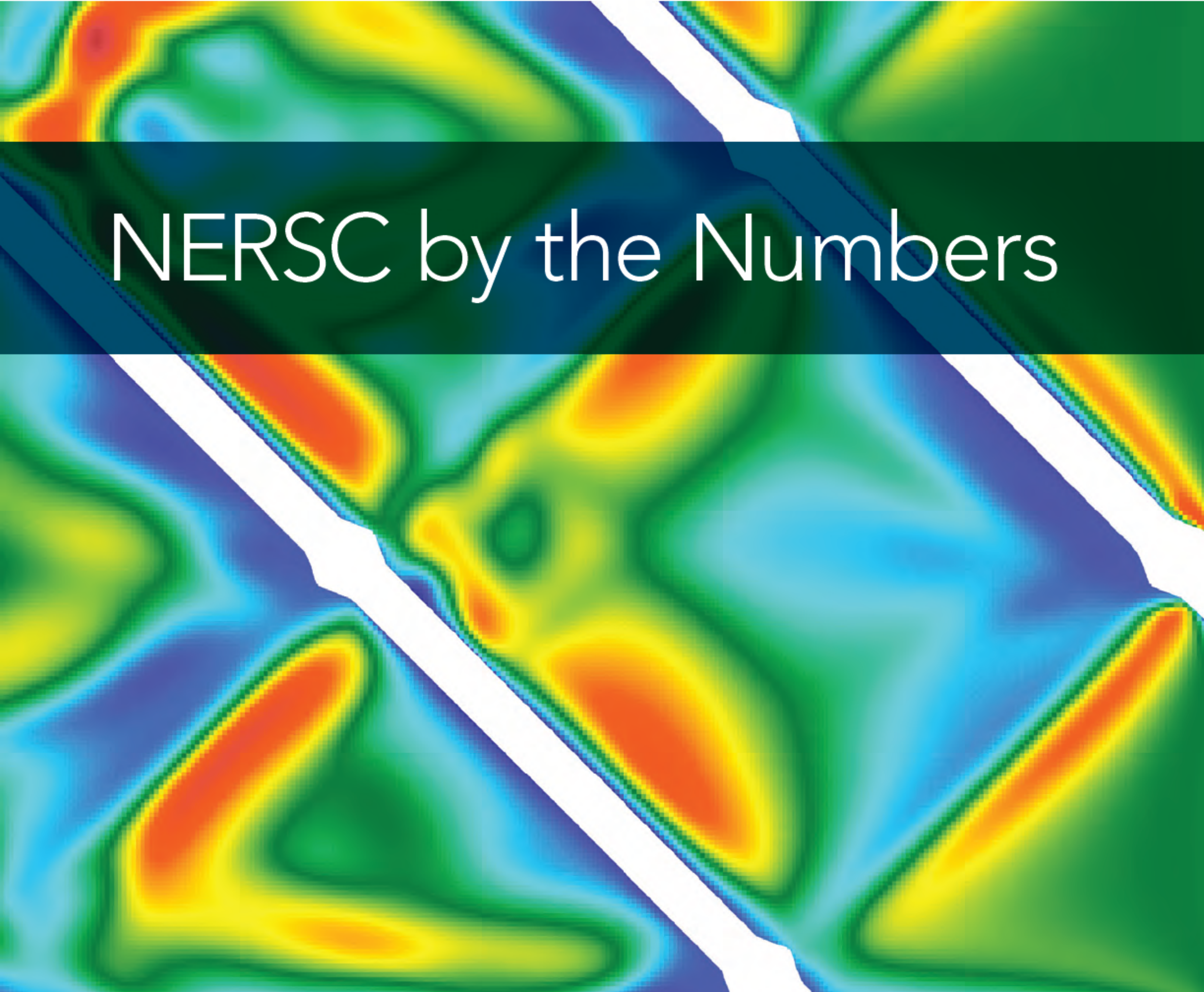
To support the Community of Practice, NERSC adopted a User Code of Conduct – one of the first at any HPC center. Based on

Berkeley Lab's five stewardship values (Team Science, Service, Trust, Innovation, and Respect), the User Code of Conduct is intended to inspire positive behavior.

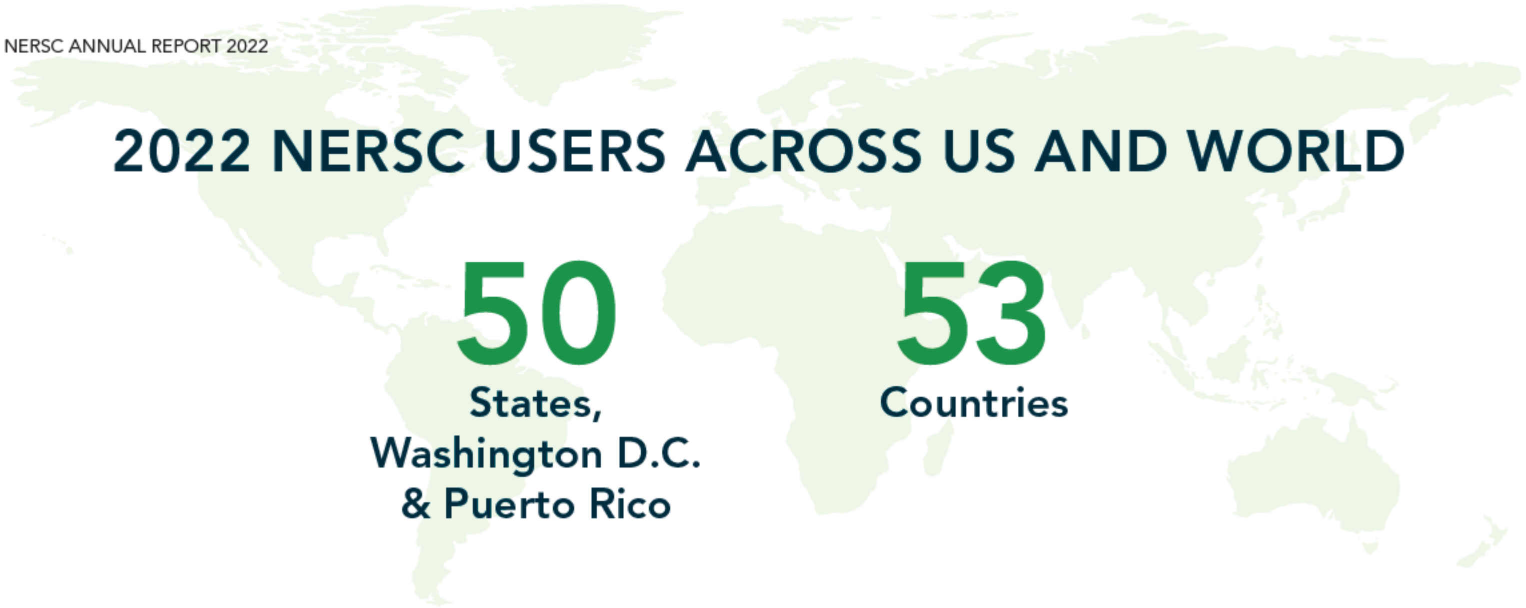
These accomplishments continue to improve NERSC's capacity to fulfill its mission of accelerating scientific discovery through HPC – both exciting destinations and valuable waystations during that pursuit. As we look to the future and prepare for new possibilities, we can be proud of the work we did this year.

— **Sudip Dosanjh**
NERSC Division Director

NERSC by the Numbers



2022 NERSC USERS ACROSS US AND WORLD



50
States,
Washington D.C.
& Puerto Rico

53
Countries

~10,000 Annual Users from **~800** Institutions + National Labs

32% Graduate Students

19% Postdoctoral Fellows

15% Staff Scientists

13% University Faculty

8% Undergraduate Students

5% Professional Staff

60% Universities

29% DOE Labs

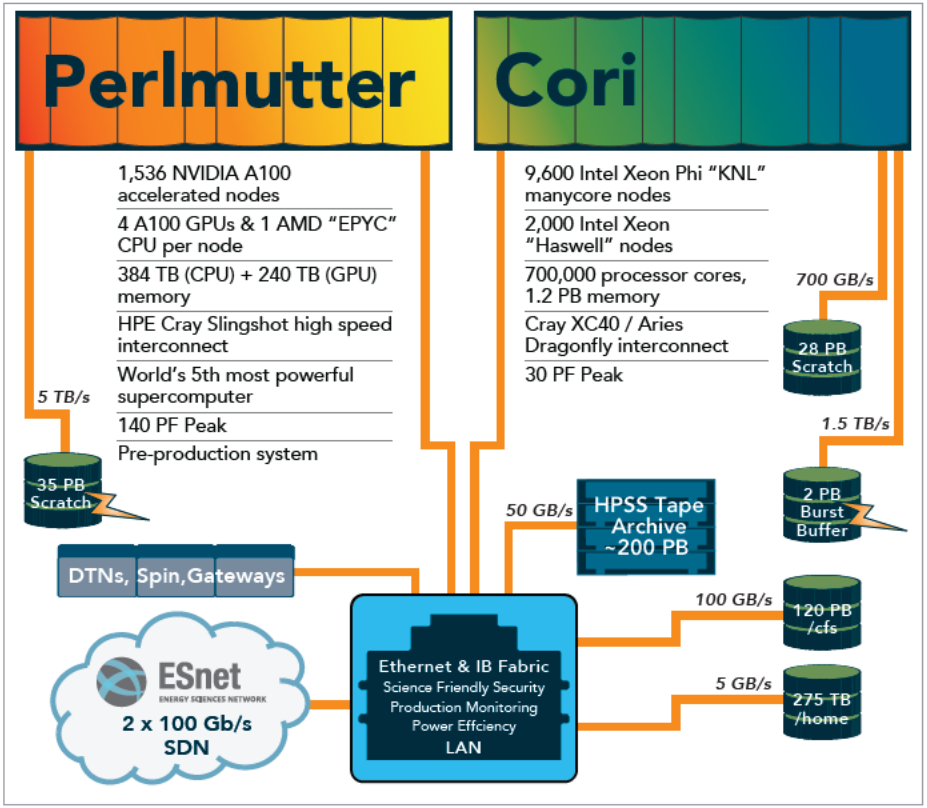
5% Other Government Labs

4% Industry

1% Small Businesses

<1% Private Labs

NERSC Systems 2022



Perlmutter

- 1,536 NVIDIA A100 accelerated nodes
- 4 A100 GPUs & 1 AMD "EPYC" CPU per node
- 384 TB (CPU) + 240 TB (GPU) memory
- HPE Cray Slingshot high speed interconnect
- World's 5th most powerful supercomputer
- 140 PF Peak
- Pre-production system

Cori

- 9,600 Intel Xeon Phi "KNL" manycore nodes
- 2,000 Intel Xeon "Haswell" nodes
- 700,000 processor cores, 1.2 PB memory
- Cray XC40 / Aries Dragonfly interconnect
- 30 PF Peak

TOP 5 DOE Labs by Users

Oak Ridge National Laboratory
Los Alamos National Laboratory
Lawrence Livermore National Laboratory
Pacific Northwest National Laboratory
Argonne National Laboratory

TOP 5 Universities by Users

Berkeley University of California
Stanford University
University of Michigan
Massachusetts Institute of Technology
Carnegie Mellon University
University of Washington

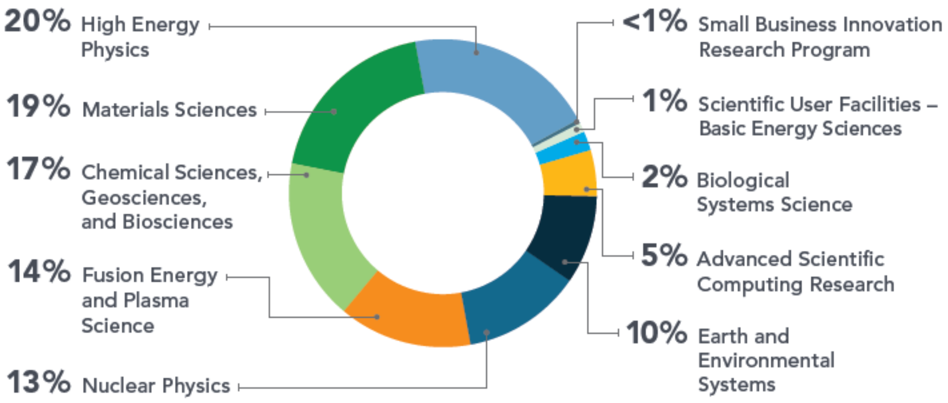
Top Science Disciplines

(By computational hours used)

Chemical Sciences
Nuclear Physics
Fusion Energy
Scientific User Facilities
High Energy Physics
Biosciences
Geosciences
Small Business Innovation Research
Materials Sciences
Plasma Science
Climate and Environmental Science
Biological Systems Science

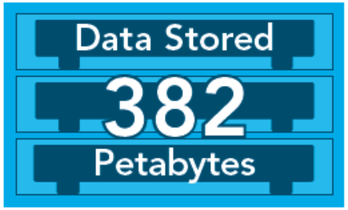
~1,000
Projects

Breakdown of Compute Used by DOE Program



110
MILLION
COMPUTE HOURS USED IN 2022

>2,000
Refereed Publications cited NERSC



Data Stored
382
Petabytes

2022 Milestones

Expanding and Upgrading *Perlmutter*

Figure 1. *Perlmutter*, the flagship supercomputer at NERSC, was fully deployed in 2022.

Perlmutter, the flagship supercomputer at NERSC, is an HPE Cray EX supercomputer that features both GPU-accelerated and CPU-only nodes and has a performance of approximately three to four times that of Cori, NERSC's previous system. Debuting in the Top500 ranking in June 2021 as the fifth-fastest supercomputer in the world, the system is designed for the future of HPC, including using machine learning for scientific analysis and discovery. Serving more than 9,000 researchers around the world, *Perlmutter* is already expanding what's possible for simulation, learning, and near-real-time data analysis, with a degree of energy efficiency that's among the best in the world.

In 2022, NERSC grew *Perlmutter* to its full size and upgraded all components to their final configuration. NERSC used a phased approach during the installation of *Perlmutter*, starting in 2021 with a set of 12 GPU-accelerated compute cabinets and integrating additional technology as it became available. During this extended installation period, NERSC took extraordinary steps to maximize the time the system was available to users for their research. In its final configuration, *Perlmutter* has 14 cabinets of GPU-accelerated compute nodes and 12 cabinets of CPU-only nodes. This translates to 1,792 GPU nodes with a total of 7,168 NVIDIA A100 GPUs and 3,072 CPU-only nodes with 6,144 AMD EPYC "Milan" processors. The custom high-speed network is known as Slingshot 11 from HPE Cray. *Perlmutter* also has a 35 PB all-flash file system connected to the high-speed network.

In addition to the expanded installation of *Perlmutter*, NERSC staff made important upgrades to improve performance and provide an even better experience for its users:

- Upgrading to the latest network technology (Slingshot 11), which quadrupled the bandwidth to the GPU nodes.
- Developing new tools to assess network health and diagnose problems as early as possible.
- Adapting a standard existing container runtime (Podman) for HPC. This change required addressing issues around security, scaling, and performance, and allows users to build and execute containers using a standards-based tool and run the resulting image at scale on *Perlmutter*.

Early Science Results Show *Perlmutter's* Promise

As *Perlmutter* Phase 2 came into production, the number of researchers using the system for science expanded with it, yielding a number of exciting early results.

Figure 2. Berkeley Lab astrophysicist and Nobel laureate Saul Perlmutter, right, poses with his family in front of his namesake machine during its installation in 2022.

One way in which NERSC supports research is through the NERSC Exascale Science Applications Program (NESAP), a collaborative effort in which NERSC partners with code teams, vendors, and library and tools developers to prepare for advanced architectures and new systems. NESAP began in 2014 to help users prepare for the Cori manycore Knights Landing/Xeon Phi architecture and has continued into the 2020s, now targeting *Perlmutter*.

Figure 3. DOE Office of Science director Asmeret Asefaw Berhe adds her signature to the mural decorating *Perlmutter* during a visit to NERSC.

NESAP collaborations were substantial in 2022, including a Gordon Bell Prize winner and a Gordon Bell Prize finalist. The Gordon Bell Prize for outstanding achievement in HPC is awarded annually at the International Conference for High Performance Computing, Networking, Storage, and Analysis (Supercomputing).

Some exciting early results from *Perlmutter's* deployment include:

Reimagining Weather Prediction with Deep Learning through FourCastNet

Accurate weather forecasts save lives by providing advance notice of extreme events, and the quality of forecasts has improved dramatically in the past century due to better atmospheric observation, models, and computing. Now, data-driven deep learning models have the potential to spur even more improvements because of their speed, low computational cost, and low energy footprint.

FourCastNet (short for Fourier ForeCasting Neural Network) is a global data-driven weather forecasting model developed in collaboration with NVIDIA to scale on *Perlmutter* and other supercomputers. The model provides accurate global predictions for the short and medium term at 0.25° resolution, considering rapidly changing variables such as surface wind speed, precipitation, and atmospheric water vapor. It can predict a weeklong forecast in less than two seconds—an order of magnitude faster than the existing numerical Integrated Forecast System (IFS). The speed, accuracy, and computational inexpensiveness of FourCastNet has important implications for resource planning and predicting and responding to extreme weather events.

HPC engineers Shashank Subramanian and Peter Harrington and NESAP for Learning postdoc Jaideep Pathak ran FourCastNet on *Perlmutter* in 2022 and demonstrated the vast performance improvement made possible by modeling on a

GPU-based system, including running a 24-hour forecast on a single GPU node in seven node-seconds and using 8 kJ of energy, compared with 984,000 node-seconds and 271 MJ for the IFS model. In a world with increasingly extreme weather, improved probabilistic forecasting will be an important tool, and FourCastNet represents the speed and efficiency that's possible with a GPU-based system.

Figure 4. NERSC staffers show students what's inside a part of the *Perlmutter* supercomputer.

Perlmutter Supports the First All-GPU Full-Scale Physics Simulation

Using *Perlmutter*, in 2022 researchers completed a simulation of a detector of neutrino interactions that's designed to run exclusively on GPUs—the first simulation of its kind and an example of using GPUs' highly parallel structure to process large amounts of physical data.

Neutrinos are the most abundant particle of matter in the universe. Efforts to measure its mass and understand its relationship to matter have been underway for decades, and this simulation could help clarify why the universe is made of matter and not antimatter (that is, particles with the same mass as matter, but opposite electrical charges and properties).

The simulation is part of the preparation for the Deep Underground Neutrino Experiment (DUNE), an international collaboration studying the neutrino that includes U.S. DOE resources. Currently under construction, the components of the DUNE experiment will consist of an intense neutrino beam produced at Fermilab in Illinois and two main detectors: a Near Detector located near the beam source and a Far Detector located at Sanford Underground Research Laboratory in South Dakota. Eventually the Near Detector will detect approximately 50 neutrino interactions per beam blast, adding up to tens of millions of interactions per year. By examining neutrinos and how they change in form over long distances, a process called oscillation, scientists hope to learn about the origin and behavior of the universe over time.

Researchers have been studying neutrinos since the 1970s, but a new method being pioneered at Berkeley Lab yields 3D images instead of 2D — an increase in information and a vast amount more data to analyze and store. Before construction of the detectors begins, the team uses digital simulations to ensure that the detectors and the workflows around them will work as planned.

That's where supercomputing resources at NERSC come in. Because GPUs are uniquely suited to executing many calculations in parallel, they represent a much faster way of dealing with large quantities of data. *Perlmutter's* thousands of GPU nodes allowed the researchers to simulate the detector over many nodes at once, greatly increasing the compute capability relative to CPU-only and speeding up the process. The simulation of the signal from each sensor took about one millisecond on the GPU compared with ten seconds on a CPU, completing the process in a fraction of the time. Overall, this approach is a major step forward for the DUNE experiment as well as for the field of high-energy physics and other scientific areas of study.

Additional Facility Upgrades to *Perlmutter*

In addition to expanding *Perlmutter*, NERSC pursued other upgrades in 2022 to provide better service to current users and prepare for future workloads. Here are some of the key changes:

- Internal- and external-facing portions of NERSC's network were upgraded to 400Gb connections, becoming one of the first ESnet customers to upgrade to 400Gb and providing 1Tb of external networking connectivity. This change significantly increased NERSC's external bandwidth to meet future network traffic growth, particularly in the area of integrated research infrastructure.
- Additionally, Spin, a platform used by more than 300 NERSC users on more than 90 projects, is a key tool for developing science gateways, APIs, databases, and more. In 2022, staff replaced the original cluster orchestrator with Kubernetes and installed new large-memory nodes, offering an enhanced management and security experience.

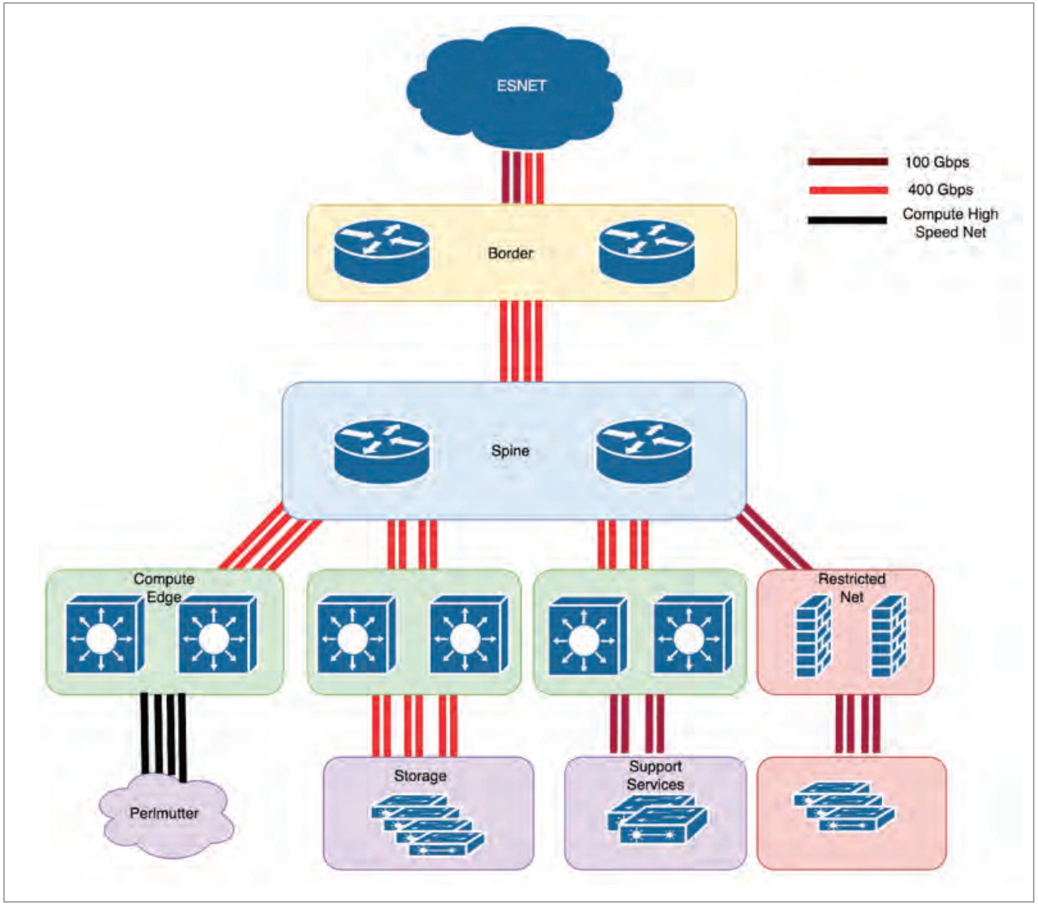


Figure 5. The 400-Gbps buildout at NERSC significantly increased NERSC's external bandwidth to meet future network traffic growth.

Cori: Continuing Contributions

With its first cabinets installed in 2015 and the system fully deployed by 2016, the Cori supercomputer was in service longer than any system in NERSC's 49-year history and enabled more than 10,000 scientific publications. Its technological innovations reflect the dynamic evolution of high performance computing (HPC) over the past decade, paving the way for the next generation of scientific computing.

With retirement planned for 2023, Cori proved reliable in its last full year of use, achieving 98.8% scheduled availability and 96.9% overall availability in 2022, and in fact experiencing fewer outages than the previous year.

In 2022, Cori continued to provide valuable compute services at the cutting edge of HPC, including additional computing for Earth systems modeling, catalyst design for environmentally friendly energy production, and many others across the fields of climate science, health research, materials science, high-energy physics, and more.

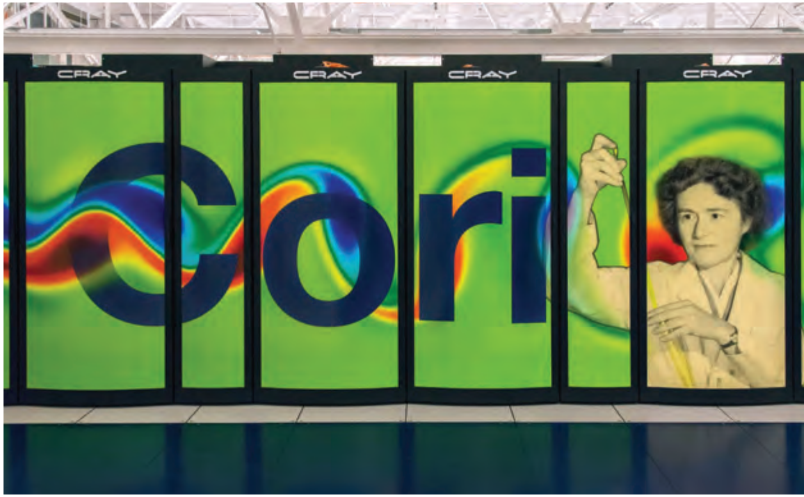


Figure 6. The Cori supercomputer was installed in 2015 and was the longest-serving supercomputer in NERSC history.

> *"I look at Cori as another step in the evolution of HPC,"* said former NERSC Deputy for Operations Jeff Broughton of the system. *"Basically, it shows that NERSC continues to be on the leading edge of deploying new and novel systems that will help our users maintain their advantage in scientific computing."*

NERSC-10: The Next Generation

Even during *Perlmutter* Phase 2 deployment, work on the next system, currently known as NERSC-10, was already underway. Due in 2026, the NERSC-10 system will accelerate end-to-end DOE Office of Science workflows and enable new modes of scientific discovery through the integration of experiment, data analysis, and simulation.

Though early in the process, the NERSC-10 project progressed through some key planning phases in 2022:

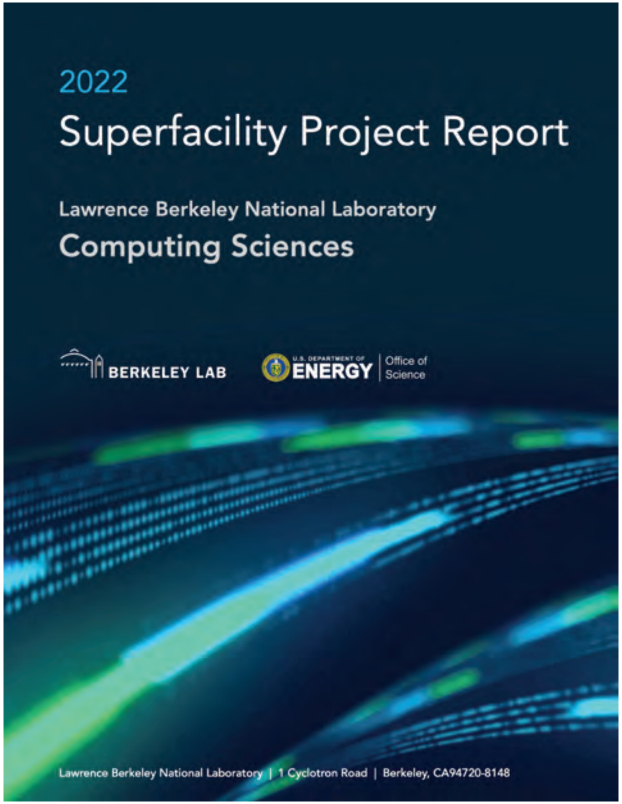
- After receiving Critical Decision 0 (Mission Need) approval in October 2021 and being formally established as a project, NERSC-10 facility design work began and continued through 2022.
- In preparation for the release of a Request for Proposals (RFP) for NERSC-10, the project team engaged in a comprehensive vendor market survey to better understand the hardware and software technology landscape. The team also drafted technical requirements and developed benchmarks required for the RFP; an Advanced Acquisition Plan was submitted for approval in June 2022.

Working Toward an Integrated Research Infrastructure

An emerging scientific use case involves the seamless connections between experimental facilities and HPC resources, offering researchers large-scale analysis capability in near-real time while they are collecting data. This model is known as superfacility — the idea of a single unified facility “overlaid” on individual facilities working in concert.

The superfacility model connects experimental facilities, including DOE user facilities, to HPC facilities via ESnet (the high-speed network for science based at Berkeley Lab and connecting DOE and international facilities). However, it requires more than physical connections to be successful. Superfacility calls for an ecosystem of tools, technologies, interfaces, and expertise to enable scientists to take advantage of these new modes of discovery. By establishing the Superfacility Project in 2019, Berkeley Lab formalized the work of supporting this model and enabling better communication across teams working on related topics. The final Superfacility Project Report was published in June 2022.

The Superfacility Project strengthened some of the processes and infrastructure that had already been implemented around integrating these workflows. To support superfacility efforts across all Office of Science research programs, a dedicated NERSC liaison works closely with each science partner to track their requirements and progress, with a particular focus on supporting their use of the tools and technologies developed



2022

Superfacility Project Report

Lawrence Berkeley National Laboratory

Computing Sciences

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY | Office of Science

Lawrence Berkeley National Laboratory | 1 Cyclotron Road | Berkeley, CA94720-8148

Figure 7. The Superfacility Report, published in June 2022, details the ways in which Berkeley Lab is supporting the superfacility model and enabling better communication across teams working on related topics.

by the superfacility team. The Superfacility Project conducted several rounds of requirements gathering throughout the experimental process and received continuous feedback from science teams as their projects evolved, building strong working relationships with the scientists and adapting NERSC projects to quickly meet their needs.

In fall 2022, members of the NERSC superfacility team also participated in the DOE Advanced Scientific Computing Research (ASCR) program's Integrated Research Infrastructure Architecture Blueprint Activity (IRI ABA). The aim of the IRI ABA was to establish the foundations of a coordinated strategy for an integrative research ecosystem.

By inviting DOE experts to discuss use cases that span SC programs and user facilities, the team was able to harness the experience of the SC community. The group developed overarching design principles and a blueprint to efficiently address the chief IRI design patterns and identified urgent priorities and "early win" opportunities. Lessons learned from the Superfacility Project informed the IRI ABA conversation, including topics like: real-time access to compute, storage, and networking; Jupyter as an emerging interface to high-performance data analysis; managing complex cross-facility workflows using standards-based REST APIs; and the attendant need for an SC-wide federated identity solution. The final IRI ABA report will be published in 2023.

Superfacility in Action: ExaFEL

One example of superfacility in use is the ExaFEL project, a NESAP project in which data is transferred from the Linac Coherent Light Source (LCLS) at the Stanford Linear Accelerator (SLAC) to NERSC via ESnet as it is being collected.

Serial femtosecond X-ray crystallography (SFX) uses X-ray free electron lasers (XFEL) to investigate material samples at the atomic level. This technique is responsible for much of our understanding of biochemistry and materials science, shining a light on unknown molecular structures and their chemical functions.

However, SFX is entering a new era, with detectors collecting data at exponentially increasing rates—a 20-minute data collection can produce up to 10 TB of X-ray scattering data. These massive datasets bring new opportunities for insight, but the only viable way to perform the necessary data analysis during an experiment is to develop and implement new algorithms for GPUs and deploy them on exascale HPC systems.

XFEL time is precious, and due to the exploratory nature of XFEL experiments, equipment operators need to operate on the fly. For scientists to make good use of their limited measurement time, SFX data analysis must happen in tandem with measurements. NERSC has been a leader at providing real-time and urgent HPC resources for fast measurement feedback, with real-time data transfer via ESnet. NERSC's interfaces and tools (such as the Superfacility API) coordinate measurement at SLAC and data analysis at NERSC. Since this workflow's first use in 2020, efforts at NERSC have focused on production hardening and expanding the available features, and in 2022 GPU-enabled nodes were incorporated for real-time feedback for the first time. As a result, NERSC is used in many of LCLS's most data-intensive measurement campaigns (on average four weeks per year).

Figure 8. Superfacility principles come into play as researchers use Stanford's Linac Coherent Light Source to pioneer a new form of X-ray crystallography. Experimental data is transferred automatically via ESnet to supercomputers at NERSC and back, yielding initial analysis in under 10 minutes—a speed record for this type of experiment. (*Image credit: Ella Maru Studio and J. Nathan Hohman*)

Building the Quantum Future

Quantum information science (QIS) and quantum computing (QC) are expected to one day provide a powerful approach to solving complex computational problems. While current quantum computers are not of sufficient quality to have an advantage over classical supercomputers in terms of time-to-solution, cost, or energy usage for useful applications that align with the NERSC workload, this balance is expected to shift over the next decade. Applications in chemistry, materials science, basic energy science, and high-energy physics are especially likely to benefit from QIS, and QC technology may open new research directions that would be impossible to pursue with classical supercomputers.

With the NERSC user base in mind, NERSC has been developing staff expertise and engaging scientists, QIS researchers, and quantum computing companies in the following ways:

- NERSC's first Quantum@NERSC virtual conference, Quantum for Science Day, took place in October 2022 and was attended by more than 100 participants. Presentations at the conference highlighted multiple QIS@Perlmutter works.
- NERSC researchers have engaged in algorithm development and codesign in collaboration with the Advanced Quantum Testbed at Berkeley Lab, which is focused on improving and developing gates and superconducting qubit technology. NERSC researchers have been part of two grants in these directions: one to develop quantum Krylov algorithms (in collaboration with the Applied Mathematics and Computational Research division) for quantum hardware and one (in collaboration with the AQT) focused on hybrid algorithm-hardware codesign.
- NERSC staff are also developing strategic collaborations with industrial partners in the quantum space, targeting both quantum hardware and software. Some of these vendors were also participants in the Quantum@NERSC conference, offering tutorials on using their latest software and algorithm developments on *Perlmutter*.
- NERSC has provided support for quantum software on *Perlmutter*, such as NVIDIA's cuQuantumSDK for GPU-accelerated quantum circuit simulations.

Quantum for Science Day 2022

On October 24, stakeholders from around the field of QIS gathered virtually to discuss the state of the discipline at the first-ever Quantum for Science Day, hosted by the Quantum@NERSC team. Quantum information science is a growing area of research and innovation at the Lab; in 2022, NERSC granted 250,000 GPU node hours to 16 QIS projects through the initiative. At Quantum for Science Day, more than 100 unique participants from across academia, research labs, and industry logged on to learn and share information.

Overall, said the organizers, the gathering was a success, and this first event certainly won't be the last: "We were very excited about the number of people who participated and the high quality of the panel and talks," said event co-organizer Katie Klymko. "We're planning to make this an annual event and hope to see many exciting new developments from the QIS-in-HPC community next year as we will continue to support QIS researchers at NERSC in 2023."

Figure 9. The QIS@Perlmutter team poses with the *Perlmutter* supercomputer.

Artificial Intelligence and Machine Learning

With growing interest and exploration in AI, machine learning, and deep learning applications, HPC centers are preparing for a shift toward new AI-enhanced workflows. The HPC community must be ready to support this workload with benchmarks that capture its computational characteristics and provide a standard basis for performance comparisons across different computational methods and HPC systems.

NERSC continues to co-lead the MLCommons HPC working group, working to extend the industry-standard MLPerf HPC benchmarks for the HPC and scientific AI workloads. The third round of benchmark development (v. 2.0), the results of which were published in November 2022, saw impressive performance improvements, with the best time-to-train results improving twofold and the best model throughput results improving by threefold. The MLPerf HPC benchmarks were used internally for extensive testing of the *Perlmutter* system in 2022, and NERSC plans to submit results with the final *Perlmutter* configuration for MLPerf HPC v3.0 in 2023.

As planning continues for the NERSC-10 system that will follow *Perlmutter*, it's clear that future NERSC supercomputers must go beyond the demand for more computational cycles from traditional simulation and modeling workloads. Instead, future systems will be required to leverage new technologies and support emerging needs in AI and experimental/observational science to accelerate workflows and enable new modes of scientific discovery through the integration of experiment, data analysis, and simulation.

The NERSC-10 system must also address the growing need for more dynamic and programmable systems to accommodate increasingly complex workflows that may require running many interdependent simulations and/or analysis tasks, moving vast amounts of complex data among storage hierarchies within NERSC and externally, both for providing interactive real-time feedback to experiments and for integrating simulations with AI.

Investing in the NERSC Community

One of NERSC's greatest resources is its people, and making NERSC an excellent place to work is an ongoing effort. In 2022, NERSC staff and management collaborated on several initiatives to make NERSC a more open and accessible place for users and staff, embracing new modes of work, evolving the ways in which the center communicates, and building the scaffolding for a stronger and more integrated NERSC community.

Building a Community of Practice

In 2022, NERSC's User Engagement Group began developing a NERSC user community of practice to develop user skills and effectiveness and foster cross-disciplinary discussion and collaboration.

A Community of Practice (CoP) is defined as a group of people who "share a concern or a passion for something they do and learn how to do it better as they interact regularly."

What makes a Community of Practice?

1. The identity of a CoP is defined by a shared domain of interest. A neighborhood could be considered a community, but it lacks a shared domain of interest. In the case of NERSC, the shared domain of interest is "using NERSC resources."
2. There is a sense of community that is actively cultivated and maintained. The community members know each other. They actively participate in community initiatives and activities and build relationships across the community. They learn and grow skills while interacting with one another.
3. The members of the community are active practitioners of the shared domain of interest. They are not just people who like supercomputers; they are people who use supercomputers and associated resources, and want to become even more effective at using them.

Implementing a User Code of Conduct

In October 2022 NERSC implemented a User Code of Conduct—a behavioral framework for users to assure a safe, welcoming environment for all.

Team science is a core value at Berkeley Lab, and as technological changes enable increasing collaboration in science research, NERSC staff anticipates an increase in interaction between users and between users and staff. With that in mind, the NERSC User Code of Conduct lays the groundwork for productive collaboration and interaction.

The NERSC User Code of Conduct came out of a careful and collaborative process with the NERSC User Engagement Group and departments across Berkeley Lab. Although all users are required to sign the User Code of Conduct before using NERSC resources, and users may report misconduct anonymously or personally through multiple channels, the User Code of Conduct is based on Berkeley Lab's five stewardship values (Team Science, Service, Trust, Innovation, and Respect) and is intended to inspire positive behavior through appeals to these values shared by the community rather than taking a punitive approach.

The User Code of Conduct promotes the values of collaboration, kindness, and respect within the NERSC user community and has thus far been met with positive feedback. It's believed that NERSC is the first HPC center to implement a user code of conduct.

NERSC Week

When Berkeley Lab's pandemic-related travel moratorium was lifted at the beginning of 2022, NERSC began thinking about in-person events. While the center functioned well in its pandemic-induced remote-work mode, leadership recognized the importance of face-to-face interaction and planned a week of events called "NERSC Week" to encourage staff to gather and make connections in person.

The first NERSC Week took place May 23–27, 2022 and featured trainings, strategic planning sessions, an ice cream social, and an optional Saturday-morning cherry-picking event at a nearby cherry orchard.

More than 80 NERSC staff participated in the first NERSC Week. Additional NERSC Week events took place during 2022, and it has evolved into a quarterly event for 2023.

Welcoming Summer Students: Supporting the Future of Science

The Computing Sciences Area's Summer Program offers high school, undergraduate, and graduate students the opportunity to work with Berkeley Lab researchers on focused projects each summer. Here are a few of the students hosted at NERSC in 2022.

James Duncan presents his poster at the 2022 Summer Student poster session *(Credit: Margie Wylie, Berkeley Lab)*

James Duncan Develops Models for a Rainy Day

Student: James Duncan

Ph.D. student in Biostatistics, UC Berkeley

Project: Developing deep generative machine-learning models for forecasting precipitation

In his own words:

"Machine-learning models are quite a blurry picture of precipitation compared to the ground truth. So one of the goals of this project was to try and better capture those fine-scale structures — but then also, hopefully, to predict some of the extreme states of the atmosphere like tropical cyclones and atmospheric rivers, because that's really important for disaster preparedness and safety."

NERSC Summer Student Muna Tageldin did her research on NERSC systems and ways to improve their performance. *(Credit: Margie Wylie, Berkeley Lab)*

Muna Tageldin Puts MPI Under the Microscope

Student: Muna Tageldin

Ph.D. student in Electrical Engineering at Marquette University

Project: Developing a microbenchmark to analyze variances in message-passing interface (MPI) performance on NERSC systems and looking for the best statistical methods to characterize the results.

In her own words:

"We find that collective performance measurements form a multimodal distribution on a system running a production workload. Frequently used summary statistics like minimum, maximum, median, and mean don't always capture the intrinsic characteristics of MPI performance variations that can have complex forms. My project is understanding MPI performance variation and also finding statistical tests that can correctly describe this variation. My dissertation is on analyzing HPC systems performance using probabilistic models, and in this internship, I'm tackling HPC performance from a statistics and coding perspective. It's interesting because MPI is an area I haven't delved into in detail."

Quantum Computing Captivates Dhilan Nag

Student: Dhilan Nag

Plano East High School

Project: Developing hybrid classical-quantum algorithms for calculating energies in models for strongly correlated materials

In his own words:

"I was essentially trying to improve the convergence of a particular quantum method called the variational quantum eigensolver, which is a hybrid (quantum-classical) algorithm to calculate energies. The bigger impact of this research is that quantum computing is the next step into chemical simulations. If we can accurately simulate behaviors of molecules, we can more efficiently discover new chemical systems. And if we do that, we can reduce the cost of creating new drugs and new materials."

On spending time in Berkeley:

"My time in Berkeley felt very immersive, and I got to experience the scientific world as a high schooler. I could walk down to the Pitzer Center and Café Strada, then go uphill to the Lab. I could be in the community like a local but also be immersed in the research like the students. It is a beautiful campus with a lot of great resources."

LeAnn Lindsey Enhances Genomic Sequencing Alignment on GPU Architectures

Student: LeAnn Lindsey

Ph.D. candidate in Computer Science, University of Utah

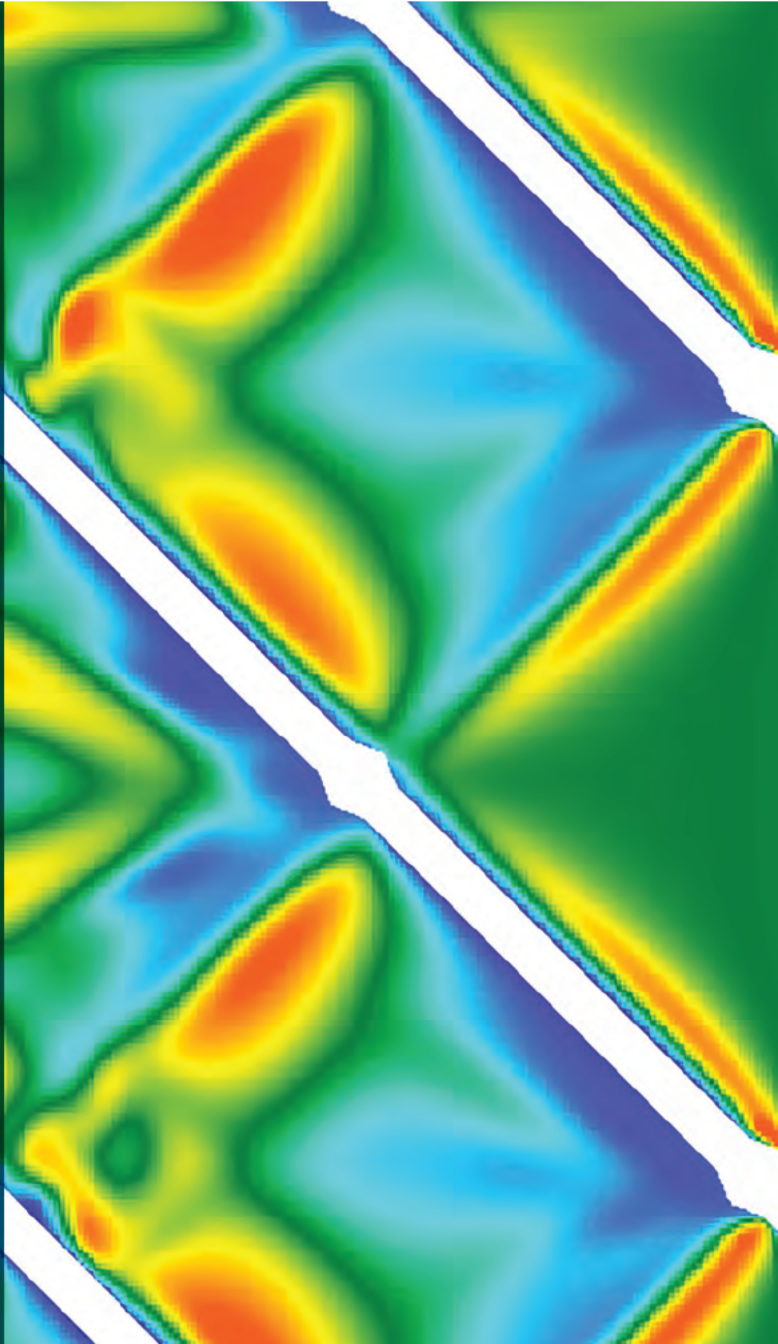
Project: Developing and applying a GPU-accelerated "traceback" computer code that addressed challenging problems in a biological sequencing alignment library, used to investigate biological sequences like DNA or RNA to understand their structure function

In her own words:

"Our goal was to make the software run faster on the Perlmutter GPU architecture because traditional assemblers on small-scale systems can take as long as two weeks to finish. There is so much data out there, if you want to study really large datasets like a cancer dataset or an environmental dataset that is terabytes large, you are going to need something that runs very, very fast."

Science Highlights

NERSC's strategic planning and engagement activities have helped enable yet another highly productive year of science for thousands of users. In this section, we document the methods and practices we use to monitor our contributions to the DOE mission via science highlights and accomplishments; describe the quality of our engagements with strategic stakeholders, including other DOE programs and partner institutions; and outline how we build understanding of the HPC landscape to enable facility and DOE mission success.



NERSC Supercomputing Sheds Light on Early Supernova Discovery

THE SCIENCE. With continuous 24-hour coverage of the night sky, astronomers were able to detect a Type Ia supernova just one hour after its explosion in 2018 — some of the earliest such observations ever made. Such a rare discovery is an opportunity to test competing theories of how Type Ia supernovae explode and what kinds of stellar systems produce them. Answering these questions is relevant to both nuclear and high-energy physics because these thermonuclear supernovae are the source of many heavy elements on the periodic table and also useful cosmological tools for probing Dark Energy. Supercomputer simulations at NERSC have been indispensable to understanding the complex physics of these events.

THE IMPACT. The results inform a deeper understanding of the Type Ia supernova explosion process. The analysis reveals a concentration of iron and other heavy elements in the outermost portion of debris from the explosion, which was revealed by observations of a rapid reddening of the supernova's light in the first few hours of data. This means that thermonuclear explosions are initiated by either nuclear burning on the surface of an exploding white dwarf star or an extreme mixing of the exploding stellar material — an important milestone in understanding how supernovae explode.

ADDITIONAL DETAILS. The observations and their physical interpretation enabled by NERSC's Cori supercomputer were published in Nature Astronomy in February 2022. The researchers used Cori for all aspects of modeling the supernova, performing hydrodynamical simulations of the actual explosion with CASTRO, an adaptive mesh simulation code for astrophysics, and SEDONA, a multi-dimensional radiative transport code, to calculate the emergent spectrum and its evolution over time.

NERSC PI: Peter Nugent

PROJECT FUNDING AND ALLOCATION AWARD: DOE Office of Science, Office of High Energy Physics (HEP); NERSC Director's Discretionary Reserve

PUBLICATION: Ni, Y.Q., Moon, DS., Drout, M.R. et al. "Infant-phase reddening by surface Fe-peak elements in a normal type Ia supernova." Nat Astron 6, 568–576 (2022). https://doi.org/10.1038/s41550-022-01603-4

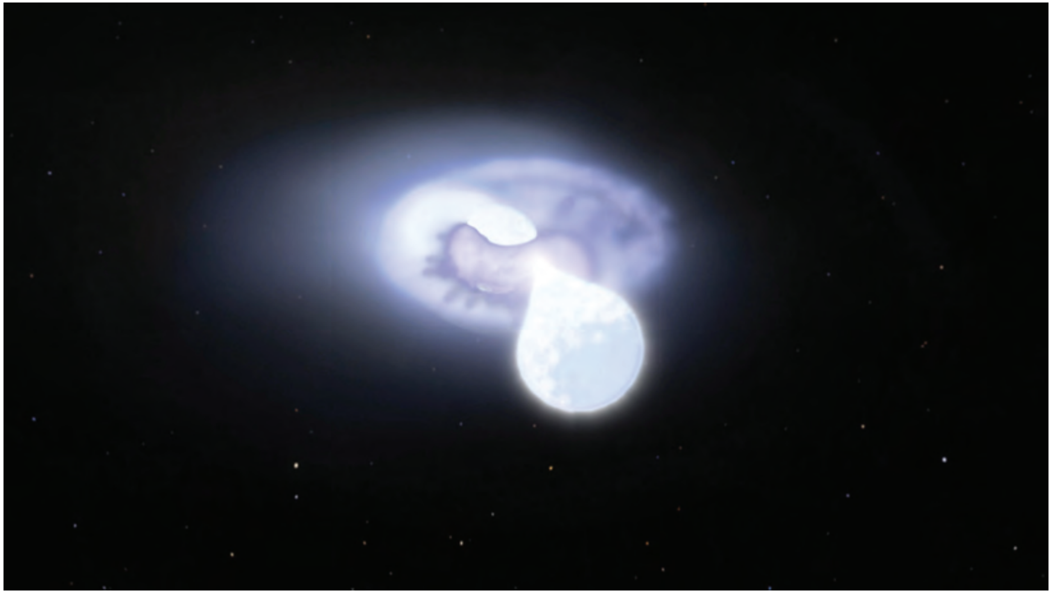


Figure 10. With continuous 24-hour coverage of the night sky, astronomers were able to detect a Type Ia supernova just one hour after its explosion in 2018 — some of the earliest such observations ever made. NERSC supercomputing resources were instrumental in several aspects of the analysis.

New Climate Models Reveal Geographical Link Between Wildfires and Extreme Weather

THE SCIENCE. A team of researchers led by scientists from Pacific Northwest National Laboratory (PNNL) ran a series of models on NERSC's Cori supercomputer to investigate an emerging causal connection between wildfires in the Western U.S. and extreme weather in the Central U.S. The results were published in the October 17, 2022 issue of the Proceedings of the National Academy of Sciences.

THE IMPACT. Typically, western wildfires and storms in the Central U.S. are separated by seasons. As Western U.S. blazes begin earlier each year, however, the two events now occur closer together. The team investigated the influence of Western US wildfires on extreme precipitation and hailstones in the Central U.S. and found that heat and tiny airborne particles produced by these wildfires intensify distant severe storms, in some cases bringing baseball-sized hail, heavier rain, and flash flooding to "downwind" states like Colorado, Wyoming, Nebraska, Kansas, Oklahoma, and the Dakotas.

ADDITIONAL DETAILS. The team used WRF-Chem (Weather Research and Forecasting model coupled with Chemistry) including a spectral-bin microphysics scheme and parameterized sensible heat models for wildfires. Python was used for post-processing and analyzing the simulations. This project, "Understanding Factors Impacting Past and Future Severe Thunderstorms in the Central United States," allocated by the BER office under Climate and Environmental Sciences, used 9,000 Haswell and 1,000 KNL node hours on Cori in 2022 in support of this publication.

NERSC PI: Jiwen Fang

PROJECT FUNDING AND ALLOCATION AWARD: DOE Office of Science, Office of Biological and Environmental Research

PUBLICATIONS: Zhang, Y., et al., "Notable impact of wildfires in the western United States on weather hazards in the central United States," Proceedings of the National Academy of Sciences, 119 (44) e2207329119 (2022). https://doi.org/10.1073/pnas.2207329119

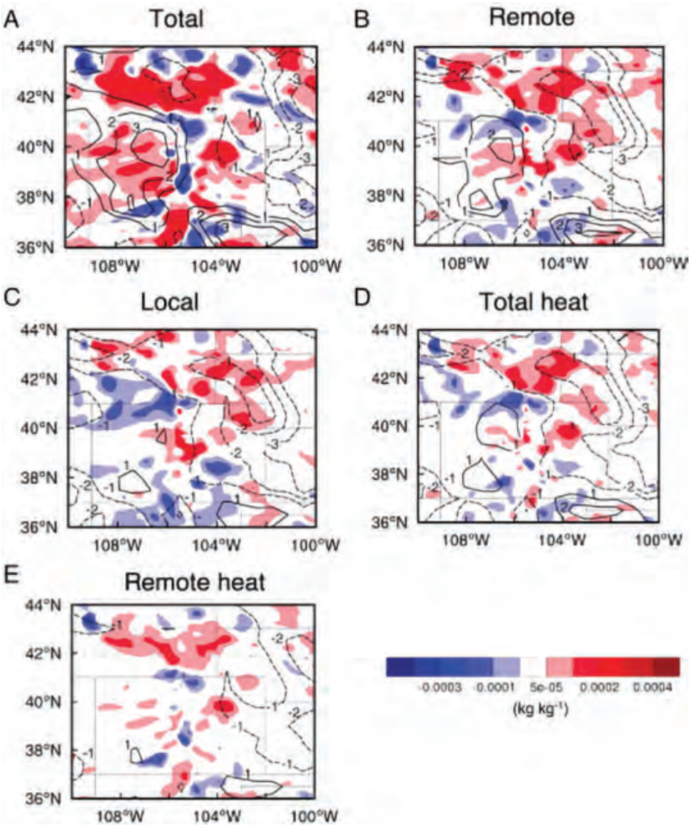


Figure 11. Figures from the paper show differences in moisture (shaded color) and wind speed (contour lines) at 850 hPa due to (A) total wildfire effect (Fire2-NoFire2), (B) remote wildfire effect (Fire2-Fire2L), (C) local wildfire effect (Fire2-Fire2R), (D) total heat effect (Fire2-Fire2_NH), and (E) remote heat effect (Fire2-Fire2_NRH) during two storm periods. Solid (dashed) contour lines denote increased (decreased) wind speeds.

Simulations at NERSC Drive Progress Toward Nuclear Fusion

THE SCIENCE. The National Spherical Torus Experiment-Upgrade (NSTX-U) at Princeton Plasma Physics Laboratory is a nearly spherical tokamak — a toroidal, or donut-shaped, nuclear fusion reactor that uses magnetic fields to contain the fusion reaction. First constructed in 1995 as NSTX, NSTX-U is currently being rebuilt and upgraded, and a series of simulations performed at NERSC show that its unique shape may help it disperse exhaust heat over a wider area on the material wall, making it a promising candidate for a pilot plant by the year 2040 in combination with other design features.

THE IMPACT. NSTX-U displays especially robust turbulent heat load spreading due to its tight aspect ratio — essentially, the hole in the donut at NSTX-U is smaller than that of a typical tokamak, compared with the width of the donut itself. The narrow space in the middle of the donut forces the confining magnetic field to be very strong as it passes through the center of the donut, and as electrons approach the hole of the reactor, the strong magnetic force field bounces some of them back into the donut, where the forces are weaker. This bouncing around of electrons causes a type of turbulence known as trapped-electron mode turbulence, which these simulations show is a particularly effective mechanism for heat load spreading. The results of these simulations are directly applicable to ITER, the International Thermonuclear Experimental Reactor currently under construction in France.

ADDITIONAL DETAILS. The simulations identifying these features at NSTX-U were performed on the Cori supercomputer at NERSC using XGC1, a particle-in-cell, five-dimensional gyrokinetic code. High-speed data transfer using ESnet allowed researchers to observe the simulations in real time and tweak them as necessary in the next simulation. Work continues using the next-generation supercomputer *Perlmutter* at NERSC, which will make the work even more efficient: switching from Cori's CPUs to *Perlmutter's* GPUs is expected to improve simulation performance by an order of magnitude.

NERSC PI: C.S. Chang

PROJECT FUNDING AND ALLOCATION AWARD: DOE Office of Science, Office of Fusion Energy Science (FES)

PUBLICATION: Guttenfelder, W., et al., "NSTX-U theory, modeling and analysis results," Nuclear Fusion, **62**, 042023 (2022). https://doi.org/10.1088/1741-4326/ac5448

Figure 12. A series of simulations performed at NERSC show that the unique shape of the NSTX-U tokamak may help it disperse exhaust heat over a wider area on the material wall, making it a promising candidate for a pilot plant by the year 2040 in combination with other design features.

Machine Learning Accelerates the Search for New Polymer Electrolytes

THE SCIENCE. Researchers at the Massachusetts Institute of Technology (MIT), in partnership with the Toyota Research Institute, have pioneered a new method for using machine learning to screen for new materials, creating a large dataset of polymer electrolytes, new classes of which are promising candidates for next-generation lithium-ion batteries. These results were published in the June 2022 issue of Nature Communications.

THE IMPACT. Scientists increasingly use automated search to screen different molecules for useful properties based on molecular structure. However, the amorphous nature of these polymers makes them difficult and computationally expensive to screen. The team accelerated the screening process by using a multi-task graph neural network that learns from a relatively large set of noisy, unconverged, short (5-nanosecond) molecular dynamics simulations and a small number of converged, long (50-nanosecond) simulations. The simulations were generated using LAMMPS, the Large-scale Atomic/Molecular Massively Parallel Simulator, on Cori.

ADDITIONAL DETAILS. This approach enabled the team to screen a search space of 6,247 polymers for four different properties — several orders of magnitude larger than previous search spaces of 20 polymers or fewer. The authors conservatively estimate that the 394,000 CPU hours needed to generate the simulations and complete the screening with machine learning was 25 times less expensive than direct "brute force" simulation-based screening.

NERSC PI: Tian Xie

PROJECT FUNDING AND ALLOCATION AWARD: DOE Office of Science, Office of Basic Energy Science

PUBLICATION: Xie, T., France-Lanord, A., Wang, Y. et al. "Accelerating amorphous polymer electrolyte screening by learning to reduce errors in molecular dynamics simulated properties." Nature Communications, 13, 3415. (2022) https://doi.org/10.1038/s41467-022-30994-1

Figure 13. Improved methods of searching for new electrolytes could be instrumental to finding alternatives to the lithium-ion batteries found in today's cars and consumer electronics.

GIGA-Lens Modeling Framework Enhances the Study of Dark Matter

THE SCIENCE. Strong gravitational lensing is the warping of spacetime due to the rare, chance alignment of two galaxies, which can be used to study both dark matter and dark energy. Strong lenses can be discovered using neural networks to search ground-based imaging surveys like the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Survey, but the computational expense of the process has previously limited its use. To address this problem, graduate student Andi Gu and a team overseen by Berkeley Lab scientist Xiaosheng Huang developed GIGA-Lens, a fast, robust, and scalable Bayesian modeling code for use on multiple GPUs. Their work was published in August 2022 in The Astrophysical Journal.

THE IMPACT. Strong gravitational lensing systems constitute a powerful tool for cosmology. They are uniquely suited to probe the low end of the dark matter mass function and test predictions of the cold dark matter model beyond the local universe. Gravitational lenses that produce multiple images of a single quasar (and supernovae in the near future) are being used to provide independent constraints on the Hubble constant. GIGA-Lens's three components — optimization using multi-start gradient descent, posterior covariance estimation with variational inference, and sampling via Hamiltonian Monte Carlo — all take advantage of gradient information through automatic differentiation and parallelization on GPUs. The robustness, speed, and scalability offered by GIGA-Lens make it possible to model the large number of strong lenses found in current surveys and present a very promising prospect for the modeling of some 105 lensing systems expected to be discovered in the era of the Rubin Observatory, the Euclid Mission, and the Roman Space Telescope.

ADDITIONAL DETAILS. Built for implementation using TensorFlow and JAX, GIGA-Lens is built to run on a single node of four A100 GPUs on *Perlmutter* or a similar system. It can model a complete strong lens system in 105 seconds, making modeling large numbers of systems within a reasonable time frame feasible for the first time.

NERSC PI: Saul Perlmutter

PROJECT FUNDING AND ALLOCATION AWARD: DOE Office of Science, Office of High Energy Physics

PUBLICATION: A. Gu, X. Huang, W. Sheu, G. Aldering, A. S. Bolton, K. Boone, A. Dey, A. Filipp, E. Jullo, S. Perlmutter, D. Rubin, E. F. Schlafly, D. J. Schlegel, Y. Shu, S. H. Suyu. "GIGA-Lens: Fast Bayesian Inference for Strong Gravitational Lens Modeling". *The Astrophysical Journal*, 935, 1. (2022) https://doi.org/10.3847/1538-4357/ac6de4

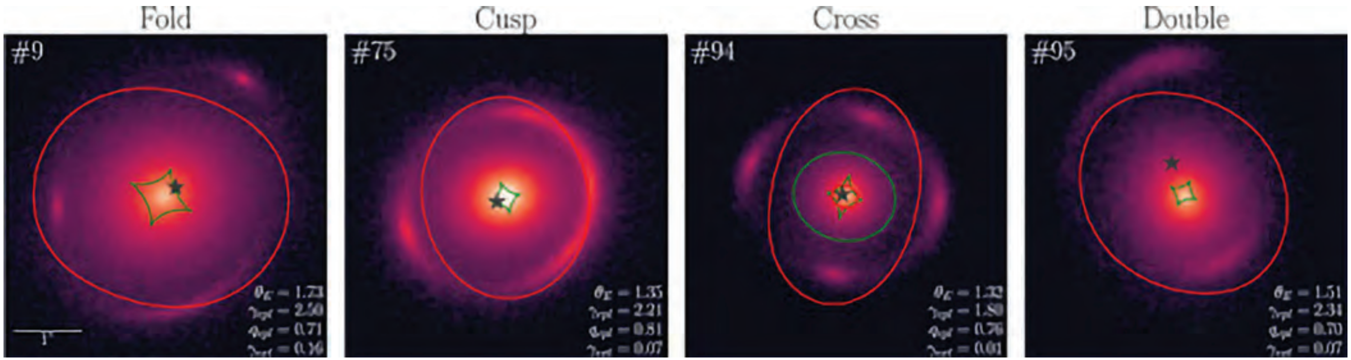


Figure 14. Four types of models of gravitational lensing yielded by GIGA-Lens. Built for implementation using either TensorFlow or JAX, GIGA-Lens can run on a single node of four A100 GPUs on *Perlmutter* or a similar system. The spots and arcs surrounding the central object are images of the lensed quasar or supernova.

Acronyms and Abbreviations

ALS
Advanced Light Source, Lawrence Berkeley National Laboratory

AMR
Adaptive Mesh Refinement

ANL
Argonne National Laboratory

API
Application Programming Interface

ASCR
Office of Advanced Scientific Computing Research

BER
Office of Biological and Environmental Research

BES
Office of Basic Energy Sciences

BNL
Brookhaven National Laboratory

CERN
European Organization for Nuclear Research

CMB
Cosmic Microwave Background

CPU
Central Processing Unit

CSCS
Swiss National Supercomputing Centre

DESI
Dark Energy Spectroscopic Instrument

DFT
Density Functional Theory

DTN
Data Transfer Node

ECP
Exascale Computing Project

FES
Office of Fusion Energy Sciences

GB
Gigabytes

Gbps
Gigabits Per Second

GPCNet
Global Performance and Congestion Network

GPU
Graphics Processing Unit

HDF5
Hierarchical Data Format 5

HEP
Office of High Energy Physics

HPC4Mfg
High Performance Computing for Manufacturing

JGI
Joint Genome Institute

KNL
Knights Landing Processors

LANL
Los Alamos National Laboratory

LCLS
Linac Coherent Light Source

LLNL
Lawrence Livermore National Laboratory

LZ Dark Matter Experiment
LUX-Zeplin Dark Matter Experiment

MFA
Multi-Factor Authentication

MHD
Magnetohydrodynamic

NCEM
National Center for Electron Microscopy

NESAP
NERSC Exascale Scientific Application Program

NIM
NERSC Information Management

NOAA
National Oceanic and Atmospheric Administration

NP
Office of Nuclear Physics

OLCF
Oak Ridge Leadership Computing Facility

OpenMP
Open Multi-Processing

OpenMSI
Open Mass Spectrometry Imaging

PB
Petabytes

PNNL
Pacific Northwest National Laboratory

PPPL
Princeton Plasma Physics Laboratory

PUE
Power Usage Effectiveness

SENSE
Software-defined Network for End-to-End Networked Science at Exascale

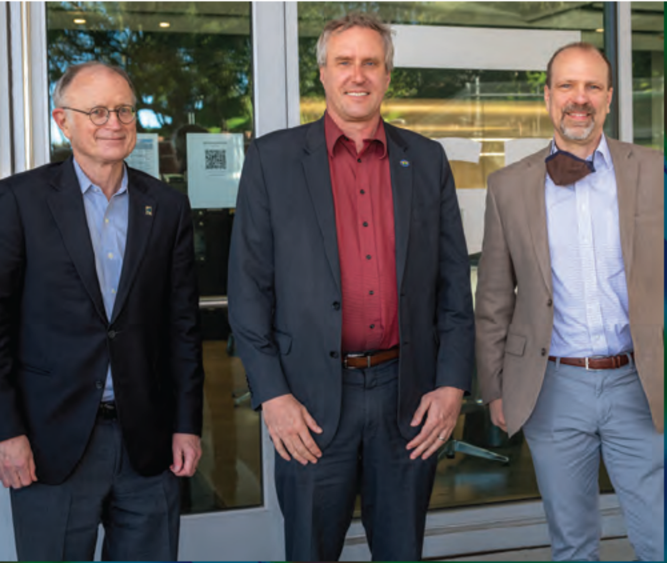
SciDAC
Scientific Discovery Through Advanced Computing

SDN
Software-defined Networking

SLURM
Simple Linux Utility for Resource Management

TAP
Trusted Access Platform

TB
Terabytes



FOR MORE INFORMATION ABOUT NERSC, CONTACT:

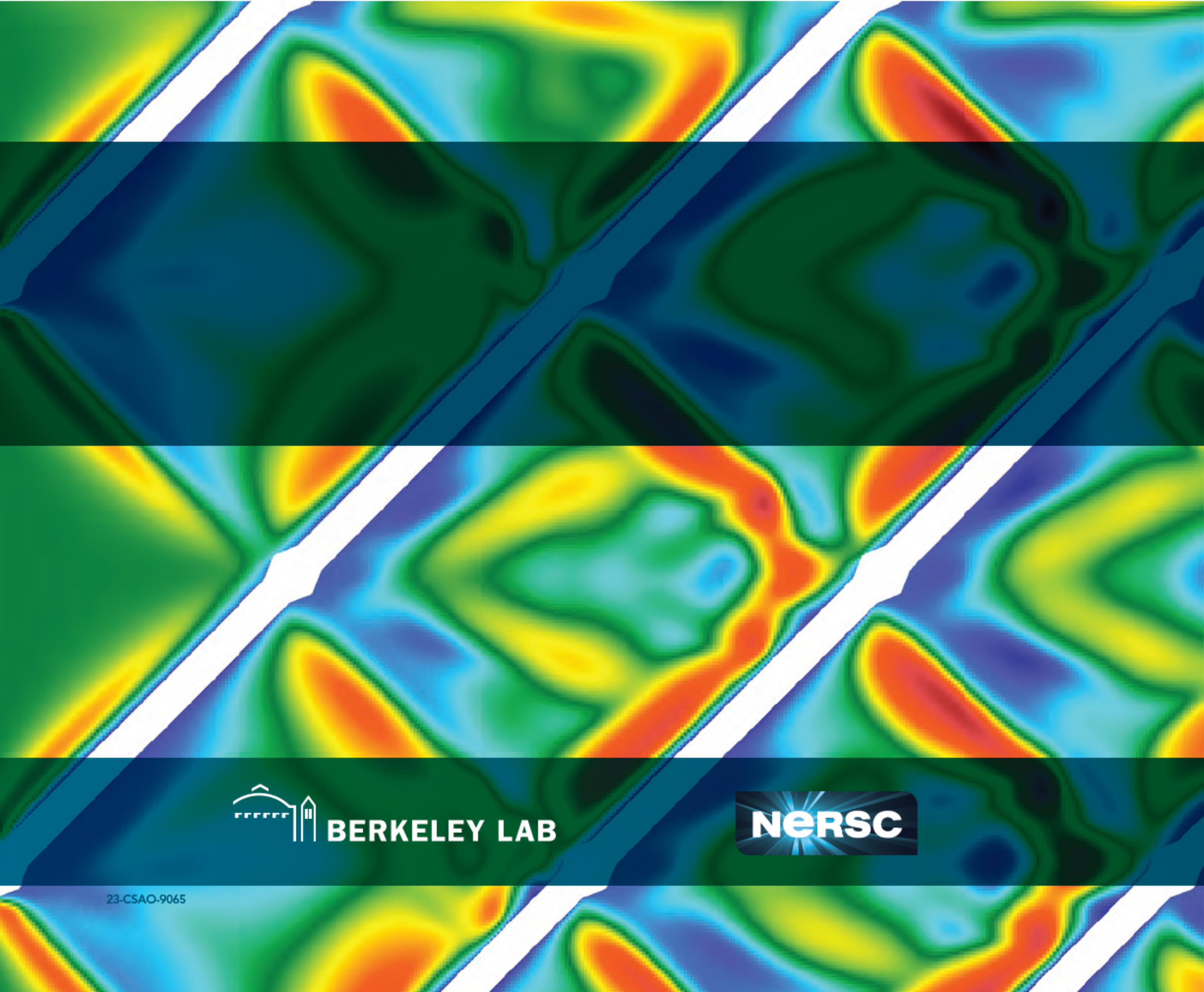
NERSC Communications
Lawrence Berkeley National Laboratory
1 Cyclotron Road
Berkeley, CA 94720-8148
Email: CSComms@lbl.gov

NERSC's Web Site
www.nersc.gov

NERSC Annual Report Editor
Elizabeth Ball

Design
Design, layout, illustration, photography, and printing coordination:
Berkeley Lab Creative Services

Cover Image
This visualization, produced by researcher David Trebotich using NERSC's *Perlmutter* supercomputer, shows a pattern of turbulent flow near a reverse-osmosis permeable membrane in a desalination device. The application code Chombo-Crunch, developed at Berkeley Lab, models reactions that can cause performance-limiting membrane fouling; researchers can then optimize treatment of the system's brine and allow it to be reused.
Credit: David Trebotich and Sergi Molins, Berkeley Lab



BERKELEY LAB

NERSC

23-CSAO-9065