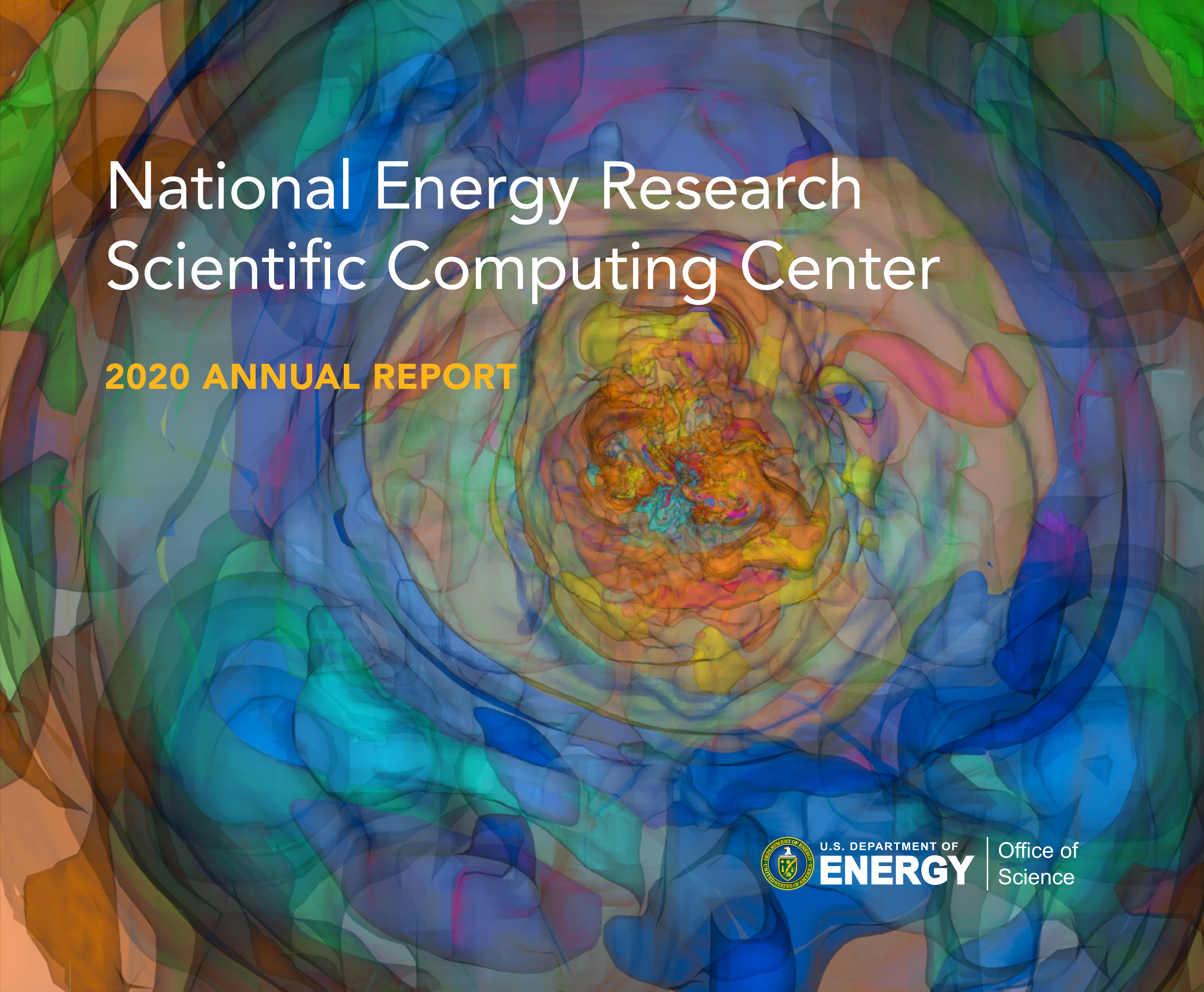
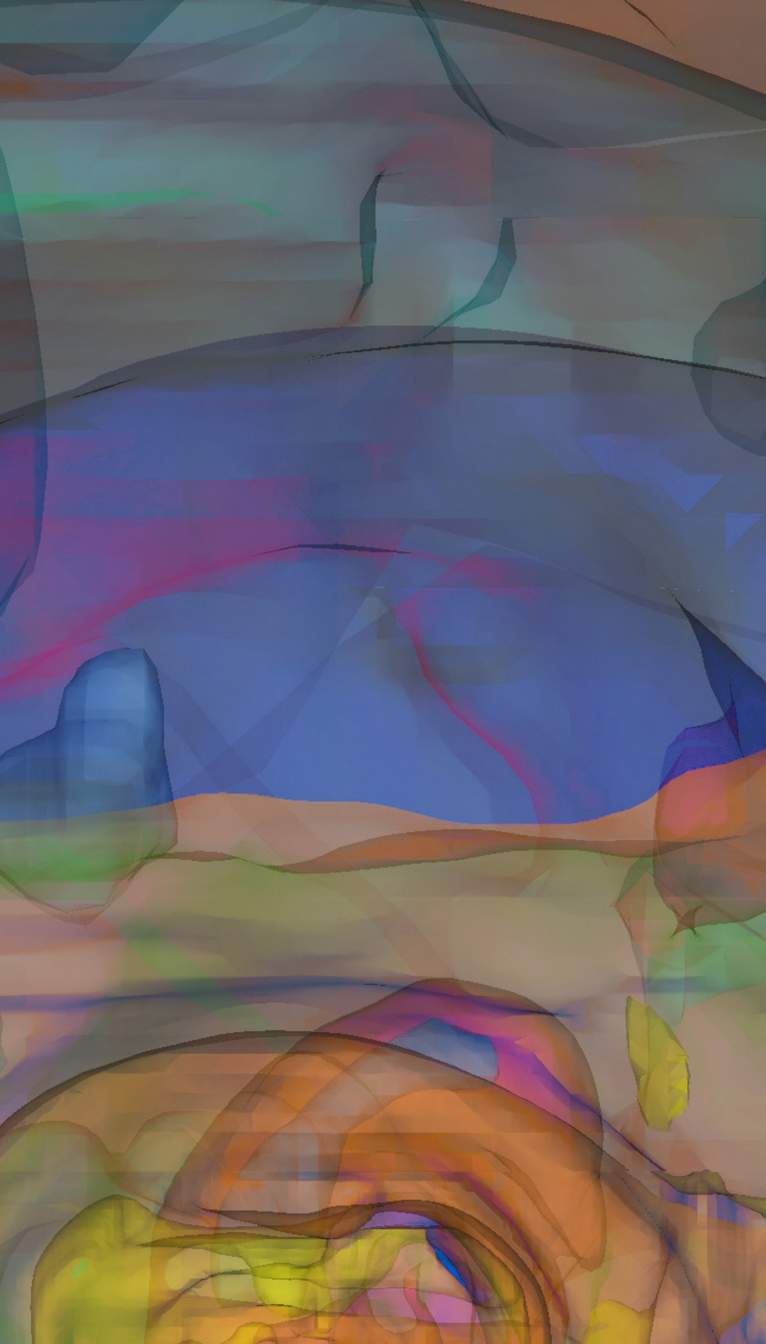


National Energy Research Scientific Computing Center

2020 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. 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

Despite the COVID pandemic and the many challenges it has created worldwide, in 2020 the National Energy Research Scientific Computing Center fulfilled its mission to deliver high-performance computing capabilities to the U.S. Department of Energy's Office of Science user community. In addition, NERSC achieved all operational metrics while maintaining very high user satisfaction and made major advances in its key strategic initiatives.

Managed by Lawrence Berkeley National Laboratory and funded by the DOE Office of Science Advanced Scientific Computing Research Office, NERSC currently serves approximately 8,000 scientists who are working on more than 900 research projects spanning the range of Office of Science scientific disciplines. In 2020 more than 1,800 refereed publications cited NERSC, demonstrating the large impact NERSC continues to have on the science community.

A major focus for NERSC in 2020 was to continue preparations for *Perlmutter*, our next-generation supercomputer that is being delivered in two phases in 2021. These efforts included upgrading the NERSC facility with adequate power and cooling, building networking and monitoring infrastructure, and designing the system administration, in addition to preparing user codes and pipelines for the system's new architecture.

As part of these efforts, the NERSC Exascale Science Applications Program (NESAP) made big advances optimizing codes targeting *Perlmutter's* GPU nodes. The NESAP applications include nearly 30 simulation, data analysis, and learning codes, plus workflows. One of the codes that was optimized in collaboration with NERSC staff was named a 2020 Gordon Bell Prize finalist. Throughout the year we also held several virtual training events on a range of topics to help users adopt best practices and use GPU resources more efficiently.

The internal Superfacility Project — which is focused on accelerating pipelines from DOE experimental and observational facilities in an automated, high-performing manner — continues to be another major focus for NERSC. In 2020 NERSC's superfacility team held a number of virtual outreach events and continues to enable collaborative abilities that improve data movement and sharing for experimental facilities.

The extraordinary circumstances of the pandemic, along with an unprecedented wildfire season in the second half of the year, demonstrated NERSC's flexibility and resilience to adversity. We went from a predominantly onsite operation in the early part of 2020 to an almost entirely remote operation in a matter of weeks. Even so, over the course of the year we developed a number of key innovations, finished our preparations for *Perlmutter*, and served our user base and the science community well. It is a year we are ultimately quite proud of.

— Sudip Dosanjh
NERSC Division Director

NERSC by the Numbers

2020 NERSC USERS ACROSS US AND WORLD

50 States + Washington D.C. and Puerto Rico

46 Countries

~8,000 ANNUAL USERS FROM ~1,750 Institutions + National Labs

29% Graduate Students

20% Postdoctoral Fellows

16% Staff Scientists

11% University Faculty

6% Undergraduate Students

6% Professional Staff

61% Universities

30% DOE Labs

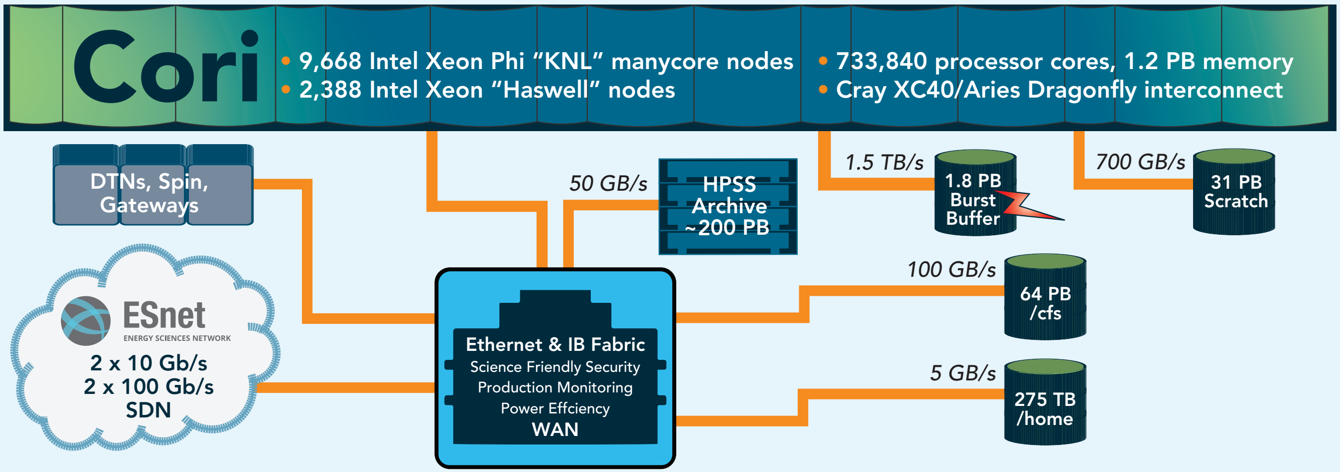
5% Other Government Labs

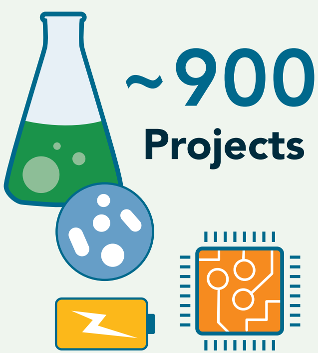
2% Industry

1% Small Businesses

<1% Private Labs

NERSC Systems





~900
Projects

Top Science Disciplines (By computational hours used)

Geoscience
Accelerator Science
Energy Technologies
Nuclear Physics
Earth Systems
Engineering
Chemistry
Mathematics
Materials Science
Fusion Energy
Astrophysics and Cosmology
Condensed Matter
Computer Science
User Facilities
High Energy Physics

2020 COVID-19 Research @ NERSC

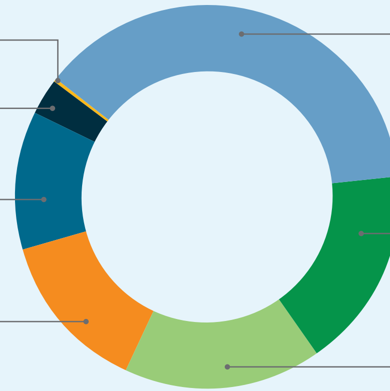
200 million

Node hours on the Cori supercomputer

21 projects

Teams include epidemiologists from the Department of Health and Human Services, plus researchers from national labs, academia, and industry, all working to develop COVID-19 diagnostics and therapeutics.

2020 DOE Office of Science Program Usage Breakdown



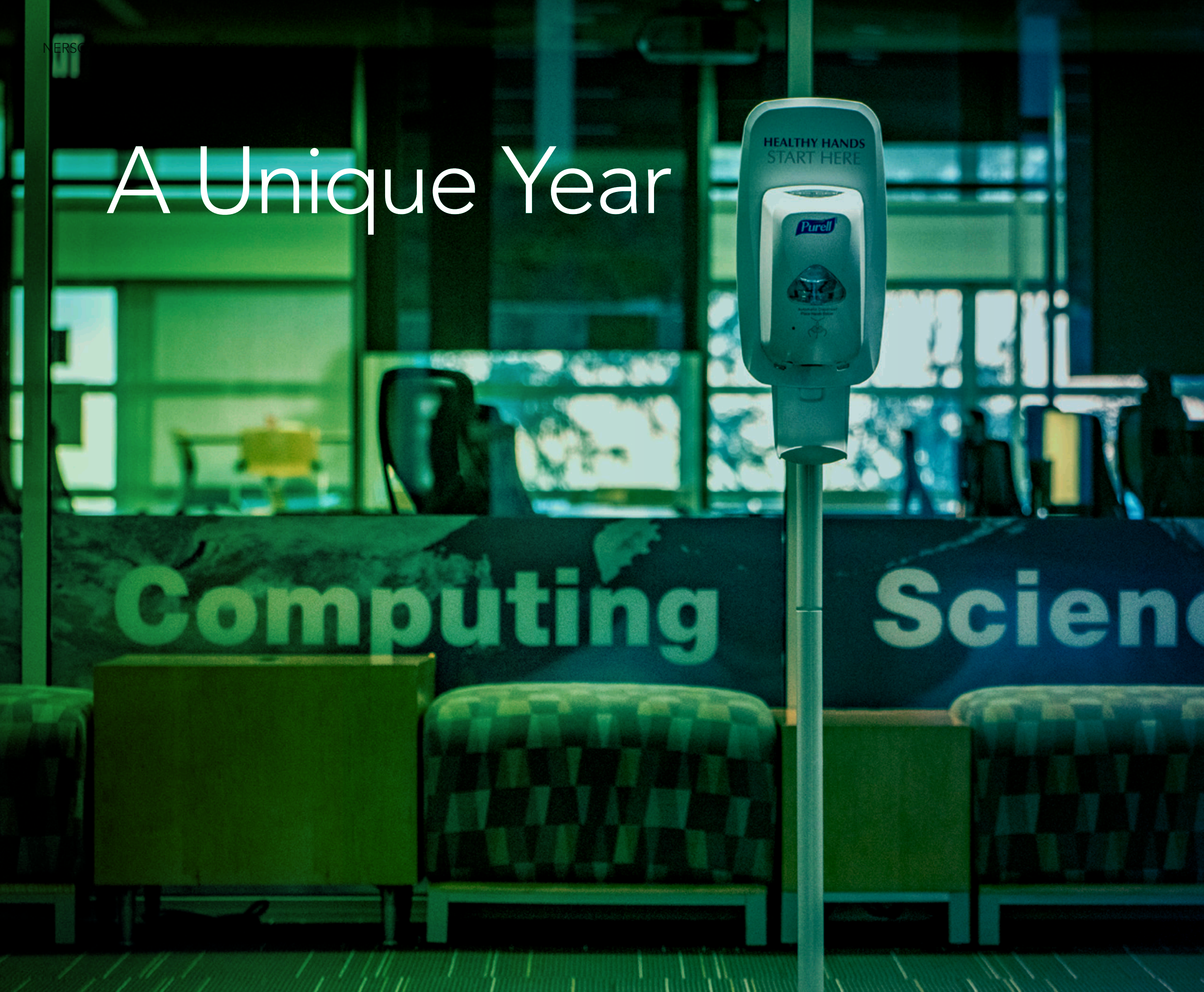
9 BILLION CORE HOURS USED IN 2020

>1,800

Refereed Publications Cited NERSC

Data Stored 200 Petabytes

A Unique Year



In 2020, the extraordinary circumstances of the COVID-19 pandemic prompted NERSC to find new and creative solutions to meet the needs of its users and staff.

We went from a predominantly onsite operation to nearly entirely remote operations in a matter of weeks. Despite the challenges this created, NERSC was able to provide high performance computing and data (HPC systems and services as usual, resulting in another very scientifically productive year by its users. This included research into COVID-19 related topics as NERSC contributed 200 million hours of compute time and expert HPC consulting towards fighting the pandemic.

NERSC staff had to make a number of adjustments to keep the center running and on track for planned upgrades and enhancements. Some staff were completely remote, while others were on site, depending on responsibilities.

From the user perspective, there was little change to how they interacted with the facility on a day to day basis because they were all remote, even before SARS-CoV-2. In fact, NERSC's computers and data systems ran with very high availability and utilization in 2020. We did make a few user-visible changes in response to the pandemic that were popular and effective. For example, we created a NERSC user

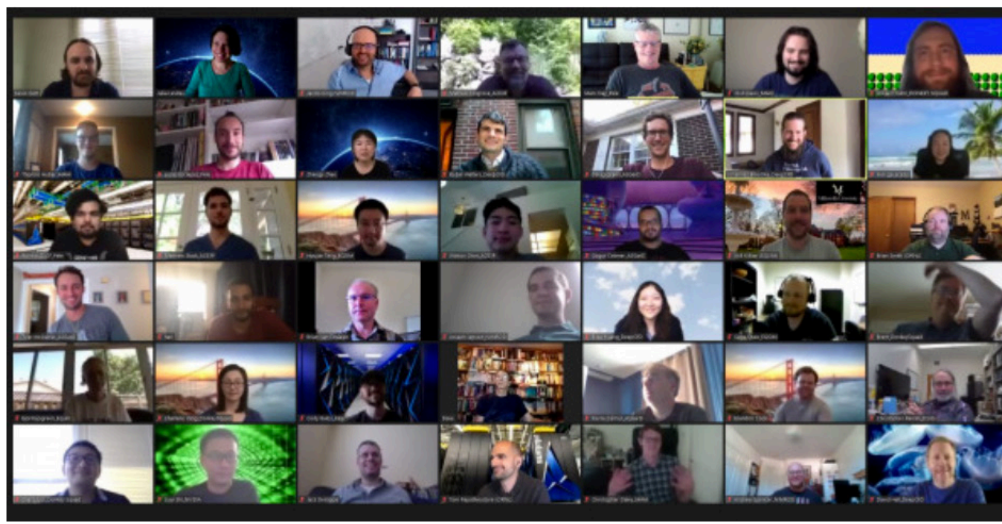


Figure 1. In July 2020, NERSC hosted a GPU hackathon in conjunction with NVIDIA, the Oak Ridge Leadership Computing Facility, and OpenACC as part of the GPU Hackathons series. Participants hailed from across the U.S. and as far away as Australia.

Slack channel for users to communicate with each other, implemented virtual “office hours” by appointment and converted our intensive in-person hackathon coding events into a series of shorter virtual events. Similarly, the format of the NESAP hackathons was transformed from intense, weeklong events into a series of events with the same total number of hours spaced across the span of 12 weeks. There were advantages to this approach, including the opportunity for more thoughtful analysis of the codes being optimized.

The pandemic also impacted how we conducted and participated in staff meetings, user trainings, hackathons, and more, with all of us learning how to interact in the virtual world primarily via Zoom. For example, we typically hold a number of in-person training events each year, with the option to participate remotely but with in-person attendance encouraged. The pandemic changed our focus, by necessity, to the remote participant experience. We realized that we often put together a full-day program to maximize the impact on in-person attendees who have disrupted their schedule for the day, but remote participation does not require such substantial overhead, meaning multiple sessions may be more convenient. We therefore began focusing on training series rather than long single-session events, and posting edited versions of the events on the NERSC YouTube channel for asynchronous learning. One advantage of an in-person event is accessibility for those who may lip-read. We thus made it a priority to provide accurate captioning on the recordings of NERSC training events to maximize accessibility for everyone.

Navigating a Facility Upgrade

Another major focus for NERSC in 2020 was getting ready for the arrival of our next-generation supercomputer, *Perlmutter.*

This included upgrading the facility with power and cooling improvements to support an additional 12.5 MW of HPC, building networking and monitoring infrastructure, designing the system administration, and preparing user codes and pipelines. NERSC operational staff also prepared software and hardware infrastructure to support the arrival of the machine.

The pandemic complicated the completion of the building power and cooling upgrades, as well as the installation of *Perlmutter* hardware and the Community File System upgrade. Extraordinary safety precautions were put in place to ensure the safety of NERSC staff, Berkeley Lab staff, system vendor staff, and construction trades. These included coordination of work areas and time of work, health checks, sanitization of tools, and use of respirators for close work. While the safety protocols sometimes slowed work, there was no on-site transmission of COVID and no one was injured.

Thus, despite some interruptions as a result of the pandemic, the NERSC-9 Facility Upgrade project reached near substantial completion in 2020, and the first phase of *Perlmutter* was installed in early 2021. This is a testament to the Project

Figure 2. As part of a facilities upgrade to Berkeley Lab's Bldg. 59, where NERSC's supercomputers are housed, three new air handling units were installed in 2020, contributing to expanded cooling capabilities to the facility.

Figure 3. In addition to the new air handling units, which take advantage of the Bay Area's temperate climate by processing air from outside the building to help cool its computing resources, NERSC also installed new power distribution panels, overhead conduit, and five new substations, adding 12.5 MW of power to the facility.

Infrastructure and Modernization Division and Facilities Division, NERSC, and XL Construction (the primary construction subcontractor), who together were able to remotely coordinate all of these activities — traditionally done face-to- face — while the crew was still in the field.

Responding to California's Wildfire Season

Following the 2019 California wildfires, NERSC made a number of changes to be better prepared to deal with future fire events.

Figure 4. The Air Quality Incident system uses Purple Air sensors to provide real-time visibility of the data center's indoor AQI and real-time AQI data for HPC operations.

NERSC implemented functionality to its Building Management System to include an Air Quality Incident (AQI) mitigation mode, allowing operators of the Building Management System to quickly change the operational configuration of the computer room, office areas, or both from using outside air to using recirculated air. NERSC utilizes multiple tiers of particle sensors, indoors and outdoors, to track air quality. Within the computer room, clean-room grade sensors are used to protect the computing equipment, including the very sensitive tape library system. Purple Air and Awair Omni sensors are used to monitor indoor air quality to help ensure the health of staff members. Outdoor sensors provide an early warning of air quality incidents, including those local to the Berkeley area, that might overwhelm the building's high-grade filtration system for intake air.

Looking ahead, NERSC will continue to make these sorts of adjustments in response to what is, unfortunately, becoming a seasonal norm in Northern California and other parts of the West. In 2020, the state experienced a record-setting wildfire season, with almost 10,000 incidents that amounted to more than 4 million acres burned across the state. Several Red Flag Warnings were issued at Berkeley Lab between the months of August and October, and the Lab was closed on October 26, 2020 as a precautionary move due to a Red Flag Warning. Despite the numerous wildfires in the area, no Public Safety Power Shutoffs occurred in 2020 at Berkeley Lab.

Supporting COVID-19 Research

Long before the image of the coronavirus virion waving its prominent protein spikes became iconic, scientists were exploring the way those spikes bind protein-to-protein to figure out ways to disrupt that binding and, ultimately, epidemic infection. As the pandemic took hold, however, researchers, epidemiologists, and computer scientists began tapping into the HPC expertise and computational might of DOE supercomputers to accelerate their studies.

In April 2020, NERSC joined in the battle against the coronavirus pandemic as a member of the COVID-19 HPC Consortium of technology companies, federal agencies, and other national labs aiming to find innovative solutions to combat COVID-19. Over the year, NERSC allotted 200 million node hours on the Cori supercomputer to Department of Health and Human Services epidemiologists and 19 other teams from national labs, academia, and industry working to develop diagnostics and therapeutics, model the spread of the virus, and understand the structure of SAR-CoV-2 and how it infects cells. NERSC has also provided dedicated HPC staff liaisons and other resources to support Consortium projects and other COVID-19 focused research.

Figure 5. In 2020, as part of the COVID-19 HPC Consortium, NERSC allotted 200 million node hours on the Cori supercomputer for projects specifically related to COVID-19 focused research.

The projects cover a wide swath of COVID-19 research, including explorations into FDA-approved drug repurposing, molecular simulations, explorations of resilience, temperature, and humidity effects on SARS-CoV2; epidemic simulation models of the US and other populations; and COVID-19 publications text mining. Here are some examples:

- **How Protein Spike Binding Hijacks Cells.** One project — a worldwide collaboration led by Professor Wai-Yim Ching at the University of Missouri, Kansas City — is seeking to understand how the COVID-19 virus enters and infects human cells in unprecedented detail at the molecular and atomic level. The project team used NERSC's Cori supercomputer and advanced algorithms to determine the precise locations of the atoms involved in protein-spike binding to ACE2, and improve the precision to less than 0.1 angstroms. The calculations run at NERSC reveal that the correct shape for attaching to ACE2 appears to be determined by a cluster of amino acids, the units that make up the spike protein, with large positive charges. This result provided insight into the infection process at the level of the molecular machinery of cells.
- **AI Plays a Role.** With the awareness that developing novel active therapeutics against coronaviruses such as the one responsible for COVID-19 can be a long, arduous process, another collaboration is looking to uncover existing FDA-approved drugs that could be repurposed to combat COVID-19. Researchers from Harvard University and the Massachusetts Institute of Technology are employing 3D machine learning techniques to accelerate the discovery of an existing small molecule that has been tested and is bioavailable. They are using highly efficient electronic structure simulations to quickly calculate molecular conformations and train 3D message-passing neural networks from existing molecular screens against the related SARS-CoV-1 and SARS-CoV-2 data as it becomes available.
- **Massive Epidemiology Simulations to Inform U.S. Government COVID-19 Policy Decisions.** A U.S. Department of Health and Human Services project is performing epidemiology simulations of the entire U.S. population to inform the White House Coronavirus Task Force about the number of COVID-19 cases that are projected to occur in various regions, the need for medical resources, and to understand the impact of social distancing and other interventions. The team is developing county-level mobility and population movement estimates and applying global circulation models, an agent-based epidemiological model used by the Federal Emergency Management Agency. NERSC worked with the team to run a COVID-19 simulation for the full U.S. population; this simulation was used to compare and validate the stay-by-state models that are run frequently for different scenarios.

In addition to supporting these and other research projects, NERSC resources were used to help create the COVID Scholar literature search portal, designed to help researchers find of-the-moment research results related to the COVID-19 pandemic. Using NERSC's "Spin" containers-as-a-service platform and expert staff assistance, COVID Scholar is powered by natural language processing models running daily on Cori and is having an impact beyond just the search capabilities provided by the portal.

Summer Students Go Virtual with Deep Learning Research

Each summer, NERSC hosts several college-level students through the Computing Sciences Area's Summer Program. In 2020, many of the students focused on applying deep learning methods to an array of science applications, from weather and climate modeling to improving I/O scaling for GPUs.

Here is a look at four students who spent the summer of 2020 involved in cutting-edge research at NERSC — virtually, that is.

Ashesh Chattopadhyay

During his summer internship, Ashesh Chattopadhyay was a third-year Ph.D. student at Rice University who obtained his undergraduate degree in mechanical engineering from the Indian Institute of Technology in Patna, India, and his masters in computational science from the University of Texas at El Paso. For his Ph.D. research, Chattopadhyay returned to his roots in mechanical engineering with a focus on deep learning for high-dimensional systems such as turbulence and generally fluid dynamics.

His research experience is in the development of physics-informed deep learning algorithms in high-dimensional spatio-temporal turbulence, especially in climate dynamics and non-linear dynamical systems. At Berkeley Lab, his focus was on developing deep learning algorithms that can be used to perform weather and climate modeling without any knowledge of the physical equations that govern the complex physics of these systems.

"A major challenge in problems such as climate/weather or generally fluid dynamics is the huge computational cost that is required for effective high-resolution simulations," he said. "Deep learning algorithms can play a major role in reducing these computational costs without compromising on the quality and accuracy of these simulations."

Shuni Li

Statistics is the core of machine learning and was the focus of Shuni Li's Ph.D. work at U.C. Berkeley and her research at Berkeley Lab in 2020. The project she worked on over the summer in conjunction with her mentor, Steve Farrell, in the DAS group, had a lot in common with her Ph.D. research, which is about modeling and designing synthetic polymer sequences in materials science: Both try to model biological sequences and have the common goal of understanding what sequence compositions/structures allow certain sequences to function. Li's undergrad work at Macalester College was in math and computer science, and she became interested in deep learning after watching a series of lectures on deep unsupervised learning by UC Berkeley's Pieter Abbeel.

"When he talked about autoregressive models, latent models, generative adversarial networks, etc. and showed how they could be used to solve complex vision and language problems, I started to have this idea of combining deep learning with my own research," she said. "And that's when I started to explore relevant literature and build my own deep learning models to study synthetic biological sequences."

Kevin Luna

As a Ph.D. student in applied mathematics at the University of Arizona, the bulk of Luna's training has focused on traditional mathematical approaches to science problems. Along the way he's discovered that, while useful, these approaches have limits in their ability to produce answers to pressing scientific problems. So during his time at Berkeley Lab, he leveraged deep learning methods to accelerate traditional partial differential equation-based simulations in real time.

"I became deeply interested in applying deep learning to science problems because of deep learning's impressive performance and flexibility on a wide variety of science problems," he said. "However, as a mathematician at heart, rather than seeing deep learning as a means to replace traditional methods, I view it as a powerful complement to existing methods that we already have. This complementary view can be seen in my summer project, where we combine deep learning with traditional methods to accelerate simulations."

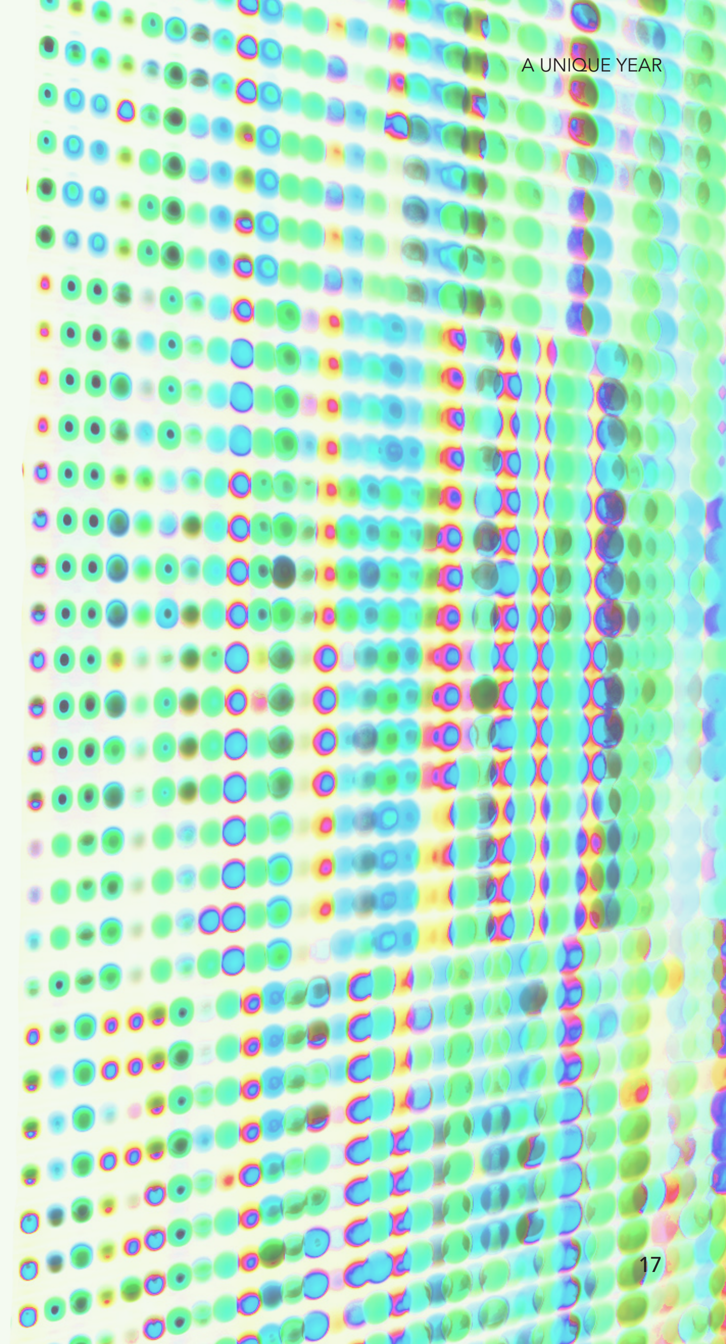
As a student with a budding research career, Luna added,

"Berkeley Lab caught my interest because it provides a unique, supportive, and collaborative environment full of opportunities to learn from and work with world-class researchers on cutting-edge research problems."

John Ravi

John Ravi made the most of his time at Berkeley Lab by working on different projects that aimed to enhance I/O scaling for HPC, with a particular focus on improving GPU I/O for applications using HDF5 (version 5 of the Hierarchical Data Format for heterogeneous data). A Ph.D. student at North Carolina State University — where he also earned his bachelors and masters degrees — Ravi's background is in computer engineering and computer science. He was drawn to working at Berkeley Lab through his Ph.D. advisor, Michela Becchi, who had collaborated in the past with Suren Byna and Quincey Koziol, who were his mentors at Berkeley Lab.

"I am working on methods to improve the capability of applications to compute on larger data sizes in a shorter amount of time, in the hope that it would unlock more interesting findings," he said. "I have been intrigued with what is possible with deep learning techniques ever since I first learned about them."



2020 Milestones

NESAP: Preparing the NERSC Workload for *Perlmutter*

Arriving in 2021, the *Perlmutter* system is an HPE Cray EX supercomputer that will initially include more than 1,500 GPU-accelerated nodes and a 35PB all-flash Lustre file system for high-bandwidth storage. Phase 2 will add more than 3,000 CPU-only nodes with 512GB of memory per CPU node.

To ensure that its users can productively use *Perlmutter*, NERSC has been working with key application development teams since 2019 to prepare codes to use the new technology in the *Perlmutter* system through the NERSC Exascale Science Applications Program (NESAP).

NESAP is 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 late 2014 to help users prepare for the Cori manycore Knights Landing/Xeon Phi architecture. Once the contract to procure *Perlmutter* was announced, the focus shifted to that system.

Throughout 2020 NESAP worked with 58 projects, broken down into two tiers, with 29 projects in each tier. Tier 1 teams receive enhanced support including, in some cases, a postdoctoral fellow working on their project. The projects span three focus areas:

- **NESAP for Simulations:** Cutting-edge simulations of complex physical phenomena
- **NESAP for Data:** Data analysis science pipelines that process massive datasets from experimental and observational science facilities

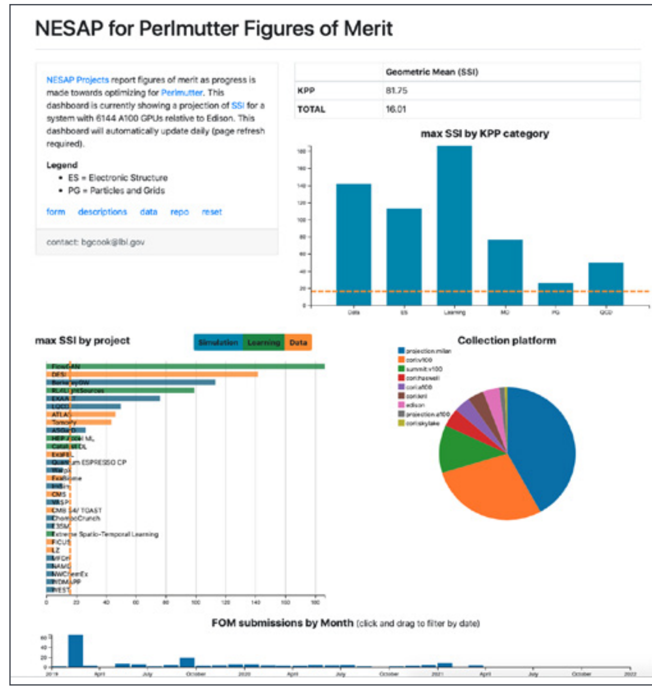


Figure 6. The NESAP-developed custom online dashboard for tracking applications progress toward ***Perlmutter*** performance goals, as measured by the scalable system improvement metric.

- **NESAP for Learning:** Machine learning and deep learning solutions to improve scientific discovery potential on experimental or simulation data or improve HPC applications by replacing parts of the software stack or algorithms with machine learning or deep learning solutions.

NESAP made substantial progress in 2020, with several projects showing large reductions in their problems' time to solution through combined efforts of NERSC postdocs, NERSC staff liaisons, vendor partners, and project developers. Progress is being tracked by a NERSC-developed custom dashboard that maintains a live view of the current status of each NESAP project and a projection of the total scientific throughput, as measured by the scalable system improvement metric, by project and by category. Below are some examples from 2020.

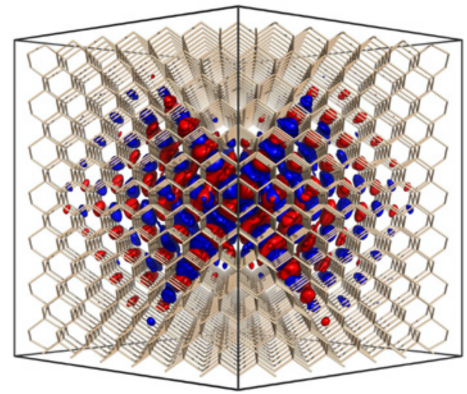


Figure 7. Isosurface enclosing 90% of the wave function for an in-gap state from a divacancy defect in silicon. Point defects in semiconductors are prototypes of solid-state qubits. This calculation cell contains 2,742 atoms and 10,968 electrons.

NESAP Efforts on BerkeleyGW Lead to Gordon Bell Finalist Honors

BerkeleyGW, one of the *Perlmutter* NESAP for Simulations projects and an important code that is funded under DOE Basic Energy Sciences' Computational Materials Science Center program, was honored as a Gordon Bell Award finalist at SC20. This work was done as part of the NESAP effort to optimize this code for GPU systems and extremely large scales. NESAP team members heavily tuned the GPU performance using absolute performance metrics and NVIDIA's NSight Roofline performance modeling feature.

The group's SC20 paper, "Accelerating Large-Scale Excited-State GW Calculations on Leadership HPC Systems," demonstrates for the first time the possibility of performing high-fidelity, excited-state calculations of complex materials at very large scales within minutes on current HPC systems, paving the way for future efficient HPC software development in materials, physical, chemical, and engineering sciences.

The paper, led by NERSC's Charlene Yang and Berkeley Lab Computational Research Division's Mauro Del Ben, presents algorithm and implementation advancements to scale calculations to over 10,000 electrons utilizing the entire Summit supercomputer at the Oak Ridge Leadership

Computing Facility. Like *Perlmutter*, Summit gets most of its computational power from GPUs. The team achieved excellent strong and weak scaling, and reached 105.9 PFLOP/s double precision performance on 27,648 NVIDIA V100 GPUs — 52.7% of the machine peak.

The parallelization and performance optimization techniques presented in this paper are widely applicable to other HPC applications and will be used as a case study when transitioning other NERSC users to the *Perlmutter* system. This work sets a new milestone for large-scale excited-state electronic-structure calculations using the GW method and opens up new possibilities for high-fidelity complex materials science studies in the exascale timeframe.

The BerkeleyGW code has also been executed on NVIDIA A100 GPUs, those that will power the *Perlmutter* system. Key computational kernels have been shown to see approximately 2x performance increases over execution on Summit's V100 GPUs due to the new double-precision Tensor Core support on A100 and the DMMA instructions.

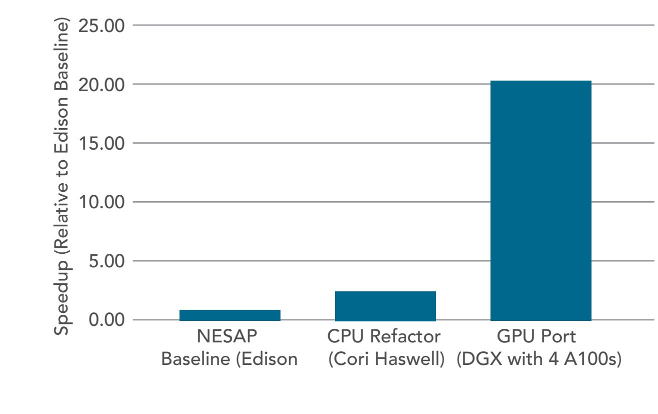
20x Speedup for DESI Spectroscopic Extraction using GPUs

As part of the NESAP for Data program, NERSC has partnered with the data management team of the Dark Energy Spectroscopic Instrument (DESI) experiment to accelerate DESI's existing spectral extraction image-processing pipeline using GPUs. During operation, DESI sends batches of images each night to be converted into spectra by the DESI spectroscopic pipeline running at NERSC. The data is processed in near-real-time to monitor survey progress and update the observing schedule for the following night. The pipeline is implemented almost entirely in Python and relies heavily on libraries, including NumPy and SciPy. The pipeline utilizes a divide-and-conquer strategy and MPI for multi-core and multi-node scaling.

In early 2020, the team participated in a NESAP hackathon to begin the GPU port of the spectral extraction code. An outcome of the hackathon was a major refactor of the CPU-based Python code that would then serve as a base for the GPU port, along with a road map for implementing the GPU port. The team followed an iterative development approach for porting and optimizing the application using a variety of tools:

- NVIDIA Nsight Systems for performance analysis
- A Shifter-based containerized workflow for recording reproducible benchmarks
- A combination of CuPy and JIT-compiled CUDA kernels via Numba for GPU acceleration
- MPI via mpi4py and CUDA Multi-Process Service to over-subscribe processes to GPUdevices and maximize use of both CPU and GPU resources.

The team contributed Python wrappers of low-level, high-throughput batched linear algebra operations in CUDA cuSOLVER (C/C++) to the open-source CuPy package. The team has also implemented interleaved I/O using excess CPU processes beyond those necessary to saturate GPU utilization.



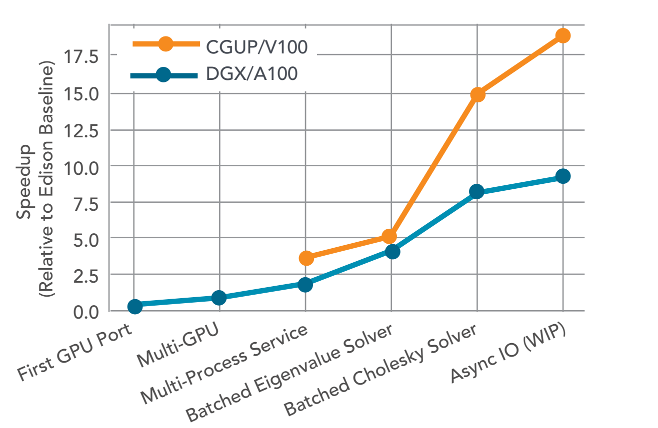


Figure 8 (left). Performance improvement relative to the Edison baseline for the number of images processed per node-hour by the spectral extraction code, due to CPU-oriented refactoring and GPU porting.

Figure 9. Cumulative performance improvement of the number of images processed per node-hour by DESI spectroscopic extraction, relative to the Edison baseline, due to each feature implementation milestone in the GPU porting effort.

Benchmarks using a *Perlmutter*-like configuration on Cori GPU (four V100 GPUs per node) and DGX (four A100 GPUs per node) systems, respectively, showed 10x and 20x speedups relative to the Cray XC30 Edison baseline in the number of images processed per node-hour by the spectral extraction code. This was a significant achievement for the DESI pipeline that puts the throughput of the spectral extraction on par or better with other steps in the data pipeline, thus exposing new targets for optimization. The lessons learned during the course of this work will help guide future efforts of the DESI team and inform other science teams looking to leverage GPU acceleration in their Python-based data-processing applications.

TurbulenceNet: Using Machine Learning to Augment Turbulent Fluid Dynamics Simulations

Numerical modeling of fluid dynamics is important for applications in areas of science and technology where fluid physics plays an important role, such as climate and weather modeling, aerodynamics, and combustion.

Most fluid flows of interest are turbulent and across a wide range of temporal and spatial scales. Modeling these fluid flows is challenging; a direct numerical simulation approach — in which the size of the computational grid is sufficient to fully resolve the full range of spatial and temporal scales exhibited by the governing equations — is computationally expensive, but simulation at lower-resolution on a coarse-grid introduces significant errors.

The goal of the TurbulenceNet NESAP for Learning project is to introduce a machine learning technique based on a deep neural network architecture to augment traditional numerical simulations. The machine learning model is used to correct the numerical errors induced by a coarse-grid simulation of turbulent flows at high-Reynolds numbers, while simultaneously recovering an estimate of the high-resolution fields.

The first phase of the NESAP project successfully demonstrated an architecture using state-of-the-art neural networks combined with a PDE solver to demonstrate error-correction and up to 4x superresolution of a 2D Rayleigh-Benard Convection system.

The project is now preparing to demonstrate the methodology at scale on *Perlmutter* for a variety of two-dimensional and three-dimensional, high-Reynolds number fluid flows, using a massively scalable, incompressible Navier-Stokes flow solver

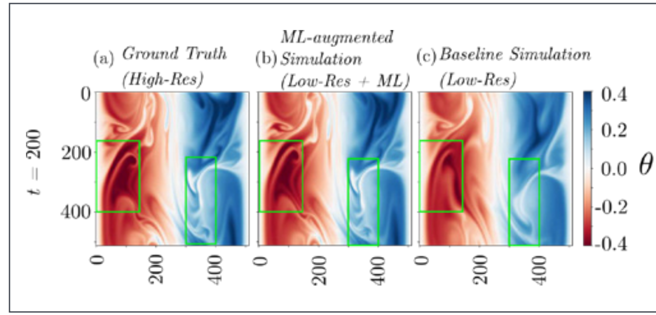


Figure 10. Temperature snapshot of a Rayleigh-Benard simulation. (a) using a high-resolution PDE solver, (b) using a machine-learning-augmented low-resolution PDE solver, and (c) baseline low-resolution PDE solver for reference. The highlighted boxes indicate the features of the ground truth flow captured by the machine-language augmented PDE solver but absent in the low-resolution baseline.

based on the AMReX package developed at Berkeley Lab. Initial timing studies indicate that the *Perlmutter* GPU resources will offer a significant speedup in training times, enabling the project to tackle full, complex 3D datasets.

The project used a 10,000-image Rayleigh-Bénard Convection dataset (about 10% of the full dataset) to measure training performance on Cori Haswell and GPU nodes. Using four Nvidia V100 GPUs reduced the training time for this dataset by a factor of more than 180x, compared with a single dual-socket Haswell compute node.

NERSC ROLLS OUT NEW COMMUNITY FILE SYSTEM

Recognizing the evolving data management needs of its diverse user community, in 2020 NERSC unveiled the Community File System (CFS), a long-term data storage tier developed in collaboration with IBM that is optimized for capacity and manageability.

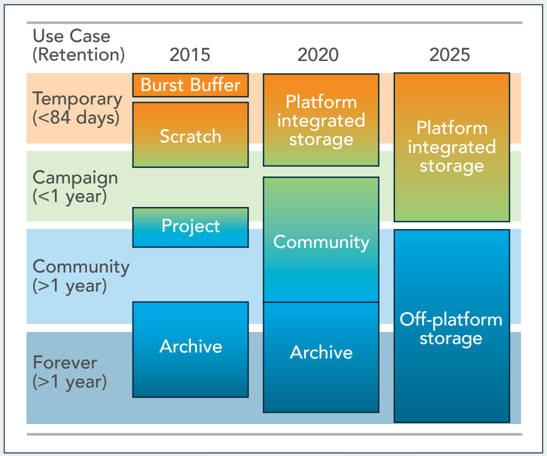


Figure 11. The NERSC storage hierarchy, 2015-2025.

The CFS replaced NERSC's Project File System, a data storage mainstay at the center for years that was designed more for performance and input/output than capacity or workflow management. But as high performance computing edges closer to the exascale era, the data storage and management landscape is changing, especially in the science community. In the next few years, the explosive growth in data coming from exascale simulations and next-generation experimental detectors will enable new data-driven science across virtually every domain. At the same time, new nonvolatile storage technologies are entering the market in volume and upending long-held principles used to design the storage hierarchy.

In 2017, these emerging challenges prompted NERSC and others in the HPC community to publish the Storage 2020 report, which outlines a multi-tiered data storage and management approach designed to better accommodate the next generation of HPC and data analytics. Thus was born the CFS, a disk-based layer between NERSC's short-term Cori scratch file system and the High Performance Storage System tape archive that supports the storage of, and access to, large scientific datasets over the course of years. CFS serves primarily as a capacity tier as part of NERSC's larger storage ecosystem, storing data that exceeds temporary needs but should be more readily available than the tape archive tier.

With an initial capacity of 60PB of data (compared to the Project File System's 11PB, and expanding to 200PB by 2025) and aggregate bandwidth of more than 100GB/s for streaming I/O, CFS is by far the largest file system NERSC currently has — twice the size of Cori scratch. With CFS, every NERSC project has an associated Community directory that provides a "snapshot" capability that gives users a seven-day history of their content and any changes made to that content. In addition, the system offers end-to-end data integrity.

Figure 12. The current NERSC Superfacility Project strategic science engagements.

The Supercharged Superfacility Project

The needs of data-oriented science are increasingly important to the Office of Science and NERSC's users.

In 2019, when users were last asked to characterize their projects, 212 projects (21% of those allocated) self-identified as primarily using their allocation to analyze data from experiments. This workload has a number of distinct characteristics:

- They use a disproportionate amount of storage at NERSC (78% of Project storage in 2019) and have a wide range of I/O patterns, data transfer and data management needs

- They require (near) real-time turn around and flexible and interactive modes of interfacing with HPC systems
- They often use Python and containers in their workflows
- Their workflows start and end outside of the HPC system and NERSC.

The 2019 Berkeley Lab Computing Sciences Area (CSA) Strategic Review recognized that this workload poses unique challenges to both scientists and HPC practitioners and identified the superfacility model as a research priority area for CSA to address these challenges. The superfacility concept links experiment, network, and compute facilities in an ecosystem that brings together algorithmic, software, data management, and hardware expertise to create a new model for conducting experimental science in conjunction with HPC. NERSC's Superfacility Project was established in 2019 in response to this concept as a means to coordinate the various technical and research projects that are making the superfacility model a reality.

The CSA Superfacility Project consists of four areas of work:

- Applications requirements and deployment, including targeted science engagements, Jupyter, policy, scalable code development via NESAP, and user outreach
- Scheduling and middleware, including Spin, federated ID, workflow resiliency, scheduling and an API interface to NERSC
- Automation, including software defined networking, self-driving facilities, and widearea network projects
- Data management, including internal data movement and the data dashboard.

Targeted science engagements are a key element of the CSA Superfacility Project. The project originally selected seven strategic partnerships to drive the team's work, chosen because they use and challenge NERSC in different ways. The project is therefore structured to ensure that our work will be generally useful across multiple science domains, so that rather than designing one-off solutions we are developing a reusable and extensible toolkit. In 2020 we added a Fusion project to our science engagements, a collaboration between PPNL, OLCF, and KSTAR,bringing the total number of superfacility science engagements to eight.

Key Accomplishments in 2020

The Superfacility Project has made significant progress in all areas. Here we list some of the highlights of capabilities that have been developed and/or deployed in support of experimental science at NERSC.

1. Federated ID, which will allow users to authenticate to NERSC using their home institution credentials, is nearing implementation. In 2020 we developed and reviewed the appropriate security policies, implemented the major technical components, and obtained approval for our deployment plan from DOE headquarters, which aligns with

the DOE's Distributed Computing and Data Ecosystem effort. Final integration work is currently under way for a pilot with select users.

2. NERSC's infrastructure for edge services, Spin, had a major upgrade in 2020 with the deployment of Rancher2 and a deployment of large-memory nodes to support a broader range of services, including COVID-19-related research. The new system offers a Kubernetes-based infrastructure with a modern web-based user interface.
3. The API for access into NERSC was developed and reviewed and has been rolled out to early users (see below).
4. A major milestone for software-defined networking was the deployment of an NRE-funded Slurm plugin to allocate bandwidth-balanced compute nodes. This is particularly useful for the NCEM project, which as of 2020, had used this capability to stream 650GB of data in about 150 seconds over eight bridge nodes using 24 compute nodes.
5. Outreach
    - In 2020 NERSC formed the NUG Special Interest Group for Experimental Facility Users, which gives users the opportunity to discuss and compare notes on how they use NERSC. This has proven to be an effective way for users to learn from one another, part of our strategy for scalable user support.
    - In April/May, we held a series of five superfacility demos, showcasing some of our work so far in the project and offering live demonstrations of how to use the tools. These were very well attended, with more than 60 users at each demo.
    - The Superfacility Project maintained a strong presence at the 2020 Supercomputing Conference (SC20), contributing two state-of-the-practice talks, two talks at the XLOOP workshop, two other talks, and a poster.
6. Jupyter supported more than 2,000 unique users in 2020, and work throughout the year included the expansion of notebook access to Cori compute; migration of Jupyter hosting to Rancher2 in Spin; and targeted work to support data management, visualization, and scalable analytics for superfacility science engagements.
7. Globus collaboration endpoints were deployed, which allows users to transfer data in/out of NERSC as a NERSC collaboration user. Over 1.5PB of data was moved in 2020 over these endpoints.
8. A new PI Toolbox was deployed as part of the Data Dashboard, enabling PIs and proxies to change groups and permissions with the click of a button.
9. As part of a jointly funded ASCR/BES grant, the LLAna pilot project is funding research into tools and technologies to support data analysis from LCLS-II. This project includes workflow optimization via profiling to improve the performance of code, data handling, and data management. This directly contributes to the successful first run of the LCLS-II beamline, which has been used this year for Covid-19 research and materials science.

10. Resiliency

- NERSC led a successful application for compute time under ALCC across the three ASCR HPC facilities, with the aim of providing compute time to develop portable workflows. This team is exploring container technologies and data management tools at the ASCR facilities.
- NERSC's Debbie Bard was awarded a LDRD grant from Berkeley Lab to develop resilient workflows for 24/7 computing. This is approaching the problem of cross-facility workflows from the user perspective, researching where the pain points are in deploying resilient workflows at multiple compute resources. A key feature of the superfacility concept is automating workflows.

The Superfacility API

Currently, meeting the supercomputing needs of the experimental science community requires the shared efforts of many people, both in advance and during the running of the experiment, due to the complexity of these workflows that include data analysis and management. Many experiments run with automated analysis pipelines, and these pipelines need a means to communicate with the HPC center without a human getting involved.

In 2020, NERSC developed the Superfacility API, which allows automation of many of the tasks that currently require a human in the loop, such as creating reservations of compute resources, moving data across the different storage systems, and monitoring and managing job submission. As a stable, standards based, and intuitive interface, the API can more readily be integrated into tools and user interfaces developed by our user community, creating a "NERSC inside" experience for users. The Superfacility API is a Python Flask app built on top of the restx library. The library provides the framework and the Swagger-compatible documentation. The Python service and other auxiliary services (such as database and HTTP proxy) are hosted in Spin, which provides resiliency, load balancing and HTTPS access.

Adoption will be simplified further by using standardized tooling (Python/REST ecosystem) that will make it easier for middleware developers to refactor their software or lower the entry barrier for new projects. We have embraced established authentication and security models such as OAuth 2, OpenID Connect and JSON Web Tokens (JWTs), which allows software developers to make use of standard libraries and programming methods and leverages the strengths of open source (broad community adoption and testing) to avoid security risks.

An evolution of its predecessor, NEWT (Cholia et al., 2010), the Superfacility API adds features designed to support complex distributed workflows, such as placing job reservations. It also allows users to offload tedious manual tasks such as data movement via simple "REST" calls. We have designed the API endpoints and operations to cover a wide range of use cases defined by our Superfacility science engagements.

This API also enables important capabilities around workflow resiliency. The API can be used to programmatically query NERSC availability and take a prearranged action, like redirecting the workflow to another location when, for example, NERSC systems are unavailable due to scheduled maintenance.

Python Fan Base Continues to Grow

In 2020, 35% of NERSC users ran compute jobs that used Python on NERSC's Cori supercomputer. How do we know that? NERSC has developed a simple, minimally invasive monitoring framework that enables us to capture key metrics from Python user processes running on Cori. This effort is part of a larger system that collects data across our entire data software stack.

Our approach is inspired by a Python monitoring framework developed for Blue Waters (C. MacLean, Cray User Group, 2017). Both frameworks leverage standard Python features (sitecustomize and atexit) to capture Python imports of interest and other job data. The NERSC collected data are analyzed regularly using GPU-enabled Python data analytics libraries (Dask + cuDF) in a Jupyter notebook, and the results are summarized using an interactive Voila dashboard.

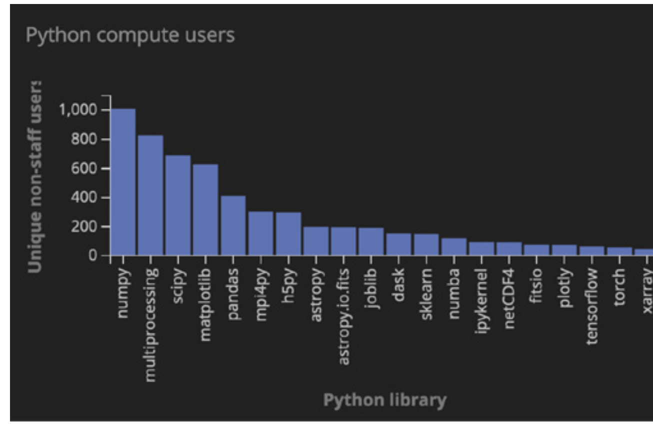


Figure 13: Plot showing Python unique user library data collected over six months on Cori's compute nodes.

The data collected gives us unprecedented visibility into how NERSC users utilize Python. Through these efforts, for example, we have quantified:

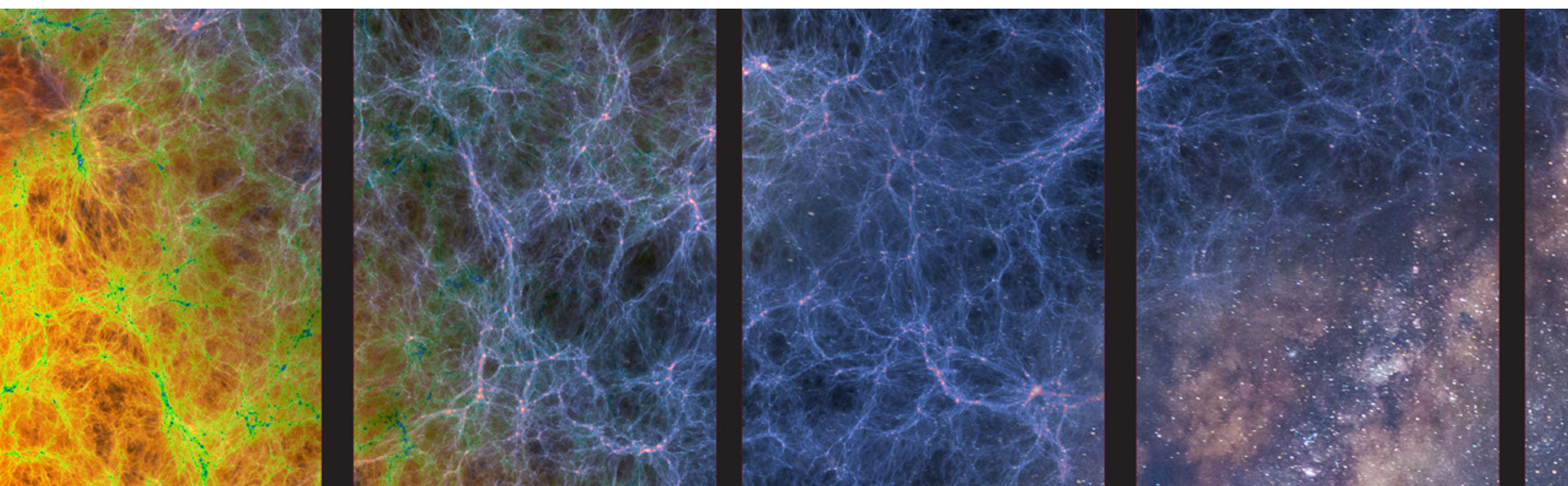
- 80% of our non-staff users are using their own Python environments rather than the Python module we provide
- We measured usage for 2,475 unique non-staff users in 2020
- On average we had 261 Python users per day in 2020
- Our most popular libraries on compute nodes

These kinds of statistics help us apply a data-driven strategy to support our Python users. They also help us advocate on behalf of our users to ASCR, NERSC management, developers, and vendors with the confidence that data provides. Usage patterns can even help influence the design and procurement of future systems and their software environments.

To date all monitoring efforts were on the Cori system, but in 2021 we will deploy this framework on *Perlmutter* and intend to use this data to study and compare Python use on both systems. A question of particular interest is whether user behavior on *Perlmutter* will change as compared to Cori. Python monitoring can tell us:

- Which libraries are used? Libraries can provide a good window into the type of scientific workflow — such as AstroPy (astrophysics and cosmology), Xarray (climate/ grids), pandas (general data science), and TensorFlow (machine learning)
- Which kinds of data do they use? h5py, netcdf4py, and fitsio can provide information
- Are they using the system interactively? We can tell by examining their QOS (interactive, Jupyter)
- Will Python users embrace GPUs? Will deep learning libraries like PyTorch and Tensorflow constitute a larger fraction of the Python workload?
- Are Python users using containers?
- Are Python users able to scale up to large job sizes? Which frameworks do they use to scale up? (mpi4py, multiprocessing, Dask)
- What fraction of Python users are using Jupyter? Data from 2020 show more than 1,200 ipykernel users (a proxy for Jupyter users).

The data collected with this framework has already helped identify that most users prefer to build their own environments (80%), and we will continue to enable and empower users to do this on Cori and *Perlmutter*.



NERSC'S EVOLVING CYBERSECURITY STRATEGY

NERSC's security strategy enables science through a focus on system-level security, in-depth monitoring, intrusion detection, and vulnerability management. NERSC's security group fosters a strong security culture by working closely with the various NERSC technical groups and through regular exchanges with Berkeley Lab security, with whom the center collaborates to monitor emerging threats and assess vulnerabilities.

NERSC provides a variety of services to its more than 8,000 users, and these services have differing risk profiles. NERSC performs frequent vulnerability scans of all resources on its networks and more targeted examination using additional tools of some systems as is warranted by their risk profile. All systems with interactive user access use heuristics to generate alerts in near real time if user activity appears malicious. These rules are updated frequently as new threat signatures emerge.

NERSC deploys service isolation as needed. An example of this is a separate public (anonymous) Science DMZ that keeps those services partitioned off from staff and administrative hosts. The NERSC network is monitored at its border with ESnet, our ISP, and at critical points within our LAN using Zeek, an intrusion detection system. Three distinct Zeek clusters are used to analyze the stream of network and log data. NERSC relies on several techniques to protect itself from Internet attacks and large-scale distributed denial-of-service attempts, keeping the facility an available and productive resource.

For users, NERSC requires multi-factor authentication for nearly all logins to systems and services, providing effective mitigation from using stolen credentials as a means of attack. Some automated workflows are granted longer-lived credentials using a NERSC service called sshproxy. Exceptions for longer-lived credentials (the default is 24 hours) are vetted and approved or denied by NERSC staff. SSH is the primary way users access NERSC, and systems that are accessible from the Internet are required to run instrumented SSH to monitor user activity.

Account activation, deactivation, resource allocation, and password policy enforcement are managed by Iris, the NERSC allocation and account management tool. Through Iris, NERSC can disable user access to all of the center's resources as needed. In addition, Iris maintains summary job and other accounting data that can be used to direct deeper analysis using additional data sources in the event of a security incident involving a NERSC user account.

Science Highlights

NERSC runs highly available, stable, complex, world-class HPC resources with extremely usable development and software environments, allowing its thousands of users to be very scientifically productive. With more than 1,800 refereed publications in 2020, we can only share a sample of NERSC users' science here. The following were chosen to represent the breadth of scientific research and data-focused projects supported by NERSC.

A New Exploration of Higgs Boson Interactions

THE SCIENCE. Using data collected by the ATLAS detector at the Large Hadron Collider (LHC) from 2015 to 2018, a team of Berkeley Lab researchers achieved high-sensitivity analysis of Higgs bosons decaying into pairs of muons. This analysis supports — at the 95% confidence level — that the muons interact with the Higgs boson, and the measured strength of the interaction is in agreement with the Standard Model prediction. A separate analysis, based on data from the LHC's CMS detector, provides a less than 1 in 700 chance that the Higgs boson does not interact with muons. The researchers used NERSC's Cori supercomputer to produce, in just two days, about 10 billion simulated particle events in which a Z boson or photon decayed into two muons. Those processes are major contributors to the background, so modeling them properly enabled researchers to get an accurate estimate of the Higgs boson decay signal they were seeking.

THE IMPACT. The Higgs particle plays a central role in the Standard Model of particle physics, which details the properties of the subatomic particles that make up our universe. Because interactions with the Higgs field are responsible for giving other elementary particles their masses, physicists are very keen to measure how strongly the Higgs boson interacts with other particles.

ADDITIONAL DETAILS. The team evaluated various methods for producing Higgs bosons at the LHC. Machine-learning techniques were used to separate each type of Higgs boson

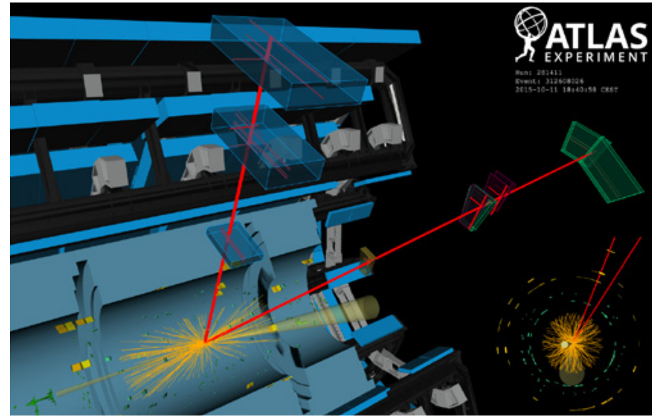


Figure 14. NERSC's Cori supercomputer produced, in just two days, about 10 billion simulated particle events in which a Z boson or photon decayed into two muons.

signature from similar-looking background events to improve the performance of the analysis. There are four production modes that account for Higgs boson production at the LHC. Evaluating these separate pathways for Higgs boson production led to a higher sensitivity to the muon decay signal the researchers were seeking. Ultimately, the team was able to improve sensitivity to this signal by a factor of about 2.5 compared to an earlier effort that used a smaller set of data. About 25% of this increased sensitivity was due to the improved background modeling and advancements in the analysis techniques.

NERSC PI: Zachary Marshall, Lawrence Berkeley National Laboratory

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

PUBLICATION: The Atlas Collaboration, "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector," *Phys. Lett.* B 812 135980, December 16, 2020, DOI: 10.1016/j.physletb.2020.135980

New Constraints on the Size of Neutron Stars

THE SCIENCE. Using the combination of state-of-the-art calculations of matter at extreme densities with gravitational-wave and electromagnetic observations of neutron stars and binary neutron star mergers, scientists have determined — with unprecedented accuracy — that a neutron star with 1.4 times the mass of our Sun is packed into a sphere 11.75 km in radius. The results were reported in the journals Nature Astronomy and Science. The theoretical nuclear-physics research that contributed to this finding relied on NERSC's Cori supercomputer to generate the results.

THE IMPACT. Neutron-star mergers teach us about the nature of the densest matter in the Universe and can now be observed in both gravitational waves and various frequencies of light. These "multi-messenger" observations together with accurate nuclear-physics calculations allow scientists to better understand neutron stars and their properties, reducing the uncertainty of the radius of a 1.4 solar mass neutron-star, which had previously been determined to be 10–14 km. The study also provided an independent measurement of H0, the precise value of which is a subject of intense research.

ADDITIONAL DETAILS. Using Cori, the research team applied "chiral effective field theory" to accurately calculate how nuclear matter behaves under the extreme conditions encountered in a neutron star, a teaspoon of which would weigh about 1 billion tons on Earth. Combining quantum Monte Carlo calculations of the "equation of state," which ultimately determines the neutron star radius, with various observational constraints, the team was able to derive a value for the radius with an accuracy of less than 10%. They used the auxiliary field diffusion Monte Carlo method, which employs statistical means to calculate the properties of dense nuclear matter. This code has very good scaling properties, which makes it ideally suited for Cori.

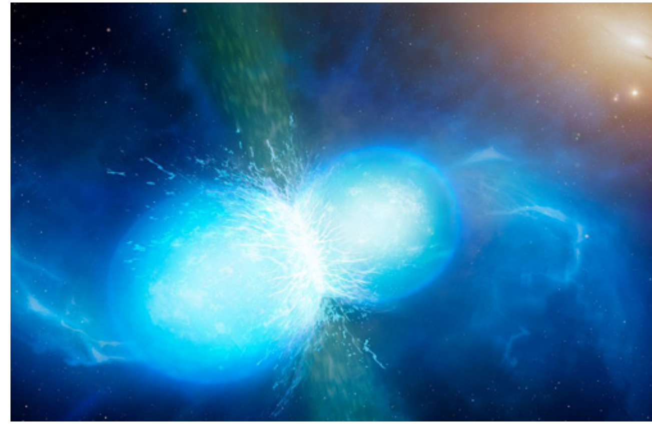


Figure 15. An artist's impression of the collision of two neutron stars. This collision causes gravitational waves, a gamma-ray burst, and a massive explosion. Image courtesy of ESO/University of Warwick/Mark Garlick.

NERSC PI: Ingo Tews, Los Alamos National Laboratory

PROJECT FUNDING AND ALLOCATION AWARD: DOE Office of Science, Office of Nuclear Physics (NP)

PUBLICATIONS: Capano, Collin D., Tews, Ingo, et al., "Stringent constraints on neutron-star radii from multimessenger observations and nuclear theory"; *Nature Astronomy*, 4:625-632; June 2020, 10.1038/s41550-020-1014-6

Dietrich, Tim, et al., "Multimessenger constraints on the neutron-star equation of state and the Hubble constant"; *Science*, 370:1450-1453; 2020 DEC 18, 10.1126/science.abb4317

Near-Real-Time Networked Analysis of Fusion Data

THE SCIENCE. Scientists and engineers have established a near real-time networked analysis of data taken at the KSTAR fusion experiment in South Korea. The framework is using machine learning and artificial intelligence algorithms running at NERSC for analysis of large data streams.

THE IMPACT. Fusion reactors have the potential to provide a huge source of clean and affordable electricity. Experiments are underway around the world to turn this promise into reality. These new capabilities allow United States fusion researchers to have broader and faster access to KSTAR (and future ITER) data, enabling faster analysis and informed steering of the fusion experiments.

ADDITIONAL DETAILS. The end-to-end Python framework DELTA streams data using ADIOS DataMan over WAN (at rates > 4 Gbps), asynchronously processes on multiple workers with MPI & multi-threading. Deep convolutional neural networks work with multi-scale fusion data, such as ECEi, to recognize events of interest. With KSTAR data streaming to NERSC, the time for an ECEi analysis was reduced from 12 hours on single-process to 10 minutes on six Cori nodes.

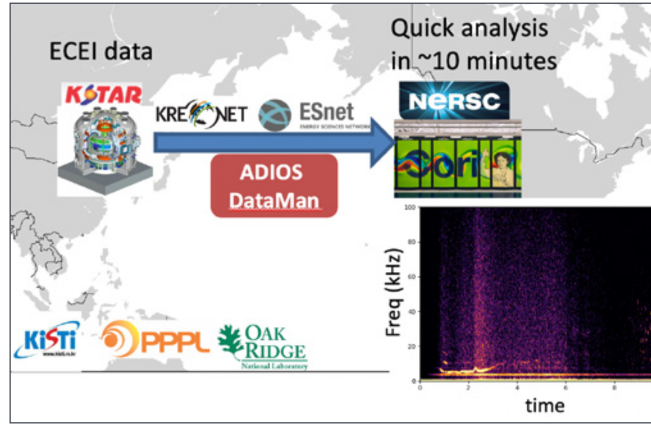


Figure 16. Machine learning algorithms at NERSC enhance analysis of large data streams.

NERSC PI: C.S. Chang, Princeton Plasma Physics Laboratory

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

PUBLICATION: R. Kube, RM Churchill, J. Choi, et al., SciPY 2020 Conference Proceedings, “Leading magnetic fusion energy science into the big-and-fast data lane,” http://doi.org/10.25080/Majora-342d178e-013

How Environmental Effects Can Steer Storms to Cities

THE SCIENCE. Using NERSC supercomputers to model destructive thunderstorms, researchers from Pacific Northwest National Laboratory found that urban landscapes and human-made aerosols — particles suspended in the atmosphere — can make wind gusts stronger, rain heavier, hail larger and even steer storms toward cities.

THE IMPACT. Urbanization has been a significant change in the earth's environment since industrialization and is expected to further expand during the coming decades. Many modeling and observational studies have shown that landscapes can impact weather and climate and it is important to understand how this will affect the lives of millions of Americans.

ADDITIONAL DETAILS. Researchers modeled two large historical storms: a supercell near Kansas City that produced hail, strong wind and a tornado; and a sea-breeze induced thunderstorm near Houston. Using a version of the Weather Research and Forecasting model that includes routines to model aerosols and detailed chemistry and physics, the team found that storms were stronger when urban landscapes and human-produced aerosols — such as from auto exhausts or farming — were included in the simulations. In addition to harnessing the computational power of NERSC's Cori supercomputer, the team also relied on a special allocation of 50 TB of data storage on Cori's high-speed file system.

Figure 17. Intensified convection is just one way urban land and anthropogenically produced aerosols may shape hazardous weather. Longer-lasting rainfall and larger hail are other potential byproducts of the interactions between cityscapes and storms, according to new research from PNNL scientists. (Photo by 12019 | Pixabay.com)

NERSC PI: Jiwen Fan, Pacific Northwest National Laboratory

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

PUBLICATION: Fan, J.; Zhang, Y.; Li, Z.; Hu, J.; Rosenfeld, D., "Urbanization-induced land and aerosol impacts on sea-breeze circulation and convective precipitation"; *Atmospheric Chemistry and Physics,* 20:14163-14182; 2020, 10.5194/acp-20-14163-2020

NERSC & LCLS Team Up on SARS-CoV-2 Research

THE SCIENCE. NERSC and researchers at the SLAC National Accelerator Laboratory connected in real time to allow researchers at the newly upgraded Linac Coherent Light Source (LCLS) to conduct analysis that informed decision-making while experimental runs were in progress. The collaboration allowed scientists to study the SARS-CoV-2 virus that causes COVID-19 in unprecedented detail.

THE IMPACT. The COVID-19 pandemic has sickened and killed many worldwide and caused widespread disruption in lifestyle and the world economy. A detailed understanding of the SARS-CoV-2 virus' structure and lifecycle is critical to developing vaccines and therapeutics.

ADDITIONAL DETAILS: A team led by Hasan DeMirci's group at Koç University used the LCLS to study two crystal forms of the SARS-CoV-2 main protease at near-physiological-temperature, which offers invaluable information for drug-repurposing studies. LCLS/SLAC, NERSC, and LBNL's Computational Research Division, built an optimized pipeline that included improved communication, I/O, seamless portability from LCLS to NERSC using containers, customized job submission for optimized node sharing and memory allocation.

Another team of researchers led by Allen Orville at Diamond Light Source (UK) studied the atomic structure, dynamics, and function of the main protease and a papain-like protease at room temperature, which scientists hope could lead to development of an anti-viral treatment. The collaboration used LCLS to determine the time-resolved atomic structure from a slurry of microcrystals to which the drugs are added. The ability to process data in near to real time with access to resources at NERSC was essential for the team's decision-making processes during the beamtime.

Figure 18. Researchers working at the Linac Coherent Light Source used X-ray crystallography to capture detailed images of the structure of the SARS-CoV-2 virus.

NERSC PI: Amedeo Perazzo, SLAC

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

PUBLICATION: https://www.nersc.gov/news-publications/nersc-news/science-news/2021/superfacility-model-brings-covid-research-into-real-time/

Regional Simulations of Building Response to Large Earthquakes

THE SCIENCE. Using the EQSIM fault-to-structure computational framework running on NERSC's Cori supercomputer, a research team from the University of Nevada-Reno (UNR), Berkeley Lab, UC Berkeley, and Livermore Lab are providing new insights into how buildings respond to large earthquakes over large regions. The simulations provide an understanding into how earthquakes stress and distort buildings depending on their location, size, and shape.

THE IMPACT. The risk to buildings and other structures over geographical areas due to major earthquakes is poorly understood. The effect on buildings depends on many complex factors, including where and how a fault ruptures, how the waves move through the Earth, and how the soil underneath the building interacts with the structure. This team's simulation framework, leveraging the computational capabilities of Cori, allows researchers to examine many scenarios of interest and inform civic planners so they can save lives and protect infrastructure.

ADDITIONAL DETAILS. The team, led by David McCallen from UNR and Berkeley Lab, simulated a magnitude 7.0 earthquake and looked in detail at the response of buildings of various sizes throughout a region. Using an allocation of computer time from the NERSC Director's Discretionary Reserve, the team ran the EQSIM framework — which is being developed with funding from the DOE Exascale Computing Project — on Cori using 2,048 nodes (139,000 processor cores) in the study. The project used 65 million hours of compute time 2020.

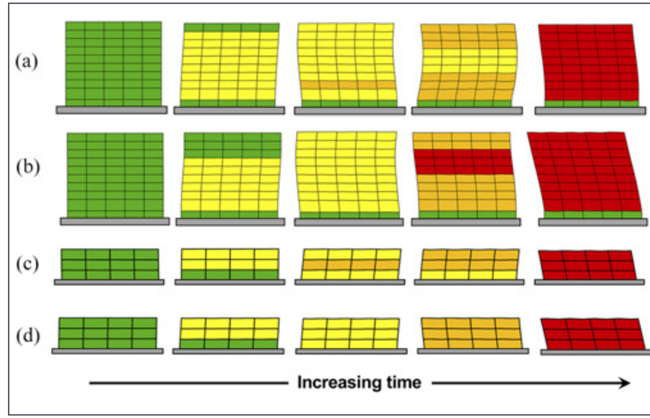


Figure 19. Effects on steel-frame buildings of various heights 2 km off fault line due to a 7.0 earthquake. Red indicates the most distortion and green the least.

NERSC PI: David McCallen, University of Nevada-Reno and Berkeley Lab

PROJECT FUNDING AND ALLOCATION AWARD: NERSC Director's Reserve, Research Funded by the Exascale Computing Project

PUBLICATION: McCallen, D.; Petrone, F.; Miah, M.; Pitarka, A.; Rodgers, A.; Abrahamson, N., "EQSIM — A multidisciplinary framework for fault-to-structure earthquake simulations on exascale computers, part II: Regional simulations of building response"; *Earthquake Spectra 2020*, DOI: 10.1177/8755293020970980

Machine Learning Tool Mines Scientific Literature for COVID-19 Insights

THE SCIENCE. A team of materials scientists at Berkeley Lab built a text-mining tool that is helping the global scientific community synthesize the mountain of scientific literature on COVID-19 being generated every day, with all of the project's infrastructure running at NERSC on the Cori supercomputer. COVIDScholar uses natural language processing techniques to quickly scan and search tens of thousands of research papers and help draw insights and connections that may otherwise not be apparent. It is doing this by text mining COVID-19 publications for new scientific insights using unsupervised textual analysis, as described in the 2019 *Nature* paper, "Unsupervised word embeddings capture latent knowledge from materials science literature."

THE IMPACT. COVIDScholar was developed in response to a March 16, 2020 call to action from the White House Office of Science and Technology Policy that asked artificial intelligence experts to develop new data- and text-mining techniques to help find answers to key questions about COVID-19. Starting in March 2020, the Berkeley Lab team got a prototype of COVIDScholar up and running in about a week. Assisted by quick access to NERSC's Cori supercomputer, the center's "Spin" internal cloud for edge services, and expert staff assistance from NERSC staff, a prototype was created in just 7 days. State-of-the-art natural language processing models run daily on Cori, and the search portal is running on the NERSC "Spin" edge services cluster.

ADDITIONAL DETAILS. On Google and other search engines, users search for what they think is relevant; COVIDScholar's objective is to do information extraction so that researchers can find nonobvious information and relationships. As a result, it is now the largest single-topic literature collection on COVID-19. All of the project's infrastructure is running at NERSC on the Cori supercomputer and related supporting resources, including the Spin cluster. COVIDScholar is available at covidscholar.org.

NERSC PIs: John Dagdelen, Kristin Persson, Gerbrand Ceder, Amalie Trewartha, Lawrence Berkeley National Laboratory

PROJECT FUNDING AND ALLOCATION AWARD: Lawrence Berkeley National Laboratory, Laboratory Directed Research and Development; C3.ai Digital Transformation Institute; National Virtual Biotechnology Laboratory; U.S. Department of Energy Office of Science with funding provided by the Coronavirus CARES Act.

PUBLICATION: Trewartha, A., Dagdelen, J., Huo, H., Cruse, K., Wang, Z, He, T., Subramanian, A., Fei, Y., Justus, B., Persson, K., & Ceder, G., 2021, "COVIDScholar: An automated COVID-19 research aggregation and analysis platform," *Journal of Artificial Intelligence Research*, submitted, https://arxiv.org/abs/2012.03891

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
Kathy Kincade

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

Cover Image
Ke-Jung (Ken) Chen
Academia Sinica Institute of Astronomy and Astrophysics (ASIAA)

21-NERSC-28028