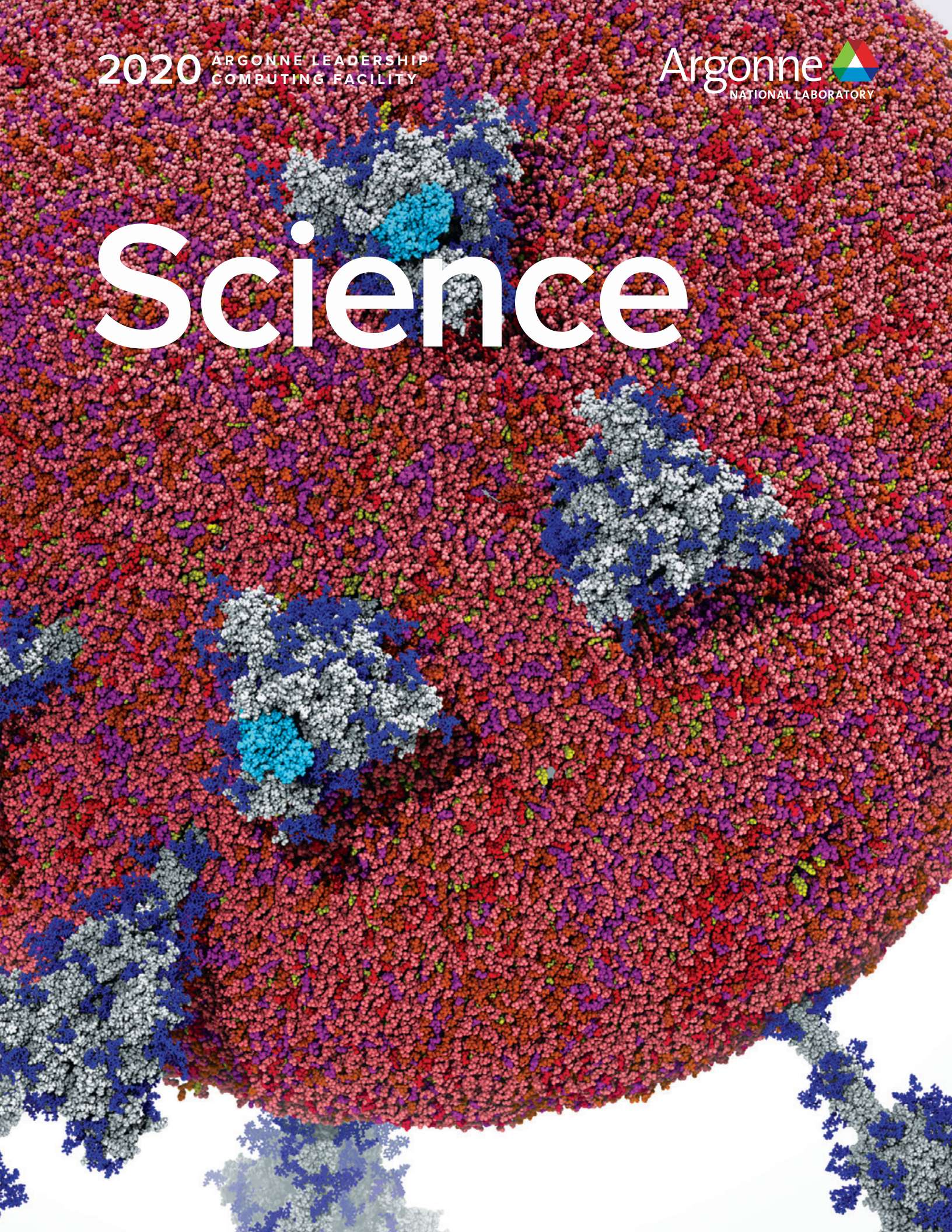
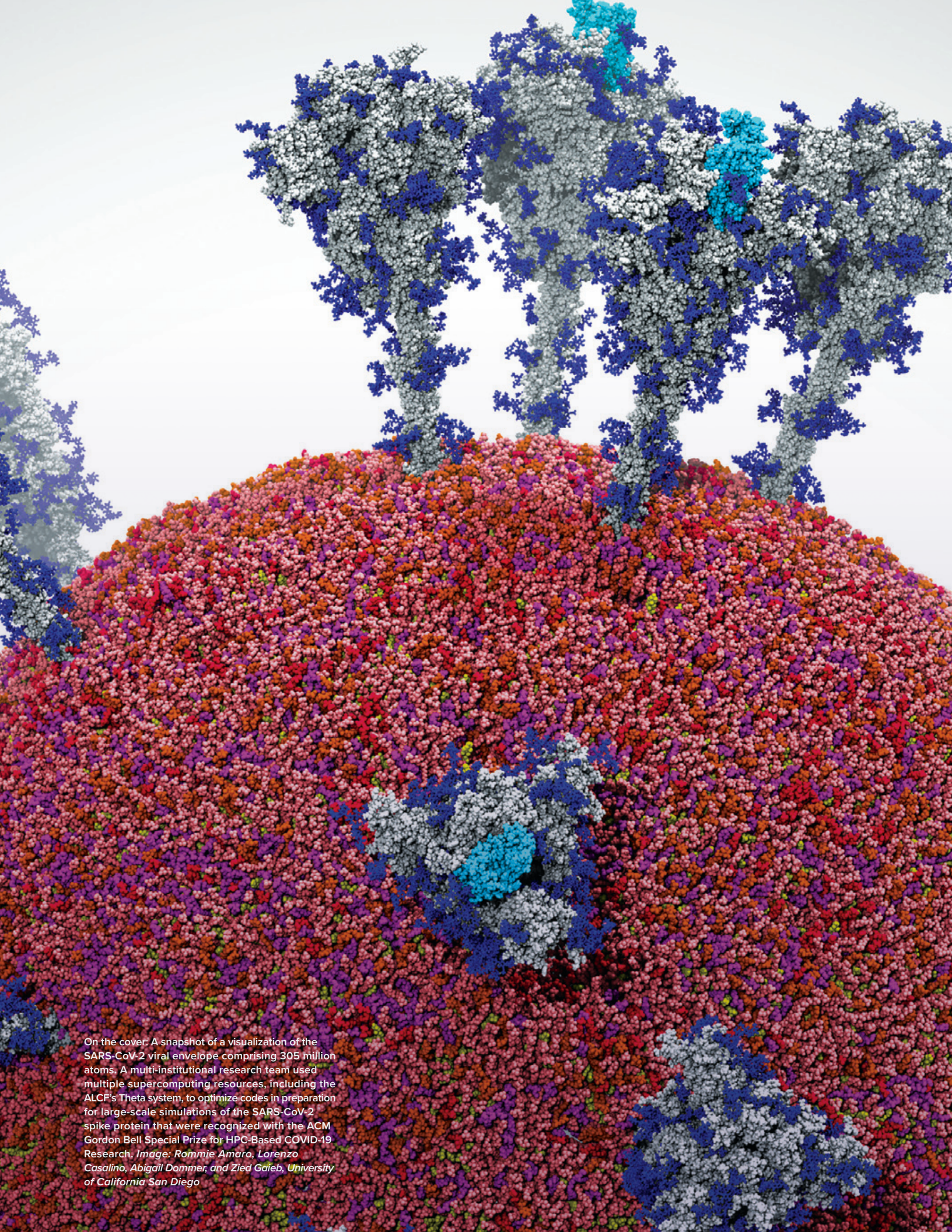
2020 ARGONNE LEADERSHIP COMPUTING FACILITY

Argonne NATIONAL LABORATORY

Science





On the cover: A snapshot of a visualization of the SARS-CoV-2 viral envelope comprising 305 million atoms. A multi-institutional research team used multiple supercomputing resources, including the ALCF's Theta system, to optimize codes in preparation for large-scale simulations of the SARS-CoV-2 spike protein that were recognized with the ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research. *Image: Rommie Amaro, Lorenzo Casalino, Abigail Dommer, and Zied Gaieb, University of California San Diego*

2020 SCIENCE CONTENTS

MICHAEL E. PAPKA
ALCF Director

JINI RAMPRAKASH
ALCF Deputy Director

KATHERINE RILEY
ALCF Director of Science

KALYAN KUMARAN
ALCF Director of Technology

MARK FAHEY
ALCF Director of Operations

SUSAN COGHLAN
ALCF-X Project Director

MESSAGE FROM ALCF LEADERSHIP

In many ways, 2020 was an extraordinary moment for science. Scientists worldwide rapidly responded to the COVID-19 crisis, aided greatly by today's technology and the pace of information sharing, all the while continuing to make progress in many fields beyond infectious diseases. Advanced computational user facilities, such as the ALCF, and major experimental research facilities, such as Argonne's Advanced Photon Source (APS), play a vital role in the nation's scientific response to crises while continuing to serve as the foundation of scientific infrastructure.

When APS users needed on-demand analysis of experimental data on the crystal structure of COVID-19 proteins, ALCF staff put a computational pipeline in place to do that. To apportion valuable computing cycles for COVID-19 research, the ALCF dedicated 10 percent of Theta to focus on supporting COVID-19 efforts. The ALCF then expanded Theta with graphics processing units (GPUs) to bring additional computational power to addressing the pandemic. (Theta's expansion, called ThetaGPU, was made possible with support from the Coronavirus Aid, Relief, and Economic Security (CARES) Act). The ALCF also joined a national effort as part of the COVID-19 High Performance Computing (HPC) Consortium to coordinate a unified computational response.

All the while, the ALCF, the APS, as well as many other facilities across the globe, have continued to produce quality and impactful science. ALCF users have always carried out their computational studies via remote access to our facility, so their scientific progress kept moving forward despite the pandemic's impact on our day-to-day lives.

Some of the most noteworthy research carried out at the ALCF this year were projects that aimed to shed light on the coronavirus, including CityCOVID, an agent-based epidemiological model used to support decision makers in COVID-19 crisis response and planning, and a finalist for the ACM Gordon Bell Special Prize for High Performance Computing-Based COVID-19 Research; and the multi-institutional collaboration that studied the complex dynamics of the spike protein, a key protein in the SARS-CoV-2 virus, an effort that won the Special Prize.

There were also many other achievements, unrelated to COVID-19—which you will read about in the pages that follow. The Last Journey, a six-month, massive simulation of the cosmos—one of the five largest cosmological simulations ever—provided data that will form the basis for sky maps used by numerous surveys. This project both capped the service life of one ALCF supercomputer and prepared the science team to run even more extensive cosmological simulations on future machines, like the ALCF's Aurora exascale system. Researchers also tapped Theta for studies ranging from modeling supernova explosions to creating automated databases for materials discovery.

We continued to prepare for Aurora. Our collaborative work through the Early Science Program (ESP) and DOE's Exascale Computing Project and available testing hardware in the Joint Laboratory for System Evaluation has helped numerous teams advance code development efforts for the upcoming system. We also hosted several workshops, webinars, and hackathons to provide details and guidance on using various aspects of the Aurora hardware and software environment.

The facility itself has been undergoing a major upgrade to make way for Aurora, which is transforming not only the machine room, but also building out new meeting and training spaces at a steady clip. Yet, even amid all the construction and regular support that ALCF staff provide our users every day, the facility was able to open up the connections that allowed the COVID work.

We are grateful to our staff's extra efforts in 2020 to keep the science flowing and to DOE's Office of Science and Argonne leadership for their ongoing support of our efforts to build out the advanced systems that will drive the next generation of discoveries. Mostly, we are grateful for your interest in our activities, for it is the scientific work done at user facilities like ours that will have immense impact on people in our nation and around the globe for decades to come.

ARGONNE LEADERSHIP COMPUTING FACILITY

The ALCF enables breakthroughs in science and engineering by providing supercomputing resources and expertise to the research community.

In 2020, the ALCF hosted multiple training events, including a hands-on workshop in February, to help researchers advance code development efforts for its upcoming exascale system, Aurora.

About ALCF

The ALCF's Theta system is designed to support research at the nexus of simulation, data analytics, and artificial intelligence.

The Argonne Leadership Computing Facility, a U.S. Department of Energy (DOE) Office of Science User Facility located at Argonne National Laboratory, enables breakthroughs in science and engineering by providing supercomputing resources and expertise to the research community.

ALCF computing resources—available to researchers from academia, industry, and government agencies—support large-scale computing projects aimed at solving some of the world's most complex and challenging scientific problems. Through awards of supercomputing time and support services, the ALCF enables its users to accelerate the pace of discovery and innovation across disciplines.

Supported by the DOE's Advanced Scientific Computing Research (ASCR) program, the ALCF and its partner organization, the Oak Ridge Leadership Computing Facility, operate leadership-class supercomputers that are orders of magnitude more powerful than the systems typically used for open scientific research.

ALCF Team

The ALCF's talented and diverse staff make the facility one of the world's premier centers for scientific computing.

Science
Computational scientists with multidisciplinary domain expertise work directly with ALCF users to maximize and accelerate their research efforts. In addition, the ALCF team applies broad expertise in data science, machine learning, data visualization and analysis, and mathematics to help application teams leverage ALCF resources to pursue data-driven discoveries.

With a deep knowledge of the ALCF computing environment and experience with a wide range of numerical methods, programming models, and computational approaches, staff scientists and performance engineers help researchers optimize the performance and productivity of simulation, data, and learning applications on ALCF systems.

Operations
The ALCF's HPC systems administrators manage and support all ALCF computing systems, ensuring users have stable, secure, and highly available resources to pursue their scientific goals. This includes the ALCF's production supercomputers, supporting system environments, storage systems, and network infrastructure. The team's software developers create tools to support the ALCF computing environment, including software for user account and project management, job failure analysis, and job scheduling.

User support specialists provide technical assitance to ALCF users and manage the workflows for user accounts and projects. In the business intelligence space, staff data architects assimilate and verify ALCF data to ensure accurate reporting of facility information.

Technology
The ALCF team plays a key role in designing and validating the facility's next-generation supercomputers. By collaborating with compute vendors and the performance tools community, staff members ensure the requisite programming models, tools, debuggers, and libraries are available on ALCF platforms. The team also helps manage Argonne's Joint Laboratory for System Evaluation, which houses next-generation testbeds that enable researchers to explore and prepare for emerging computing technologies.

ALCF computer scientists, performance engineers, and software engineers develop and optimize new tools and capabilities to facilitate science on the facility's current and future computing resources. This includes the deployment of scalable machine learning frameworks, in-situ visualization and analysis capabilities, data management services, workflow packages, and container technologies. In addition, the ALCF team is actively involved in programming language standardization efforts and contributes to cross-platform libraries to further enable the portability of HPC applications.

Outreach
ALCF staff members organize and participate in training events that prepare researchers for efficient use of leadership computing systems. They also participate in a wide variety of educational activities aimed at cultivating a diverse and skilled HPC community for the future. In addition, staff outreach efforts include facilitating partnerships with industry and communicating the impactful research enabled by ALCF resources to external audiences.

ALCF Systems

Theta

KNL nodes

- Intel-Cray XC40 architecture
- 11.69 petaflops
- 64-core, 1.3-GHz Intel Xeon Phi 7230 processor per node
- 4,392 nodes
- 281,088 cores
- 843 TB of memory
- 70 TB of high-bandwidth memory
- Aries interconnect with Dragonfly configuration
- 24 racks

GPU nodes

- NVIDIA DGX A100 architecture
- 3.9 petaflops
- AMD EPYC 7742
- 24 nodes
- 24 TB of DDR4 memory
- 7, 680 GB of GPU memory
- 7 racks

Iota

Test and development platform

- Intel-Cray XC40 architecture
- 117 teraflops
- 64-core, 1.3-GHz Intel Xeon Phi 7230 processor per node
- 44 nodes
- 2,816 cores
- 12.3 TB of memory
- 1 TB of high-bandwidth memory
- Aries interconnect with Dragonfly configuration
- 1 rack

Cooley

Data analysis and visualization cluster

- Intel Haswell architecture
- 293 teraflops
- Two 6-core, 2.4-GHz Intel E5–2620 processors per node
- 1 NVIDIA Tesla K80 GPU per node
- 126 nodes
- 1,512 cores
- 47 TB of memory
- 3 TB of GPU memory
- FDR InfiniBand interconnect
- 6 racks

DATA STORAGE SYSTEMS

Eagle

Storage system for data sharing

- HPE ClusterStor E1000
- 100 petabytes of usable capacity
- 8,480 disk drives
- Lustre filesystem
  - 160 Object Storage Targets
  - 40 Metadata Targets
- HDR Infiniband network
- Sustained 650 GB/s rate on data transfers

Grand

Storage system for project science campaigns

- HPE ClusterStor E1000
- 100 petabytes of usable capacity
- 8,480 disk drives
- Lustre filesystem
  - 160 Object Storage Targets
  - 40 Metadata Targets
- HDR Infiniband network
- Sustained 650 GB/s rate on data transfers

Data Storage

At the ALCF, disk storage provides intermediate-term storage for researchers, offering a means to access, analyze, and share computational and experimental data. Tape storage is used to archive data from completed projects.

DISK STORAGE

The Theta storage system consists of 30 I/O nodes that connect to a storage array that controls 2,300 disk drives with a total useable capacity of 9 PB and a maximum aggregate transfer speed of 240 GB/s. Theta uses Lustre to access this storage. The ALCF also utilizes a 10 PB file system based on an IBM Elastic Storage Server (ESS) to host data for science running on the Theta and Cooley systems. The ESS system is a software defined storage system based on IBM's GPFS file system and consists of 60 I/O nodes controlling 7,260 disk drives.

The ALCF supports two 100 PB globally accessible Lustre filesystems named Grand and Eagle. Each storage array controls 8,480 disk drives with a sustained transfer speed of 650 GB/s in the current environment. Eagle storage allocations will be granted to projects focused primarily on data sharing and will be accessible by non-ALCF users via Globus. Grand storage will be provided with standard compute allocations and will only be accessible to ALCF users who are project members.

A new home filesystem, homefs, will reside on an all-flash storage array with a capacity of 220 TB and a transfer rate up to 24 GB/s. A new filesystem named Swift will reside on an all flash storage array with a capacity of 123 TB and transfer rate up to 48 GB/s. It is intended to be targeted by GPU-based workflows.

TAPE STORAGE

The ALCF has three 10,000-slot libraries. The tape technology is currently undergoing an upgrade to replace LTO-6 tape drives with LTO-8 tape drives. The upgrade should ultimately provide up to 300 PB of effective storage (approximately five times the amount provided by the LTO-6 tapes).

Networking

Theta has an internal proprietary network for communicating between nodes. InfiniBand enables communication between the I/O nodes and the storage system. Ethernet is used for external user access, and for maintenance and management of the systems.

The ALCF connects to other research institutions using up to 100 Gb/s of network connectivity. Scientists can transfer datasets to and from other institutions over fast research networks, such as ESnet and Internet2.

Testbeds

Argonne's Joint Laboratory for System Evaluation (JLSE) enables researchers to assess and improve next-generation computing platforms of interest to DOE. Established by Argonne's computing divisions and run by the ALCF, JLSE centralizes the laboratory's research activities aimed at evaluating future extreme-scale computing systems, technologies, and capabilities. JLSE users leverage existing infrastructure and next-generation hardware and software to explore low-level experimental computer and computational science, including operating systems, messaging, compilers, benchmarking, power measurements, I/O, and new file systems. Some notable JLSE testbeds include:

Arcticus, Yarrow: Intel discrete GPU testbeds for ECP and ESP projects to develop, optimize, and scale applications and software for Aurora

Atos Quantum Learning Machine: Platform for testing and developing quantum algorithms and applications

Aurora Software Development Kit: Frequently updated version of the publicly available Intel oneAPI toolkit for Aurora development

HPE Comanche Prototype: ARM64 platform for exploring the Marvell Arm architecture

Intel Cooper Lake: Intel Xeon cluster for testing data types for AI and learning applications

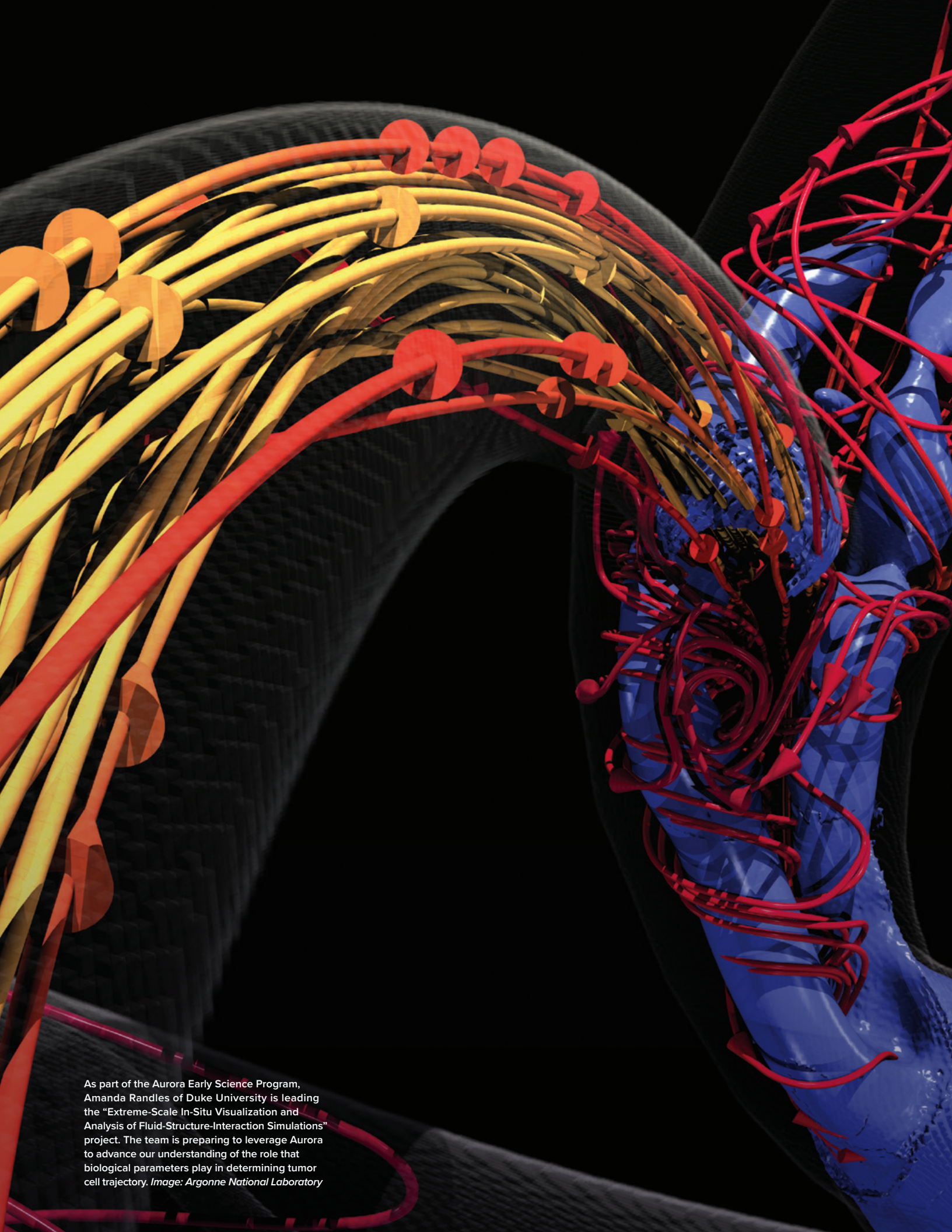
Iris: Intel integrated Gen9 GPUs for ECP and ESP projects to develop, optimize, and scale applications and software for Aurora

NVIDIA GPUs: Clusters of P100, V100, and A100 GPUs for preparing applications for heterogenous computing architectures

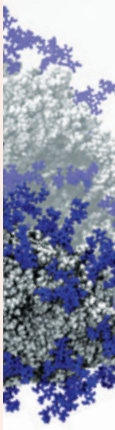
Presque: Intel DAOS nodes for testing the Aurora storage system

ADVANCING SCIENCE WITH HPC

The ALCF's high-performance computing (HPC) resources and expertise help scientists tackle some of the world's most complex and challenging scientific problems.



As part of the Aurora Early Science Program, Amanda Randles of Duke University is leading the "Extreme-Scale In-Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations" project. The team is preparing to leverage Aurora to advance our understanding of the role that biological parameters play in determining tumor cell trajectory. *Image: Argonne National Laboratory*



ALCF Resources Contribute to Fight Against COVID-19

Argonne's HPC systems and expertise are accelerating research aimed at developing treatments and strategies to combat the COVID-19 pandemic.

When the World Health Organization declared coronavirus disease 2019 (COVID-19) a pandemic in mid-March 2020, the U.S. Department of Energy (DOE)—aiming for a rapidity proportionate to the scale of the situation and with the support of the Coronavirus Aid, Relief, and Economic Security (CARES) Act passed by Congress—launched a series of efforts to bolster the global research community's ability to advance our understanding of the virus.

At Argonne, fast delivery of the ALCF's Theta upgrade was made an imperative. Upon deployment, the system—built from a core comprising 24 NVIDIA DGX A100 nodes so as to combine graphics processing unit (GPU) and central processing unit (CPU) capabilities—was immediately leveraged in the fight against the coronavirus. The upgraded Theta system enables accelerated training and learning of artificial intelligence (AI) datasets through integration with data analytics, alongside GPU-specific and -enhanced HPC applications for modeling and simulation.

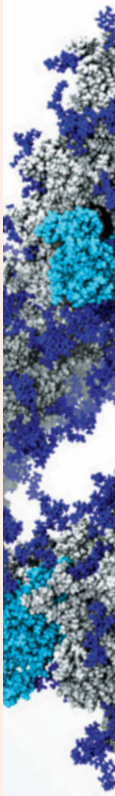
Furthermore, the global research community undertook an immense collaboration seeking to develop and evaluate vaccines and therapies, and to predict possible trajectories of future outbreaks—to understand, treat, and control COVID-19.

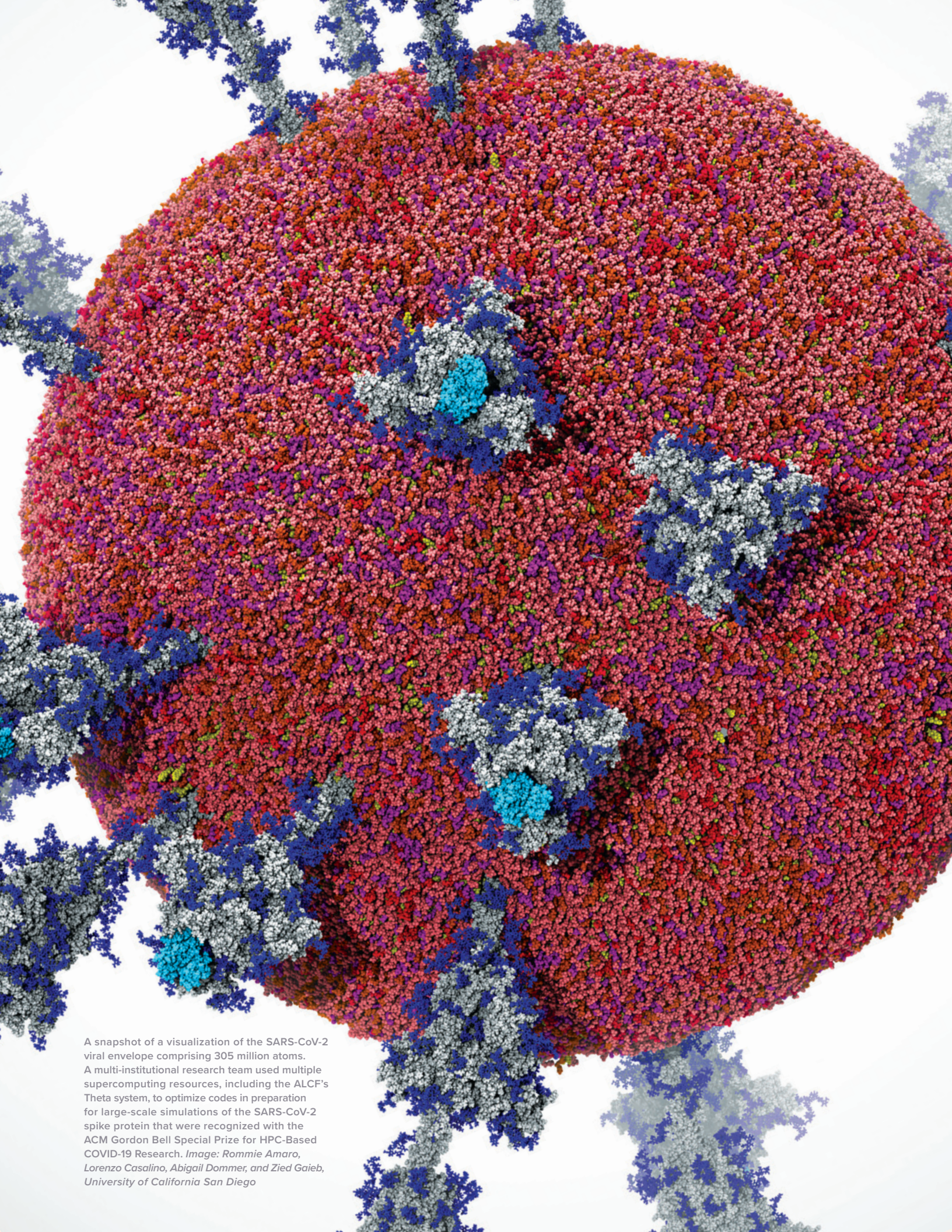
Led by DOE, the White House Office of Science and Technology Policy, and IBM, a unique private-public effort was established so as to bring together federal government, industry, and academic leaders to make powerful computational resources available for COVID-19 research. The goal of the COVID-19 High Performance Computing (HPC) Consortium was to help accelerate the pace of scientific discoveries in COVID-19 research projects. The consortium brings together more than 30 partners from national labs, federal agencies, technology companies, and academic institutions to provide resources ranging from small clusters and cloud and web services to high-end supercomputing resources.

Among those resources is the ALCF's Theta supercomputer, which has been crucial in enabling multiple branches of research. These include the development of epidemiological models to simulate the spread of COVID-19 throughout the population, analysis of the crystal structure of a protein structure associated with the virus, and identification of the virus's biological mechanisms and drug treatments.

Epidemiological simulation

Argonne researchers Jonathan Ozik and Charles Macal are leading the laboratory's epidemiological modeling efforts. Their team has developed CityCOVID, an agent-based model capable of tracking detailed COVID-19 transmission. Agent-based modeling is an approach for capturing the dynamics of heterogeneous, interacting, adaptive agents at an individual, granular level of detail. When applied to a city like Chicago, CityCOVID includes a synthetic population representing the 2.7 million residents of Chicago and the 1.2 million geo-located places where they can co-locate.





A snapshot of a visualization of the SARS-CoV-2 viral envelope comprising 305 million atoms. A multi-institutional research team used multiple supercomputing resources, including the ALCF's Theta system, to optimize codes in preparation for large-scale simulations of the SARS-CoV-2 spike protein that were recognized with the ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research. *Image: Rommie Amaro, Lorenzo Casalino, Abigail Dommer, and Zied Gaieb, University of California San Diego*

Throughout a simulated day, each individual, or agent, moves from place-to-place, hour-by-hour, engaging in social activities and interactions with co-located agents, where COVID-19 exposure events can occur. The COVID-19 disease progression is modeled within each individual, including differing symptom severities, hospitalizations, and age-dependent probabilities of transitions between disease stages. Running on ALCF computing resources, CityCOVID is being used to calibrate unobserved model parameters, such as the time-varying degree of individual self-protective behaviors across the population, and to simulate a variety of interventions and future scenarios. While the Argonne team has been using Chicago as a testbed for developing these capabilities, CityCOVID is being extended to other regions as well.

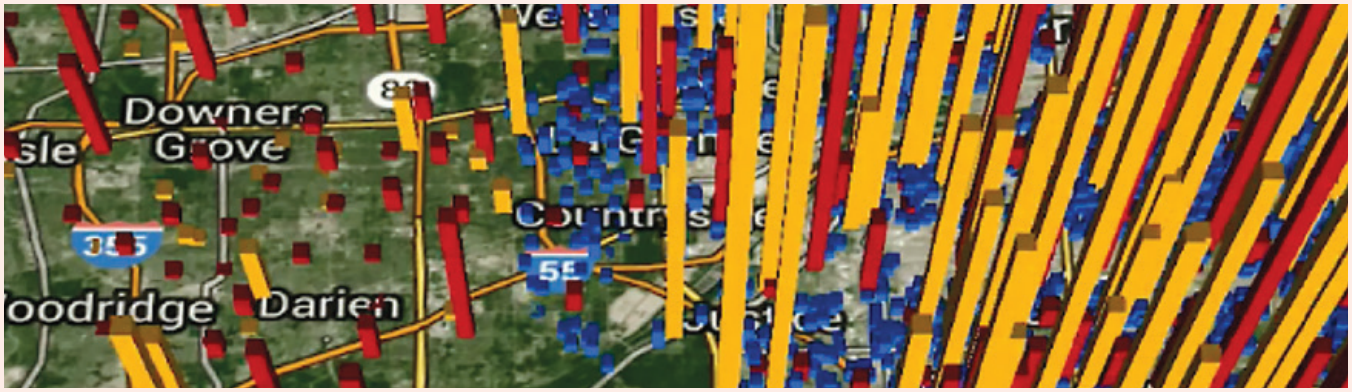
Analysis of virus structure

Scientists from Argonne's Structural Biology Center at the Advanced Photon Source (APS), including Argonne Distinguished Fellow Andrzej Joachimiak, and working with Mateusz Wilamowski of the Center for Structural Genomics of Infectious Diseases (CSGID), performed experiments on the Nsp10+Nsp16 protein complex of SARS-CoV-2 created at CSGID. The experiment provided the first low-dose, room-temperature insight into the structure of this protein complex. Its results will give the community a greater understanding of the complex than is possible with traditional crystallography techniques. The experiment is designed to determine the metal activation of the complex and later lead to time-resolved experiments, which will be the first dynamic structural experiment of a protein related to SARS-CoV-2. Time-resolved structural dynamics will help elucidate the electrochemistry of this protein function and give insights into the virus. To support the rapid processing requirements, a team led by Argonne computer scientist Ryan Chard and APS beamline scientist Darren Sherrell deployed an automated data acquisition, analysis, curation, and visualization pipeline, leveraging the Theta supercomputer for high-speed, on-demand analysis. The pipeline reactively analyzes data as it is collected, moving images of the sample from the APS to the ALCF, where they are rapidly analyzed and visualized. The same automated pipeline then moves results to a repository and extracted metadata for publication in a data portal, which scientists can monitor during an experiment.

Discovery of drug treatments

An Argonne-based team is seeking to address both the fundamental biological mechanisms of the virus and the disease, while simultaneously targeting the entire viral proteome to identify potential therapeutics. Argonne computational biologist Arvind Ramanathan and Argonne Associate Laboratory Director Rick Stevens are leveraging Argonne and Oak Ridge leadership supercomputers to design novel therapeutics against SARS-CoV-2 using AI approaches that integrate information from experimental observations, and rigorous, physics-based virtual screening and molecular simulations, to identify viable drugs that can inhibit viral proteins. These AI approaches, based on advances in deep learning and reinforcement learning, are capable of predicting how strongly a small molecule will bind to a protein as well as exploring the structural space of compounds that are predicted to bind to find more suitable variants.

The database of potential drug candidates for COVID-19 is immense, including millions to billions of potential compounds. So far, computational screening of molecules has resulted in identifying small molecules that can potentially inhibit viral function in wet-lab experiments. These experiments involve live human lung cell cultures being exposed to small molecules followed by measurements that



Occupancy patterns in the Chicago area as generated using Argonne's CityCOVID model. *Image: Charles Macal and Jonathan Ozik, Argonne National Laboratory*

monitor viral replication. The molecules are being further refined to optimize them for binding to specific viral target proteins. Using AI techniques, the team has screened over 6 million small molecules and are validating them at Argonne for activity against the virus and rapidly expanding it to screen billions of compounds. The potential impact of this work is the design of new generative models based on reinforcement learning for both small molecules and antibodies; and development of large-scale, AI-driven simulations of the entire viral particle and drugs bound to the various viral targets, as a better pathway to an antiviral drug.

Repurposing existing drugs

Led by Johns Hopkins School of Medicine biophysicist Albert Lau, a team of scientists is using Theta to identify existing drugs that can be repurposed to disrupt the function and viral viability of SARS-CoV-2. Concomitant with this is the identification of intracellular catalytic SARS-CoV-2 protein targets. The approach of inhibiting catalytic viral enzymes is established as a successful therapeutic strategy, with viral systems such as HIV and hepatitis C providing strong precedent.

The researchers begin their workflow by screening, with Theta, a library of more than 10,000 FDA-approved compounds, as well as intracellular catalytic SARS-CoV-2 protein targets, to reduce their number to an experimentally manageable size. After high-potential candidates are identified, a select group of target proteins are expressed and purified in a lab, and activity assays are developed. Finally, prioritized compounds from the Johns Hopkins Drug Library are screened against the ten purified target proteins to distinguish those that exhibit viral activity.

Beyond docking compounds to small-molecule binding sites, the researchers are using atomistic molecular dynamics simulations to search for hidden (“cryptic”) binding sites in their targets. Cryptic pockets open as a result of protein dynamics and may be related to protein function; having been shown to be druggable, they present opportunities to target proteins and protein complexes resistant to standard drug-discovery strategies. To optimize the analysis of cryptic pockets, the team developed their own software package, TACTICS (Trajectory-based Analysis of Conformations to Identify Cryptic Sites), which predicts via machine learning small-molecule binding sites in proteins and protein complexes. This workflow has identified at least two previously unknown cryptic sites.

Edge Services Propel Data-Driven Science

The ALCF is working to provide tools and technologies that enable researchers to share and analyze increasingly large scientific datasets.

Edge computing runs counter to the notion of colocating data with compute; instead it moves the compute to the data. This is only possible today because of advances in computing capability and the integration of AI. At the same time, within the traditional HPC computing facility, efforts are underway to expose new services at the edge. The development of edge services—driven by demands of data curation, exposure, and management, as well as the tools that enable and shape such practices—has allowed the ALCF to continue its evolution beyond its traditional role as a place for the simple batch submission and static datasets, and has expanded the possibilities for exploring the meaning and usability of data.

The notion of an edge service, from the ALCF perspective, is one that blurs the boundaries separating an experimental facility from a computing facility: in the context of interpreting output data with high-performance computing, it is no longer so easy to say where the laboratory ends and where analysis begins.

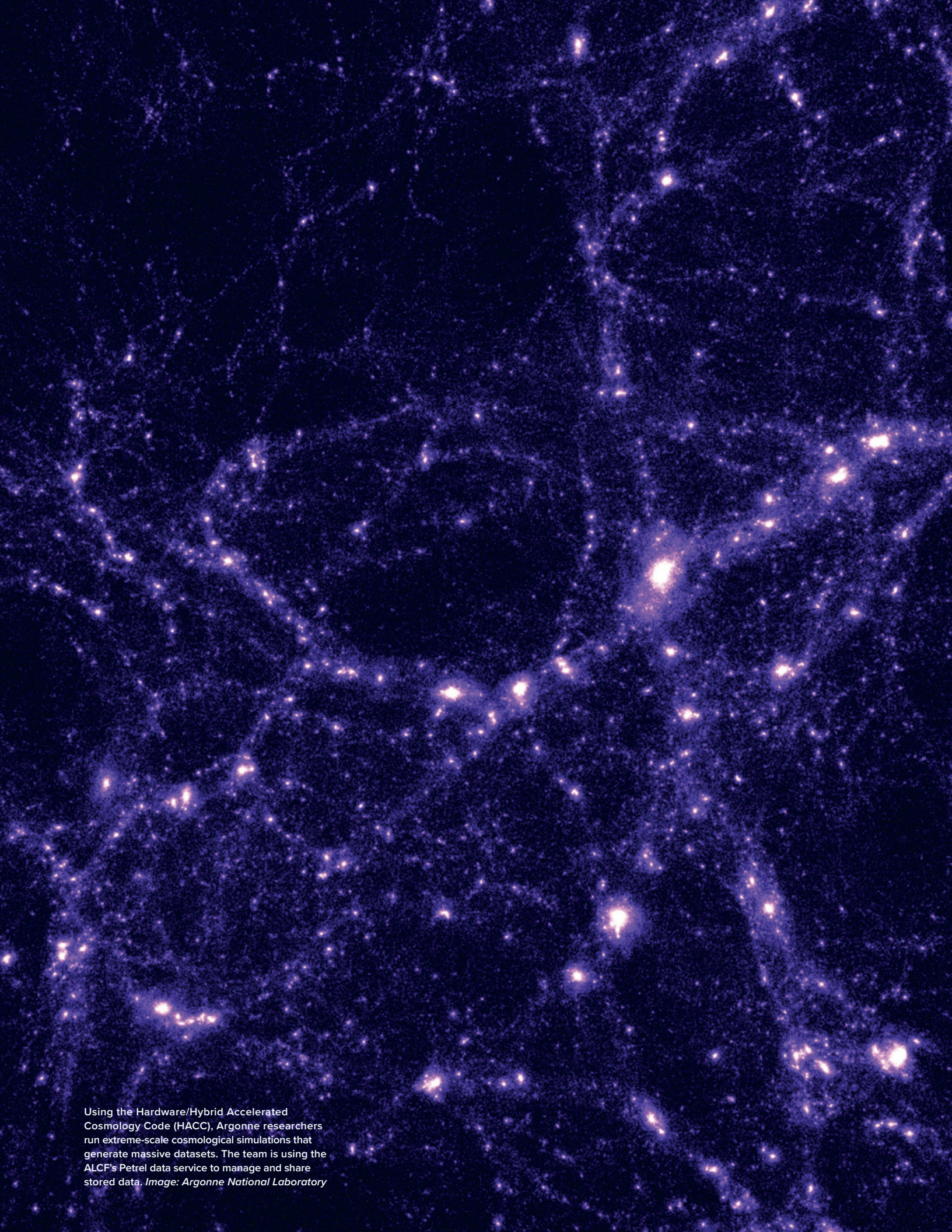
The constellation of edge services being exposed at the ALCF began taking shape in 2015 with the introduction of Petrel, a data service designed to facilitate straightforward sharing with outside collaborators via rapid data transfers.

New services are being deployed in response to a shifting landscape of high-performance computing and its associated challenges, many of these challenges resulting from the increasing prominence of data-driven discovery. While the basic problems of data management are not necessarily new—researchers have struggled for years to determine how and where to store, as well as how to make sense of, data—what has changed is the sheer scale of data, and there is now a need for more creative solutions. Exemplifying the latter emergence is the work of an Argonne cosmologist who has leveraged Petrel to distribute simulated universes and associated cosmological data generated on ALCF systems.

Advances in feeds from scientific instrumentation such as beamlines, colliders, and space telescopes—among many technologies, including artificial intelligence and deep learning—have increased data ingestion substantially at the facilities and in turn caused the traditional HPC facility to revisit how it serves its user community. In fact, the composition of its user community has begun to diversify beyond the traditional computational scientist.

In something of a feedback loop, the proliferation of data drives support for the AI that is responsible, in part, for the proliferation of data.

The next generation of experimental devices—the Large Synoptic Survey Telescope (LSST) and Argonne's Advanced Photon Source (APS) upgrade, for example—will continue to advance these trends, while also making more exotic demands of leadership computing. Meeting these



Using the Hardware/Hybrid Accelerated Cosmology Code (HACC), Argonne researchers run extreme-scale cosmological simulations that generate massive datasets. The team is using the ALCF's Petrel data service to manage and share stored data. *Image: Argonne National Laboratory*

demands and thereby realizing the potential of these devices will require new ways of interacting and engaging with the data they generate.

Moreover, there is a growing need for exposure of services at the edge of the facility as computational science becomes more ubiquitous throughout different research disciplines and attracts a steadily broader audience.

Beyond expanding the possibilities of user-facility data science, these services are intended to evolve the ALCF into a more multifaceted entity and develop its ability to tackle a wider array of problems while serving an even wider audience. They are intended to strengthen the ALCF: the ability to broadly disseminate useful data makes for a stronger facility, as demonstrated at the onset of the current pandemic.

In 2020, services at the edge have enabled large-scale data-sharing among the community of researchers leveraging HPC resources to probe the coronavirus and associated COVID-19 disease and, in the case of the workflow management tool Balsam, have accelerated research itself.

Originally developed as part of a collaboration between Argonne and CERN's Large Hadron Collider to simulate particle collision events in the ATLAS detector, Balsam, which facilitates large-scale data transfers, has recently been used by ALCF computational scientists to enable the near-realtime analysis of experimental data collected at the APS.

Their process subsequently was extended to COVID-19 research, incurring speedups in molecular simulations of the SARS-CoV-2 coronavirus and, to investigate strategies to mitigate the spread of the associated COVID-19 disease. COVID-19 research was further boosted and accelerated by

Argonne's Advanced Photon Source provides ultra-bright, high-energy storage ring-generated x-ray beams for research in almost all scientific disciplines.

DOE's rapid delivery of the GPU-based upgrade to Theta. Upon its arrival, ALCF staff quickly devised and executed an installation solution to meet the nation's needs.

Services at the edge have led to successes in other domains as well. For example, Petrel has enabled a publicly accessible repository of reconstructed visual data built from mouse neuroanatomy. Leveraging APS beamlines, the project, led by Argonne and University of Chicago computational scientists and neuroscientists, seeks to create a mouse connectome—a dataset detailing the brain's every neural connection. Thanks to the repository, biologists outside the ALCF are able to take advantage of this work.

Similarly, to make crystal data available to chemists and materials scientists, an experiment in serial crystallography conveys crystal information from an APS beamline into Petrel to feed ALCF web portals, dynamically updating with each new sample. On an even larger scale, Petrel has enabled the Kubernetes-powered Materials Data Facility (MDF). Comprising and streamlining hundreds of terabytes of materials science data from numerous sources, it automates data ingestion and indexing and simplifies publication. Machine learning methods are easily accessible and executed in the MDF, its datasets available for PetrelKube-based analysis.

The arrival of Eagle, a new global filesystem at the ALCF, will bring still larger and more capable production-level file sharing to facility users. Compared to Petrel, Eagle will represent a space for even broader distribution of reassembled data acquired from various experiments originating with the ALCF and the greater scientific community. Thanks to the service, a wider community will be able to access uploaded data, and ALCF users will be able to directly access the data for analysis. Eagle is also designed to foster experimentation: analysts will be able to write new algorithms to attempt analyses that have never been performed.

Eagle's fully supported production environment will be the next step in the expansion of services at the edge, whose use and prominence at the ALCF are only expected to increase as they become more integral to the facility's ability to deliver computational power and performance.

Preparing for Science in the Exascale Era

As the future home to Aurora, the ALCF has been ramping up efforts to ready researchers and developers for science at exascale.

Science on day one. That's the goal for each new supercomputer introduced at the ALCF, including its upcoming exascale system, Aurora. But how do you prepare for a machine before it is even built?

With Aurora's arrival drawing closer, the ALCF has been ramping up its efforts to lay the groundwork for science in the exascale era.

The facility's preparatory activities include continued collaborations with Aurora vendors Intel and HPE on the testing and development of various components to ensure they meet the requirements of the ALCF user community; partnering with DOE's Exascale Computing Project (ECP) to develop a robust exascale computing ecosystem; facilitating the Aurora Early Science Program (ESP) to prepare key applications and software for the system; deploying early hardware and software tools to provide an environment for exascale code development; and hosting training events to educate the research community on various exascale tools and technologies.

The Aurora system's exaflop of performance combined with an ability to handle both traditional HPC and emerging AI workloads will enable researchers to address some of the world's most challenging scientific problems. Early science projects range from mapping the brain to simulating the universe at extreme scales to accelerating the discovery of new drugs and functional materials.

Aurora's innovative design will be based on Intel's Xeon Scalable processors, high-performance Intel X^{e} GPU compute accelerators, and Optane DC persistent memory. The system will rely on HPE's Shasta exascale-class architecture and Slingshot interconnect technology, which can provide concurrent support for advanced simulation and modeling, AI, and analytics workflows. Aurora will leverage historical advances in software investments along with increased application portability via Intel's oneAPI. The supercomputer will also be equipped with a new I/O platform called Distributed Asynchronous Object Storage (DAOS) to meet the needs of exascale workloads.

Exascale Collaborations

The process of planning and preparing for new leadership-class supercomputers takes years of collaboration and coordination. It requires partnerships with vendors and the broader HPC community to develop hardware, software, and storage technologies, as well as making facility enhancements to ensure the infrastructure is in place to power and cool the massive systems.

With Aurora, the ALCF has also benefitted from a close collaboration with the ECP, which has opened the door to working with a broader software portfolio and more extensive partnerships with other DOE labs than have been afforded in the past. Launched in 2016, the ECP is a multi-lab initiative aimed at accelerating the delivery of a capable exascale computing ecosystem that

The Argonne-Intel Center of Excellence (COE) has held multiple workshops to provide details and instruction on various aspects of the Aurora hardware and software environment.

encompasses applications, system software, hardware technologies, architectures, and workforce development. Argonne—one of the six ECP core labs—has a strong presence on the ECP leadership team and has several researchers engaged in ECP projects and working groups in the areas of application development, software development, and hardware technology.

ALCF staff members are working closely with more than 20 ECP application and software projects to prepare a diverse set of scientific applications to run on Aurora. Efforts have focused on porting various codes, mini-apps, frameworks, and libraries to Intel Gen9, DG1, and XeHP GPU hardware to evaluate and optimize performance. ALCF researchers are also developing algorithms, benchmarking tools, and use cases to help ensure applications can leverage the full potential of Aurora.

Another area of collaboration involves deploying and integrating ECP software, vendor software, and facility-based software environments to ensure exascale software stacks meet application requirements while allowing for optimal facility operations. Activities include developing a common Continuous Integration strategy to drive automation of recurring building and testing across DOE's exascale site environments; using the Spack package manager as a tool for build automation and final deployment of software; managing the ECP allocation program that awards time at DOE computing facilities for the testing and development of new features and functionality at scale; exploring and potentially enhancing the installation, upgrade, and monitoring capabilities of HPE's Shasta software stack; and enabling container support and tools on Aurora and other exascale systems.

Working in concert with the ECP, Argonne researchers are also contributing to the advancement of programming models (OpenMP, SYCL, Kokkos, Raja), language standards (C++), and compilers (Clang/LLVM) that are critical to developing efficient and portable exascale applications.

Furthermore, the ALCF continues to work closely with Intel and HPE on the testing and development of various components to ensure they can be leveraged effectively by the scientific computing community. By analyzing the performance of key benchmarks and applications on early hardware, ALCF researchers are developing a broad understanding of the system's architecture and capabilities. This effort helps to identify best practices for optimizing codes, and, ultimately, a roadmap for future users to adapt and tune their software for the new system. In addition, the team is providing feedback to Intel about the alpha/beta compilers and software stack for Aurora, including the gold release of its oneAPI toolkits.

Early Science

The Aurora Early Science Program (ESP), an ALCF initiative dedicated to preparing key applications for the scale and architecture of the exascale machine, is another critical piece of the puzzle. Through open calls for proposals, the ESP awarded pre-production computing time and resources to a diverse set of projects that are employing emerging data science and machine learning approaches alongside traditional modeling and simulation-based research. The research teams also field-test compilers and other software, helping to pave the way for other production applications to run on the system.

ESP research teams are investigating a wide range of computational research areas that will help prepare Aurora for primetime. This includes mapping and optimizing complex workflows; exploring new machine learning methodologies; stress testing I/O hardware and other exascale technologies; and enabling connections

AURORA/EXASCALE TRAINING EVENTS

Aurora Hackathons

06/22–06/23
High-Fidelity Simulation of Fusion Reactor Boundary Plasmas

08/17–08/18
Exascale Computational Catalysis

09/17–09/18
Metascalable Layered Materials Genome

09/25 AND 09/28
Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials

12/07
Simulating and Learning in the ATLAS Detector at the Exascale

12/08–12/10
Aurora COE Dungeon Session 1: Intensive Development on Pre-Aurora GPUs (Four Simulation Applications)

Aurora Webinars

03/25
DAOS: Next-Generation Data Management for Exascale

06/02
Preparing Applications for Aurora Using the Intel DPC++ Compatibility Tool

06/24
OpenMP Offload Capabilities in the oneAPI HPC Toolkit

09/30
Overview of the New Intel oneAPI Math Kernel Library

Aurora Workshops

02/25–02/27
Aurora COE Workshop 2: Aurora Software Development Kit and Hardware Overview

10/21–10/22
Aurora COE Workshop 3: Intel Hardware Update and Pre-Aurora GPU Development

Additional Events

07/15–07/17
ALCF-ECP CMake Workshop

09/01–09/02
2020 Performance, Portability, and Productivity at HPC Forum

10/13–10/16
Intel eXtreme Performance Users Group (IXPUG) Annual Meeting

Scheduled to arrive at Argonne in 2022, Aurora will provide researchers with a powerful tool for advancing scientific discovery and innovation.

to large-scale experimental data sources, such as CERN's Large Hadron Collider, for analysis and guidance.

With access to the early Aurora software development kit (a frequently updated version of the publicly available oneAPI toolkit) and Intel Gen9, DG1, and XeHP GPUs through Argonne's Joint Laboratory for System Evaluation, researchers participating in the ESP are able to test code performance and functionality using the programming models that will be supported on Aurora.

User Training

From workshops to webinars, the ALCF has hosted hundreds of researchers for a multitude of training opportunities designed to help them prepare for Aurora.

The facility launched its new Aurora Early Adopter webinar series late last year to introduce attendees to programming models, exascale technologies, and other tools available for testing and development work. Open to the public, the quarterly webinars have covered topics like Aurora's DAOS I/O platform, oneAPI's OpenMP offload capabilities, and the new oneAPI Math Kernel Library (oneMKL).

The Argonne-Intel Center of Excellence (COE) hosted two Aurora workshops this year aimed at ESP and ECP research teams. In February, around 100 attendees visited the ALCF for a three-day workshop focused on applications and software development for Aurora. The event included substantial hands-on time for attendees to work with ALCF and Intel experts on developing, testing, and profiling their codes, as well as presentations on the Aurora software development kit, the system's memory model, and available machine learning tools and frameworks. Attendees also had the opportunity to share programming progress, best practices, and lessons learned.

In October, the COE held a virtual Aurora workshop for ESP and ECP teams, providing another opportunity for researchers to receive updates on the latest Aurora hardware and software developments, connect with ALCF and Intel staff for assistance in developing and testing their codes, and share their experiences working with various exascale tools and programming models.

The COE also continued to host its series of intensive, hands-on sessions called "hackathons." These multi-day collaborative events, held virtually this year due to the COVID-19 pandemic, pair individual ESP teams with Argonne and Intel researchers to help advance their application development efforts using the Aurora software development kit.

As part of a broader effort to groom a new generation of supercomputer users, ALCF staff members organize and manage the annual Argonne Training Program on Extreme-Scale Computing (ATPESC). Supported by the ECP, this intensive, two-week program teaches attendees the key skills, approaches, and tools needed to design, implement, and execute computational science and engineering applications on current supercomputers and the exascale systems of the future. ATPESC has hosted more than 500 participants since it began in 2013.

SCIENCE

The ALCF is accelerating scientific discoveries in many disciplines, ranging from chemistry and engineering to physics and materials science.

Researchers from Princeton University are using the ALCF's Theta supercomputer for large-scale simulations aimed at advancing our understanding of supernova explosions. This image is snapshot from a simulation of the core of a massive star that has just undergone collapse, is about to launch a supernova explosion, and has birthed a proto-neutron star that will remain after the mantle has been ejected and the debris has cleared. *Image: Argonne National Laboratory and Princeton University*

Accessing ALCF Resources for Science

As a national user facility dedicated to open science, any researcher in the world with a large-scale computing problem can apply for time on ALCF computing resources.

Researchers gain access to ALCF systems for computational science and engineering projects through competitive, peer-reviewed allocations programs supported by the DOE and Argonne.

The ALCF also hosts competitive, peer-reviewed application programs designed to prepare key scientific applications and innovative computational methods for the architecture and scale of DOE supercomputers.

Application Programs

ADSP
The ALCF Data Science Program (ADSP) supports big data projects that require the scale and performance of leadership computing resources. ADSP projects focus on developing and improving data science techniques that will enable researchers to gain insights into very large datasets produced by experimental, simulation, or observational methods.

ESP
As part of the process of bringing a new supercomputer into production, the ALCF conducts its Early Science Program (ESP) to prepare applications for the architecture and scale of a new system. ESP projects represent a typical system workload at the ALCF and cover key scientific areas and numerical methods.

Allocation Programs

INCITE
The Innovative Novel Computational Impact on Theory and Experiment (INCITE) program aims to accelerate scientific discoveries and technological innovations by awarding ALCF computing time and resources to large-scale, computationally intensive projects that address grand challenges in science and engineering.

ALCC
The ASCR Leadership Computing Challenge (ALCC) program allocates ALCF computing resources to projects that advance the DOE mission; help to broaden the community of researchers capable of using leadership computing resources; and serve the national interests for scientific discovery, technological innovation, and economic competitiveness.

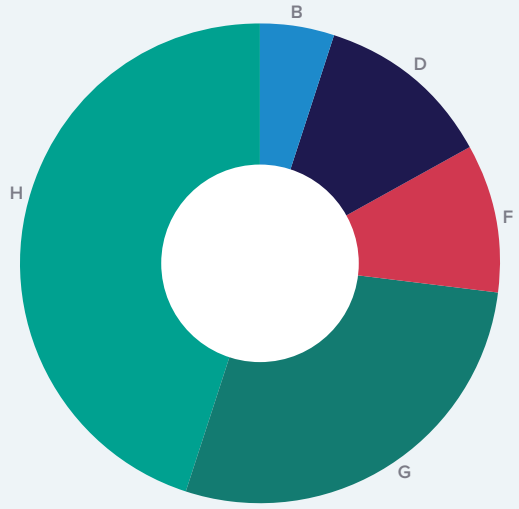
Director's Discretionary
Director's Discretionary projects are dedicated to leadership computing preparation, INCITE and ALCC scaling, and efforts to maximize scientific application efficiency and productivity on leadership computing platforms.

INCITE/ALCC BY DOMAIN

2020 INCITE

17.8M NODE HOURS

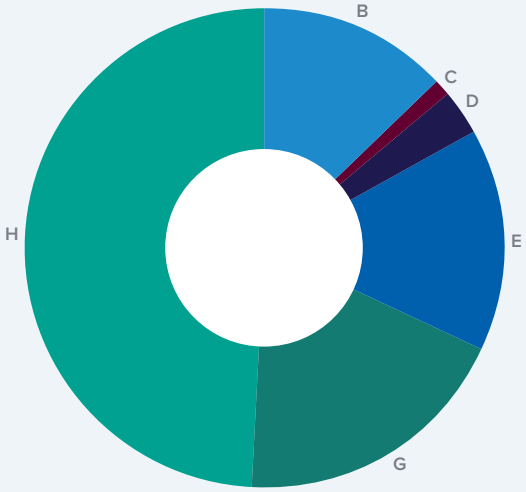
A	Biological Sciences	– %
B	Chemistry	5
C	Computer Science	–
D	Earth Science	12
E	Energy Technologies	–
F	Engineering	10
G	Materials Science	28
H	Physics	45



2020 ALCC

5.87M NODE HOURS

A	Biological Sciences	– %
B	Chemistry	13
C	Computer Science	1
D	Earth Science	3
E	Energy Technologies	15
F	Engineering	–
G	Materials Science	19
H	Physics	49



ALCC data are from calendar year 2020.

2020 Science Highlights

The ALCF user community pushes the boundaries of science across disciplines, advancing our knowledge of the universe at all scales.

With hundreds of thousands of processors working in parallel, supercomputers allow users to achieve scientific breakthroughs that would not otherwise be possible. From detailed atomic-level simulations to massive cosmological studies, researchers can investigate extremely complex physical systems and processes that are too small or large, costly, or dangerous to study in a laboratory.

Each year, ALCF users produce impressive results, whether they are developing and demonstrating novel computational methods or publishing papers in high-impact scientific journals.

In the following pages, we present a selection of notable results from projects supported by the ALCF's various allocation programs.

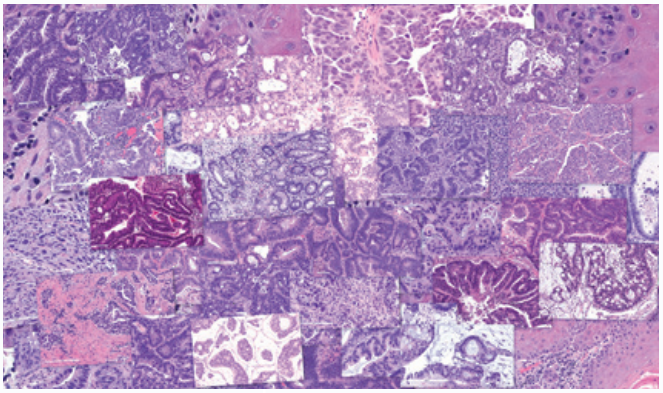
Among this year's highlights is a project that used ALCF computing resources to perform simulations that helped rule out a known quantum effect as the cause of the muon's unexpectedly strong magnetic moment. Another project is employing data mining and large-scale simulations to produce auto-generated databases designed to accelerate the design and discovery of new materials.

You will also read about a project that used Theta to carry out the largest-ever simulation of flow inside an internal combustion engine; a research team performing a massive collection of 3D supernova simulations to elucidate the physics behind the collapse of massive stars; and several other projects pursuing breakthroughs in diverse scientific areas, including fusion energy, climate modeling, and experimental data analysis.

Biological Sciences | Data, Learning

Scalable Reinforcement-Learning-Based Neural Architecture Search for Cancer Research

PI Prasanna Balaprakash, Argonne National Laboratory
AWARD Director's Discretionary
HOURS Theta: 78,125 Node-Hours



Predicting cancer type and drug response using histopathology images from the National Cancer Institute's Patient-Derived Models Repository. *Image: Argonne National Laboratory*

The complexity of cancer increasingly necessitates the design and development of technology that advances our ability to predict, detect, and diagnose the disease. To this end, researchers from Argonne National Laboratory have established a neural architecture search (NAS) that, for a class of representative cancer data, automates the development of predictive models based in deep learning—that is, machine learning algorithms capable extracting science from unstructured data.

CHALLENGE The manual development of deep learning models is accomplished via time-consuming trial and error. Determination of, for example, correct architectures and suitable parameters is therefore heuristic; moreover, cancer research and discovery require numerous diverse types of datasets, many of which are unstructured and lack images or text, threatening to bottleneck researchers.

APPROACH Leveraging the ALCF's Theta supercomputer and DeepHyper automated machine learning package, the research team studied and demonstrated the scalability of its unique reinforcement-learning-based NAS to automate the generation of predictive models for certain cancer data, constructing custom building blocks that permit domain experts to incorporate cancer-data-specific characteristics.

RESULTS The researchers demonstrated multiple benefits to their approach. The machine-designed model, with just one million parameters, achieved the same degree of accuracy as the corresponding manually designed model, which featured ten million parameters. (Fewer parameters translates into a more robust model—and one that requires less data.) Furthermore, automated design improved training speed nearly fourfold.

IMPACT Due to the broad applicability of the developed method, this work, beyond advancing our ability to predict cancer growth and treatment, stands to impact numerous disciplines, including weather forecasting, fluid dynamics, materials science, and chemistry.

PUBLICATIONS

Balaprakash, P., R. Egele, M. Salim, S. Wild, V. Vishwanath, F. Xia, T. Brettin, and R. Stevens. "Scalable Reinforcement-Learning-Based Neural Architecture Search for Cancer Deep Learning Research." *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (November 2019), ACM.

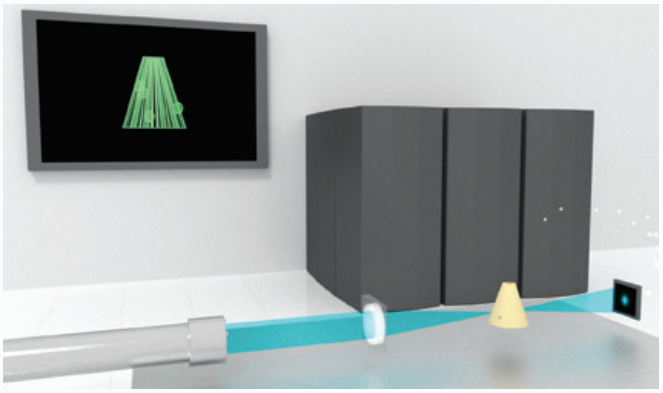
Biological Sciences | Data, Learning

X-ray Microscopy of Extended 3D Objects: Scaling Towards the Future

PI Chris Jacobsen, Argonne National Laboratory and Northwestern University

AWARD ADSP

HOURS Theta: 187,500 Node-Hours



This illustration shows a coherent x-ray beam focused on a large-scale specimen while recording far-field diffraction patterns as the specimen is scanned and rotated. In the background is a computing system using automatic differentiation approaches to reconstruct a 3D image. *Image: Ming Du, Argonne National Laboratory*

The upcoming upgrade of Argonne's Advanced Photon Source (APS) will increase the brightness of its x-ray beams by up to 500 times. This will allow for research projects that are impossible at the current intensity, such as tracing the neural connections inside a mouse's brain to learn more about neurological disorders. But the enhanced experimental capabilities will also increase the need for more advanced reconstruction tools. A research team from Argonne National Laboratory and Northwestern University is leveraging ALCF supercomputing resources to develop a novel approach that can help complete a 3D reconstruction of x-ray images with more flexibility and less human effort than traditional computational methods.

CHALLENGE The APS Upgrade project will enable researchers to image centimeter-thick samples with sub-100-nanometer resolution. However, the multiple scattering of x-ray inside a thick specimen's volume becomes increasingly problematic, as this is not accounted for by most conventional x-ray 3D reconstruction algorithms. This problem must be addressed before one can fully utilize the power of the upgraded light source.

APPROACH To address this issue, the team is leveraging ALCF computing resources to develop a new approach that employs the "multislice tomography" method to simulate x-ray multiple scattering in a continuous 3D volume, and uses automatic differentiation toolboxes to implement the mathematics for solving the reconstruction problem. In preparation for the reconstruction of large-sized specimens, the team has also explored scalable wavefield simulation algorithms to enable the reconstruction software to take full advantage of hundreds of nodes on Theta.

RESULTS In a paper published in *Science Advances*, the team detailed how the new algorithm can provide high-quality 3D reconstructions of heavily scattering samples measured under complicated imaging scenarios. The approach was shown to provide good reconstruction results that are free of diffraction-induced artifacts as seen in traditional reconstruction methods. With very minimal adaption, the code can work for both holography and ptychography. Moreover, in another paper recently accepted by *Optics Express*, the team demonstrated two scalable methods for x-ray wave propagation, namely the "tiling-based Fresnel multislice" method and the "finite-difference" method, both working efficiently while running over 16,000 processes on Theta.

IMPACT By providing a new computational approach for reconstructing x-ray images of thick, real-life materials, this research aims to advance the full range of future nanoscale imaging activities, including cell and brain research, at Argonne's APS and other DOE light sources. Furthermore, the power of machine learning-enabled automatic differentiation makes their software highly versatile and adaptable to many variants and setups of x-ray microscopes.

PUBLICATIONS

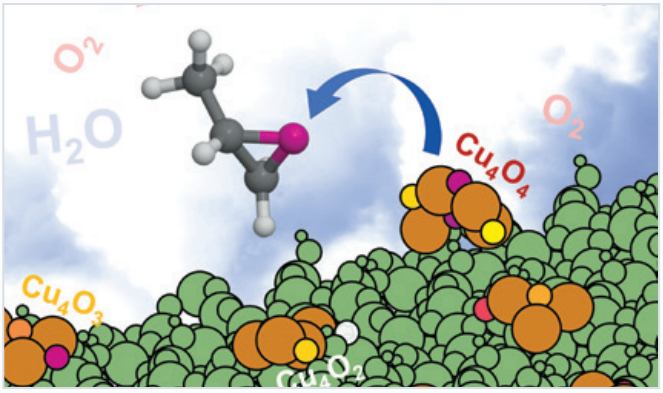
Du, M., Y. S. G. Nashed, S. Kandel, D. Gürsoy, and C. Jacobsen. "Three Dimensions, Two Microscopes, One Code: Automatic Differentiation for X-Ray Nanotomography Beyond the Depth of Focus Limit." *Science Advances* (March 2020), AAAS.

Ali, S., M. Du, M. Adams, B. Smith, and C. Jacobsen. "A comparison of distributed memory algorithms for x-ray wave propagation in inhomogeneous media." *Optics Express* (September 2020), OSA Publishing.

Chemistry | Simulation

Dynamic Nanocluster Catalysis

PI Anastassia Alexandrova, University of California, Los Angeles
AWARD INCITE
HOURS Mira: 20,000,000 Node-Hours



A variety of CuO cluster shapes and compositions are available on the catalytic interface in reaction conditions, and the reactivity is determined by the metastable and particularly active sites on some of these cluster forms. *Image: Anastassia Alexandrova, University of California, Los Angeles*

Heterogeneous catalysis is an essential process for everything from the production of chemicals, materials, and food to pollution control, medical applications, and the development of sustainable energy solutions. Due to the demand to catalyze more complex and highly selective reactions, scientists are working to gain a fundamental understanding of the processes involved in a catalytic reaction to enable the development of novel catalysts with optimized properties. In one such effort, researchers from the University of California, Los Angeles have leveraged ALCF computing resources to advance the design of new catalysts based on surface-mounted small clusters of transition metals.

CHALLENGE Small clusters of noble metals secured on supporting surfaces can be remarkable catalysts, with properties tunable through their size, nature of the support, and reaction conditions. The exploration of these catalytic interfaces, which are incredibly complex and best represented as ensembles of many geometric states, presents a formidable modeling challenge that requires the power and scale of DOE's leadership-class supercomputers.

APPROACH The team's work focuses on the development and consolidation of protocols and efficient methods to enable the computational description of surface-mounted cluster catalysts in a statistical ensemble representation in realistic conditions. The properties of cluster catalysts (e.g., reaction rates, stabilities, spectral signatures) are then represented as ensemble-averages, which can be used to explain existing experiments, or predict outcomes of new experiments.

Using this approach, the team leveraged the ALCF's Mira supercomputer to computationally manipulate catalyst systems by varying temperature, partial pressures of gases, and cluster size to achieve the exposure of the most active binding sites and desired catalytic efficiency. This work was completed using ab initio methods and density functional theory (DFT) with and without periodic boundary conditions, as well as in-house methods for global optimization of dynamic interfaces in conditions relevant to their practical use.

RESULTS In one study, the team used DFT and the grand canonical basin hopping method to explore the structures and reducibility of Cu_4 oxide clusters deposited on amorphous alumina. Their findings, published in *ACS Catalysis*, demonstrated the unique structural and electronic properties of sub-nano Cu oxide clusters and illustrated the critical roles of configuration ensembles and rearrangement to highly reactive metastable cluster isomers in nanocatalysis.

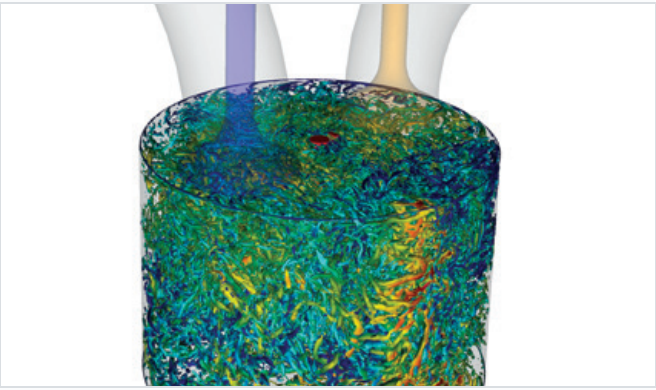
IMPACT The team's novel approach to treat catalytic properties as ensemble averages will help inform experimental efforts to improve the performance of catalysts. Ultimately, their research aims to accelerate the design of novel catalysts that are powerful, selective, stable, and economical.

PUBLICATIONS
Sun, G., A. N. Alexandrova, and P. Sautet, "Structural Rearrangements of Subnanometer Cu Oxide Clusters Govern Catalytic Oxidation," *ACS Catalysis* (April 2020), American Chemical Society.

Chemistry | Simulation

Towards Exascale Internal Combustion Engine Simulation with In-Situ Analysis

PI Muhsin Ameen, Argonne National Laboratory
AWARD ALCC
HOURS Theta: 630,000 Node-Hours



Snapshot of turbulent eddies in the compression stroke during the third engine cycle (1700.5 CAD). Eddies colored by velocity magnitude in the TCC-III internal combustion engine designed by GM. *Image: Argonne National Laboratory*

Numerical simulations play a crucial role in helping automotive manufacturers design more efficient engines, improving the understanding of the in-cylinder combustion process as well as the impact of design changes on engine performance. With this ALCC project, a team of researchers from Argonne National Laboratory are leveraging ALCF supercomputers to develop high-fidelity engine simulation capabilities for DOE's upcoming exascale systems.

CHALLENGE Engine manufacturers currently use commercial codes that have inherent limitations in predictive capability and running at large scale on leadership computing machines. To prepare for the future supercomputing resources, the Argonne team is working to develop the open-source Nek5000 computational fluid dynamics (CFD) code into an effective exascale application that can provide insights into complex turbulent flow, fuel spray, and combustion processes at an unprecedented level of detail.

APPROACH The Argonne-developed Nek5000 code is a fast and scalable high-order solver for accurate modeling of fluid turbulence. In collaboration with ALCF staff, the team is working to optimize nekRS, a GPU-based version of Nek5000, for future exascale supercomputers, including the ALCF's Aurora system. Their work includes using Theta to implement state-of-the-art submodels for combustion and spray, and integrate an in-situ analysis framework that can simultaneously reduce data storage requirements and expedite scientific discoveries. The team is validating code performance against several benchmark engine experiments using General Motors' Transparent Combustion Chamber (TCC-III) optical engine.

RESULTS As part of this effort, the team has used Theta to perform direct numerical simulations aimed at understanding the causes of cycle-to-cycle variability, which can contribute to unevenness in the running of the engine, excessive engine noise and emissions, and potentially damaging engine knock. Their simulations showed that interaction of the oncoming intake jet with the spark plug enables vortex shedding and turbulent flow structures at the start of the intake stroke. The team has hypothesized that the vortex structures, which linger for the duration of the intake stroke, reduce the energy contained in the large-scale coherent structures at the start of compression, and, ultimately, affects how efficiently the engine can burn fuel.

In addition, the researchers performed the largest-ever simulation of flow inside an engine using 51,328 cores of Theta. The simulation showed how conventional wall models fall short of capturing wall shear stress and heat flux. In collaboration with Stanford University, the simulation data is being utilized to improve wall models for low-order techniques.

IMPACT Ultimately, the development of high-fidelity engine simulation capabilities for exascale supercomputers will help automotive manufacturers advance efforts to improve the energy efficiency of combustion engines, thereby reducing the carbon footprint of the transportation sector.

PUBLICATIONS

Chatterjee T., S. Patel, and M. M. Ameen, "Understanding the Behavior of Large-Scale Structures in Highly Scalable Spectral Element Simulations of Internal Combustion Engines in the Framework of Large Eddy Simulation." *72nd Annual Meeting of the APS Division of Fluid Dynamics* (November 2019), American Physical Society.

Computer Science | Simulation, Data, Learning

Balsam: Near Real-Time Experimental Data Analysis on Supercomputers

PI Michael Salim, Argonne National Laboratory
AWARD Director's Discretionary
HOURS Theta: 781,250 Node-Hours

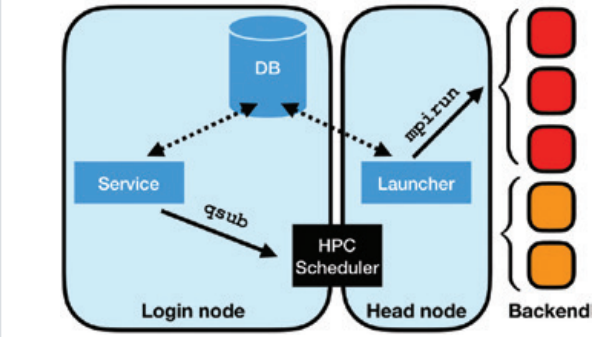


Diagram of Balsam components. *Image: Michael Salim, Argonne National Laboratory*

Users of DOE experimental facilities require increasingly powerful computers to keep pace with accelerating data rates. As facilities like the Advanced Photon Source (APS) undergo significant upgrades in the coming years, these data rates will continue to speed up, making supercomputers necessary to perform effective data analysis.

Originally developed as a workflow management tool and edge service, researchers from Argonne National Laboratory are applying the ALCF's Balsam project to APS workloads to enable realtime and near real-time analysis of the associated datasets.

CHALLENGE Once data become available, the beamlines at most experimental facilities require computational resources for processing almost immediately. For APS workloads in particular, the researchers identified three core criteria necessary to perform realtime analysis of experimental data on leadership-class computing resources: high-speed networking between the experimental and computing facilities, infrastructure capable of handling a diverse array of jobs, and a service to manage incoming datasets and the applications that operate on them.

APPROACH Balsam is designed to optimally apportion available compute nodes so as to maximize efficacy and throughput. Its task queue can grow almost without limit; when data rates exceed node availability, Balsam buffers pending tasks and dispatches resources as they become available, orchestrating data transfers and interface to local job schedulers and application launchers.

To demonstrate Balsam's ability to facilitate large-scale data transfers between the ALCF and the APS, Argonne researchers performed two experiments using real x-ray photon correlation spectroscopy (XPCS) datasets and Theta.

RESULTS In the first experiment, APS hardware submitted an XPCS analysis task to ALCF resources every 15 seconds, with the hardware and network successfully keeping pace.

The researchers subsequently stress-tested their configuration. Data submissions were scheduled such that the pipeline was at maximum capacity for the entire duration. While each dataset had a much longer runtime on ALCF hardware, necessitating a much greater number of compute nodes for processing to keep pace with submission, Balsam still successfully enabled on-the-fly data analysis. Results were detailed in a paper presented at the SC19 International Conference for High Performance Computing, Networking, Storage, and Analysis.

IMPACT Since the demonstration, Balsam has been leveraged for at least two projects to accelerate research of the SARS-CoV-2 coronavirus and the associated disease, COVID-19. The researchers are looking to deploy the service in other settings, including ptychography and electron microscopy.

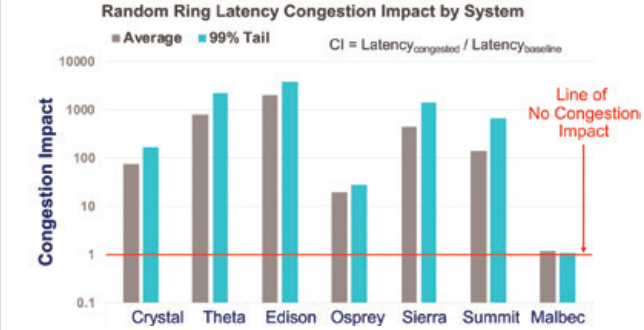
PUBLICATIONS

Salim, M., T. Uram, J. T. Childers, V. Vishwanath, and M. Papka. "Balsam: Near Real-time Experimental Data Analysis on Supercomputers," *2019 IEEE/ACM 1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing* (November 2019), IEEE.

Computer Science | Simulation, Data

GPCNeT: Designing a Benchmark Suite for Inducing and Measuring Contention in HPC Networks

PI Sudheer Chunduri, Argonne National Laboratory
AWARD Director's Discretionary
HOURS Theta: 200,000 Node-Hours



Mean and 99th percentile congestion impact (defined as the latency under congested conditions divided by the latency under isolated network conditions) across a range of systems and architectures for latency. *Image: Sudheer Chunduri, Argonne National Laboratory*

Network congestion is one of the biggest problems facing high-performance computing (HPC) systems today, affecting system throughput, performance, user experience, and reproducibility. Despite its significance, current network benchmarks fail to proxy the real-world network utilization seen on congested systems. In a collaboration between the ALCF, the National Energy Research Scientific Computing Center (NERSC), and HPE-Cray, a team of researchers has developed a new open-source benchmark suite called the Global Performance and Congestion Network Tests (GPCNeT) to serve as a proxy to the real-world network utilization on congested HPC systems and to provide meaningful characterization of congestion management for system interconnects.

CHALLENGE In HPC systems, performant networks are essential to join resources efficiently together to enable scientific discovery at unprecedented scales. While simple metrics such as latency, injection bandwidth, and bisection bandwidth are useful indicators of a network's peak performance, they typically do not correspond to realized performance on production machines, particularly when diverse workloads contend for the same network resources. They often overemphasize best-case configurations on quiet systems and fail to expose behavioral differences between network topologies and architectures.

APPROACH GPCNeT is a generic, topology-agnostic benchmark suite that captures the complex workloads anticipated on multitenant HPC networks. It uses multiple configurable congestion patterns concurrently to emulate multiple separate jobs and uses concurrent canary jobs that emulate boundary exchange algorithms to measure the impact of congestion. To demonstrate the utility of GPCNeT, the team used the benchmark to evaluate the impact of congestion across topologies and architectures, including Theta (ALCF), the now-retired Edison system (NERSC), Summit (OLCF), and Sierra (Lawrence Livermore National Laboratory).

RESULTS Their results showed that systems of all technologies and scales are susceptible to congestion, illustrating the need for congestion control in next-generation networks. As detailed in a paper presented at SC19, the team's studies revealed that latency-bound communication can be very sensitive to congestion, bandwidth is generally less sensitive to congestion, and congestion has a scale-dependent component. The researchers also found that congestion manifests in different ways on the same machine depending on how an MPI library implements different interfaces.

In addition, the team used GPCNeT to effectively showcase the advanced congestion management feature available in the HPE-Cray Slingshot network, which will be used in DOE's three upcoming exascale systems. Furthermore, the benchmark helped the DOE computing facilities provide feedback to vendors on limitations of current networks.

IMPACT The GPCNeT benchmark suite provides a new tool for gaining insights into the congestion management efficiency of a system interconnect and its effectiveness in reducing or avoiding tail latency effects on scientific application performance. The benchmark can also be used to define concrete network performance expectations for future system procurements.

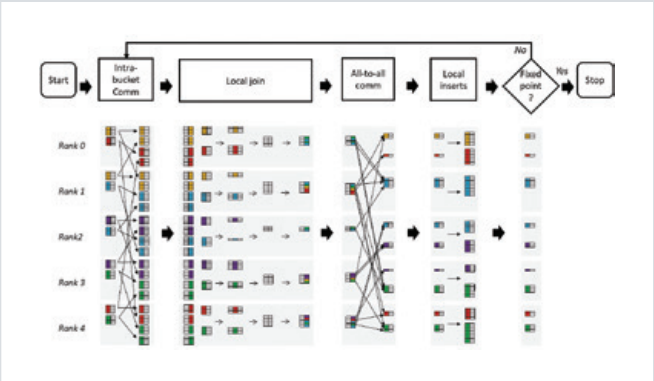
PUBLICATIONS

Chunduri, S., T. Groves, P. Mendygral, B. Austin, J. Balma, K. Kandalla, K. Kumaran, G. Lockwood, S. Parker, S. Warren, N. Wichmann, and N. Wright. "GPCNeT: Designing a Benchmark Suite for Inducing and Measuring Contention in HPC Networks." *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (November 2019), ACM.

Parallel Relational Algebra for Logical Inferencing at Scale

PI Sidharth Kumar and Thomas Gilray, University of Alabama at Birmingham

AWARD Director's Discretionary
HOURS Theta: 15,625 Node-Hours



A diagram of the distributed hash-tree join algorithm highlighting the different stages and multiple levels of partitioning used to balance the workload. *Image: Sidharth Kumar, University of Alabama at Birmingham*

Relational algebra (RA) is a widely used procedural query language that forms a basis of primitive operations suitable for AI applications in graphs and networks, program analysis, deductive databases, and constraint logic programming. While RA has been a subject of great interest in computer science research, its exploration on supercomputers has been limited. Researchers from the University of Alabama at Birmingham are leveraging ALCF supercomputing resources to develop strategies to advance the use of scalable distributed RA for solving massive logical inference problems, graph problems, and more on supercomputers.

CHALLENGE Despite its expressive power, RA has not received the same attention in HPC research as more common primitives like stencil computations, floating-point operations, numerical integration, and sparse linear algebra. Specific challenges in addressing representation and communication among distributed portions of a relation have previously thwarted successful scaling of RA applications to supercomputers. Furthermore, implementing RA poses unique challenges when distributed over many-node, networked computing systems—especially regarding how the workload is balanced across available compute resources.

APPROACH The team's work is focused on developing algorithms and strategies to effectively parallelize and scale key RA primitives on HPC systems. The researchers performed their experiments on the ALCF's Theta supercomputer, using the system's state-of-art network and infrastructure to run some of the largest data-parallel logical inference tasks performed to date.

RESULTS In one study, the researchers developed a hybrid approach to representing relations and performing efficient distributed operations on networked machines. The team's approach is the first truly scalable distributed RA implementation to address load balancing and inter-process communication, permitting fixed-point iteration and laying the groundwork for more broad use on supercomputers. A paper detailing their research, "Distributed Relational Algebra at Scale," garnered the best paper award at the 2019 IEEE International Conference on High Performance Computing, Data and Analytics.

In a follow-up paper, "Load-Balancing Parallel Relational Algebra," the researchers presented strategies to address both spatial and temporal load imbalance. The techniques demonstrated fully dynamic load balancing for RA that is robust to changes across time. Their paper received the Hans Meuer Award for the most outstanding research paper at the ISC High Performance 2020 Conference.

IMPACT The team's research demonstrates the potential effectiveness of using relational algebra to extract data-level parallelism automatically for key HPC applications. The ability to iterate load-balanced data parallel operations is opening up new horizons for the scalability of AI applications and a wide range of analytics problems on supercomputers.

PUBLICATIONS

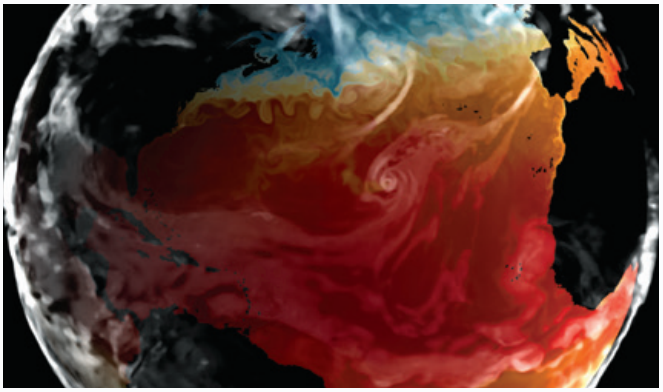
Gilray, T., and S. Kumar. "Distributed Relational Algebra at Scale." *2019 IEEE 26th International Conference on High Performance Computing, Data, and Analytics* (HiPC) (December 2019).

Kumar S., and Gilray T. "Load-Balancing Parallel Relational Algebra." *High Performance Computing. ISC High Performance 2020. Lecture Notes in Computer Science* (June 2020).

Earth Science | Simulation

Energy Exascale Earth System Model

PI Mark Taylor, Sandia National Laboratories
AWARD INCITE
HOURS Theta: 1,800,000 Node-Hours



A snapshot of the atmosphere and ocean conditions in the high-resolution E3SM v1 simulation being run on ALCF Theta. Cloud fraction (grey-scale) shown on top of the sea surface temperature (color contours) with land masses in black. *Image: Mark Bolstad, Sandia National Laboratories*

With the coming paradigm shift in computer architectures and programming models as capability moves to the exascale era, the Energy Exascale Earthy System Model (E3SM) project aims to develop a cutting-edge climate and earth system that can tackle the most demanding climate research imperatives. Harnessing ALCF supercomputing resources, the E3SM project is addressing questions concerning the water cycle and cryosphere systems.

CHALLENGE The research team is focusing on two grand challenge questions: (1) In the water cycle, how will more realistic portrayals of its important features (e.g., resolution, clouds, aerosols) affect river flow and associated freshwater supplies at the watershed scale? (2) In cryosphere systems, could a dynamic instability in the Antarctic Ice Sheet be triggered within the next 40 years? For (1), the team's objective is to simulate changes in the hydrological cycle with a specific focus on precipitation and surface water in orographically complex regions, such as the western United States and Amazon headwaters. For (2), in the first fully coupled global simulations to include ocean circulation and dynamic sub-ice shelf melting, the team aims to examine the near-term risk of initiating the West Antarctic Ice Sheet instability or collapse due to rapid melting by adjacent warming waters. Moreover, the simulations will focus on assessing the potential for significant increase in sub-ice shelf melt rates in response to past, ongoing, and future climatic changes.

APPROACH The team makes extensive use of Theta to run the ES3M code, which comprises component models for atmosphere, ocean, sea ice, and land. Typically 64 tasks are assigned to every node, each with two OpenMP threads so as to establish intranodal parallelism. The majority of the cores are allocated to the atmosphere model, a subset of which also run the land and sea models; the remaining cores are allocated to the ocean model, which runs concurrently with the atmosphere model.

E3SM v1's high-resolution coupled confirmation runs atmosphere and ocean components on a grid with 25-kilometer average spacing at the Equator with 72 vertical layers. The ocean and ice run on a variable resolution grid, with horizontal resolution ranging from 18 to 6 kilometers and 100 vertical layers.

RESULTS Leveraging Theta, the researchers have completed 125 years of a 150 year pre-industrial control simulation. A transient or historical simulation representing the years 1970-2010 is also in progress.

IMPACT In addition to further advancing the predictive power of climate models and providing insight into the climatic effects of rapid changes in the earth's ice content, the ES3M simulations have the potential to answer how water resources and the hydrological cycle interact with the climate system on both local and global scales. Hurricane hindcast simulations performed for this project demonstrated the high fidelity with which extreme weather can be modeled, while exposing parametric weaknesses that need improvement.

PUBLICATIONS

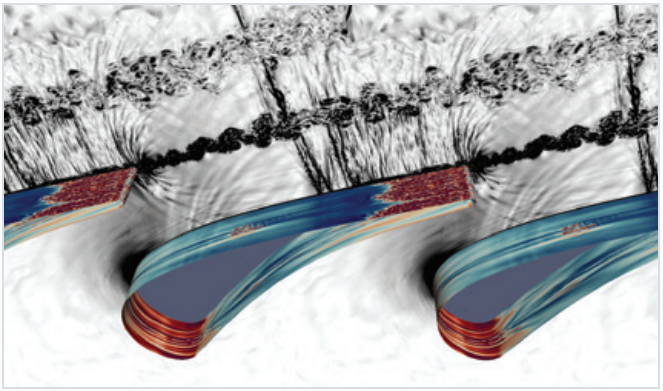
Bertagna, L., O. Guba, M. A. Taylor, J. G. Foucar, J. Larkin, A. M. Bradley, S. Rajamanickam, and A. G. Salinger. "A Performance-Portable Nonhydrostatic Atmosphere Dycore for the Energy Exascale Earth System Model Running at Cloud-Resolving Resolutions," *Multiphysics Applications, Supercomputing 2020* (to be published).

Caldwell, P., A. Mametjanov, Q. Tang, L. P. Van Roekel, J.-C. Golaz, W. Lin, D. C. Bader et al. 2019. The DOE E3SM Coupled Model Version 1: Description and Results at High Resolution," *Journal of Advances in Modeling Earth Systems* (November 2019), John Wiley and Sons.

Engineering | Simulation, Data

DNS Reference Data for Turbulence Model Development on the Bachalo-Johnson ATB

PI Koen Hillewaert, Cenaero
AWARD INCITE
HOURS Mira: 7,875,000 Node-Hours



Flow in the transonic turbine inlet guide vane LS89 at Re=1.2 million, computed using 4th-order-accurate DG. The acoustic waves and shock system are visualized via numerical Schlieren, whereas the instantaneous wall heat flux is shown on the suction side. *Image: J. S. Cagnone, M. Rasquin, and K. Hillewaert, Cenaero.*

Today, the design and optimization of most external and internal flows rely on low-fidelity statistically averaged turbulence models. The limited availability of high-quality reference data for sufficiently complex flows remains a significant obstacle to the development of more sophisticated turbulence models. To overcome this hurdle, this project, led by researchers from Cenaero and Stanford University, aims to generate exhaustive reference databases for the development of turbulence models.

CHALLENGE Low-fidelity, statistically averaged turbulence models are known as RANS; they are the most accurate approach to flow design and optimization affordable to industry. The limited availability of high-quality reference data on complex flows, however, hinders their improvement. The generation of such data requires numerous computationally demanding simulations be performed, necessitating leadership-class resources.

APPROACH To generate an exhaustive database that can be used for not only the validation of turbulence models but also for their development and calibration, the researchers perform direct numerical simulations (DNS) of the Bachalo-Johnson axisymmetric transonic bump at a Reynolds number of one million. This test case has been used to validate turbulence models in the presence of shock-boundary layer interaction and shock-induced separation, the modeling of which phenomena is critical for aeronautical and aircraft engine companies. An extensive DNS dataset, capable of reconstructing all RANS turbulence modeling terms, is being cross-validated utilizing two contrasting codes—a discontinuous Galerkin method named Argo developed at Cenaero and a finite volume method CharLESx developed at Stanford—and an extensive span-dependence study is being performed up to the full circumference.

RESULTS A presentation at the 72nd Annual Meeting of the APS Division of Fluid Dynamics detailed the researchers' efforts. Using the Argo method, low dispersion and dissipation commensurate with DNS and large-eddy simulations are maintained on an unstructured mesh. To create the database, the researchers store all RANS-relevant terms and, to improve modeling for shock-boundary layer interaction conditions, time-resolved data in the boundary layer.

IMPACT The database that results from this work is crucial for aeronautical and aircraft engine companies, and will be published online to contribute to the development and improvement of the next generation of lower-fidelity models.

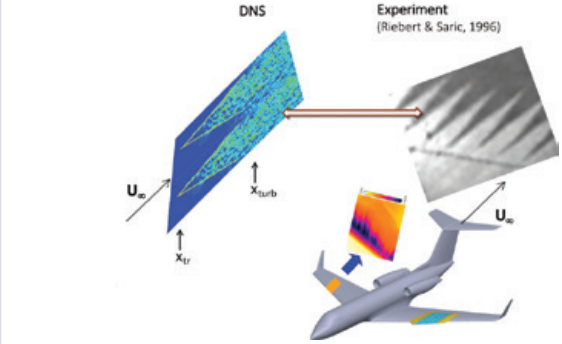
PUBLICATIONS

Hillewaert, K., M. Rasquin, and T. Toulorge. "Using DNS to Improve Wall-Modeled LES of Turbomachinery Flows," *72nd Annual Meeting of the APS Division of Fluid Dynamics* (November 2019), APS.

Engineering | Simulation, Data

Laminar-Turbulent Transition in Swept-Wing Boundary Layers

PI Lian Duan, The Ohio State University
AWARD INCITE
HOURS Mira: 6,000,000 Node-Hours



DNS captures swept-wing boundary layer transition over a transonic aircraft with laminar flow technology. Near-wall flow visualizations confirm the sawtooth nature of transition front as a generic feature of transition due to stationary crossflow vortices regardless of the type of secondary instability. *Image: NASA*

Skin-friction drag accounts for approximately one half of the total drag for long-haul transport aircraft. Due to the substantial increase in drag in the course of laminar-turbulent transition, delay of crossflow-influenced transition in swept-wing boundary layers via laminar flow technology is a leading contender for reducing the aircraft fuel burn.

Researchers from The Ohio State University and NASA Langley Research Center harnessed ALCF supercomputers to enable simulations of the entire crossflow transition process—from the laminar regime to the fully turbulent regime—over a transonic natural-laminar-flow wing with high chord Reynolds numbers relevant to the transport aircraft.

CHALLENGE In particular, the work focused on generating a detailed knowledge base that addresses the existing gaps in crossflow transition prediction, including (1) the excitation and control of crossflow instabilities via discrete roughness elements, and (2) the laminar breakdown of crossflow vortices preceding the onset of turbulent flow and the properties of developed turbulence in the post-transitional region.

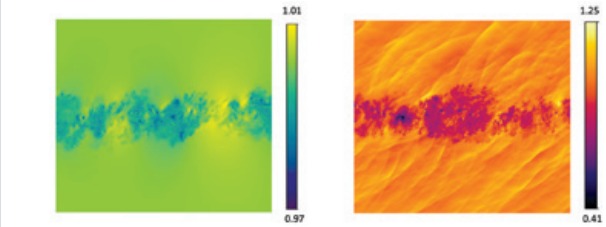
APPROACH Laminar-turbulent boundary layer transition, including the receptivity and breakdown phases, was studied via direct numerical simulations (DNS) based on the notoriously intractable Navier-Stokes equations. Full resolution of the receptivity and breakdown phases necessitated use of leadership-class computing systems. The primary code employed was the high-order finite-difference solver HyperWENO. HyperWENO solves the compressible Navier-Stokes equations in generalized curvilinear coordinates that describe the evolution of the flow's density, momentum, and total energy. ALCF researchers were instrumental in efforts to scale HyperWENO for use on Theta and Mira.

RESULTS The team executed DNS of complete crossflow transition to developed turbulence without discrete roughness elements (DRE) control, with stationary crossflow breakdown initiated by three types of secondary instabilities: Z-mode, Y-mode, and mixed mode. Following this, the team carried out the more complicated case of DNS of complete crossflow transition to developed turbulence with DRE control, in addition to large-eddy simulations and assessment of subgrid-scale models.

IMPACT This project will significantly advance our ability to predict and control laminar-turbulent transitions in 3D boundary layers. It will also help delineate the limitations of the current predictive tools, providing a basis to develop alternate models for cases in which those tools are either inapplicable or lead to unacceptable errors. Beyond aerospace engineering, the work will more broadly impact the design of devices involving 3D transitional and turbulent boundary layers, such as wind turbines and turbomachinery.

Shock-Induced Multi-Material Mixing

PI Sanjiva Lele, Stanford University
AWARD INCITE
HOURS Mira: 4,500,000 Node-Hours



These images show the density fields at a low convective Mach number (left) and a higher convective Mach number (right). At high convective Mach number significant acoustic emission, including eddy shocklets, result from the turbulent mixing. *Image: Sanjiva Lele, Stanford University*

A detailed understanding of shock-induced turbulent mixing is critical to many engineering and scientific applications, including high-speed propulsion in hypersonic air-breathing engines for aerospace vehicles and energy generation through inertial confinement fusion. However, the turbulent flow physics involved these applications remains poorly understood. With this INCITE project, researchers from Stanford University have used ALCF computing resources to explore important aspects of the flow physics that are inaccessible to experiments.

CHALLENGE Shock-induced turbulent mixing arises from a fluid flow phenomenon known as the Richtmyer-Meshkov instability, which occurs at the interface of two fluids when a shock passes through a fluid. The fast time scales and extreme conditions that characterize the flow physics in the vicinity of the shock-interface interaction region limit the scope of experimental measurements. Numerical simulations therefore play an important role in advancing understanding of compressible turbulent mixing.

APPROACH Using the ALCF's Mira supercomputer, the team carried out high-resolution simulations of shock-induced turbulence resulting from the Richtmyer-Meshkov instability, with a focus on compressible mixing with shear and variable density effects. The simulations were performed on a domain with a spanwise extent that was four times larger than previous simulations. The larger span simulations have made it possible to obtain reliable results that are free from domain bounding effects until late times. The high-resolution data also makes it possible to quantify higher order statistics better and to make stronger quantitative claims about the novel findings that were suggested by previously obtained lower-resolution data from experiments and simulations.

RESULTS The team's simulations have established a new regime of mode-coupling in shock-induced mixing and helped to quantify the effect of compressibility and the fundamental compressible energy transfer mechanisms. The researchers are using the simulation results to develop a benchmark-quality database for the model problem of a temporal mixing layer that is validated against experiments and for which the effects of the numerical modeling errors are thoroughly evaluated and demonstrated to be small. Having such a database on turbulence in the compressible regime will help advance the development of models for lower-fidelity methods.

IMPACT The team's findings will advance the state of the art in simulations of shock-induced turbulent mixing and help improve engineering models of variable density turbulence. The enhanced models that result from this work could help enable innovations in new combustion devices and propulsion systems.

PUBLICATIONS

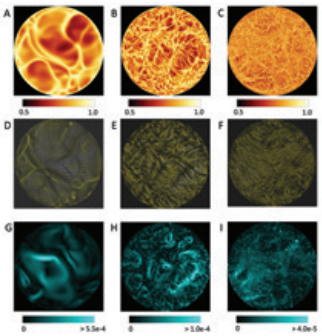
Jeun, J., G. J. Wu, and S. K. Lele. "Large Eddy Simulations of Screeching Twin Rectangular Jets." *AIAA Scitech 2020 Forum* (January 2020), AAIA.

Matsuno, K., and S. K. Lele. "Compressibility Effects in High Speed Turbulent Shear Layers – Revisited." *AIAA Scitech 2020 Forum* (January 2020), AAIA.

Engineering | Simulation

Towards Ultimate Rayleigh-Bénard Convection

PI Janet Scheel, Occidental College
AWARD INCITE
HOURS Mira: 12,500,000 Node-Hours



Convection flow inside the boundary layers, with snapshots of temperature and velocity fields displayed for three different Rayleigh numbers. *Image: Joerg Schumacher, Ilmenau University of Technology*

While the heat transport law in turbulent Rayleigh-Bénard convection remains central to research, our knowledge of heat transport behavior in high-strength turbulent convection remains inconclusive. Led by researchers from Occidental College, Ilmenau University of Technology, and New York University, this project performed well-resolved, direct numerical simulations of fluid convection for a new level of turbulence, and demonstrated that turbulent heat transport adheres to a classical scaling law.

CHALLENGE Understanding of the heat transport law breaks down for Rayleigh numbers (dimensionless parameters that describe the vigor of convective turbulence) greater than 10^{12}. The researchers leveraged massively parallel simulations of 3D Rayleigh-Bénard convection in a slender cylinder with Rayleigh numbers as high as 10^{15}, for which laboratory experiments have reported different outcomes.

APPROACH Direct numerical simulations (DNS) created with the highly scalable spectral element fluid dynamics solver Nek5000 were run on the ALCF's Mira supercomputer. Porting and code establishment were completed by ALCF staff.

RESULTS This work has advanced DNS of turbulent RBC to Rayleigh numbers never before numerically achieved. The simulations resolved velocity gradients inside thin boundary layers and showed that the turbulent heat transport continues to follow the classical 1/3 scaling law without transition to an ultimate state. The research suggests that boundary layers remain marginally stable and continue to act as the bottleneck for global heat transport.

IMPACT The turbulent RBC model is crucial for understanding turbulent convection, itself an important area within fluid dynamics. Applications include improved chip-cooling devices and power-plant heat exchangers, while greater elucidation deepens our understanding of convection as it occurs in the Earth's atmosphere, core, and oceans, as well as in stars.

PUBLICATIONS

Iyer, K. P., J. D. Scheel, J. Schumacher, and K. R. Sreenivasan. "Classical 1/3 Scaling of Convection Holds Up to Ra = 10^{15}," *PNAS* (March 2020), United States National Academy of Sciences.

Constructing and Navigating Polymorphic Landscapes of Molecular Crystals

PI Alexandre Tkatchenko, University of Luxembourg
AWARD ADSP
HOURS Theta: 8,000,000 Node-Hours

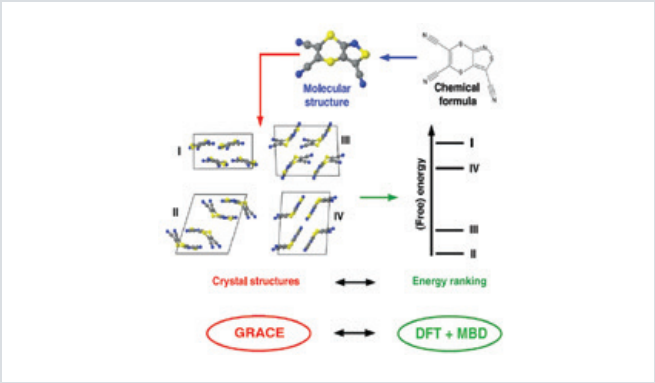


Diagram of the established workflow for crystal prediction. *Image: Alexandre Tkatchenko, University of Luxembourg*

Molecular properties are a fundamental issue in computer-aided drug design. These properties are determined by the molecular arrangements assumed during the formation of solids (crystal polymorphs). Despite its paramount importance, accurately predicting the structure of molecular crystals remains a challenge due to the large number of potential configurations and the computational cost of precise quantum-chemical simulations, which this project—led by University of Luxembourg researchers—sought to overcome by combining state-of-the-art techniques drawn from atomistic simulation and data science.

CHALLENGE The prediction of crystal polymorphs is a difficult task due to the problem's high dimensionality; the differences in relative free energy between polymorphs are usually within 1 kilojoule per mole, which is too slight for even state-of-the-art quantum chemistry methods. Leveraging leadership-class computers, this project sought to unite aspects of two of the most successful prediction methods in a blind-test competition so as to produce a hierarchical workflow for reliable crystal polymorphs that would create accurate datasets for data-driven discovery.

APPROACH This project combined ad-hoc force fields (created using a Monte Carlo parallel tempering algorithm) with reduced-scale electronic-structure calculations, aiming to sample conformational space of molecular crystals at a fast rate. The most promising of the identified conformations were analyzed via all-electron density functional theory calculations, computing the many-body dispersion energies using a wavelike model for long-range interactions (e.g., London and van der Walls forces) as well as a vibrational correction.

ALCF staff contributed by providing a tailored workflow manager to orchestrate a massive number of quantum-mechanical calculations. ALCF staff also assisted in the production of binaries for efficient use of the supercomputer architecture.

RESULTS This procedure enabled prediction of the correct relative-stability ranking of numerous molecular polymorphs seen in real laboratory setups. The predictions include the contributions associated with atomic vibrations arising from finite temperature effects within the crystals. Additionally, the large number of possible conformations found during this process are an excellent source for producing accurate datasets to further accelerate materials discovery with data-driven techniques.

IMPACT This work will help drive predictions of novel molecular crystals, with numerous applications ranging from organic photovoltaics, molecular electronics, hydrogen sources, and pharmaceuticals, to agrochemicals, dyes, and food science. The resulting database is an invaluable resource to the greater research and high-performance computing communities, enabling breakthroughs in computational materials discovery.

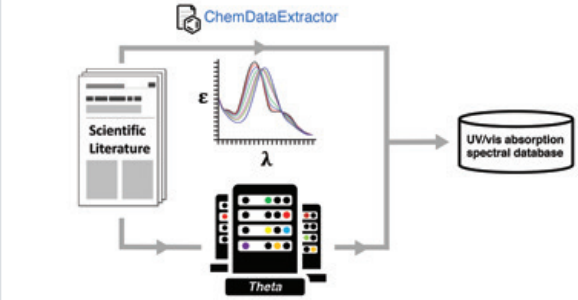
PUBLICATIONS

Hoja, J., H.-Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio, and A. Tkatchenko. "Reliable and Practical Computational Description of Molecular Crystal Polymorphs," *Science Advances* (January 2019), AAAS.

Materials Science | Simulation, Data, Learning

Data-Driven Materials Discovery for Optoelectronic Applications

PI Jacqueline Cole, University of Cambridge
AWARD ADSP
HOURS Theta: 468,750 Node-Hours



Auto-generating a UV/vis absorption spectral database via a dual experimental and computational chemical data pathway using the Theta supercomputer. *Image: Jacqueline Cole and Ulrich Mayer, University of Cambridge*

New materials are needed to drive advances in solar cells, data storage technologies, and other optoelectronic applications. With this ADSP project, researchers have developed data-driven, materials-by-design capabilities to accelerate the discovery of novel materials for targeted applications.

CHALLENGE A majority of functional materials are discovered by trial-and-error methods, but this unpredictable approach can present a bottleneck to technological innovation. To overcome this issue, the ADSP team is using systematic molecular design and engineering strategies to develop algorithms that mine massive chemical datasets to enable property prediction and materials discovery.

APPROACH The workflow for this project involves four key steps: data extraction, data enrichment, materials prediction, and experimental validation. Using text-mining tools, the researchers extract data from scientific literature to build a comprehensive repository of materials properties. To enrich the dataset, the team uses ALCF supercomputing resources to perform high-throughput electronic structure calculations, which produce paired quantities of experimental and computational data. The researchers then employ analytical methods to determine patterns in data that can be used to predict structure-function relationships for new materials. Finally, the team carries out experimental work to validate the candidate materials.

RESULTS As detailed in a paper in *Scientific Data* (2019), the researchers developed a pipeline to auto-generate a comprehensive database of ultraviolet-visible (UV-vis) absorption spectra. They used their text-mining tool, ChemDataExtractor, on more than 400,000 scientific documents to identify 18,309 experimentally determined UV-vis absorption maxima. The team found large numbers of paired quantities of experimental and computational physical properties, laying the path for reliable *in silico* calculations of additional optical properties.

The team recently extended the application of their optoelectronics workflow to additional materials, including magnetic and superconducting compounds and battery chemicals. In a paper published in *npj Computational Materials*, the researchers successfully reconstructed the phase diagrams of well-known magnetic and superconducting materials, and demonstrated the ability to predict the phase-transition temperatures of novel compounds. For the battery materials study, they published a paper in *Scientific Data* (2020), detailing the creation of a database that contains over 250,000 records for a vast range of chemicals.

IMPACT The team's use of data mining, in conjunction with large-scale simulations and experiments, offers a novel approach to advance the design and discovery of new functional materials. In addition, the project's development of open-source databases and data-extraction software tools will help accelerate materials discoveries by removing the hurdle of manual database creation.

PUBLICATIONS

E. J. Beard, G. Sivaraman, Á. Vázquez-Mayagoitia, V. Vishwanath, and J. M. Cole. "Comparative Dataset of Experimental and Computational Attributes of UV/vis Absorption Spectra." *Scientific Data* (December 2019), Springer Nature.

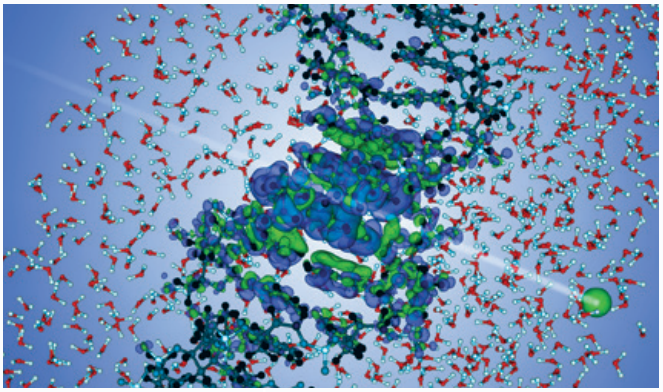
C. J. Court, and J. M. Cole. "Magnetic and Superconducting Phase Diagrams and Transition Temperatures Predicted Using Text Mining and Machine Learning." *npj Computational Materials* (March 2020), Springer Nature.

Huang, S., and J. M. Cole. "A database of battery materials auto-generated using ChemDataExtractor." *Scientific Data* (August 2020), Springer Nature.

Materials Science | Simulation

Electronic Stopping Simulation of Complex Systems

PI Yosuke Kanai, University of North Carolina at Chapel Hill
AWARD INCITE
HOURS Theta: 2,808,000 Node-Hours



Rendering of solvated DNA undergoing proton irradiation. *Image: Christopher C. Shepard, University of North Carolina at Chapel Hill*

Electronic stopping describes the dynamical transfer of kinetic energy from energetic charged particles (e.g. protons) to electrons in a target matter, consequently inducing massive electronic excitations therein. Elucidation of this phenomenon as it occurs in various systems under ion irradiation contributes to impactful breakthroughs in a number of modern technologies. A team of researchers is developing and using predictive simulation methods to model and understand electronic stopping dynamics in DNA at the molecular level due to their importance in various applications such as proton-beam cancer therapy.

CHALLENGE Understanding the electronic excitation dynamics of DNA under ion irradiation such as those of protons, alpha-particles, and carbon ions in the context of emerging ion beam cancer therapy and comparing to the electronic excitations under typical photon irradiation require a detailed atomistic model of water-solvated DNA as in the physiological condition. While quantum-mechanical simulations can provide such detailed description, complexity of the chemical system requires new innovations in massively parallel time-dependent electron dynamics simulation.

APPROACH This project continues to develop new approaches for a highly-scalable implementation of real-time, time-dependent density functional theory using the massively parallelized Qb@ll electronic-structure code. Many thousands of processors in the Theta system are used to simulate the quantum-mechanical electronic response of complex systems to the ion irradiation.

RESULTS In a recent paper published in *Journal of Physical Chemistry Letters*, the researchers describe recent progresses they have made in simulating and understanding electronic stopping phenomena in a variety of systems. They describe how their quantum-mechanical method is used to investigate electronic stopping dynamics in relatively simple solids like silicon and aluminum, and how the work has been extended further to study complex systems like solvated DNA under ion irradiation in recent years.

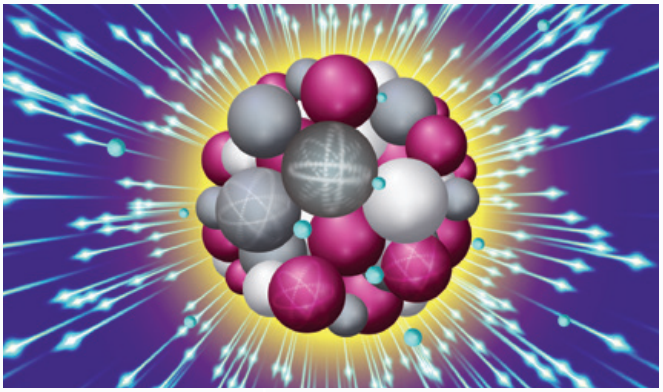
IMPACT Electronic stopping of protons and other light ions in solvated DNA, particularly in the context of ion-beam cancer therapies, represents a problem of great societal significance. The simulations performed in this work will provide much-needed insights into how DNA-damaging electronic excitation might differ under proton irradiation and traditional x- and gamma-ray irradiation in radiation oncology.

PUBLICATIONS
Yost, D. C., Y. Yao, and Y. Kanai. "First-Principles Modeling of Electronic Stopping in Complex Matter under Ion Radiation," *Journal of Physical Chemistry Letters* (December 2019), ACS Publications.

Materials Science | Simulation

Imaging and Controlling Elemental Contrast of Nanocluster in Intense X-ray Pulses

PI Phay Ho, Argonne National Laboratory
AWARD ALCC
HOURS Mira: 5,625,000 Node-Hours



An intense x-ray pulse scatters off a sucrose cluster (red, white and gray balls are oxygen, carbon and hydrogen atoms) resulting in ejected electrons (blue balls) and structural deformation. *Image: Stacy Huang*

Scientists have long pursued the ability to see the structure of a single, free-form molecule at atomic resolution, what many call the "holy grail" of imaging. One potential method involves aiming extremely short, highly intense x-ray free-electron laser (XFEL) pulses at a sample material, but this imaging technique can also destroy its target very quickly. With help from ALCF supercomputers, an Argonne-led research team is working to better understand how XFEL pulses interact with their targets and to discover optimal parameters for ultrafast imaging experiments.

CHALLENGE While intense XFEL pulses hold great promise for imaging processes in nanoscale and biological systems with atomic resolution, reaching nanometer resolution with single-particle imaging experiments remains a challenge. Models of radiation-induced damage in single-shot experiments have included significant approximations and have been limited to small system sizes. Using a combination of experimental and computational methods, the researchers produced calculated and experimental diffractive imaging data that revealed clear deviations from the commonly applied linear scattering models.

APPROACH The cross-disciplinary team studied single nanoclusters of sucrose at the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory to determine how different parameters can affect an XFEL experiment's outcome. The researchers then compared the experimental data with calculations run on the ALCF's Mira supercomputer using a parallel Monte Carlo-molecular dynamics approach. This involved performing a large ensemble of simulations that tracked 185,193 sucrose molecules (42 million particles) interacting with an XFEL pulse.

RESULTS In a paper published in *Nature Communications*, the researchers illuminated the complex interplay of the imaging process with the rapidly changing transient electronic structures in XFEL experiments. The team's large-scale dynamic scattering calculations, with and without radiation damage as a function of photon energy, showed that x-ray induced processes have a significant impact on the scattering cross section even at high energies that cannot be neglected. The ability to model complex imaging experiments of organic (biological) samples on the atomistic level, as demonstrated for the first time here, will allow for more precise and efficient experiments at XFEL facilities.

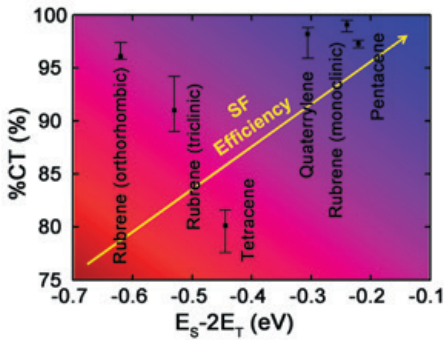
IMPACT The team's research demonstrates how supercomputing resources and computational models can help facilitate the optimization of parameters for ultrafast imaging experiments and predict optimal imaging conditions. This work is essential in guiding future XFEL experiments with heterogeneous materials and establishing the applied methodology as an effective large-scale computational tool for the new research frontier of ultrafast x-ray science.

PUBLICATIONS

Ho, P. J., B. J. Daurer, M. F. Hantke, J. Bielecki, A. Al Haddad, M. Bucher, G. Doumy, K. R. Ferguson, L. Flückiger, T. Gorkhover, B. Iwan, C. Knight, S. Moeller, T. Osipov, D. Ray, S. H. Southworth, M. Svenda, N. Timneanu, A. Ulmer, P. Walter, J. Hajdu, L. Young, F. R. N. C. Maia, and C. Bostedt. "The Role of Transient Resonances for Ultra-Fast Imaging of Single Sucrose Nanoclusters." *Nature Communications* (January 2020), Springer Nature.

Materials and Interfaces for Organic and Hybrid Photovoltaics

PI Noa Marom, Carnegie Mellon University
AWARD INCITE
HOURS Theta: 5,626,000 Node-Hours
Mira: 21,875,000 Node-Hours



Known and predicted singlet fission materials ranked with respect to a 2D descriptor based on maximizing the energy conservation criterion and the degree of charge transfer character of the singlet exciton. *Image: Xingyu Liu, Noa Marom, and Xiaopeng Wang, Carnegie Mellon University; Cameron Cook and Bohdan Schatschneider, California State Polytechnic University, Pomona*

Organic and hybrid organic-inorganic photovoltaics are a promising alternative to silicon solar cells for certain applications. The discovery and design of such materials pose steep demands, making predictive computer simulations an ideal tool for development. An international research team from Carnegie Mellon University, Duke University, Aalto University, University of Potsdam, Technical University Graz, and Technical University Munich leveraged ALCF supercomputers to revolutionize not only solar energy, but materials discovery at large via new computational paradigms.

CHALLENGE This project aimed to advance—via computer-aided discovery and design of new materials, crystal forms, and enhanced functional interfaces—the efficiency of organic and hybrid organic-inorganic photovoltaic and electronic devices. Moreover, it sought to advance the general state of the art in computational materials science through the development of a general first-principles framework for electronic structure simulations, structure prediction, and computational design of materials and interfaces.

APPROACH To guide material discovery, the INCITE research team paired the density-functional code FHI-aims, which determines electronic structure, with a set of optimization algorithms that predict materials structure, including the genetic algorithm (GA) GAtor and its associated structure-generation package, Genarris, which generates random molecular crystal structures with physical constraints. GAs generally operate by mimicking the evolutionary process. The target property is mapped onto a fitness function. Structures with a high fitness have an increased probability to "mate," whereby a crossover operator combines the structural "genes" of the parent structures to produce a child structure, which is then added to the population. The process repeats iteratively until an optimum is found. The calculations ran on Mira and Theta.

RESULTS A paper published in *Nature Chemistry* demonstrated the modulation of the electronic structure (and, as a result, the optical properties) of organic semiconducting optoelectronic building blocks incorporated between layers of perovskites. Energy and charge transfers between adjacent organic and inorganic layers were shown to be fast and efficient due to an atomically flat interface and the ultrasmall interlayer distance of the perovskite materials.

IMPACT This project is helping create breakthroughs in the ways we generate solar energy, while also broadly advancing materials and data sciences. It has fundamentally deepened our understanding of atomic-level structural property relations and helped prepare versatile, parallelized codes ready for the next generation of supercomputers. Moreover, the publicly available datasets produced establish a bedrock for further simulation-based materials discovery.

PUBLICATIONS

Gao, Y., E. Shi, S. Deng, S. B. Shiring, J. M. Snaider, C. Liang, B. Yuan, R. Song, S. M. Janke, A. Liebman-Peláez, P. Yoo, M. Zeller, B. W. Boudouris, P. Liao, C. Zhu, V. Blum, Y. Yu, B. M. Savoie, L. Huang, and L. Dou. "Molecular Engineering of Organic-Inorganic Hybrid Perovskites Quantum Wells," *Nature Chemistry* (November 2019), Springer Nature.

Tom, R., T. Rose, I. Bier, H. O'Brien, Á. Vázquez-Mayagoitia, and N. Marom. "Genarris 2.0: A Random Structure Generator for Molecular Crystals," *Computer Physics Communications* (May 2020), Elsevier.

Materials Science | Simulation

Metascalable Layered Materials Genome

PI Aiichiro Nakano, University of Southern California
AWARD Aurora ESP
HOURS Theta: 78,125 Node-Hours

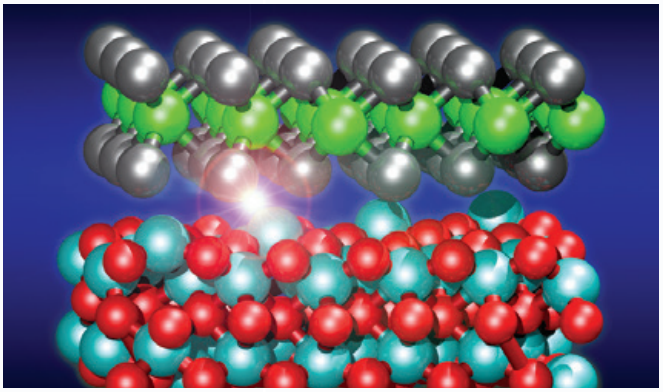


Photo-excitation of WSe2 monolayer on Al2O3 substrate. Green, gray, cyan and red spheres represent W, Se, Al and O atoms, respectively. *Image: Thomas Linker, Ken-ichi Nomura, and Subodh Tiwari, University of Southern California*

Functional layered materials possess outstanding electronic, optical, magnetic, and chemical properties that can be fine-tuned to design new materials for targeted applications. The promise of novel layered materials helped inspire the Materials Genome Initiative, a multi-agency effort to create a new era of policy, resources, and infrastructure that support the work of U.S. institutions to discover, manufacture, and deploy advanced materials faster but at a reduced cost. Researchers from the University of Southern California are preparing to use the ALCF's upcoming Aurora system to advance this initiative by using exascale-powered simulations to aid in the synthesis of layered materials and elucidate new property relationships.

CHALLENGE Layered materials are free from the lattice matching constraints that severely limit conventional materials design, leading to the possibility of tuning materials properties in desired ways by building van der Waal heterostructures composed of unlimited combinations of layers. Exploring this vast combinatorial materials design space poses an enormous challenge that necessitates the use of future exascale supercomputers.

APPROACH The team will use their scalable computational synthesis and characterization software to provide function-property-structure relationships and time-resolved information to guide the experimental synthesis of new layered materials. The researchers will validate their nonadiabatic quantum molecular dynamics (NAQMD) and reactive molecular dynamics (RMD) simulations against ultrafast pump-probe experiments at SLAC National Accelerator Laboratory's Linac Coherent Light Source (LCLS). The team is using current ALCF supercomputers and early exascale hardware and software tools to prepare for future simulation runs on Aurora.

RESULTS As part of the team's effort to scale their software for Aurora, they developed a series of algorithmic and computational techniques to speed up their linear-scaling quantum molecular dynamics simulation code based on divide-and-conquer density functional theory. Running on Theta, the optimized simulation engine allowed them to establish the significant effect of dielectric substrate on the dynamics of a layered material upon electronic excitation for the first time. Quantitative understanding enabled by such simulations will be catapulted to the next level on Aurora, likely filling a fundamental knowledge gap in more complex functional materials and emergent phases. A paper detailing the team's findings received the best paper award at the HPC Asia 2020 Conference.

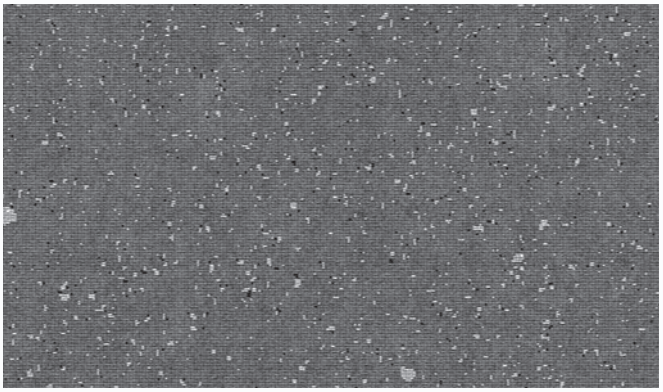
IMPACT The team's exascale simulations will provide critical information on the function-property-structure relationships of novel layered materials. Ultimately, this research will help guide the synthesis of new materials engineered for targeted applications, such as batteries, catalysts, and solar cells.

PUBLICATIONS

Tiwari, S. C., A. Krishnamoorthy, P. Rajak, P. Sakdhnagool, M. Kunaseth, F. Shimojo, S. Fukushima, A. Nakano, Y. Luo, R. K. Kalia, K. Nomura, and P. Vashishta. "Quantum Dynamics at Scale: Ultrafast Control of Emergent Functional Materials" *Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region* (January 2020), ACM.

Modeling the Response of Fusion Plasma Facing Components

PI Brian Wirth, Oak Ridge National Laboratory
AWARD ALCC
HOURS Mira: 5,000,000 Node-Hours



Snapshot from a simulation of helium atoms beneath a tungsten surface to 100 eV helium plasma for 840 ns (fluence of 8.4 x 10^{18} m^{-2}). *Image: K. D. Hammond, University of Missouri, et al., Nuclear Fusion (2019) vol. 59, no. 6.*

Fusion as a practical energy source requires greater knowledge of plasma-surface interactions and materials engineering design of components to withstand the extreme heat and particle flux exposure conditions of a fusion power plant. This project, led by researchers from Oak Ridge National Laboratory, leveraged large-scale molecular dynamics (MD) simulations to model difficult-to-study phenomena that have limited our understanding of fusion.

CHALLENGE A key complication to understanding fusion is the formation of "fuzz" on the surface of tungsten (that is, the primary plasma-facing divertor material in ITER) following exposure to material-destabilizing energetic helium ions. The mechanism driving this formation is challenging to probe, both experimentally and computationally. Large-scale MD simulations—relatively free of finite-size effects—were thus used to the mechanisms of helium transport and tungsten surface deformation.

APPROACH MD simulations, rendered with the LAMMPS code, were used to model helium flux effects for two surface orientations. Lateral dimensions of up to 100 nanometers were considered to reduce finite-size effects, such as the formation of bubbles spanning the full width of the supercell. Four distinct helium fluxes were considered, ranging from 10^{25} to 10^{28} $m^{-2}s^{-1}$, with simulation times ranging from 500 nanoseconds up to 2.5 microseconds, depending on the system size. A temperature of 933 Kelvin was chosen so as to be at the low end of the fuzz-forming regime. The helium insertion depth was sampled from a distribution consistent with 100 electronvolt incident energy, which results in 90 percent of helium atoms being implanted within 5 nanometers of the surface.

RESULTS Pronounced changes across three orders of magnitude of helium flux were observed for helium retention and helium bubble depth distribution, with deep bubbles tending to grow larger and create more prominent surface features over time. These results suggest that nearly all prior reported molecular simulations suffer from finite-size effects and that helium flux is a very important parameter in determining the behavior of helium in plasma-facing components.

IMPACT The simulations provide much-needed benchmarks for larger-scale models, which are necessary to better understand the mechanisms of helium transport, retention, and agglomeration. This, in turn, will help to identify material design strategies and potentially mitigate issues related to the extreme conditions of a fusion energy environment.

PUBLICATIONS

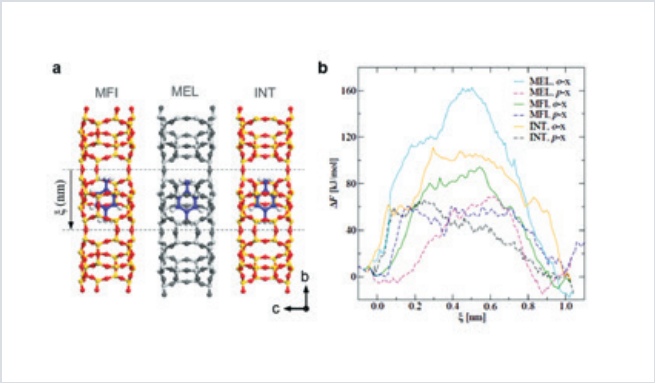
Hammond, K. D., D. Maroudas, and B. D. Wirth. "Theoretical Model of Helium Bubble Growth and Density in Plasma-Facing Metals," *Scientific Reports* (February 2020), Springer Nature.

Hammond, K. D., I. V. Naeger, W. Widanagamaachchi, L.-T. Lo, D. Maroudas, and B. D. Wirth. "Helium Flux Effects on Bubble Growth and Surface Morphology in Plasma-Facing Tungsten from Large-Scale Molecular Dynamics Simulations," *Nuclear Fusion* (May 2019), IOP Publishing.

Materials Science | Simulation, Learning

Predictive Modeling and Machine Learning for Functional Nanoporous Materials

PI J. Ilja Siepmann, University of Minnesota
AWARD ALCC
HOURS Theta: 620,000 Node-Hours



Diffusivity of *p*-xylene and *o*-xylene molecules through zeolite pores. (a) Projection of MFI, MEL and MEL/MFI interfacial pores with *p*-xylene at the center. (b) Free energy barriers for *p*-xylene and *o*-xylene transport. *Image: Evgenii Fetisov and J. Ilja Siepmann, University of Minnesota*

Nanoporous materials, such as zeolites, tailored for specific separation and catalytic processes have the potential to significantly improve energy efficiency in the chemical, biorenewable, and petrochemical industries. An interdisciplinary team led by researchers from the University of Minnesota is using predictive modeling and machine learning techniques on ALCF computing resources to accelerate the discovery and design of nanoporous materials for a variety of energy-related applications.

CHALLENGE While commercially available zeolite membrane technologies are already enabling improved energy efficiency in gas separations and other industrial uses, efforts are underway to explore their application to hydrocarbon separations. In particular, mixtures of hydrocarbon isomers, like the three xylene isomers, are difficult to separate by distillation and their production could become more efficient by employing membranes that can operate with sufficiently high flux and selectivity at the temperatures and pressures required for membrane–reactor configurations. The zeolite MFI, a framework for widely used industrial catalysts, has shown promise as a thin-film membrane, but the crystal structure of 2D-MFI nanosheets and their relationship to separation performance remain elusive.

APPROACH In a recent study, the multi-institutional team carried out transmission electron microscopy experiments that revealed intergrowths of the zeolite MEL exist within 2D-MFI. To assess the role of MEL intergrowths, the researchers performed first-principles molecular dynamics simulations on ALCF supercomputers to model the transport of xylene isomers through MFI-MEL nanosheets.

RESULTS The team's atomistic simulations revealed that commensurate knitting of 1D-MEL within 2D-MFI creates more selective pores compared to pristine MFI nanosheets. Due to slight variations in pore cross section, transport of o-xylene through MEL and MEL-MFI interface pores can only be accomplished via significant framework distortions, whereas transport of p-xylene is less hindered which, in turn, leads to dramatic differences in permeability. In permeation experiments using an industrially relevant feed, the researchers achieved a separation factor of 60, which far exceeds previous experimental results. The observation of these intergrowths opens the door to the development of ultra-selective zeolite membranes. The team's findings were published in *Nature Materials*.

IMPACT The team's research into novel zeolite materials is helping to advance the development of membrane technologies that can provide more energy-efficient separations processes for industry. More broadly, their work to improve the understanding and selection of nanoporous materials for energy applications, such as the production of biofuel and petroleum products, could lead to significant economic and environmental benefits.

PUBLICATIONS

Kumar, P., D. W. Kim, N. Rangnekar, H. Xu, E. O. Fetisov, S. Ghosh, H. Zhang, et al. "One-Dimensional Intergrowths in Two-Dimensional Zeolite Nanosheets and Their Effect on Ultra-Selective Transport." *Nature Materials* (February 2020), Springer Nature.

Physics | Simulation

Hadronic Light-by-Light Scattering and Vacuum Polarization Contributions to the Muon Anomalous Magnetic Moment

PI Thomas Blum, University of Connecticut
AWARD ALCC
HOURS Mira: 10,125,000 Node-Hours

A typical diagrammatic representation of the hadronic light-by-light scattering contribution with Argonne's Mira supercomputer in the background. *Image: Luchang Jin, University of Connecticut*

The Standard Model has been the definitive description of particle physics for the past 50 years, but it is being challenged by the discrepancy between the measured and predicted values of the magnetic moment of a particle called the muon. A multi-institutional research team has used Argonne's Mira supercomputer to quantify and reduce the largest uncertainties associated with this value in order to obtain the most precise calculation of the anomaly. Their results, combined with data soon to come from Fermilab's Muon g-2 experiment, may lead to the discovery of new physics beyond the Standard Model.

CHALLENGE The muon's magnetic moment describes how this fundamental particle interacts with a magnetic field; it depends on all particles that can couple to the muon—including as-yet-undiscovered particles. The muon moment has been both measured in experiments and calculated theoretically, but those two values do not quite match up—hinting at the existence of new physics. A higher-precision measurement could erase the discrepancy, so minimizing uncertainties in the theoretical calculation is necessary to reach a more definite conclusion. The largest uncertainty in the calculation comes from particles that interact through the strong force, known as hadronic contributions. Since these contributions cannot be solved with conventional perturbative methods at low-energy scales, researchers previously resorted to experimental data or approximations with large uncertainties.

APPROACH Over the course of four years and multiple ALCC allocations, the researchers leveraged the ALCF's Mira supercomputer to compute hadron contributions, namely light-by-light scattering and vacuum polarization, from first principles using large-scale lattice quantum chromodynamics simulations. The team's computations of hadronic light-by-light scattering removed the dominant systematic uncertainties by analyzing the continuum limit and the infinite volume limit for the quantum electrodynamics portion (muon and photons) of the calculations.

RESULTS In a paper published in *Physical Review Letters*, the researchers detailed the first-ever result for the hadronic light-by-light scattering contribution to the muon anomaly, controlling for all errors. The team produced a range of estimates for the effect and found that its contribution was much too small to explain the anomaly.

IMPACT These results suggest that the hadronic light-by-light contribution is not the cause of the muon's unexpectedly strong magnetic moment. Comparison with measurements from Fermilab's Muon g-2 experiment could potentially lead to the discovery of new physics.

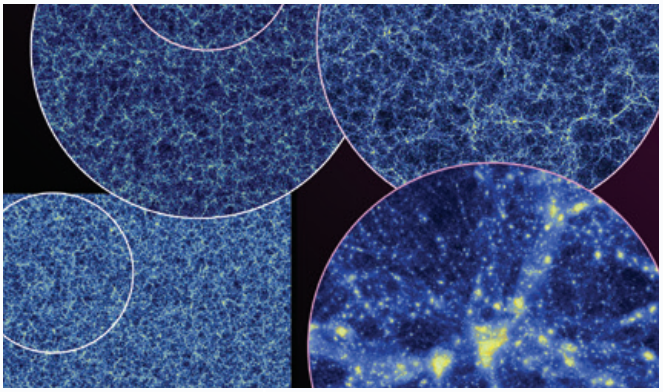
PUBLICATIONS

Blum, T., N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner. "Hadronic Light-by-Light Scattering Contribution to the Muon Anomalous Magnetic Moment from Lattice QCD." *Physical Review Letters* (April 2020), APS.

Physics | Simulation, Data

The Last Journey

PI Katrin Heitmann, Argonne National Laboratory
AWARD ALCC
HOURS Mira: 50,000,000 Node-Hours



Dark-matter-dominated halos from a small region of the simulation. The radius of the spheres is proportional to halo mass; the dominant halo is the simulation's largest, at ~6×10^{15} solar masses. *Image: Argonne National Laboratory*

This project, led by Argonne National Laboratory researchers, is one of the world's five largest cosmological simulations of sufficient resolution and volume to permit the generation of detailed sky maps across multiple wavebands that are targeted to upcoming cosmological surveys. Running on the entirety of the Mira system, the simulation modeled the lifespan of the universe to help answer some of science's deepest questions, including the origins of dark matter and dark energy and what it means for the universe to expand at an accelerating pace.

CHALLENGE The research team structured the simulation to begin 50 million years after the Big Bang, with conditions that agree with the most up-to-date cosmological theories. Billions of years of evolution between then and now were subsequently modeled in order to create a high-resolution model of what a large portion of the universe should look like at present day.

APPROACH The project was implemented using the Hardware/Hybrid Accelerated Cosmology Code (HACC) simulation and analysis framework on the full Mira system. It evolved more than 1.24 trillion particles to resolve the cosmological structures which host faint galaxies that will be observed by the Legacy Survey of Space and Time (LSST) project when it is carried out at the Vera Rubin Observatory. Cosmological parameters chosen to be consistent with the results from the Planck satellite. Analysis outputs were generated such that synthetic galaxy catalogs could be constructed using a semi-analytic modeling approach in post-processing.

RESULTS The detailed history of the evolution of cosmological structures is now being processed to create synthetic sky maps for optical and cosmic microwave background surveys. Results are being documented in a sequence of papers to be published in *The Astrophysical Journal Supplement Series*.

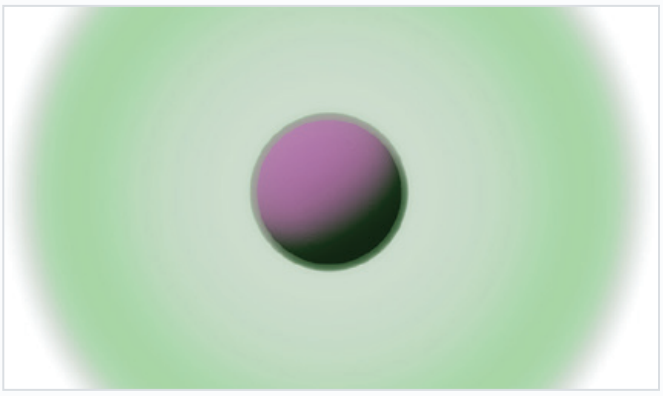
IMPACT This simulation was designed to address numerous fundamental questions in cosmology; the data produced are essential for enabling the refinement of existing predictive tools and aid the development of new models, impacting both ongoing and upcoming cosmological surveys, including the Dark Energy Spectroscopic Instrument (DESI), the LSST, SPHEREx, and the "Stage-4" ground-based cosmic microwave background experiment (CMB-S4).

PUBLICATIONS

Heitmann, K., N. Frontiere, E. Rangel, P. Larsen, A. Pope, I. Sultan, T. Uram, S. Habib, H. Finkel, D. Korytov, E. Kovacs, S. Rizzi, and J. Insley. "The Last Journey. I. An Extreme-Scale Simulation on the Mira Supercomputer," *The Astrophysical Journal Supplement Series* (submitted), IOP Publishing.

Lattice QCD

PI Paul Mackenzie, Fermilab
AWARD INCITE
HOURS Mira: 24,000,000 Node-Hours
Theta: 1,600,000 Node-Hours



An antiquark (magenta) inside a cloud of gluons (green). USQCD researchers discovered a new and unambiguous way to decide how much of the cloud's energy should be considered part of the quark mass. This idea was applied to simulation data generated with Mira to compute quark massed from first principles. *Image: Argonne National Laboratory*

Lattice quantum chromodynamics (QCD) addresses fundamental questions in high energy and nuclear physics critical for a number of DOE Office of Science milestones. This project, led by researchers from the U.S. Lattice Quantum Chromodynamics (USQCD) Collaboration, leveraged ALCF leadership-class computing systems to perform calculations that are necessary in the search for physics beyond the Standard Model.

CHALLENGE A central objective of this work was to generate gauge configurations that will enable QCD researchers to push the search for new effects in flavor physics at the highest energy levels every considered.

Additionally, this work calculated the fluctuations in conserved quantities in the quark gluon plasma produced at Brookhaven National Laboratory's Relativistic Heavy Ion Collider (RHIC) and CERN's Large Hadron Collider.

APPROACH This project leveraged DOE leadership computing resources to advance research in QCD. Beginning with lattice QCD, a non-perturbative approach to QCD, the team operated two kinds of workflows: gauge-field generation via use of hybrid molecular dynamics, and evaluation of physical observables via use of a sparse-matrix conjugate-gradient solver. A variety of C- and C++-based codes, including MILC and CPS, were employed on and optimized for Mira and Theta.

RESULTS The muon is an elementary particle that can be likened to a comparatively massive electron. Its magnetic moment is an intrinsic property that has been measured experimentally and computed with a high degree of accuracy. However, experimental measurement and computed value disagree by three standard deviations, which is substantial enough to suggest the need for a new theory. As detailed in a paper published in *Physical Review Letters*, the project yielded the first result for the hadronic light-by-light scattering contribution to the muon magnetic moment. Relatedly, the researchers calculated the contribution to the muon anomalous magnetic moment hadronic vacuum polarization from the connected diagrams of up and down quarks, omitting electromagnetism. These results were published in *Physical Review D*.

IMPACT In addition to advancing our knowledge of QCD, this work generated data required throughout DOE's Intensity Frontier program, which includes Fermilab's Muon g-2 and neutrino-scattering experiments, while also informing experiments at RHIC and CERN.

PUBLICATIONS

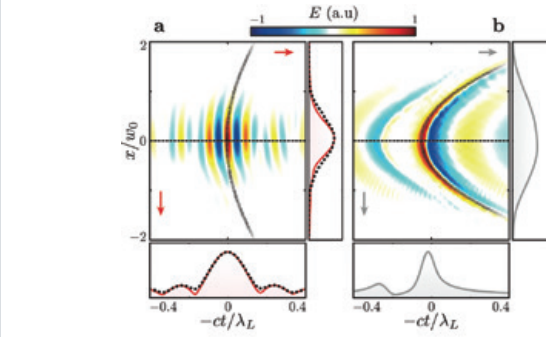
Blum, T., N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner. "Hadronic Light-by-Light Scattering Contribution to the Muon Anomalous Magnetic Moment from Lattice QCD," *Physical Review Letters* (April 2020), APS.

Davies, C. T. H., C. DeTar, A. X. El-Khadra, E. Gámiz, S. Gottlieb, D. Hatton, A. S. Kronfeld, J. Laiho, G. P. Lepage, Y. Liu, P. B. Mackenzie, C. McNeile, E. T. Neil, T. Primer, J. N. Simone, D. Toussaint, R. S. Van de Water, and A. Vaquero. "Hadronic-Vacuum-Polarization Contribution to the Muon's Anomalous Magnetic Moment from Four-Flavor Lattice QCD," *Physical Review D* (February 2020), APS.

Physics | Simulation

PLASM-IN-SILICO: HPC Modeling of High-Intensity Laser-Solid Interaction

PI Jean-Luc Vay, Lawrence Berkeley National Laboratory
AWARD INCITE
HOURS Theta: 600,000 Node-Hours



Spatio-temporal field of an attosecond pulse produced by a relativistic plasma mirror. (a) Experimental spatio-temporal reconstruction of the electric field of an attosecond pulse generated from the superposition of Doppler harmonics 9 to 14, in the plane of the PM surface. (b) Electric field obtained from PIC simulation with laser intensity of 10^{22}W/cm^2. *Image: Henri Vincenti, CEA Saclay*

High-power lasers have laid the groundwork for a new branch of physics—ultra-high intensity (UHI) physics—to facilitate the study of light-matter interactions at extreme light intensities. Nowadays, petawatt (PW) lasers can deliver intensities exceeding 10^{22}W/cm^2, at which point matter turns into a strongly relativistic and out-of-equilibrium plasma. Beyond the fundamental interest presented by the study of these complex plasma regimes, laser-plasma interactions can produce very compact and ultra-short particle/light sources with promising applications for industry and medicine.

This project, led by researchers from Lawrence Berkeley National Laboratory in the US and Commissariat à l'Energie Atomique in France, uses massively parallel particle-in-cell (PIC) simulations on ALCF supercomputers to understand complex light-plasma interactions and overcome major challenges facing UHI physics.

CHALLENGE The researchers intend to show that relativistic plasma mirrors (PMs) irradiated by PW lasers can provide a solution to three major challenges in UHI physics: (1) Can we produce high-charge compact electron accelerators with high beam-quality that will push forward the horizons of high-energy science? (2) Can we produce efficient and compact high-energy ion accelerators to democratize cancer hadron-therapy? (3) Can we overcome the current limitations of optical laser technology and reach extreme light intensities approaching the Schwinger limit (10^{29}W/cm^2), beyond which novel regimes of quantum electrodynamics arise?

APPROACH The researchers are devising solutions in-silico with massively parallel simulations on the ALCF's Theta system using the PIC WarpX code (developed under the auspices of DOE's Exascale Computing Project) and the high-performance library PICSAR (for the QED modules). To this end, their work leverages recent transformative developments in the first-principles simulation of UHI laser-plasma interactions that enabled the 3D, high-fidelity modeling of plasma mirror sources. Close collaborations with teams at PW laser facilities help enable experimental validation of the devised solutions.

RESULTS Ptychography is a powerful lensless spatial imaging technique that fails to provide any information about the temporal phase and amplitude of a PM. To overcome this, the research team proposed a transient plasma optical grating, which acts as a reflection function that varies in a predictable manner at the attosecond scale. This is achieved via a frequency-doubled secondary laser pulse focused at a slightly different angle; the spatio-temporal amplitude and phase of the source field can be obtained from the resulting dynamical diffraction pattern. As detailed in a paper under review at Nature Physics, the researchers used PIC simulations performed on Theta to assess the validity of this proposed "dynamical ptychography" method.

IMPACT The team provided the very first spatio-temporal measurements of the electromagnetic field of a high-power laser reflected on a curved relativistic PMs. These measurements show that PMs are indeed promising candidates to reach the Schwinger limit, and their method will be instrumental in the quest to achieve the highest intensities.

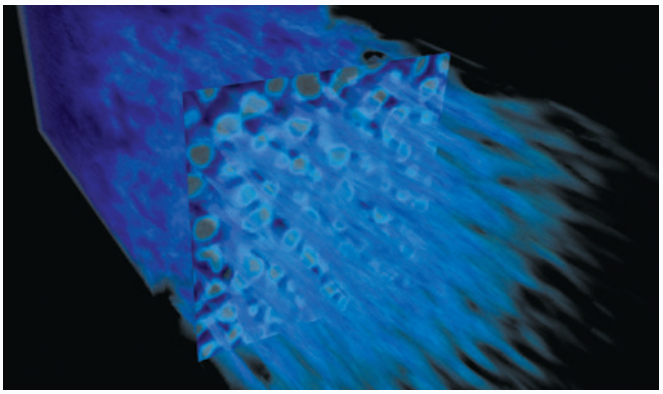
PUBLICATIONS

Chopineau, L., A. Denoeud, A. Leblanc, E. Porat, P. Martin, H. Vincenti, and F. Quéré. "Plasma Mirrors as a Path to the Schwinger Limit," *Nature Physics* (under review), Springer Nature.

Physics | Simulation, Data

Studying Astrophysical Particle Acceleration in HED Plasmas

PI Frederico Fiuza, SLAC National Accelerator Laboratory
AWARD ALCC, Director's Discretionary
HOURS Mira: 3,125,000 Node-Hours
Theta: 130,000 Node-Hours



Laser-driven experiments and numerical simulations now show that small-scale turbulence produced at the shock can be key in the initial acceleration of electrons from supernovas. *Image: Frederico Fiuza, SLAC National Accelerator Laboratory*

When stars explode as supernovas, they produce powerful shock waves in the high-energy density (HED) plasmas that surround them. These powerful shocks can act as particle accelerators that blast streams of particles, called cosmic rays, out into the universe at nearly the speed of light. With help from ALCF supercomputers, researchers from SLAC National Accelerator Laboratory have devised a new way to study the inner workings of these mysterious astrophysical shock waves by creating a scaled-down version of the shock in a laboratory.

CHALLENGE With the closest supernovas thousands of light years away, it is very difficult to study astrophysical shocks directly from observations. However, with access to world-class supercomputers at the ALCF and the powerful lasers at the National Ignition Facility (NIF) at Lawrence Livermore National Laboratory, the SLAC team developed a new experimental setup that could mimic the shock wave conditions of supernova remnants.

APPROACH For their work at the ALCF, the researchers used 3D, massively parallel, particle-in-cell (PIC) simulations to explore the physics of particle acceleration in HED plasmas. In particular, based on laser-ablated plasma profiles obtained from data generated at NIF, the team was able to perform simulations to characterize the formation of collisionless shocks and study particle injection that occurs during these shocks. The team's computations were critical to informing and validating their experimental work.

RESULTS In their laser-driven experiments at NIF, the team recreated for the first time in the lab the phenomenon that kick-starts particle acceleration in supernova shocks. The scientists observed that the shocks accelerated a group of electrons to more than 100 times the average energy of the particles, with the fastest electrons reaching relativistic velocities. The maximum electron energy the scientists observed was consistent with estimates for acceleration produced by scattering off moving magnetic fields based on the measured shock properties. Numerical simulations performed at the ALCF confirmed this model. They further revealed that these magnetic fields were associated with small-scale turbulence produced within the shock wave itself. The researchers concluded that this mechanism can help accelerate electrons in supernova shocks. Their findings were detailed in a paper published in *Nature Physics*.

IMPACT The team's results shed new light on the long-standing question of how cosmic accelerators work. Researchers can use these findings to produce and test better models of acceleration in supernova shocks. This understanding can also help researchers develop better particle accelerators on Earth for applications in science, industry, and medicine.

PUBLICATIONS

Fiuza, F., G. F. Swadling, A. Grassi, H. G. Rinderknecht, D. P. Higginson, D. D. Ryutov, C. Bruulsema, et al. "Electron Acceleration in Laboratory-Produced Turbulent Collisionless Shocks." *Nature Physics* (June 2020), Springer Nature.

Physics | Simulation

Towards a Definitive Model of Core-Collapse Supernova Explosions

PI Adam Burrows, Princeton University
AWARD INCITE
HOURS Theta: 2,000,000 Node-Hours

Fractured surface of the exploding core of a 16 solar-mass star computed on Theta. Bluish outer veil is the shock wave traveling ~ 10,000 kilometers per second. The surface is colored by electron fraction: Purple is proton rich matter; blue is neutron rich. *Image: Argonne National Laboratory*

Core-collapse supernovae dramatically announce the death of massive stars and the birth of neutron stars. These violent explosions, which produce the highest densities of matter and energy in the universe, are responsible for creating most of the elements in nature. A fundamental theoretical understanding of such explosions is needed to advance research in nuclear and particle physics, and inform the interpretation of data from large-scale experiments. To shed light on this mysterious cosmological phenomenon, a research team led by Princeton University is using ALCF supercomputers to address whether and how 3D supernova explosion models differ from their 2D counterparts.

CHALLENGE 2D simulations of supernovae have supported the theory that capturing a small fraction of the neutrinos emitted during collapse powers supernova explosions, but detailed 3D calculations proving this paradigm are lacking. With the power of leadership-class supercomputers and continued advances in software, researchers now have the capabilities to tackle this longstanding challenge in nuclear astrophysics.

APPROACH To carry out this research, the team is using FORNAX, their highly scalable, 3D radiation-hydrodynamics code. By addressing the transport operator with an explicit method, the code significantly reduces the computational complexity and communication overhead of traditional multidimensional radiative transfer solutions by bypassing the need for global iterative solvers. The team is running FORNAX on Theta to perform large-scale simulations aimed at determining if the neutrino mechanism is a robust explanation for supernova explosions and the stellar progenitor dependence of their observables.

RESULTS After completing and analyzing the largest suite of 3D simulations ever performed, spanning a broad range of progenitor masses and structures, the team is now focusing on a few long-term simulations to witness the achievement of the asymptotic explosion state. Simulating a full-physics 3D model for more than two seconds of physical time has never before been done, but the more massive star models seem to require such durations to reach the final explosion energy. In addition, the team has completed a comprehensive study of proto-neutron star (PNS) convection in the deep interior of the newly born neutron star. Such core convection may be implicated in the generation of pulsar magnetic fields and affects the neutrino-cooling rate of the residue. Hence, an exploration of the variety of convective behaviors experienced in the core can speak volumes about the hand-off between the explosive and the PNS phases. Both studies have been published in the *Monthly Notices of the Royal Astronomical Society*.

IMPACT The team's efforts to advance the fundamental theoretical understanding of supernova explosions will benefit ongoing research efforts to determine the origin of the elements in the universe, measure gravitational waves, and interpret laboratory nuclear reaction rate measurements in light of stellar nucleosynthesis.

PUBLICATIONS

Burrows, A., D. Radice, D. Vartanyan, H. Nagakura, M. A. Skinner, and J. C. Dolence. "The Overarching Framework of Core-Collapse Supernova Explosions as Revealed by 3D Fornax Simulations." *Monthly Notices of the Royal Astronomical Society* (January 2020), Oxford University Press.

Nagakura, H., Burrows, A., Radice, D., and Vartanyan, D. "A Systematic Study of Proto-Neutron Star Convection in Three-Dimensional Core-Collapse Supernova Simulations." *Monthly Notices of the Royal Astronomical Society* (January 2020), Oxford University Press.

Snapshot of turbulent eddies in the compression stroke during the third engine cycle (1700.5 CAD). Eddies colored by velocity magnitude in the TCC-III internal combustion engine designed by GM.
Image: Argonne National Laboratory

ALCF Projects

2020 INCITE

BIOLOGICAL SCIENCES

Computational Physical Genomics: Exploring Potential Novel Cancer Therapies

PI Allen Taflove, Northwestern University
HOURS ALCF: 1,000,000 Node-Hours

EARTH SCIENCE

Energy Exascale Earth System Model

PI Mark Taylor, Sandia National Laboratories
HOURS ALCF: 1,800,000 Node-Hours
OLCF: 1,000,000 Node-Hours

Extreme-Scale Simulations for Advanced Seismic Ground Motion and Hazard Modeling

PI Christine Goulet, University of Southern California
HOURS ALCF: 500,000 Node-Hours
OLCF: 300,000 Node-Hours

ENGINEERING

High-Speed Turbulence with Shocks over Non-Adiabatic and Flexible Walls

PI Johan Larsson, University of Maryland
HOURS ALCF: 2,000,000 Node-Hours

MATERIALS SCIENCE

Electronic Stopping Simulation of Complex Systems

PI Yosuke Kanai, University of North Carolina at Chapel Hill
HOURS ALCF: 2,000,000 Node-Hours

Towards Predictive Simulations of Functional and Quantum Materials

PI Paul Kent, Oak Ridge National Laboratory
HOURS ALCF: 1,500,000 Node-Hours
OLCF: 400,000 Node-Hours

Ultrafast Control of Functional Materials

PI Priya Vashishta, University of Southern California
HOURS ALCF: 2,000,000 Node-Hours

PHYSICS

Ab-initio Nuclear Structure and Nuclear Reactions

PI Gaute Hagen, Oak Ridge National Laboratory
HOURS ALCF: 1,000,000 Node-Hours
OLCF: 550,000 Node-Hours

Extreme-Scale Simulation of Supernovae and Magnetars from Realistic Progenitors

PI Sean Couch, Michigan State University
HOURS ALCF: 1,000,000 Node-Hours

High-Fidelity Gyrokinetic Simulation of Tokamak and ITER Edge Physics

PI Choongseock Chang, Princeton Plasma Physics Laboratory
HOURS ALCF: 1,500,000 Node-Hours
OLCF: 970,000 Node-Hours

Petascale Simulations of Kinetic Effects in IFE Plasmas

PI Frank Tsung, University of California, Los Angeles
HOURS ALCF: 1,000,000 Node-Hours

PLASM-IN-SILICO: HPC Modeling of High-Intensity Laser-Solid Interaction

PI Jean-Luc Vay, Lawrence Berkeley National Laboratory
HOURS ALCF: 600,000 Node-Hours
OLCF: 110,000 Node-Hours

Radiation Hydrodynamic Simulations of Massive Stars with Rotation

PI Lars Bildsten, Kavli Institute for Theoretical Physics
HOURS ALCF: 1,800,000 Node-Hours

Towards a Definitive Model of Core-Collapse Supernova Explosions

PI Adam Burrows, Princeton University
HOURS Theta: 2,000,000 Node-Hours

ALCC 2019–2020

CHEMISTRY

Accelerated Catalyst Discovery from First Principles Simulations and Machine Learning

PI Rajeev Surendran Assary, Argonne National Laboratory
HOURS ALCF: 240,000 Node-Hours

Towards Exascale Internal Combustion Engine Simulation with In-Situ Analysis

PI Muhsin Ameen, Argonne National Laboratory
HOURS ALCF: 630,000 Node-Hours

ENERGY TECHNOLOGIES

High-Fidelity Physics Simulations for DOE and Industry Fast Spectrum Nuclear Reactor Systems

PI Emily Shemon, Argonne National Laboratory
HOURS ALCF: 880,000 Node-Hours

Nuclear Energy Industry Validation of Nek5000: ALAIN and HYMERES

PI Aleksander Obabko, Argonne National Laboratory
HOURS ALCF: 340,000 Node-Hours

MATERIALS SCIENCE

Predictive Modeling and Machine Learning for Functionally Nanoporous Materials

PI J. Ilja Siepmann, University of Minnesota
HOURS ALCF: 620,000 Node-Hours
NERSC: 200,000 Node-Hours

Supercomputing for Automotive High-Temperature Alloy Design

PI Dongwon Shin, Oak Ridge National Laboratory
HOURS ALCF: 350,000 Node-Hours
OLCF: 391,000 Node-Hours

PHYSICS

Low Energy Neutrino-Nucleus Interactions

PI Saori Pastore, Washington University in St. Louis
HOURS ALCF: 390,000 Node-Hours

Modeling the Response of Fusion Plasma Components

PI Brian Wirth, Oak Ridge National Laboratory
HOURS ALCF: 200,000 Node-Hours
OLCF: 250,000 Node-Hours

Neutrino Flux, Energy Deposition and Radiological Studies for the DUNE-LBNF Beamline

PI Igor Rakhno, Fermilab
HOURS ALCF: 450,000 Node-Hours

The Next Leap Forward in LSST Sky Simulations

PI Katrin Heitmann, Argonne National Laboratory
HOURS ALCF: 400,000 Node-Hours

Particle Acceleration in Plasma Jets: From Astrophysics to the Laboratory

PI Paulo E. Alves, SLAC National Accelerator Laboratory
HOURS ALCF: 1,000,000 Node-Hours

Semileptonic B- and D-Meson Form Factors with High Precision

PI Aida El-Kahdra, University of Illinois at Urbana-Champaign
HOURS ALCF: 400,000 Node-Hours
NERSC: 330,000 Node-Hours

ALCC 2020–2021

COMPUTER SCIENCE

Enabling Resilient and Portable Workflows from DOE's Experimental Facilities

PI Katie Antypas, Lawrence Berkeley National Laboratory
HOURS ALCF: 100,000 Node-Hours
OLCF: 20,000 Node-Hours

CHEMISTRY

Benchmarking Many-Body Perturbation Theory

PI Olle Heinonen, Argonne National Laboratory
HOURS ALCF: 100,000 Node-Hours

Interpretable Machine Learning Force Fields for Accurate Chemical Reactive Dynamic

PI Olexandr Isayev, Carnegie Mellon University
HOURS ALCF: 359,000 Node-Hours

Stochastic A Priori Dynamics for Complex Reactive Chemical Environments

PI Ahren Jasper, Argonne National Laboratory
HOURS ALCF: 100,000 Node-Hours

Understanding the Role of Hierarchical Correlations in Solution-Phase Chemical Separations

PI Lynda Soderholm, Argonne National Laboratory
HOURS ALCF: 100,000 Node-Hours

EARTH SCIENCE

Variable-Resolution Earth System Modeling of the Cryosphere with E3SM

PI Darin Comeau, Los Alamos National Laboratory
HOURS ALCF: 400,000 Node-Hours
NERSC: 500,000 Node-Hours

ENERGY TECHNOLOGIES

Automatic Building Energy Modeling

PI Joshua New, Oak Ridge National Laboratory
HOURS ALCF: 300,000 Node-Hours

MATERIALS SCIENCE

Data-Driven Molecular Engineering of Advanced Functional Materials

PI Jacqueline Cole, University of Cambridge
HOURS ALCF: 100,000 Node-Hours

High-Temperature Material Properties from First Principles

PI Mark Messner, Argonne National Laboratory
HOURS ALCF: 200,000 Node-Hours

Many-Body Perturbation Theory Meets Machine Learning to Discover Materials for Organic Photovoltaics

PI Noa Marom, Carnegie Mellon University
HOURS ALCF: 100,000 Node-Hours

Plasma Surface Interaction Modeling

PI Brian Wirth, University of Tennessee
HOURS ALCF: 318,000 Node-Hours
OLCF: 155,000 Node-Hours
NERSC: 30,000 Node-Hours

Predictive Modeling of Nanoporous Materials and Multiphase Systems

PI Joern Siepmann, University of Minnesota
HOURS ALCF: 220,000 Node-Hours

Supercomputing for Automotive High-Temperature Alloy Design

PI Dongwon Shin, Oak Ridge National Laboratory
HOURS ALCF: 100,000 Node-Hours

PHYSICS

Toward the Future: High-Fidelity Simulation for Next-Generation Nuclear Reactors

PI Yiqi Yu, Argonne National Laboratory
HOURS ALCF: 208,000 Node-Hours
OLCF: 300,000 Node-Hours

Reconstructing Neutrino Data with the MicroBooNE Liquid Argon Detector

PI Andrzej Szelc, University of Manchester
HOURS ALCF: 200,000 Node-Hours

Optimization Studies of the LBNF - PIP-II Complex for Megawatt Beams on Target

PI Igor Rakhno, Fermi National Accelerator Laboratory
HOURS ALCF: 450,000 Node-Hours

DNS Simulations of Coolant Flow in the High-Flux Isotope Reactor

PI Emilian Popov, Oak Ridge National Laboratory
HOURS ALCF: 220,000 Node-Hours

Chiral Nuclear Interactions from Nuclei to Nucleonic Matter

PI Maria Piarulli, Washington University in St. Louis
HOURS ALCF: 200,000 Node-Hours

Nucleon Axial Charge with All-Staggered Lattice QCD

PI Andreas Kronfeld, Fermi National Accelerator Laboratory
HOURS ALCF: 200,000 Node-Hours
NERSC: 870,000 Node-Hours

Distributed Large Wavefield Propagation and 3D Reconstruction Beyond the Depth of Focus Limit

PI Ming Du, Argonne National Laboratory
HOURS ALCF: 250,000 Node-Hours

Field-Reversed Configuration Stability and Transport

PI Sean Dettrick, TAE Technologies Inc.
HOURS ALCF: 64,000 Node-Hours

Improving Direct-Drive Laser Fusion Predictive Capability with a Simulation Database

PI Duc Cao, Laboratory for Laser Energetics
HOURS ALCF: 318,000 Node-Hours

High-Luminosity LHC Detector Upgrade Studies by the ATLAS and CMS Collaborations

PI Douglas Benjamin, Argonne National Laboratory
HOURS ALCF: 950,000 Node-Hours

Multiphase Flow Simulations of Reactor Flows

PI Igor Bolotnov, North Carolina State University
HOURS ALCF: 192,000 Node-Hours
NERSC: 250,000 Node-Hours

ALCF DATA SCIENCE PROGRAM

Advanced Materials Characterization with AI-Informed Computation

PI Marco Govoni, Argonne National Laboratory

Deep Learning at Scale for Multimessenger Astrophysics Through the NCSA-Argonne Collaboration

PI Eliu Huerta, University of Illinois at Urbana-Champaign

Developing High-Performance-Computing Applications for Liquid Argon Neutrino Detectors

PI Andrzej Szelc, University of Manchester

Dynamic Compressed Sensing for Real-Time Tomographic Reconstruction

PI Robert Hovden, University of Michigan

Machine Learning Magnetic Properties of van der Waals Heterostructures

PI Trevor Rhone, Harvard University

X-ray Microscopy of Extended 3D Objects: Scaling Towards the Future

PI Chris Jacobsen, Argonne National Laboratory and Northwestern University

AURORA EARLY SCIENCE PROGRAM

Accelerated Deep Learning Discovery in Fusion Energy Science

PI William Tang, Princeton Plasma Physics Laboratory

Dark Sky Mining

PI Salman Habib, Argonne National Laboratory

Data Analytics and Machine Learning for Exascale Computational Fluid Dynamics

PI Kenneth Jansen, University of Colorado Boulder

Enabling Connectomics at Exascale to Facilitate Discoveries in Neuroscience

PI Nicola Ferrier, Argonne National Laboratory

Exascale Computational Catalysis

PI David Bross, Argonne National Laboratory

Extending Moore's Law Computing with Quantum Monte Carlo

PI Anouar Benali, Argonne National Laboratory

Extreme-Scale Cosmological Hydrodynamics

PI Katrin Heitmann, Argonne National Laboratory

Extreme-Scale In-Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations

PI Amanda Randles, Duke University and Oak Ridge National Laboratory

Extreme-Scale Unstructured Adaptive CFD

PI Kenneth Jansen, University of Colorado Boulder

High-Fidelity Simulation of Fusion Reactor Boundary Plasmas

PI C.S. Chang, Princeton Plasma Physics Laboratory

Machine Learning for Lattice Quantum Chromodynamics

PI William Detmold, Massachusetts Institute of Technology

Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials

PI Noa Marom, Carnegie Mellon University

NWChemEx: Tackling Chemical, Materials, and Biochemical Challenges in the Exascale Era

PI Theresa Windus, Iowa State University and Ames Laboratory

Simulating and Learning in the ATLAS Detector at the Exascale

PI Walter Hopkins, Argonne National Laboratory

2020 DIRECTOR'S DISCRETIONARY

The following list provides a sampling of the many Director's Discretionary projects at the ALCF.

BIOLOGICAL SCIENCES

Agent-Based Model Called CityCOVID Capable of Tracking Detailed COVID-19 Transmission

PI Jonathan Ozik, Argonne National Laboratory

AI-Driven Integrative Biology for Accelerating Therapeutic Discovery Against SARS-CoV-2

PI Arvind Ramanathan, Thomas Brettin, Argonne National Laboratory

Compound Screening to Repurpose FDA-Approved Drugs Against SARS-CoV-2 Catalytic Enzymes

PI Albert Y. Lau, Johns Hopkins University

Machine Learning of Drug Binding and Toxicity Based on High-Throughput Free Energy Computations

PI Wei Jiang, Johns Hopkins University

Modeling Coronavirus

PI Zhangli Peng, University of Illinois at Chicago and University of Notre Dame

Understanding the Mechanism of Ligand-Induced Conformational Dynamics of HIV-1 Protease and the Effects of Mutations

PI Ao Ma, University of Illinois at Chicago

CHEMISTRY

INtegrated and Scalable PredictIon of REsistanace (INSPIRE)

PI Peter Coveney, University College London

Ionic Liquid as a Potential Electrolyte of High Performance Lithium Ion Battery

PI Zhengcheng Zhang, Argonne National Laboratory

COMPUTER SCIENCE

Collaboration with Cray on Interconnect-Related Studies

PI Sudheer Chunduri, Argonne National Laboratory

Distributed Relational Algebra at Scale

PI Sidharth Kumar, University of Alabama at Birmingham

Energy Efficient Tradeoff Among Execution Time, Power, and Resilience of Two ECP Applications

PI Xingfu Wu, Argonne National Laboratory

EARTH SCIENCE

Linking Climate to Water: Implementing a 4km Regional Climate Model with Hydrologic Model Coupling (WRF-Hydro)

PI Jiali Wang, Argonne National Laboratory

ENERGY TECHNOLOGIES

High-Fidelity CFD Simulations of Multi-Mode Combustion

PI Pinaki Pal, Argonne National Laboratory

ENGINEERING

Direct Numerical Simulation of Three-Dimensional Turbulence

PI Ramesh Balakrishnan, Argonne National Laboratory

High-Fidelity Simulation of Supersonic Turbulent Flow-Structure Interaction and Mixing

PI Ivan Bermejo Moreno, University of Southern Californi

MATERIALS SCIENCE

Metastable Phase Diagram of Material

PI Subramanian Sankaranarayanan, Argonne National Laboratory

Rational Design of Ultrastrong Composites

PI Hendrik Heinz, University of Colorado Boulder

Structure and Properties of Grain Boundaries in Materials for Energy Applications

PI Wissam Saidi, University of Pittsburgh

PHYSICS

Exploring Astrophysical Particle Acceleration in HED Laboratory Plasmas

PI Frederico Fiuza, SLAC National Accelerator Laboratory

Artificial Intelligence Assisted Safety Modeling and Analysis of Advanced Reactors

PI Rui Hu, Argonne National Laboratory

About the Argonne Leadership Computing Facility

Argonne's Leadership Computing Facility Division operates the Argonne Leadership Computing Facility (ALCF) as part of the U.S. Department of Energy's effort to provide leadership-class computing resources to the scientific community. The ALCF is supported by the DOE Office of Science, Advanced Scientific Computing Research (ASCR) program.

About Argonne National Laboratory

Argonne is a U.S. Department of Energy Laboratory managed by UChicago Argonne, LLC, under contract DE-AC02-06CH11357. The Laboratory's main facility is outside of Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne and its pioneering science and technology programs, visit www.anl.gov.

Availability of this Report (ANL/ALCF-20/07)

Online Access: U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available for free via DOE's SciTech Connect (http://www.osti.gov/scitech/).

Reports not in digital format may be purchased by the public from the National Technical Information Service (NTIS):

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd.
Alexandra, VA 22312

phone | 800.553.NTIS (6847) or 703.605.6000
fax | 703.605.6900
orders@ntis.gov
www.ntis.gov

Reports not in digital format are available to DOE and DOE contractors from the Office of Scientific and Technical Information (OSTI):

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062

phone | 865.576.8401
fax | 865.576.5728
reports@adonis.osti.gov
www.osti.gov

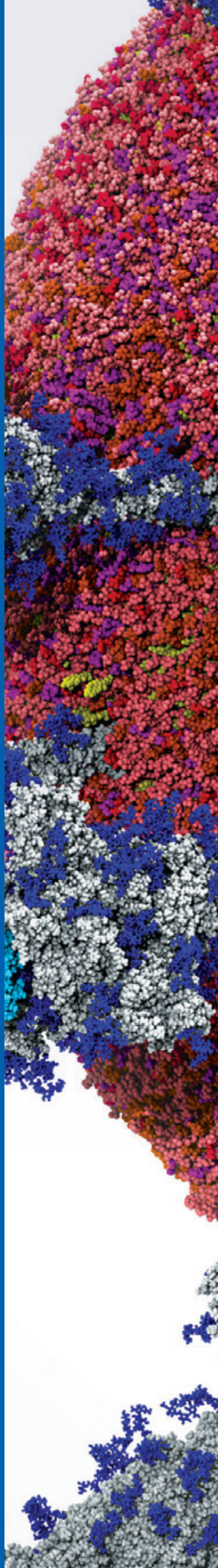
ALCF Leadership: Michael E. Papka (Division Director), Jini Ramprakash (Deputy Division Director), Katherine Riley (Director of Science), Kalyan Kumaran (Director of Technology), Mark Fahey (Director of Operations), and Susan Coghlan (ALCF-X Project Director)

Editorial Team: Beth Cerny, Jim Collins, Nils Heinonen, and Laura Wolf

Design and production: Sandbox Studio, Chicago

Argonne NATIONAL LABORATORY

CONTACT
media@alcf.anl.gov
alcf.anl.gov



U.S. DEPARTMENT OF ENERGY

Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.