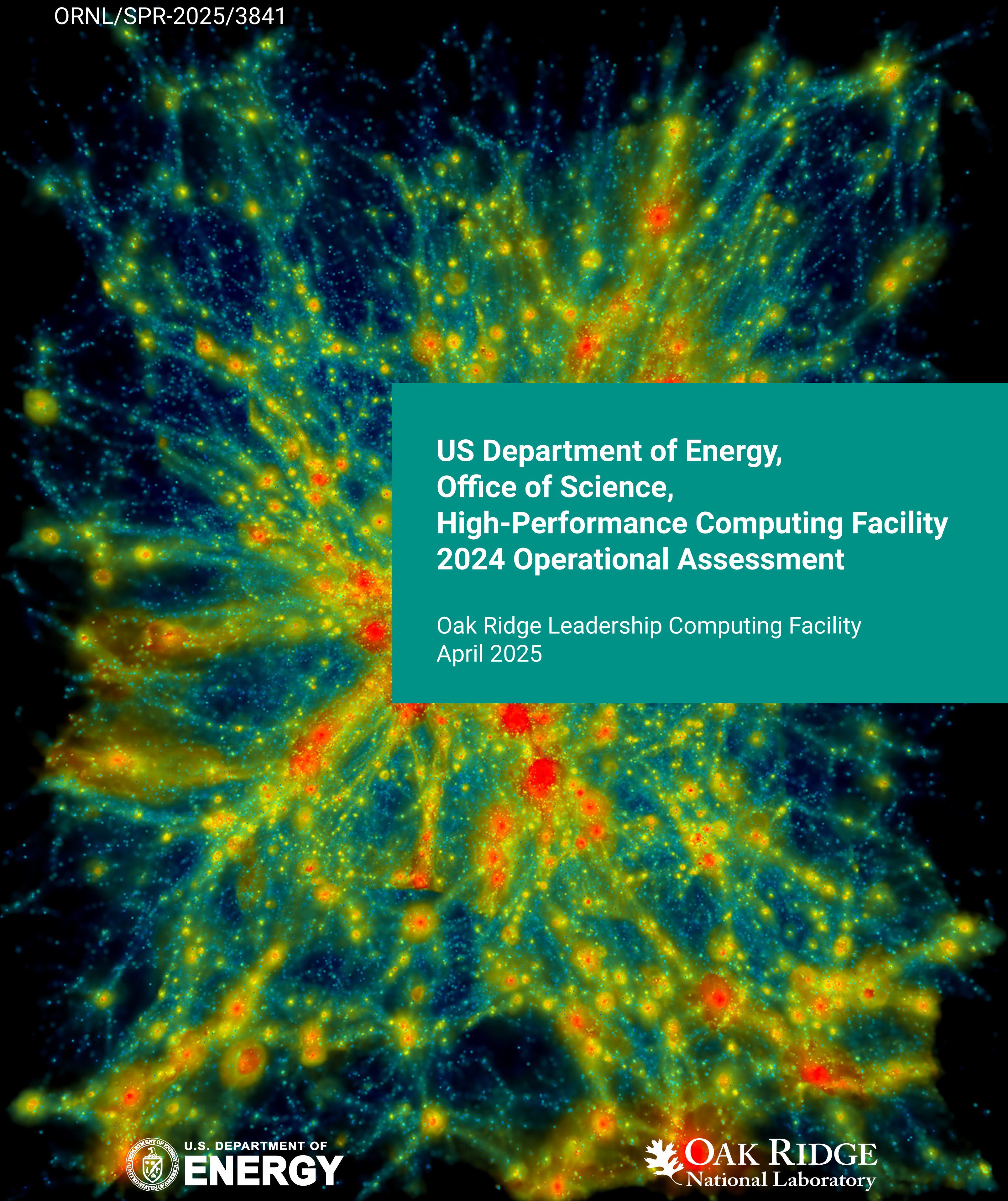
ORNL/SPR-2025/3841



US Department of Energy, Office of Science, High-Performance Computing Facility 2024 Operational Assessment

Oak Ridge Leadership Computing Facility
April 2025

U.S. DEPARTMENT OF ENERGY

Oak Ridge National Laboratory

ORNL/SPR-2025/3841

Oak Ridge Leadership Computing Facility

US DEPARTMENT OF ENERGY, OFFICE OF SCIENCE, HIGH-PERFORMANCE COMPUTING FACILITY 2024 OPERATIONAL ASSESSMENT OAK RIDGE LEADERSHIP COMPUTING FACILITY

Subil Abraham
J. Paul Abston
Ryan M. Adamson
Valentine Anantharaj
Dilip Asthagiri
Ashley Barker
Aaron Barlow
Tom Beck
Katie L. Bethea
Wes Brewer
Michael Brim
Dmytro Bykov
Josh Cunningham
Rafael Ferreira da Silva
Christopher Fuson
Gustav Jansen
Olga Kuchar
Ryan Landfield
Jack Lange
Kellen Leland
Dustin Leverman
Don Maxwell
Murali Gopalakrishnan Meena
Verónica Melesse Vergara
Bronson Messer
Ross Miller
Sheila Moore
James Morgan
Sarp Oral
Tim Osborne
Suzanne Parete-Koon
Paul Peltz
Fernando Posada
Suzanne Prentice
William Renaud
Tori Robinson
David Rogers
James Rogers
Michael Sandoval
Mallikarjun Shankar
Scott Simmerman
Christopher Stahl
In-Saeng Suh
Kevin G. Thach
Suzy Tichenor
Coury Turczyn
Feiyi Wang
Spencer Ward
Junqi Yin
Christopher Zimmer

April 2025

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACE	Advanced Computing Ecosystem
ACCEL	Accelerating Competitiveness through Computational ExceLlence
ACM	Association of Computing Machinery
AFW	Air Force Weather
ALCC	ASCR Leadership Computing Challenge
ALCF	Argonne Leadership Computing Facility
Argonne	Argonne National Laboratory
ARM	Atmospheric Radiation Measurement
ASCR	Advanced Scientific Computing Research
BES	Basic Energy Sciences
CAAR	Center for Accelerated Application Readiness
CCSD	Computing and Computational Sciences Directorate
CFD	computational fluid dynamics
CNMS	Center for Nanophase Materials Science
CRM	customer relationship management software
CSED	Computational Sciences and Engineering Division
CSEEN	Computational Scientists for Energy, the Environment, and National Security
DARTs	Days Away Restricted or Transferred
DD	Director's Discretionary program
Deleria	Distributed Event-Level Experiment Readout and Integrated Analysis
DOE	US Department of Energy
DOI	digital object identifier
DS2HPC	Data Streaming to HPC
E4S	Extreme-Scale Scientific Software Stack
ECP	Exascale Computing Project
ES&H	environment, safety, and health
EXESS	Extreme-scale Electronic Structure System
FAIR	findability, accessibility, interoperability, and reusability
GPFS	General Parallel File System
GRETA	Gamma-Ray Energy Tracking Array
HPC	high-performance computing
HPL	High-Performance Linpack
HPSS	High-Performance Storage System
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
IRI	Integrated Research Infrastructure Program
ISM	Integrated Safety Management
ITAR	International Traffic and Arms Regulations
JH	Joule heating
JIF	journal impact factor
KAUST	King Abdullah University of Science and Technology
LANL	Los Alamos National Laboratory
LBNL	Lawrence Berkeley National Laboratory
LCLS	Linac Coherent Light Source
LDES	long-duration energy storage

LDRD	Laboratory Directed Research and Development
LLM	large language model
LLNL	Lawrence Livermore National Laboratory
LVA	Live Visual Analytics
ML4NSE	Machine Learning for Neutron Scattering Experiment
MM	molecular mechanics
MOSSAIC	Modeling Outcomes using Surveillance data and Scalable Artificial Intelligence for Cancer
MPI	Message Passing Interface
MTTF	mean time to failure
MTTI	mean time to interrupt
NAIRR	National Artificial Intelligence Research Resource
NAM	not a metric
NCCS	National Center for Computational Sciences
NCI	National Cancer Institute
NCRC	National Climate-computing Research Center
NERSC	National Energy Research Scientific Computing Center
NGP	Next Generation Pathways Summer School
NIH	National Institutes of Health
NOAA	National Oceanic and Atmospheric Administration
NSDS	National Secure Data Service
NSP	NCCS Software Provisioning
NTCL	Nuclear Tensor Contraction Library
OA	overall availability
OAR	Operational Assessment Review
OCR	Operations Control Room
OLCF	Oak Ridge Leadership Computing Facility
ORNL	Oak Ridge National Laboratory
OUG	OLCF User Group
PAS	Personnel Access System
PEARC	Practice and Experience in Advanced Research Computing
PHI	protected health information
PI	principal investigator
PNNL	Pacific Northwest National Laboratory
QC	quantum computing
QCUP	Quantum Computing User Program
QM	quantum mechanics
RAPS	Resource Allocator and Power Simulator
RATS	Resource and Allocation Tracking System
ROVI	Rapid Operational Validation Initiative
RPS	Restricted Party Screening
RSS	research safety summary
S3M	Secure Scientific Service Mesh
SA	scheduled availability
Sandia	Sandia National Laboratories
SBMS	Standards Based Management System

SC	International Conference for High Performance Computing, Networking, Storage, and Analysis
SEER	Surveillance, Epidemiology, and End Results program
SLIME	Secure LLM Inspection and Malware Evaluation
SNS	Spallation Neutron Source
SPI	Scalable Protected Infrastructure
SSH	Secure Shell
SU	system utilization
TFT	Temporal Fusion Transformer
TTU	Tennessee Technological University
UIUC	University of Illinois at Urbana-Champaign
UT	University of Tennessee
VA	US Department of Veterans Affairs
WES	Workflow and Ecosystem Services
WDTS	Workforce Development for Teachers and Scientists

HIGH-PERFORMANCE COMPUTING FACILITY
2024 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY

April 2025

EXECUTIVE SUMMARY

The Oak Ridge Leadership Computing Facility (OLCF) was established to accelerate scientific discovery by providing world-leading computational performance and advanced data infrastructure to the US Department of Energy (DOE) computing community. As a DOE Office of Science user facility, the OLCF has managed the successful deployment and operation of a succession of leadership-class resources dedicated to open science. In addition to these resources, the OLCF staff continually strive to develop innovative processes and technologies, improve security, and empower users through effective allocation management and comprehensive user support and training. These efforts support the advancement of science by the OLCF users and benefit high-performance computing (HPC) facilities around the world.

In calendar year (CY) 2024, the OLCF supported 2,072 users and 639 projects and exceeded all targets for user satisfaction. The facility received an average satisfaction score of 4.53 out of 5 on the annual user survey, and 95% of respondents reported a high satisfaction rate with the OLCF overall. Of the 3,562 user tickets submitted in CY 2024, OLCF staff resolved 97% within 3 business days, underscoring the facility's commitment to responsive and effective user support. .

The facility's Frontier exascale supercomputer continued to enable unprecedented achievements in scientific computing. Notably, two projects conducted on Frontier received the Association for Computing Machinery's Gordon Bell Prize and the Gordon Bell Special Prize for Climate Modeling in 2024. The Gordon Bell Prize is awarded each year at the International Conference for High Performance Computing, Networking, Storage and Analysis (SC) to recognize researchers who have made significant strides toward applying HPC systems to scientific applications. A project on Earth System Predictability with an AI foundation model earned an HPCwire Top Supercomputing Achievement award. Meanwhile, the facility's previous flagship machine, Summit, extended its operations through 2024 with the SummitPLUS program, in part to help provide resources to the Integrated Research Infrastructure projects and the National Artificial Intelligence Research Resource pilot program. Summit was decommissioned in November 2024 after 6 years and over 200 million node hours of service. The Alpine file system was also decommissioned at this time. In preparation for the decommissioning of the High-Performance Storage System in January 2025, the OLCF launched Kronos, a new multitiered storage system that utilizes both disk and tape.

Throughout the year, OLCF staff maintained exceptional system reliability, ensuring user access to key HPC resources: the HPE Cray EX Frontier, the IBM AC922 Summit, the General Parallel File System (Alpine), the Lustre file system Orion, and the archival storage system (the High-Performance Storage System and its replacement, Kronos). Overall availability for both the compute resources and file systems exceeded 98% for CY 2024. The facility also successfully delivered on the 60/30/10 allocation split of roughly 60% of core-hours offered to the Innovative and Novel Computational Impact on Theory and Experiment program, 30% to the Advanced Scientific Computing Research Leadership Computing Challenge program, and 10% to the Director's Discretionary program (see Section 2). Table ES-1 summarizes the 2024 OLCF metric targets and the associated results. More information about each OLCF resource is provided in Section 2.

Table ES-1. 2024 OLCF metric summary.

Metric description	CY 2024 target	CY 2024 actual
Overall OLCF score on the user survey will be 3.5 based on a statistically meaningful sample.	3.5	4.53
Time between Receipt of User Query (Jira Ticket) and Center Response: 80% of OLCF problems will be addressed within 3 working days (72 hours) by either resolving the problem or informing the user how the problem will be resolved.	80%	97%
CAPABILITY JOBS: *For the CY following a new system/upgrade, at least 30% of the consumed node hours will be from jobs requesting 20% or more of the available nodes. In subsequent years, at least 35% of consumed core hours/node hours will be from jobs requiring 20% or more of cores/nodes available to the users.*		
Scientific and Technological Research and Innovation – Demonstrate Leadership Computing, Frontier	35%	57.55%
SCHEDULED AVAILABILITY (COMPUTE): *For the CY following a new system/upgrade, the scheduled availability (SA) target for an HPC compute resource is at least 85%. For year 2, the SA target for an HPC compute resource increases to at least 90%, and for year 3 through the end of life for the associated compute resource, the SA target for an HPC compute resource increases to 95%. Consequently, SA targets are described as 85/90/95.*		
SA, Summit: Sustain scheduled availability to users, measured as a percentage of maximum possible scheduled.	95%	98.82%
SA Frontier: Sustain scheduled availability to users, measured as a percentage of maximum possible scheduled.	90%	99.81%
OVERALL AVAILABILITY (COMPUTE): *For the CY following a new system/upgrade, the overall availability (OA) target for an HPC compute resource is at least 80%. For year 2, the OA target increases to at least 85%, and for year 3 through the end of life for the associated compute resource, the OA target increases to 90%. Consequently, OA targets are described as 80/85/90.*		
OA, Summit: Sustain availability to users measured as a percentage of maximum possible.	90%	98.53%
OA, Frontier: Sustain availability to users measured as a percentage of maximum possible.	85%	98.49%
OVERALL AVAILABILITY (FILE SYSTEMS): *For the CY following a new system/upgrade, the OA target for an external file system is at least 85%. For year 2 through the end of life of the asset, the OA target for an external file system increases to at least 90%. OA targets are thus described as 85/90.*		
OA, external file system Alpine2: Sustain availability to users measured as a percentage of maximum possible.	90%	98.49%
OA, external file system Orion: Sustain availability to users measured as a percentage of maximum possible.	90%	98.92%

OA, archive storage: Sustain availability to users measured as a percentage of maximum possible.	90%	98.37%
SCHEDULED AVAILABILITY (FILE SYSTEMS): *For the CY following a new system/upgrade, the SA target for an external file system is at least 90%. For year 2 through the end of life of the asset, the SA target for an external file system increases to at least 95%. SA targets are thus described as 90/95.*		
SA, Alpine2: Sustain scheduled availability to users measured as a percentage of maximum possible scheduled.	95%	98.96%
SA, Orion: Sustain scheduled availability to users measured as a percentage of maximum possible scheduled.	95%	99.69%
SA, High Performance Storage System: Sustain scheduled availability to users measured as a percentage of maximum possible scheduled.	95%	99.39%

CY 2024 marked another year of sustained technological development and innovation for the OLCF.

A team at the OLCF played a pivotal role in implementing an optimized High-Performance Linpack (HPL) benchmark code, enabling Frontier to achieve 1.353 exaflops on 9,604 nodes for an increase of over 12%. This breakthrough was made possible through a collaboration with AMD, which introduced a new technique that delivered 65.8% efficiency—the highest ever recorded for an exascale system, exceeding the performance of any exascale supercomputer. Additionally, an improved version of the mixed-precision HPL code achieved 11.39 exaflops on an expanded number of nodes, marking a nearly 12% increase from the previous result of 10.2 exaflops. This reinforces Frontier's position as a leader in exascale computing, with remarkable advancements in both raw performance for scientific computing, AI, and compute-efficient operations.

To enhance user support and mitigate risks associated with reliance on a single vendor's software stack, the OLCF advanced alternative communication library support for Frontier. This effort included extending open-source software projects and optimizing intranode communication mechanisms via AMD's high-speed XGMI interconnect. The successful upstreaming of these features to libfabric and OpenMPI has broadened accessibility and strengthened system capabilities.

The OLCF also developed, refined, and implemented other technical innovations that will continually improve operations. These innovations include PoliMOR, which is an automated data management policy engine for data migration, purging, data collection, and telemetry for file systems (e.g., Lustre); the more efficient and streamlined *hsi_xfer* data transfer tool; the AI-specific FORGE benchmark, which emphasizes low-precision compute, communication bandwidth, and energy efficiency in training large language models (LLMs) using scientific datasets; and the OLCF-6 Workflow Benchmark, which evaluates HPC system capabilities for handling dynamic data stream workloads through Machine Learning for Neutron Scattering Experiment applications.

The scientific accomplishments of OLCF users and staff members are strong indications of long-term operational success and broad scientific impact, and this year's user projects and programs have continued to advance DOE's strategic objectives. Users published 50 papers in high-impact journals such as *Science, PNAS, Nature Physics*, *ACS Nano*, *Journal of the American Chemical Society*, *Physical Review Letters*, and *Advanced Functional Materials*.

The OLCF supported scientific accomplishments for a broad community of researchers in 2024, from traditional modeling and simulation projects to studies that leverage AI, ML, and big data analytics techniques. AI methods and mixed-precision approaches are becoming essential to the world's most significant computational campaigns. Frontier's leadership capability was critical to six teams who were finalists in the prestigious Gordon Bell Prize competition, including both winning teams who used leadership-class computing, innovative machine learning, and low-precision techniques in their campaigns.

The titular Gordon Bell Prize winners were members of a team led by the University of Melbourne, Australia. By using the Extreme-scale Electronic Structure System code, the team successfully conducted a quantum molecular dynamics simulation 1,000× greater in size and speed than any previous simulation of its kind. The other winning team, which was from the King Abdullah University of Science and Technology in Saudi Arabia, won the Gordon Bell Prize for Climate Modeling by leveraging Frontier's GPUs and mixed-precision arithmetic to establish a climate emulator that offers a remarkable resolution of 3.5 km (approximately 2.2 miles) and can replicate local conditions on a timescale from days to hours.

In addition to the Gordon Bell Prize winners, OLCF systems supported a wide range of projects that tackled major scientific challenges. For example, a multi-institutional team led by Argonne National Laboratory used Frontier to train an AI model to design proteins. The team trained an LLM on a dataset of proteins as part of a high-speed digital workflow with four other supercomputing systems. The workflow included Argonne's Aurora exascale machine, the Swiss National Supercomputing Centre's Alps, the CINECA data center's Leonardo, and NVIDIA corporation's PDX. The workflow's simulations broke the exascale barrier at a peak speed of roughly 5.57 exaflops using mixed precision. The study earned the team a finalist nomination for the Association for Computing Machinery's Gordon Bell Prize.

Other groundbreaking projects include NASA's use of both Summit and Frontier to simulate a human-piloted spacecraft landing on Mars, the University of South Florida's simulation on Frontier of a material harder and tougher than diamond, Tennessee-based Whisper Aero's feasibility studies of its ultraquiet electric aircraft design on Summit, and Argonne National Laboratory's use of Frontier to create the largest astrophysical simulation of the universe ever conducted.

In 2024, the OLCF advanced quantum computing (QC) integration within its high-performance computing (HPC) infrastructure, collaborating across ORNL's Quantum Science Center, Quantum Information Sciences, and NCCS. A team published a strategic roadmap outlining hardware, software, workflows, and benchmarking for seamless quantum-HPC integration. Securing LDRD funding, OLCF launched initiatives in quantum frameworks, computational chemistry applications, and quantum networking. A new Quantum-HPC group was established in early 2025 to drive integration efforts. Ongoing work includes developing a hardware-agnostic testbed, with prototype Quantum Brilliance hardware slated for installation in mid-2025. In addition, the OLCF continued the advancement of the Quantum Computing User Program (QCUP) and added IQM to its vendor roster, bringing the total to four vendors: Quantinuum, IonQ, IBM Quantum, and IQM. QCUP supported 88 projects in CY24 and have tracked 64 publications that resulted from the use of QCUP resources.

In preparation to release the Request For Proposals (RFP) for OLCF-6, the OLCF team met with multiple system integrators, cloud vendors, processor vendors, interconnect providers, and storage vendors. The team met with each vendor multiple times to refine the RFP requirements, answer questions, and learn about potential proposal responses. This extensive engagement by the OLCF team led to a record number of proposals submitted in response to the RFP and retired one of the project's highest risks: that only one vendor or no vendors would respond to the RFP.

As a follow-on to the DOE Advanced Computing Ecosystem Request For Information in 2022, the OLCF developed a Request For Proposal (RFP) for an OLCF-funded vendor research and development program. Before the release of the RFP, the OLCF team drafted an initial set of requirements. The OLCF team met with dozens of potential vendors to ensure that the requirements covered their areas of interest, and based on feedback from these interactions, finalized the RFP, released it, and then evaluated the responses. The evaluation team was composed of over 35 representatives from ORNL, NERSC, ANL, PNL, and LLNL. The interest in the program was very high with 14 vendors proposing 43 work packages — far exceeding the New Frontiers initial budget.

The OLCF's Advanced Computing Ecosystem (ACE) initiative significantly contributed to the DOE's Integrated Research Infrastructure (IRI) program in FY24, delivering unique and critical infrastructure and capabilities for cutting-edge scientific computing development and evaluations. Key accomplishments include the deployment of a robust testbed environment integrated into the Energy Sciences Network (ESnet), which has advanced computing evaluation and development collaborations across DOE facilities such as SLAC, Argonne, NERSC, and Jefferson Lab. ACE has successfully supported diverse science IRI pilots, notably for development projects like LCLStream and DELERIA, accelerating data processing and real-time analysis capabilities essential for solving complex scientific challenges. Technological innovations in data movement, interface design, and workflow orchestration — including tools like the OLCF Facility API (S3M), Data Streaming technologies (DS2HPC), and Zambeze orchestration — have significantly enhanced computational efficiency and flexibility. Strategic outreach activities have further strengthened ACE's role in fostering collaboration and driving scientific discovery, positioning it as an adaptable and future-ready testbed platform integral to DOE's leadership in HPC test and development efforts.

Committed to workforce development and community engagement, the OLCF pursued a a multipronged pipeline approach, starting with student programs and followed by on-the-job training programs. The OLCF expanded student training programs, recruitment efforts, and hands-on HPC engagement through conferences, hackathons, and outreach initiatives.

In 2024, the OLCF strengthened strategic collaborations with DOE Office of Science programs, applied office initiatives, and sister federal agencies. Notably, the OLCF supported Atmospheric Radiation Measurement (ARM) by enhancing data transformation workflows and leveraging Slurm's Multi-Category Security feature to enable limited node sharing. This optimization improved cluster throughput and reduced job completion times, enhancing ARM's HPC workflows. Key external partnerships included the National Institutes of Health's National Cancer Institute, the U.S. Department of Veterans Affairs, Air Force Weather, and the National Oceanic and Atmospheric Administration. These collaborations fostered broader community engagement, drove software and capability development, and advanced operational innovation for both the OLCF and its partners.

Maintaining a culture of operational excellence, the OLCF prioritized risk management, workplace safety, and cybersecurity, ensuring smooth operations and a zero-accident record despite large-scale workloads and supercomputer installations.

The OLCF's success is driven by the dedication and expertise of its staff, who develop and deploy innovative technologies that advance science and benefit the global HPC community and boosts economic competitiveness. In 2024, the facility made remarkable strides in exascale computing, AI-driven research, and system optimization, reinforcing its leadership in computational science. Through sustained innovation, strategic collaborations, and a commitment to excellence, the OLCF empowered researchers to tackle critical challenges. Looking ahead, the OLCF remains committed to driving scientific discovery, refining computational methodologies, and shaping the future of HPC with world-class resources and an engaged user community.

ES.1 COMMUNICATIONS WITH KEY STAKEHOLDERS

ES.1.1 Communication with the Program Office

The OLCF communicates with the Advanced Scientific Computing Research (ASCR) Program Office through a series of regularly occurring events. These include weekly Integrated Project Team calls with the local DOE ORNL Site Office and the ASCR Program office, monthly highlight reports, quarterly reports, the annual Operational Assessment Report, and an annual "Budget Deep Dive." Through a team of communications specialists and writers working with users and management, the OLCF produces a steady flow of reports and highlights for sponsors, current and potential users, and the public. See APPENDIX B for a list of science highlights submitted to ASCR.

ES.1.2 Communication with the User Community

The OLCF's user-targeted communications are designed to relate science results to the HPC community and to help users more efficiently and effectively leverage OLCF systems. The OLCF offers many training and educational opportunities throughout the year for current facility users and the next generation of HPC users, as outlined in Section 1.

The impact of OLCF communications is assessed as part of an annual user survey. In the 2024 annual user survey, OLCF communications received a mean rating for users' overall satisfaction of 4.56 on a scale of 5.0. The OLCF uses a variety of methods to communicate with users, including the following:

- Weekly email message
- General email announcements
- Automated notifications of system outages
- OLCF website
- Monthly conference calls
- OLCF User Council and Executive Board meetings
- One-on-one interactions with liaisons and analysts
- Office hours
- Social media
- Annual OLCF User Meeting
- Targeted training events (e.g., GPU hackathons or tutorials)

ES.2 SUMMARY OF 2024 METRICS

In consultation with the DOE program manager, a series of metrics and targets was identified to assess the operational performance of the OLCF in CY 2024. The 2024 metrics, target values, and actual results as of December 31, 2024, are noted throughout this report. The OLCF exceeded all agreed-upon metric targets.

ES.3 RESPONSES TO RECOMMENDATIONS FROM THE 2023 OPERATIONAL ASSESSMENT REVIEW

There were no recommendations resulting from last year's Operational Assessment.

2024 1 – User Support Results

HIGH-PERFORMANCE COMPUTING FACILITY
2024 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY

April 2025

1. USER SUPPORT RESULTS

CHARGE QUESTION 1: Are the processes for supporting users, resolving users' problems, and conducting outreach to the user population effective?

OLCF RESPONSE: Yes. The Oak Ridge Leadership Computing Facility (OLCF) effectively supports its user community through a well-established, continuously improving support model. In CY 2024, the OLCF provided resources and assistance to 2,072 users across 639 projects, including quantum computing initiatives. Our approach prioritizes responsiveness, technical expertise, and proactive engagement, ensuring users receive timely and effective support.

User satisfaction remains a key performance indicator, measured through an annual survey in which users rate their overall experience on a scale of 1 to 5. In 2024, the OLCF received a mean satisfaction rating of 4.53, with 95% of respondents reporting they were "satisfied" or "very satisfied."

The OLCF's performance is also tracked through quantifiable service metrics to ensure rapid issue resolution and continuous process improvement. In 2024, the facility exceeded all user support benchmarks, with 97% of support tickets resolved within three business days.

Beyond direct user support, the OLCF remains committed to expanding technical training, fostering collaborations, and conducting outreach to enhance accessibility and engagement in high-performance computing (HPC), artificial intelligence (AI), and quantum computing (QC). These efforts help equip both new and experienced users with the knowledge and tools needed to maximize their productivity on OLCF systems.

Through data-driven assessment, targeted improvements, and strategic outreach, the OLCF ensures its user support processes remain effective, responsive, and aligned with the evolving needs of the global scientific computing community.

1.1 USER SUPPORT RESULTS SUMMARY

The OLCF's user support model comprises customer support interfaces, including user satisfaction surveys, formal problem-resolution mechanisms, user assistance engineers, and scientific and data liaisons; multiple channels for stakeholder communication, including the OLCF User Group Executive Board; and training programs, user workshops, and tools to reach and train both current facility users and the next generation of computer and computational scientists. The success of these activities and identification of areas for development are tracked by the annual OLCF user survey.

To promote continual improvement at the OLCF, users are sent surveys that solicit feedback on support services and user experiences at the facility. The 2024 survey was launched on October 9, 2024, and remained open for participation through December 2, 2024. The survey was sent to 1,637 users of the Innovative and Novel Computational Impact on Theory and Experiment (INCITE), Advanced Scientific

Computing Research (ASCR) Leadership Computing Challenge (ALCC), SummitPlus, Exascale Computing Project (ECP), and Director's Discretionary (DD) programs. These metrics include all users of these programs who logged into an OLCF system between January 1, 2024, and September 30, 2024. Although the ECP allocations come from the OLCF DD program, those responses were once again tracked separately from the DD responses. OLCF staff members were excluded from participation. A total of 860 users completed the survey for an overall response rate of 52.5%, as shown in Table 1-1. The results of the 2024 survey can be found on the OLCF website: https://www.olcf.ornl.gov/wp-content/uploads/OLCF-2024-User-Survey-Report.pdf.

Table 1-1. Annual survey response rate for 2024.

Survey response	2023 target	2023 actual	2024 target	2024 actual
Number of users surveyed	N/A	1,508	N/A	1,637
Total number of respondents	N/A	786	N/A	860
Percent responded	N/A	52.1%	N/A	52.5%

Users were asked to rate their satisfaction on a 5-point scale, in which a score of 5 indicates a rating of "very satisfied," and a score of 1 indicates a rating of "very dissatisfied." The metrics were agreed on by the US Department of Energy (DOE) and OLCF program manager, who defined 3.5 as "satisfactory." The effectiveness of the processes for supporting customers, resolving problems, and conducting outreach is measured, in part, by the key survey responses for User Support shown in Table 1-2.

Table 1-2. Key survey OLCF user responses for 2024.

Survey prompt	2023 target	2023 actual	2024 target	2024 actual
Overall OLCF Satisfaction Score on the User Survey	3.5/5.0	4.5/5.0	3.5/5.0	4.5/5.0
Overall Satisfaction with OLCF Support	3.5/5.0	4.5/5.0	3.5/5.0	4.5/5.0
Overall Satisfaction with OLCF Services	3.5/5.0	4.4/5.0	3.5/5.0	4.4/5.0
Overall Satisfaction with the Website	3.5/5.0	4.4/5.0	3.5/5.0	4.5/5.0
Overall Satisfaction with Communications	3.5/5.0	4.5/5.0	3.5/5.0	4.6/5.0
Overall Satisfaction with Problem Resolution	3.5/5.0	4.6/5.0	3.5/5.0	4.6/5.0
Show improvement on results that scored below satisfactory in the previous period	Results will show improvement in at least half of questions that scored below satisfactory (3.5) in the previous period.	No question scored below satisfactory (3.5/5.0) on the 2023 survey.	Results will show improvement in at least half of questions that scored below satisfactory (3.5) in the previous period.	No question scored below satisfactory (3.5/5.0) on the 2024 survey.

1.2 USER SUPPORT METRICS

The OLCF exceeded all user support metrics for 2024. The OLCF metric targets and actual results for user support by CY are shown in Table 1-3.

Table 1-3. User support metric targets and CY results.

Survey area	CY 2023 target	CY 2023 actual	CY 2024 target	CY 2024 actual
Overall OLCF satisfaction rating	3.5/5.0	4.6/5.0	3.5/5.0	4.5/5.0
Average of all user support services ratings	3.5/5.0	4.5/5.0	3.5/5.0	4.4/5.0

1.2.1 Average Rating across All User Support Questions

The calculated mean of answers to the user support services–specific questions on the 2024 survey was 4.4/5.0, thereby indicating that the OLCF exceeded the 2024 user support metric target and that users have a high degree of satisfaction with user support services. Users were asked to provide ratings of their satisfaction with support received from the wide variety of OLCF support and services. Respondents were highly satisfied with training (96%), communications (93%), documentation (97%), the projects and accounts team (97%), user assistance (95%), and the INCITE liaisons (92%). Included below are select open-ended responses to "What are the best qualities of the OLCF?" These responses highlight various aspects of user support:

- "In every part of the organization, I find interaction with OLCF staff to be incredibly rewarding. The OLCF truly demonstrates a culture of selflessness and genuine interest in user success."
- "Over the many years I have worked with the OLCF, I have found the staff and facilities to be very efficient. Most likely one of the best facilities in the world in terms of both the hardware and support. I make this statement based on significant experience with many other HPC facilities."
- "The infrastructure for applying for access to OLCF facilities (project proposals), managing projects, tracking computing allocations, and gaining access to OLCF systems—in other words all the functions of myOLCF, is exceptional. The documentation of OLCF systems like Frontier is also excellent in comparison to some other computational centers."
- "The OLCF provides leadership-class computing capability that is crucial for enabling the science we do. The OLCF's systems are very robust with limited downtimes, supported by extensive documentation and training materials, and include top-notch user support."
- "The OLCF manages world-class computational facilities, which have been second to none for the past several years. In addition to the hardware, the best asset for the OLCF is the technical and support staff. They have been instrumental to our research program for the last several years. They are responsive, professional, and thoughtful when assistance is required. I cannot thank the OLCF enough."
- "Great machines; supportive and helpful staff; great website; very detailed user guides; and I really liked the hackathon!"
- "My interactions with the account management team as well as support staff have been exceptional. The myOLCF site is also well organized and thorough."

1.2.2 Improvement on Past Year Unsatisfactory Ratings

Each year, the OLCF works to show improvement on no less than half of any questions that scored below satisfactory (3.5) in the previous year's survey. However, all questions scored above 3.5 on both the 2023 and 2024 surveys. Although no results scored below satisfactory on the 2024 survey, a thorough review of the survey identified areas that could be improved or in which new services could be added to enhance the user experience at the OLCF. In some cases, the issue had already been addressed or a solution is in development and forthcoming.

- Queue: Based on user feedback, a new partition was added to Frontier to allow for extended wall times for batch jobs that need fewer nodes. The extended partition is designed with smaller, long-running jobs in mind and features a 24-hour max wall time on jobs up to 64 nodes.
- Scratch file system interaction: On Lustre, file striping can be key to performance. Similarly, incorrect striping can have adverse effects on file system performance. To help prevent inadvertent striping across Orion's performance tier, members of the User Assistance team created a wrapper for the Lustre lfs stripping tool. The wrapper is called each time the lfs tool is executed and provides immediate feedback for cases of non-optimal striping requests. The wrapper also helps prevent the use of Orion's performance tier, thereby helping prevent negative impacts for *all* users.
- High-Performance Storage System (HPSS) interaction: The center's new longer-term storage resource, Kronos, is mounted directly on the Data Transfer Nodes. Users can interact with Kronos by using standard UNIX commands in addition to Globus.
- Based on user feedback in 2023, JAX documentation (https://docs.olcf.ornl.gov/software/analytics/jax.html) was created for users interested in using JAX on Frontier.
- In the 2023 user survey, users requested a more efficient way to manage Python virtual environments on Frontier (specifically requesting the Conda package manager). Based on this feedback, the open-source "Miniforge" tool was added to Andes, Summit, and Frontier. This tool provides users with the ability to manage Conda environments.

1.3 PROBLEM RESOLUTION METRICS

The following Operational Assessment Review (OAR) metrics were used for problem resolution.

- At least 80% of user problems are addressed (i.e., the problem is resolved or the user is told how the problem will be handled) within 3 business days.
- According to the user survey, average satisfaction ratings for problem resolution are "satisfactory" or better.

1.3.1 Problem Resolution Metric Summary

In most cases, the OLCF resolves reported problems directly by identifying and executing the necessary corrective actions. Occasionally, the facility receives problem reports that it may not be able to resolve because of factors beyond the facility's control. In such scenarios, addressing the problem requires OLCF staff to identify and carry out all corrective actions at their disposal for the given situation. For example, if a user reports a suspected bug in a commercial product, then prudent measures might be to recreate the issue; open a bug or ticket with the product vendor; provide the vendor with the necessary information about the issue; provide a workaround to the user, if possible; and track the issue to resolution with the product vendor, which may resolve the issue with a bug fix or workaround acknowledgment.

The OLCF uses Jira to track user-reported issues to ensure response goals are met or exceeded. Users can submit tickets in a variety of ways, including by email, telephone, and an online request form. Email remains the most common method. During CY 2024, the OLCF issued 3,562 tickets in response to user inquiries. The OLCF resolved 97% of issues within 3 business days, as shown in Table 1-4.

Nearly three-quarters (73.3%) of survey respondents submitted between one and five queries to the OLCF (via phone or email) in 2024, and 94% of respondents were satisfied or very satisfied with the OLCF's response to reported issues, with similarly high ratings for the *timeliness of responses to reported issues* (93% satisfied) and the *quality of technical advice given* (94% satisfied).

Table 1-4. Problem resolution metric summary.

Survey area	CY 2023		CY 2024	
	Target	Actual	Target	Actual
Percentage of problems addressed in 3 business days	80%	97%	80%	97%
Average of problem resolution ratings	3.5/5.0	4.6/5.0	3.5/5.0	4.6/5.0

1.4 USER SUPPORT AND ENGAGEMENT

The following sections discuss key activities and contributions in the areas that the OLCF recognizes as pillars of user support and engagement, including the following:

- A user support staff comprising account management liaisons, User Assistance engineers, Scientific Engagement Group liaisons, and data liaisons;
- Multiple pathways to communicate with users, sponsors, and vendors;
- Training for current and future users; and
- Strong outreach to engage the next generation of HPC users, the external media, and the public.

1.4.1 User Support

The OLCF recognizes that users of HPC facilities have a wide range of needs that require diverse solutions—from immediate, short-term, trouble ticket–oriented support (e.g., assistance with debugging and optimizing code) to more in-depth support that requires total immersion and collaboration on projects.

The facility provides complementary user support vehicles that include user assistance and outreach staff; liaisons in scientific, data, and visualization areas; and computer scientists who assist with issues related to the programming environments and tools. The following sections detail some of the high-level support activities during CY 2024 and the specific OLCF staff resources available to assist users.

1.4.2 User Support Team

The OLCF addresses user queries; acts as user advocates; covers frontline ticket triage, resolution, and escalation; provides user communications; develops and delivers training and documentation; and installs third-party applications for use on the computational and data resources. The team also manages the OLCF's Resource and Allocation Tracking System (RATS), which is the authoritative source for most of the system, user, and project data at the OLCF. In addition to some of the initiatives already mentioned in the section above, some examples of OLCF initiatives in 2024 that helped improve the overall user experience are provided below, although some of them are very much behind the scenes.

1.4.2.1 MyOLCF

The OLCF provides users with a web-based self-service application called myOLCF (https://my.olcf.ornl.gov), which offers principal investigators (PIs) and project members timely and accurate data to empower decision-making for OLCF projects and self-service tools for project administration. MyOLCF makes relevant information about projects, users, project membership applications, project applications, resource allocations, help tickets, and project usage analytics available to PIs and project members.

MyOLCF also allows PIs and project members to perform administrative tasks without contacting the OLCF Accounts Team. Documentation on myOLCF can be found on the publicly available OLCF user documentation site. The software application is under continuous development, and new features and user experience improvements are deployed twice a month on average.

Highlights for myOLCF in 2024 include the following:

- Developed a project application for the INCITE program to simplify collection of information from INCITE 2025 projects
- Deployed additional questions in the project application to simplify workflow when reviewing projects that plan to use protected health information (PHI)
- Developed an Advanced Computing Ecosystem (ACE) test bed proposal front end
- Integrated Defiant testbed information into application workflow
- Added various other improvements to help with user interaction with the facility
- Expanded internal unit tests for MyOLCF
- Added a status tracker for Business Process Steps to allow users to view the status of their applications
- Added a new My Account dropdown menu
- Added a new center status to the login page
- Added a troubleshooting FAQ page
- Added a list of changes for each release (i.e., changelog) for users to learn about new features and bug fixes
- Upgraded MyOLCF from Vue 2 to 3

1.4.2.2 RATS CRM

The OLCF's customer relationship management software, called RATS CRM, is under continuous development. RATS CRM provides the ability for the OLCF Accounts Team to process user and project applications, authorize user access to systems, track project allocations on systems, and manage certain configurations for HPC resources.

Notable updates to RATS CRM in 2024 include the following:

- Integrated new file system policies for directory trees in support of the OLCF's Kronos nearline storage system
- Improved page and API performance across the application by using caching (specifically to the home and video conference pages)
- Added a feature to manage data streams to be used by the OLCF's Analytics and Monitoring team
- Added RATS CRM endpoints to support new myOLCF features including a center status endpoint to view downtimes for computing systems and an endpoint for security classifications
- New features for the Quantum Computing User Program (QCUP), such as adding a new quantum vendor and further integrating with vendor APIs to automate account processes
- Various other improvements to business processes and system integrations to support the facility
- Implemented auto-generation of weekly DOE Integrated Project Team reports
- Transitioned infrastructure from virtual machine environment to OpenShift kubernetes cluster with new continuous integration and continuous deployment (CI/CD) pipeline
- Implemented aggregated daily usage stats for programs with very large numbers of small jobs
- Provided new metrics and charts for Air Force project usage

1.4.2.3 RATS Report

The RATS Report is vital to the OLCF's internal reporting and utilization tracking capabilities. It consists of various data pipelines, a data store, and a web-based front end for collecting information and providing dashboards for OLCF leadership. The RATS Report also tracks system availability and allocation usage across programs for all OLCF compute resources and file systems.

Notable RATS Report highlights for 2024 include the following:

- Implemented collection and display of usage metrics for the new nearline storage system, Kronos, which is the IBM Spectrum Archive system for the National Center for Computational Sciences (NCCS) moderate enclave
- Incorporated ACE test bed metrics
- Made various improvements to the data pipeline infrastructure

1.4.2.4 Rapid Operational Validation Initiative

The OLCF is collaborating on the Rapid Operational Validation Initiative (ROVI) t to develop and host data ingress, storage, and egress services for operational data from long-duration energy storage (LDES) systems. This effort involves partnerships with vendors and staff from six other national laboratories: Pacific Northwest National Laboratory (PNNL), Argonne National Laboratory (Argonne), Idaho National Laboratory, Sandia National Laboratories, and the National Renewable Energy Laboratory. A key component of the ROVI project is the implementation of a DataHub designed to capture operational data from LDES technologies. This data will support ML and AI applications for tasks such as life-cycle prediction and performance analysis. The DataHub accommodates various data types and time intervals, including bulk uploads of previously recorded data, smaller periodic updates, and real-time data streams. A proof-of-concept of the ROVI DataHub was successfully completed, with a live demo presented to stakeholders in January 2025. The team is now transitioning the service into a production environment and onboarding external collaborators to further test and refine the platform. Additionally, efforts are focused on developing and integrating advanced ML and AI workflows to maximize the potential of the DataHub and meet the requirements for life-cycle prediction and performance analysis of the LDES in the future.

1.4.2.5 OLCF Quantum Computing User Program

In 2024, the OLCF enhanced its strategic capability to offer users access to cutting-edge QC resources to meet growing demand and support discovery and innovation in scientific computing. The program helps users understand QC's unique aspects and challenges, thereby enabling them to develop and test quantum algorithms on the increasingly capable systems that are now available. The initiative, which includes research from all DOE Office of Science programs such as ASCR, Basic Energy Sciences (BES), Biological Environmental Research, High-Energy Physics, Fusion Energy Science, and Nuclear Physics, reflects a collaborative effort to expand the OLCF's with significant staff contributions from many areas inside and outside of the OLCF.

The following lists some of the notable 2024 QCUP efforts:

- In 2024, the team renewed its contract with Quantinuum, extending user access to their systems through 2025. Additionally, the QCUP added IQM to its vendor roster, bringing the total to four vendors: Quantinuum, IonQ, IBM Quantum, and IQM.
- The external documentation site (https://docs.olcf.ornl.gov/quantum/index.html) was maintained and frequently updated to provide user support documentation.

- The fifth annual Quantum Computing User Forum was held in-person at ORNL in August and brought together over 120 attendees to discuss QC advancements and allow the opportunity for QCUP users to highlight their projects. The event featured 23 talks on QC-enabled research; panels; expert-led workshops from Quantinuum, IBM Quantum, IonQ, and Amazon Braket; and a poster session with international participation.
- The inaugural Southeastern Quantum Conference was held in October to facilitate collaboration between ORNL and partner institutions across the quantum community and highlighted QCUP.
- In collaboration with NVIDIA, the OLCF hosted a standalone training workshop on how to use the CUDA Quantum tool on the OLCF Ascent and Summit HPC systems (https://www.olcf.ornl.gov/calendar/cuda-q-2024/). This workshop demonstrated a pathway for how QCUP and OLCF HPC users can utilize multiple CPUs, GPUs, and compute nodes to run a quantum simulator, using both OLCF systems and NVIDIA JupyterLab resources.
- OLCF staff members are working closely with the QCUP vendors to establish the infrastructure needed to further support hybrid QC+HPC computations at the center. By collaborating in software development of a given vendor's quantum library, the OLCF can guide how the vendor's API may interface with an HPC environment. Documentation on HPC quantum libraries was expanded to include new QCUP vendors and libraries (https://docs.olcf.ornl.gov/quantum/quantum_software/hybrid_hpc.html).
- QCUP supported 88 approved projects during the reporting period and has supported 151 distinct projects over the lifetime of the program and has tracked 68 publications that resulted from the use of these resources.
- To provide a streamlined approach for QCUP users trying to connect to the vendors, "Hello QCUP" tutorials were created (https://docs.olcf.ornl.gov/quantum/hello_qcup.html) for each QCUP vendor.

1.4.2.6 Jupyter at the OLCF

The OLCF's Jupyter provides a critical function that allows OLCF users to perform data analysis and visualization tasks at a scale that does not require Summit's or Frontier's compute resources. Jupyter allows a user to simply browse to a URL and, by using their SSO token, launch a lab instance that grants them access to a prebuilt environment of various AI/ML/programming tools, a terminal for OLCF network access, and the ability to run data visualizations through a notebook. Jupyter exists on both the OLCF's open and moderate enclaves, allowing users to access the file systems and computing resources associated with each one.

Notable Jupyter highlights include the following:

- Overall in 2024, 473 unique users launched over 13,000 sessions on both the moderate and open enclaves
- 306 sessions were used in support of OLCF training events

1.4.2.7 Summit and Alpine2 Decommission

After almost 6 years of production service and providing over 200 million node hours to researchers around the world, Summit was decommissioned on November 15, 2024. Summit's General Parallel File System (GPFS), Alpine2, became read-only on November 15 and was decommissioned January 31, 2025. The Summit and Alpine2 decommissions required users to carefully manage their allocations and migrate all needed data off Alpine2 prior to each resource's decommissioning. To ensure all users were aware of the migration timeline, a web page was created to advertise the details of the timeline and migration procedures. The link was provided in concise email messages sent to the user community. Email messages were sent over time to all users through the weekly email communication. Direct emails were also sent to specific user groups (e.g., users with large amounts of data and users with purge protection).

The notification campaign and procedures allowed the user community to successfully migrate data and utilize allocations before the decommission deadlines.

See https://docs.olcf.ornl.gov/systems/2024_olcf_system_changes.html for more details.

1.4.2.8 High-Performance Storage System Decommissioning and Kronos Transition to Operations

After decades in service and having served hundreds of users that have archived over 160 PB, the center's long term storage resource, HPSS, is reaching end of its life and will be decommissioned early in 2025. In 2024, the OLCF created a data migration strategy and user communication campaign to ensure all users were aware of key dates in the HPSS decommissioning process. In addition, the OLCF deployed a new nearline storage resource, Kronos, to serve as the new archival file system after HPSS is decommissioned. To ensure all users are aware of the migration timeline, a web page was created to advertise the details of the timeline and migration procedures. The link was provided in concise email messages sent to the user community. Email messages were sent over time to all users through the weekly email communication. Direct emails were also sent to specific user groups (e.g., users with large amounts of data and users with purge protection). The notification campaign and procedures allowed the center and user community to successfully migrate data from HPSS to Kronos in preparation for HPSS decommissioning in early 2025.

See https://docs.olcf.ornl.gov/systems/2024_olcf_system_changes.html for more details.

1.4.2.9 Scientific Liaisons

Scientific Liaisons—experts in scientific domains and computation—partner with facility users to obtain optimal scientific results from the OLCF's computational resources and systems. The Scientific Liaisons include experts in chemical and materials sciences; nuclear physics such as nuclear structure and quantum chromodynamics; high-energy physics such as particle physics; astrophysics such as stellar evolution and cosmology; and climate science, geophysics, biology, biomedical sciences, and engineering. Some unsolicited user feedback on the Scientific Liaisons is provided below:

- "Our INCITE liaison Isaac Lyngaas is a co-PI and has been instrumental in getting the team in the right place to use our allocation appropriately. He is responsive, productive, and a pleasure to work with."
- "Vassili Mewes has been extremely helpful. As mentioned, he has offered invaluable advice and comments about simulation techniques and robustness strategies. He has also been quick to offer advice about the Frontier queue and policies to help us run our somewhat unique workload efficiently."
- "Our INCITE liaison, Reuben Budiardja, has been wonderful. He's always helpful and responsive and is interested in our project being successful. It's extremely beneficial to have OLCF scientists directly connected with the ongoing INCITE programs."
- "Our liaison, Reuben Budiardja, has been extremely beneficial for our project. He has been very responsive and helpful in answering questions, providing advice on portability/optimizations on Frontier, and quickly diagnosing issues. Importantly, he has provided us with up-to-date build of the essential Hipfort libraries. He has also been invaluable in diagnosing and aiding in finding a solution to a recent file system I/O performance issue and in helping to make the code more portable to other AMD-GPU systems (CSC LUMI)."
- "It has always been a pleasure to interact with Stephen Nichols; we greatly appreciate the continuity he has provided as our liaison in recent years."
- "We continue to be very grateful for the effectiveness and helpfulness of our Center for Accelerated Application Readiness (CAAR)/INCITE liaison Dr. Stephen Nichols. Dr. Nichols has been highly

engaged every week for more than 4 years now. He has played a vital role in helping us understand the details of HIP interfaces (for our Fortran codes), of rocfft work buffers, CEE and rocm library versions, and of memory monitoring on the GPUs. More recently, he also conducted MPI (Message Passing Interface) scaling experiments in a very insightful manner that is now integrated into our main joint manuscript. The quality of Dr Nichols's work and the depth of his commitment to our project demonstrates the importance of the OLCF's and the DOE's investment in this type of PI-staff relationship for demanding projects."

- "Liaison Murali Meena has been fantastic with connecting us with resources and people at the center and with leading part of our analysis effort related to his expertise in data science."
- "Phil Roth has been great in helping us with the scheduler and srun peculiarities."
- "Kalyan Gottiparthi has been very helpful in porting the code onto Frontier."
- "The center was very helpful in helping us get our VASP simulations working. Markus Eisenbach and Spencer Ward were vital in us getting our simulations running. We thank them both for their help."
- "David Rogers (our INCITE Liaison) is very efficient and supportive in advancing the execution of our scientific needs."
- "Our INCITE liaison, Dr. Antigoni Georgiadou, has been highly beneficial to our team. Computational progress has not been trivial. There have been some hanging and node crash issues. Our liaison has been very helpful in getting these issues resolved. Dr. Georgiadou participates in our team conference call every week."
- "Our liaison, Dr. Swarnava Ghosh, assisted us in compiling LAMMPS with GPU acceleration. He also provided examples of Bash scripts to run LAMMPS efficiently, thereby accelerating our progress. During our discussions, Dr. Ghosh clarified many questions that we had about policies and processes for our INCITE allocations that are different from other resources that we have used previously. During these meetings, he recommended that we set up an automatic framework that copies simulation data via globes and back up important data files to HPSS. We find that the advice from Dr. Ghosh was very helpful, and we are setting up the automatic data transfer pipeline for high throughput simulations to be performed in the next quarters. We are now working with Dr. Ghosh to set up a better framework for ensemble simulations."
- "The help from the liaison, Dr. Van Ngo, has been instrumental in moving the project forward."

1.4.3 OLCF User Group and Executive Board

1.4.3.1 OLCF User Group Monthly Webinar Series

PIs and users on approved OLCF projects are automatically members of the OLCF User Group (OUG) and remain members for 3 years following the conclusion of their OLCF project. The OUG meets 10 times a year via the OLCF User Conference Call webinar series. The OLCF User Call provides users with a forum to discuss issues and ask questions about OLCF resources and offers training on timely topics to start the discussion. Table 1-5 contains a list of meetings that occurred in CY 2024.

Table 1-5. OLCF User conference call webinar attendance.

Topic	Date	Speaker(s)	Participants
First Experiences at the Exascale with Parthenon	January 31, 2024	Philipp Grete (Hamburg Observatory)	101
Large-scale DFT simulations commensurate with quantum accuracy	February 28, 2024	Sambit Das, Bikash Kanungo (UMICH), and David Rogers (ORNL)	112
Unlocking the power of LLMs with NVIDIA NeMo	March 27, 2024	Zahra Ronaghi and Janaki Vamaraju (NVIDIA)	102

Topic	Date	Speaker(s)	Participants
An Overview of the Fortran Standard: Fortran 2023 and Beyond	April 24, 2024	Reuben Budiardja (OLCF)	96
Omnitrace by Example	May 29, 2024	Gina Sitaraman (AMD)	102
Linaro Forge	June 26, 2024	Beau Paisley (Linaro)	69
ADIOS	July 31, 2024	Norbert Podhorszki (ORNL)	88
Containers on Frontier	August 28, 2024	Subil Abraham (OLCF)	116
Kronos Overview and Data Migration	October 30, 2024	Gregg Gawinski and Jake Wynne (OLCF)	94
Velocity	December 04, 2024	Asa Rentschler (OLCF)	65

1.4.3.2 OUG Executive Board

The OUG is represented by the OUG Executive Board. This board meets monthly for in-depth discussions with OLCF staff to provide feedback and guidance on training topics as well as the facility's resources and policies. The OUG Executive Board's terms are 3 years and are staggered so that 3 new members are elected each year. Additionally, an outgoing chair remains on the board for 1 year as an ex officio member if the term as chair is their third year on the board. Emily Belli (General Atomics) replaced Steven Gottlieb (Indiana University) as chair for the 2024–2025 term. Sara Isbill (ORNL), Min Xu (ORNL), Mia Li (University of Oklahoma), and Steve de Bruyn Kops (University of Massachusetts) began new 3-year terms that will conclude in 2027. The current Executive Board is listed at https://www.olcf.ornl.gov/about-olcf/oug/.

1.4.3.3 OLCF User Meeting

The 20th annual OLCF User Meeting was held September 10–11, 2024, for the OLCF user community, HPC researchers, and industry collaborators. Attendees gathered to learn about the latest updates on OLCF systems and software, receive hands-on training, and discuss new research opportunities. Presentations and discussions covered topics such as the Frontier supercomputer, HPC containers, performance analysis tools, and QC integration. This annual event served as a venue for exchanging ideas, offering user support, and guiding the future direction of HPC efforts at the OLCF. In 2024, the OLCF hosted the first Data Visualization Showcase to complement the Posters track. Furthermore, the OLCF opened submissions for user-contributed talks. The talks were reviewed by the OUG Executive Board and were chosen to be presented at the 2024 OLCF User Meeting.

See https://www.olcf.ornl.gov/calendar/2024-olcf-user-meeting/for more information.

1.4.4 Training

In 2024, OLCF users had the choice of attending 40 training events, many of which were focused on performance portability across heterogeneous HPC systems while also expanding user support for emerging technologies such as QC and AI and offering hands-on hackathons and practical workshops to enhance user proficiency on Frontier and other advanced platforms. In most cases, the training events are recorded, and the slides, recordings, and hands-on materials are made available to users through the OLCF training archive: https://docs.olcf.ornl.gov/training/training_archive.html. The OLCF continued to collaborate on training series and workshops with other HPC centers such as the National Energy Research Scientific Computing Center (NERSC) and the Argonne Leadership Computing Facility (ALCF) to maximize the benefit for the often-shared user bases and share the load of course development. Many training events were provided or enhanced by contributions from HPE and AMD Frontier Center of Excellence staff.

Survey respondents in 2024 rated the training at 4.5 out of 5.0, and 96% were highly satisfied with the training overall. Usefulness of the online training archive and the quality of the training content received a high rating (93%), whereas the lowest rated aspect was the breadth of training events offered at 90%.

Notable highlights from training sessions are provided below.

1.4.4.1 Training Series

The OLCF expanded several 2023 series into 2024, including the OUG Monthly Webinar Series, AI Training Series, and Performance Portability Series. Four AI-focused events provided insights and hands-on experience with AI and HPC integration. Highlights included scaling large language models (LLMs) with NVIDIA NeMo, optimizing PyTorch on Frontier, advanced parallelization for deep learning, and integrating AI with scientific simulations. See Table 1-6 for details.

The Performance Portability Series, a joint effort by the OLCF, NERSC, and the ALCF, included 12 events—most notably a seven-part, hands-on OpenMP training series hosted by NERSC, with OLCF offering two office hours to support participants. The series also featured a Julia for Science training session developed in collaboration with ORNL's Computer Science and Mathematics Division, Neutron Scattering Division and LBL's NERSC. This event prepared experimental and HPC User facility attendees to use Julia for HPC and Science.

The OLCF hosted two Quantum computing training events and they are described in Table 1-6 below.

The OLCF aimed to ensure that users onboarded throughout the year were never far from a training session that would help them get started on Frontier and at the OLCF. Seven events throughout the year focused on teaching users how to use tools and software on Frontier and other OLCF resources. The OLCF held regular Monday and Wednesday office hours throughout the year. These office hours were staffed by experts from the OLCF and the HPE/AMD Frontier Center of Excellence, and users could sign up to work on any problems they encountered while using any of the OLCF's systems. The OLCF also held a new user training workshops for Summit users and Frontier users. New this year was a more hands-on New User Training that actively engaged participants with frequent programming exercises every few minutes, guiding them through running on Summit and Frontier, managing data, and navigating user guides.

The agenda, recordings, and slides are available at https://docs.olcf.ornl.gov/training/training_archive.html.

Table 1-6. All user training events.

Category	Event Title	Description	Host	Date	Attendees
OLCF User Conference Calls	January 2024 OLCF User Conference Call: First experiences at the exascale with Parthenon	First experiences at the exascale with Parthenon.	OLCF	January 31,2024	101
Performance Portability Events	Performance Portability for Next-Generation Heterogeneous Systems	Training on writing high-performance applications optimized for diverse supercomputing architectures.	OLCF, NERSC, and ALCF	February 26, 2024	79

Category	Event Title	Description	Host	Date	Attendees
How to Use OLCF	OLCF New User Training	Introductory training for new users of OLCF systems, focusing on best practices and system use.	OLCF	February 28, 2024	50
OLCF User Conference Calls	February 2024 OLCF User Conference Call: Gordon Bell User Experience Talk	Insights into large-scale DFT simulations on the Frontier supercomputer.	OLCF	February 28, 2024	112
How to Use OLCF	AMD & HPE Profiler Tutorial	Profiling and performance analysis using AMD's Omniperf and HPE tools.	OLCF	March 1, 2024	44
Hackathons	Frontier Hackathon – March 2024	A virtual 4-day hackathon for collaborative coding and performance optimization on Frontier.	OLCF, HPE, and AMD	March 4–8, 2024	50
Performance Portability Events	Performance Portability Series: AMRex Tutorial	Training on AMReX for developing portable applications on HPC architectures.	OLCF and NERSC	March 14, 2024	77
OLCF User Conference Calls	March 2024 OLCF User Conference Call: NVIDIA NeMo	Discussion on using NVIDIA NeMo for scaling large language models (LLMs).	OLCF	March 27,2024	102
AI-Focused Events	AI Training Series: Enhancing PyTorch Performance on Frontier with the RCCL/OFI-plugin	A seminar on optimizing PyTorch performance on the Frontier supercomputer using the AWS-OFI-RCCL plugin. This session covers RCCL basics, plugin integration with PyTorch, distributed multi-node usage, and profiling techniques.	OLCF	April 17, 2024	44
OLCF User Conference Calls	April 2024 OLCF User Conference Call: Fortran 2023 and Beyond	An overview of the latest Fortran 2023 standard and future developments.	OLCF	April 24, 2024	96
Performance Portability Events	Performance Portability Series: Kokkos Training	A 2-day training on the Kokkos programming model for developing performance-portable applications across HPC systems with AMD, NVIDIA, and Intel GPUs.	OLCF, NERSC, and ALCF	April 25–26, 2024	122
How to Use OLCF	Kokkos 4.3 Release Briefing	A briefing on Kokkos 4.3 updates, including new features and development changes.	Kokkos Team	April 29, 2024	60

Category	Event Title	Description	Host	Date	Attendees
Performance Portability Events	OpenMP Training Series: OpenMP Introduction	Introduction to OpenMP and parallel programming for HPC systems in the Performance Portability Training Series.	NERSC, OLCF, and ALCF	May 6, 2024	105
Performance Portability Events	OpenMP Series Office Hour	A dedicated office hour session for OLCF users participating in the NERSC-hosted OpenMP training series, offering personalized guidance on homework and hands-on tasks related to Frontier and other OLCF systems.	OLCF	May 9, 2024	4
OLCF User Conference Calls	May 2024 OLCF User Conference Call: Omnitrace on Frontier	Introduction to AMD's Omnitrace profiling tool for tracing and analyzing applications on Frontier.	OLCF	May 29,2024	102
Performance Portability Events	OpenMP Training Series: Tasking	Training on task-based parallelism in OpenMP.	NERSC, OLCF, and ALCF	June 10, 2024	65
Performance Portability Events	Julia for HPC and Intro to Julia for Science	A 2-day workshop for HPC users and neutron scattering scientists about how to utilize Julia and HPC via Julia for their science.	OLCF, NERSC, NScD, CSMD	June 18,2024	128
Performance Portability Events	OpenMP Training Series OLCF Office Hours	A dedicated office hour session for OLCF users participating in the NERSC-hosted OpenMP training series, offering personalized guidance on homework and hands-on tasks related to Frontier and other OLCF systems.	OLCF	June 20,2024	3
Performance Portability Events	Portable SYCL Code Using oneMKL on AMD, Intel, and NVIDIA GPUs	Writing portable SYCL code with oneMKL for heterogeneous GPU systems.	NERSC, OLCF, and ALCF	June 20, 2024	81
OLCF User Conference Calls	June 2024 OLCF User Conference Call: Linaro Forge	Overview of Linaro Forge for debugging and optimizing parallel applications.	OLCF	June 26, 2024	69
How to Use OLCF	New User Training Hands-On 2024	Hands-on introduction to OLCF systems for new users.	OLCF	June 27, 2024	10

Category	Event Title	Description	Host	Date	Attendees
Performance Portability Events	OpenMP Training Series: Optimization for NUMA & SIMD	A session focused on optimizing applications for NUMA (Non-Uniform Memory Access) architectures and SIMD (Single Instruction, Multiple Data) processing in OpenMP to improve performance on modern HPC systems.	NERSC, OLCF, and ALCF	July 8, 2024	58
AI-Focused Events	AI Training Series: AI for Science at Scale – Part 3	Hands-on training for scaling and fine-tuning deep learning models on Frontier using advanced parallelization techniques.	OLCF	July 11, 2024	58
OLCF User Conference Calls	July 2024 OLCF User Conference Call: The ADIOS I/O Framework	An overview of the ADIOS (Adaptable I/O System) framework, designed to optimize data movement and storage for high-performance scientific computing.	OLCF	July 31, 2024	88
Performance Portability Events	OpenMP Training Series: What Could Possibly Go Wrong Using OpenMP	Common pitfalls and how to avoid them when using OpenMP.	NERSC, OLCF, and ALCF	August 5, 2024	41
Quantum Computing Events	2024 Quantum Computing User Forum	Discussions on quantum computing applications and software development.	OLCF	August 12–15, 2024	120
Quantum Computing Events	Cuda-Q 2024	Training about a software development kit for quantum and integrated quantum-classical programming.	OLCF	August 22, 2024	36
Hackathons	Frontier Hackathon August 2024	Hybrid hackathon focused on advancing applications on Frontier.	OLCF, HPE, and AMD	August 23, 2024	55
How to Use OLCF	Kokkos 4.4 Release Briefing	Updates on the Kokkos 4.4 release and future development.	Kokkos Team	August 27, 2024	65
OLCF User Conference Calls	August 2024 OLCF User Conference Call: Containers on Frontier	Overview of container support on Frontier using Apptainer.	OLCF	August 28, 2024	116
How to Use OLCF	ParaView on Frontier	Training on using ParaView for data analysis and visualization on Frontier.	OLCF	August 29, 2024	20

Category	Event Title	Description	Host	Date	Attendees
Performance Portability Events	OpenMP Training Series: Introduction to Offloading with OpenMP	Techniques for offloading computations using OpenMP.	NERSC, OLCF, and ALCF	September 5, 2024	45
How to Use OLCF	2024 OLCF User Meeting	Annual event sharing OLCF user achievements and future plans, including hands-on Containers and HPCToolKit training.	OLCF	September 10–11, 2024	173
AI-Focused Events	SmartSim 2024	Workshop on integrating AI with scientific simulations.	OLCF	September 26, 2024	20
Performance Portability Events	OpenMP Training Series: Advanced OpenMP Offloading Tips	Advanced techniques for efficient OpenMP offloading.	NERSC, OLCF, and ALCF	October 7, 2024	40
Performance Portability Events	OpenMP Training Series: Selected/Remaining Topics	Final session covering remaining OpenMP topics.	NERSC, OLCF, and ALCF	October 29, 2024	29
OLCF User Conference Calls	October 2024 OLCF User Conference Call: Kronos Overview and Data Migration	Introduction to Kronos, OLCF's new nearline storage system.	OLCF	October 30, 2024	94
OLCF User Conference Calls	December 2024 OLCF User Conference Call: Velocity	Overview of Velocity for managing container build scripts.	OLCF	December 4, 2024	65
How to Use OLCF	MPI Best Practices for OLCF	Best practices for using HPE Cray MPI on Frontier.	OLCF	December 5, 2024	55
How to Use OLCF	Kokkos 4.5 Release Briefing	A briefing on updates in Kokkos 4.5 and upcoming development plans.	Kokkos Team	December 10, 2024	65

1.4.5 Community Engagement

With STEM careers projected to outpace non-STEM fields in growth over the next decade, fostering interest in computing and advanced technologies is essential. OLCF staff support this effort by organizing seminars, workshops, and tutorials that expand HPC and AI training while engaging students in STEM.

In 2024, the OLCF sustained its community engagement efforts from 2023 by providing challenge problems, HPC training, and resources for the Winter Invitational Classic Student Cluster Competition and the Data Challenge at the Smoky Mountains Computational Sciences and Engineering Conference; by mentoring for the Faculty Hackathon; and by implementing improvements to the hands-on HPC Crash Course.

Additionally, OLCF staff further developed the Next Generation Pathways (NGP) Summer School curriculum, a 5-week program in which high school students collaborate with ORNL researchers and explore HPC and AI.

Notable highlights from these initiatives are outlined below.

1.4.5.1 Next Generation Pathways to Computing Summer School

In the summer of 2023, ORNL launched the NGP, a 5-week summer school designed to inspire students to choose computing careers. Building on its initial success, the program more than doubled its student capacity in 2024, bringing together 24 high school students from East Tennessee, representing a broad range of schools. The program was managed by a project team that included staff from ORNL's Office of Research Education, the Spallation Neutron Source (SNS), and the OLCF, which also contributed to curriculum development.

Students began by learning Python using NVIDIA JetBots and Jupyter notebook-based coding workbooks developed by NVIDIA, with additional contributions from SNS and OLCF staff. Each pair of students was matched with an ORNL mentor from the research or technical staff to work on small projects throughout the program. In addition, students participated in a modified version of the OLCF's 4-week Hands-On HPC Crash Course. To support student learning, OLCF staff provided supplemental instruction for each module to ensure a smoother transition into the material. For modules presented virtually, an in-person instructor guided students through the coursework. This year's curriculum expanded to include additional AI lessons taught by ORNL researchers.

OLCF staff played a key role in working with the Office of Research Education to secure Workforce Development for Teachers and Scientists (WDTS) funding for the program. They also led curriculum development, set up and tested the NVIDIA JetBots, taught many of the programming sessions, and mentored research projects for six students. By the end of the program, ten students presented posters to showcase their research projects.

1.4.5.2 HPC Crash Course

In 2024, the OLCF continued to build its Hands-On HPC Crash Course, adding a beginner-friendly Jupyter notebook-based AI challenge and a QC simulator challenge. The course maintained its certification program, awarding certificates to students who successfully completed at least 7 of the 17 programming challenges.

The course was given four times this year at computing conferences or at ORNL, providing HPC, AI, and QC instruction to a total of 318 students.

The largest offering of the HPC Crash Course was a 4-week summer program tailored to ORNL summer interns and the NGP high school students. Although most students participated virtually, high school students attended in person. The summer session resulted in 69 certificates awarded out of the 144 participants.

The OLCF continued its collaboration with Purdue's Rosen Center for Advanced Computing, adapting the Frontier version of the course for Purdue's Anvil Supercomputer and co-teaching the course at the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC24), where 44 of the 60 participants earned certificates for the 1-day version of the course.

To support ongoing engagement, OLCF staff continued its LinkedIn page for course alumni, providing a platform for continued training, workshop opportunities, and career tracking.

1.4.5.3 Faculty Hackathon

For the third year in a row, the OLCF helped plan and participated in the Faculty Hackathon, a Science Gateways Community Initiative event designed to introduce faculty to HPC concepts, tools, and techniques applicable in their courses. This in turn exposes students to HPC, preparing them for further studies or careers in the field. This program broadens participation in HPC and creates an environment for collaboration among HPC educators from different institutions and industries.

In 2024, a total of nine faculty teams, including nine faculty members from Albany State University, Claflin University, Elizabeth City State University, Jackson State University, Lincoln University of Pennsylvania, Mississippi Valley State University, and Our Lady of the Lake University, participated. ORNL mentored five of the faculty teams. See https://hackhpc.github.io/facultyhack-gateways24/teams.html.

1.4.5.4 PEARC24 and SC24 Student Program

OLCF staff members participated in the 2024 Practice and Experience in Advanced Research Computing (PEARC24) student program by serving as the lead organizers for three workshops (the Resume Clinic workshop, PitchIt! Perfecting your Elevator Speech workshop, and Negotiation Workshop) as well as the Student Volunteers program. For the resume workshop, professionals from national labs, industry, and universities reviewed student resumes and provided best practices and tips for how to tailor their resumes to different fields. For the PitchIt! workshop, OLCF staff and university faculty showed students how to present a short sales pitch to introduce themselves and their research. Students then broke into teams to practice their pitches with volunteer mentors. In the Negotiation Workshop, the students were taught how to navigate the job interview process and what factors to consider when comparing job offers. Following the workshop, staff organized a panel of professionals to answer the students' questions. Lastly, OLCF staff coordinated the PEARC Student Volunteers Program, working with the broader PEARC committee to ensure presentations went smoothly and gathering data such as session attendance.

1.4.5.5 Smoky Mountains Computational Sciences and Engineering Conference, Data Challenge session

In 2024, OLCF staff members organized the annual Smoky Mountains Computational Sciences and Engineering Conference's Data Challenge Session. This year's challenge took the form of a "Trustworthy AI for Science" essay contest, inviting participants to explore the ethical and scientific implications of using AI training datasets or AI processes in biomedical science, geospatial information science, or climate science for scientific discovery. From ten finalists, five were selected to deliver TEDx-style talks at the conference in September 2024. To prepare them for these presentations, the finalists received training through ORNL's "Science in a Nutshell" program, which focuses on creating engaging, public-facing science talks.

1.4.5.6 DOE National Laboratory Day at the 2024 Richard Tapia Computing Conference

The OLCF has been a key player in the DOE Labs Day at the Richard Tapia Conference for the past few years. The annual Richard Tapia Conference gathers IT professionals and students with the goal of promoting leadership among students, faculty, and professionals. As a result, the DOE Labs Day aims to educate and train students on the operations of the DOE laboratories. Staff from Lawrence Livermore

National Laboratory (LLNL), Los Alamos National Laboratory (LANL), PNNL, UCAR, and ORNL, among others, represented the national laboratories at DOE Labs Day.

OLCF staff introduced students to GPU programming by using OpenACC, OpenMP Offload, and CUDA in the 2024 cohort of the conference. Students participated in hands-on exercises in which they offloaded applications to the GPUs of the OLCF's Frontier supercomputer. A total of 140 students attended the GPU programming session at the event in San Diego, CA, in September of 2024.

1.4.5.7 SC24

In 2024, OLCF staff members maintained a high level of participation in SC24. In addition to technical content, staff members served in various roles and leadership positions in the conference's organizing committee. Staff members were involved in every aspect of conference planning, decision-making on the executive committee, managing finances, organizing student programs, and technical planning of SCINet, to name a few. This allowed staff members to engage and work with the broader HPC community and with staff members from other institutions.

1.4.5.8 USRSE24

In addition to presenting technical presentations at USRSE24 (US Research Software Engineering 2024), a conference that encourages collaboration between software engineers and researchers writing scientific code, OLCF staff members served on the executive conference committee. Staff members also worked with other committee members on communication and outreach to broaden participation in the conference and served on the Code of Conduct Committee.

2. OPERATIONAL PERFORMANCE

CHARGE QUESTION 2: Did the facility's operational performance meet established targets?

OLCF RESPONSE: Yes. The OLCF provides highly capable and reliable systems for the user community. The 2024 reporting period covers production operations during the CY for the following HPC resources: the HPE Cray EX Frontier, the IBM AC922 Summit (January 24–November 15), the Alpine2 GPFS (January 23–December 31), the Lustre file system Orion, and the archival storage system (HPSS). In 2024, the OLCF once again delivered all the compute hours committed to the three major allocation programs: INCITE, ALCC, and DD. The operational performance demonstrates that the OLCF delivered another prominent operational year of reliable and technically sufficient resources to the scientific research community.

2.1 RESOURCE AVAILABILITY

The OLCF exceeded all resource availability metrics in 2024 for the Frontier (Table 2-1) and Summit (**Error! Reference source not found.**) computational resources, HPSS data resources (Table 2-3), the Orion resources (Table 2-4) and the Alpine2 resources (Table 2-5). Supporting systems such as EVEREST, Andes, data transfer nodes, Themis (nearline storage), and Kronos (archival storage) were also offered. Metrics for these supporting systems are not provided. See APPENDIX C for more information about each of these systems. The following tables describe the availability of OLCF resources. More details on the definitions and formulae that describe the scheduled availability (SA), overall availability (OA), mean time to interrupt (MTTI), and mean time to failure (MTTF) are provided in APPENDIX D.

Operational performance metrics are provided for OLCF's Frontier, Summit, the HPSS archive system, Orion, and the Alpine external GPFS (Table 2-1 -- Table 2-5).

Table 2-1. OLCF operational performance summary for Frontier.

	Measurement	2023 target	2023 actual	2024 target	2024 actual
HPE Cray EX Frontier	SA	85%	99.65%	90%	99.81%
	OA	80%	97.64%	85%	98.49%
	MTTI (hours)	NAM[a]	397	NAM	665
	MTTF (hours)	NAM	1,621	NAM	2,192

[a] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

Table 2-2. OLCF operational performance summary for Summit (January 24–November 15, 2024).

	Measurement	2023 target	2023 actual	2024 target	2024 actual
IBM AC922 Summit	SA	95%	99.84%	95%	98.92%
	OA	90%	99.41%	90%	98.53%
	MTTI (hours)	NAM [a]	1,396	NAM	1,171
	MTTF (hours)	NAM	4,205	NAM	2,351

[a] NAM = Not a metric. No defined metric or target exists for this system. Data provided as reference only.

Table 2-3. OLCF operational performance summary for HPSS.

	Measurement	2023 target	2023 actual	2024 target	2024 actual
HPSS	SA	95%	99.65%	95%	99.39%
	OA	90%	96.10%	90%	98.37%
	MTTI (hours)	NAM	443	NAM	411
	MTTF (hours)	NAM	2,910	NAM	970

NAM data is provided as reference only.

Table 2-4. OLCF operational performance summary for the Orion external Lustre file system.

	Measurement	2023 target	2023 actual	2024 target	2024 actual
Orion	SA	90%	99.08%	95%	99.69%
	OA	85%	98.42%	90%	98.92%
	MTTI (hours)	NAM	397	NAM	621
	MTTF (hours)	NAM	545	NAM	973

NAM data is provided as reference only.

Table 2-5. OLCF operational performance summary for the Alpine2 external GPFS.

	Measurement	2023 target	2023 actual	2024 target	2024 actual
Alpine	SA	N/A	N/A	90%	98.96%
	OA	N/A	N/A	85%	98.49%
	MTTI (hours)	NAM	NAM	NAM	1,016
	MTTF (hours)	NAM	NAM	NAM	2,043

NAM data is provided as reference only.

In CY 2024, Summit saw a significant decrease in MTTF due to instability of the new Alpine2 GPFS. When the Alpine2 filesystem was deployed to support Summit during Year 6 (SummitPLUS), the high-speed network technology previously utilized by Summit and Alpine was no longer available for purchase. As a result, it was necessary to procure next-generation technology. This shift in technology led to interoperability issues in communication and took time to stabilize. Additionally, it triggered other software bugs that required resolution with the vendor. These two issues led to a couple of significant unscheduled downtimes for Summit. Frontier saw significant improvements in MTTI and MTTF due to increased stability in the SlingShot fabric and other systems software. Frontier saw significant improvements in MTTI and MTTF due to increased stability in the SlingShot fabric and other systems software. For a period of 1 year following either system acceptance or a major system upgrade, the SA target for an HPC compute resource is at least 85%, and the OA target is at least 80%. For year two, the SA target for an HPC compute resource increases to at least 90%, and the OA target increases to at least 85%. For year three through the end of life for the associated compute resource, the SA target increases to 95%, and the OA target increases to 90%. Consequently, for HPC compute resources, SA targets are described as 85%/90%/95%, and OA targets are described as 80%/85%/90%.

For a period of 1 year following either system acceptance or a major system upgrade, the SA target for an external file system is at least 90%, and the OA target is at least 85%. For year two through the end of life of the asset, the SA target for an external file system increases to at least 95%, and the OA target increases to at least 90%. Consequently, for an external file system, SA targets are described as 90%/95%, and OA targets are described as 85%/90%.

2.1.1 Scheduled Availability

SA is the percentage of time that a designated level of resource is available to users, excluding scheduled downtime for maintenance and upgrades. The user community must be notified of a maintenance event no less than 24 hours in advance of the outage (emergency fixes) for it to be considered scheduled downtime. Users will be notified of regularly scheduled maintenance in advance: no less than 72 hours prior to the event and preferably as many as 7 calendar days prior. If that regularly scheduled maintenance is not needed, then users will be informed of the cancellation of that maintenance event in a timely manner. Any interruption of service that does not meet the minimum notification window is categorized as an unscheduled outage.

A significant event that delays the return to scheduled production by more than 4 hours will be counted as an adjacent unscheduled outage, as an unscheduled availability, and as an additional interrupt.

SA is described by Eq. (2.1). The OLCF has exceeded the SA targets for the facility's computational resources for 2023 and 2043 (Table 2-6).

$$SA = \left(\frac{\text{time in period} - \text{time unavailable due to outages in period}}{\text{time in period} - \text{time unavailable due to scheduled outages in period}}\right) * 100 \quad (2.1)$$

Table 2-6. OLCF operational performance summary: SA.

	System	2023 target	2023 actual	2024 target	2024 actual
SA	IBM AC922	95%	99.84%	95%	98.92%
	Frontier	85%	99.65%	90%	99.81%
	HPSS	95%	99.65%	95%	99.39%
	Alpine2	N/A	N/A	90%	98.96%
	Orion	90%	99.08%	95%	99.69%

2.1.1.1 OLCF Maintenance Procedures

Preventative maintenance is exercised only with the concurrence of the vendor hardware and software teams, the OLCF HPC Systems groups, and the OLCF Resource Utilization Council. Typical preventative maintenance activities include software updates, application of field notices, and hardware maintenance to replace failed components. Without concurrence, the systems remain in their respective normal operating conditions. Preventative maintenance is advertised to users a minimum of 2 weeks in advance if the maintenance activities include changing default software and a minimum of 1 week in advance if default software is not being changed.

2.1.2 Overall Availability

OA is the percentage of time that a system is available to users. Outage time reflects both scheduled and unscheduled outages. The OA of OLCF resources is derived by using Eq. (2.2).

$$OA = \left(\frac{\text{time in period} - \text{time unavailable due to outages in period}}{\text{time in period}}\right) * 100 \quad (2.2)$$

As shown in Table 2-7, the OLCF exceeded the OA targets of the facility's resources for 2023 and 2024.

Table 2-7. OLCF operational performance summary: OA.

	System	2023 target	2023 actual	2024 target	2024 actual
OA	IBM AC922	90%	99.41%	90%	98.53%
	Frontier	80%	97.64%	85%	98.49%
	HPSS	90%	96.10%	90%	98.37%
	Alpine2	N/A	N/A	90%	98.49%
	Orion	85%	98.42%	90%	98.92%

2.1.3 Mean Time to Interrupt

MTTI is, on average, the time to any outage of the full system, whether unscheduled or scheduled. It is also known as MTBI. The MTTI for OLCF resources is derived by Eq. (2.3), and a summary is shown in Table 2-8.

$$\text{MTTI} = \left(\frac{\text{time in period} - (\text{duration of scheduled outages} + \text{duration of unscheduled outages})}{\text{number of scheduled outages} + \text{number of unscheduled outages} + 1}\right) \quad (2.3)$$

Table 2-8. OLCF operational performance summary: MTTI.

	System	2023 actual	2024 actual
MTTI [a] (hours)	IBM AC922	1,396	1,171
	Frontier	397	665
	HPSS	443	411
	Alpine2	N/A	1,016
	Orion	397	621

[a] MTTI is not an assigned metric. Data provided as reference only.

2.1.4 Mean Time to Failure

MTTF is the time, on average, to an unscheduled outage of the full system. The MTTF for OLCF resources is derived from Eq. (2.4), and a summary is provided in Table 2-9.

$$MTTF = \frac{\text{time in period} - (\text{duration of unscheduled outages})}{\text{number of unscheduled outages} + 1} \quad (2.4)$$

Table 2-9. OLCF operational performance summary: MTTF.

	System	2023 actual	2024 actual
MTTF [a] (hours)	IBM AC922	4,205	2,351
	Frontier	1,621	2,192
	HPSS	2,910	970
	Alpine2	N/A	2,043
	Orion	545	973

[a] MTTI is not an assigned metric. The data provided as reference only.

2.2 TOTAL SYSTEM UTILIZATION IN 2024

2.2.1 Resource Utilization Snapshot

During the operational assessment period (January 1, 2024–December 31, 2024), 23,735,169 Summit node hours were used outside of outage periods from an available 32,363,750 node hours. The total system utilization (SU) for the IBM AC922 Summit was 73.34%. On Frontier, 72,453,223 node hours were used outside of outage periods from an available 83,607,613 node hours. The total SU for the HPE Cray EX Frontier was 86.66%.

2.2.2 Total System Utilization

2.2.2.1 2024 Operational Assessment Guidance

SU is the percentage of time that the system's computational nodes run user jobs. The SU for OLCF resources is derived from Eq. (2.5). No adjustment is made to exclude any user group, including staff and vendors.

$$SU = \left(\frac{\text{core hours used in period}}{\text{core hours available in period}}\right) * 100 \quad (2.5)$$

The measurement period is for 2024, regardless of the prescribed allocation period of any single program. As an example, the INCITE allocation period follows a CY schedule. The ALCC program follows an allocation cycle that runs for 12 months beginning July 1 of each year. The OLCF tracks the consumption of Summit and Frontier node hours by job. This method is extended to track the consumption of Summit and Frontier node hours by program, project, user, and system with high fidelity. Figure 2-1 shows the IBM AC922 Summit utilization by month and by program for CY 2024. Figure 2-2 shows the HPE Cray EX Frontier utilization by month and by program for CY 2024. The three major OLCF user programs and usage by the ECP are represented, but the graph does not include consumed node hours from staff or vendor projects. For the 6th production year of Summit, utilization was 73.34%. For the 2nd production year of Frontier, utilization was 86.66%.

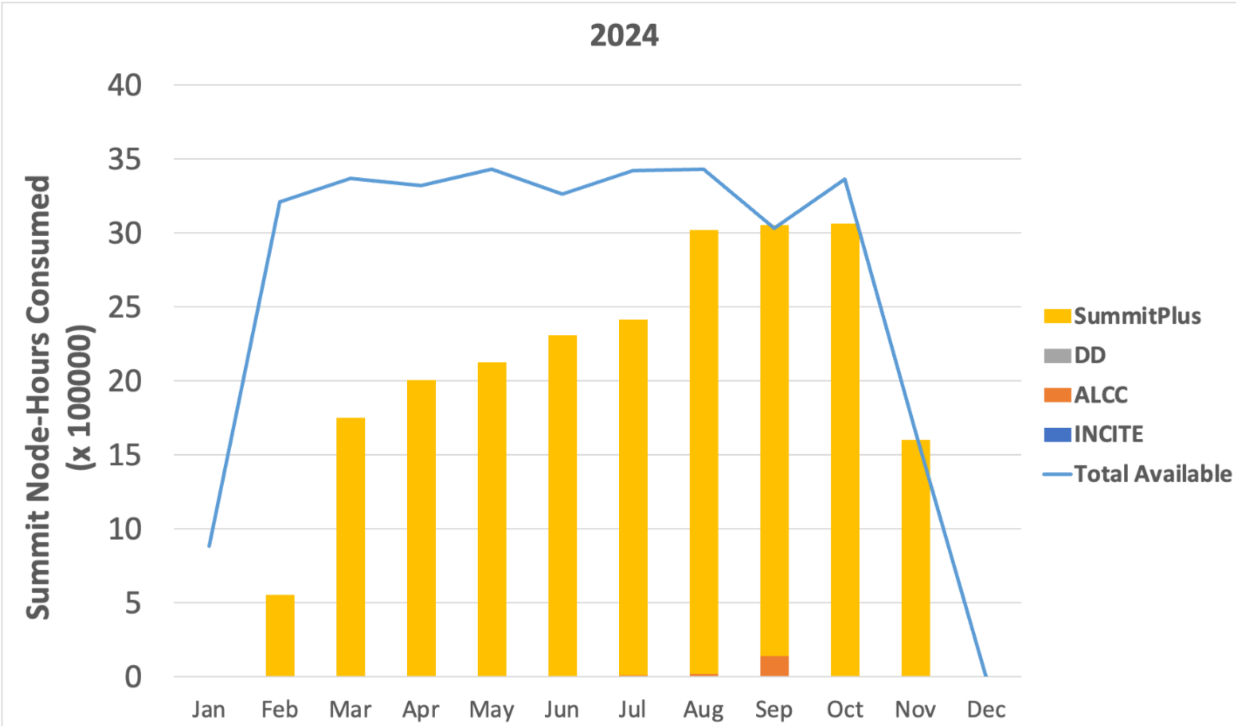


Figure 2-1. IBM AC922 resource utilization: Summit node hours by program for 2024.

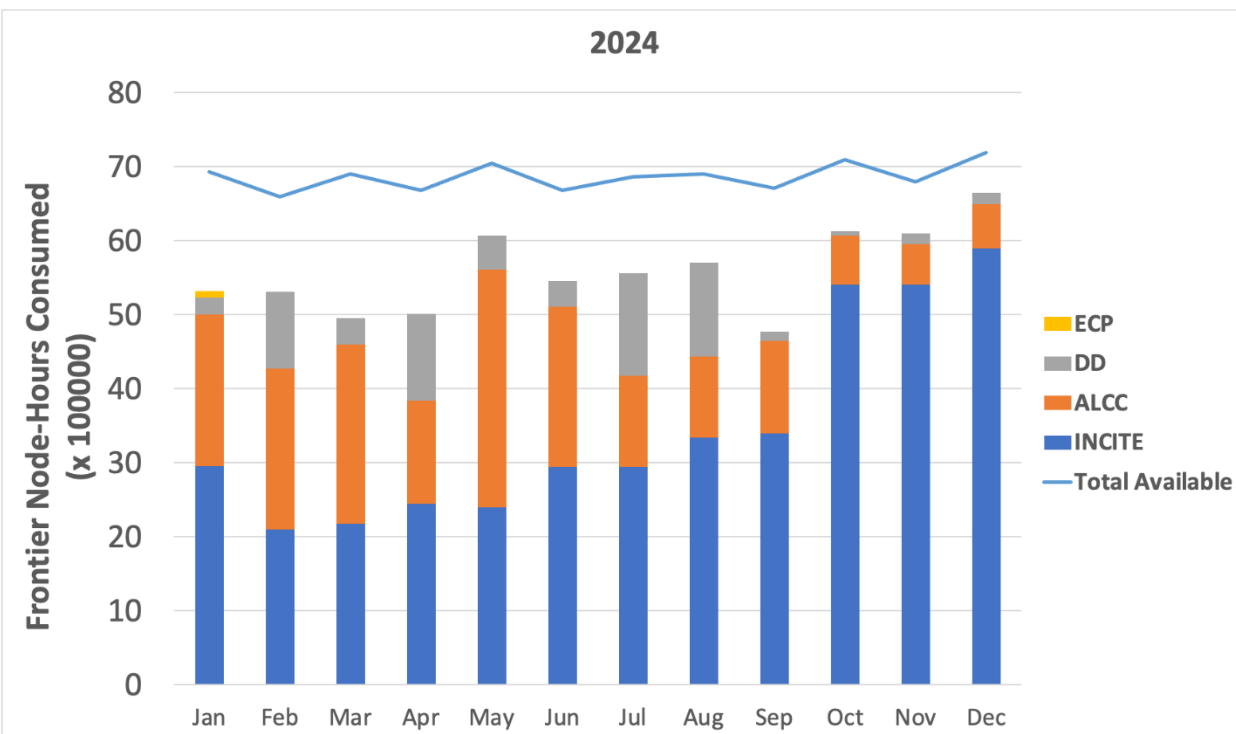


Figure 2-2. HPE Cray EX resource utilization: Frontier node hours by program for 2024.

2.2.2.2 Performance of the Allocation Programs

All allocation programs, including INCITE, ALCC, and DD, are aggressively monitored to ensure that projects within these allocation programs maintain appropriate consumption rates. The 2024 INCITE allocation program was the largest program in 2024, with a commitment of 38.8 million Frontier node hours. For Summit, the largest allocation was to the SummitPLUS allocation program with a commitment of 19.2 million Summit node hours. The consumption of these allocation programs is shown in (Summit) Table 2-10 and Table 2-11(Frontier). As shown, all 2024 commitments were exceeded for each allocation program on Summit and Frontier. This programmatic overachievement is due in part to the high uptime and diligent work of the OLCF staff.

Table 2-10. The 2024 allocated program performance on Summit.

Program	Allocation	Hours consumed	Percent of total
SummitPLUS	19,183,767	21,359,812	99%
INCITE [a]	–	–	–
ALCC [b]	Allocation spans multiple CY	178,290	1%
DD	–	–	–
Total [c]		29,932,793	100%

[a] Includes all 2024 INCITE program.
[b] Includes all ALCC program usage for CY 2024.
[c] Does not include usage outside of the three primary allocation programs.

Table 2-11. The 2024 allocated program performance on Frontier.

Program	Allocation	Hours consumed	Percent of total
INCITE [a]	38,755,000	43,394,129	63%
ALCC [b]	Allocation spans multiple CY	18,785,422	27%
DD	–	6,773,090	10%
Total [c]		68,952,641	100%

[a] Includes all 2024 INCITE program.
[b] Includes all ALCC program usage for CY 2024.
[c] Does not include usage outside of the three primary allocation programs.

Non-renewed INCITE projects from 2024 continued running through January 2025 under the OLCF's 13th month policy. This policy permits an additional, final month for completion and was recognized as a best practice during a previous OAR. It also serves to maintain high utilization while new projects establish a more predictable consumption routine. ALCC projects from the 2024 allocation period (ending June 30, 2024) were also granted extensions as appropriate.

2.3 CAPABILITY UTILIZATION

To be classified as a *capability job*, any single job must use at least 20% of the leadership system's available nodes. For the CY following a new system/upgrade, at least 30% of the consumed node hours will be from jobs that request 20% or more of the available nodes. In subsequent years, at least 35% of the node hours consumed will be from jobs that require 20% or more of the nodes available to users. The metric for capability utilization describes the aggregate number of node hours delivered by capability jobs. The metric for CY 2024 was 35% for Frontier based on years of service as described above. For the sixth, extended service year of Summit no metric was assigned for capability usage. The OLCF continues to exceed expectations for capability usage of its HPC resources.

Keys to successful demonstration of capability usage include the liaison roles provided by Science Engagement members who work hand-in-hand with users to port, tune, and scale code and the OLCF support of the application readiness efforts (i.e., CAAR) to actively engage with code developers to promote application portability, suitability for hybrid systems, and performance. The OLCF also aggressively prioritizes capability jobs in the scheduling system (see Table 2-12 and Table 2-13 and Figure 2-3 and Figure 2-4).

Table 2-12. OLCF capability usage on the IBM AC922 Summit system.

Leadership usage	CY 2023 target	CY 2023 actual	CY 2024 target	CY 2024 actual
INCITE	NAM [a]	53.86%	NAM	–
ALCC	NAM	26.51%	NAM	74.02%
SummitPLUS	NAM	–	NAM	32.37%
All projects	35%	56.51%	NAM	36.94%

[a] NAM means no defined metric or target exists for this system. Data provided as reference only.

Table 2-13. OLCF capability usage on the HPE Cray EX Frontier system.

Leadership usage	CY 2023 target	CY 2023 actual	CY 2024 target	CY 2024 actual
INCITE	NAM [a]	77.87%	NAM	56.18%
ALCC	NAM	90.50%	NAM	53.57%
All projects	30%	79.08%	35%	57.55%

[a] NAM means no defined metric or target exists for this system. Data provided as reference only.

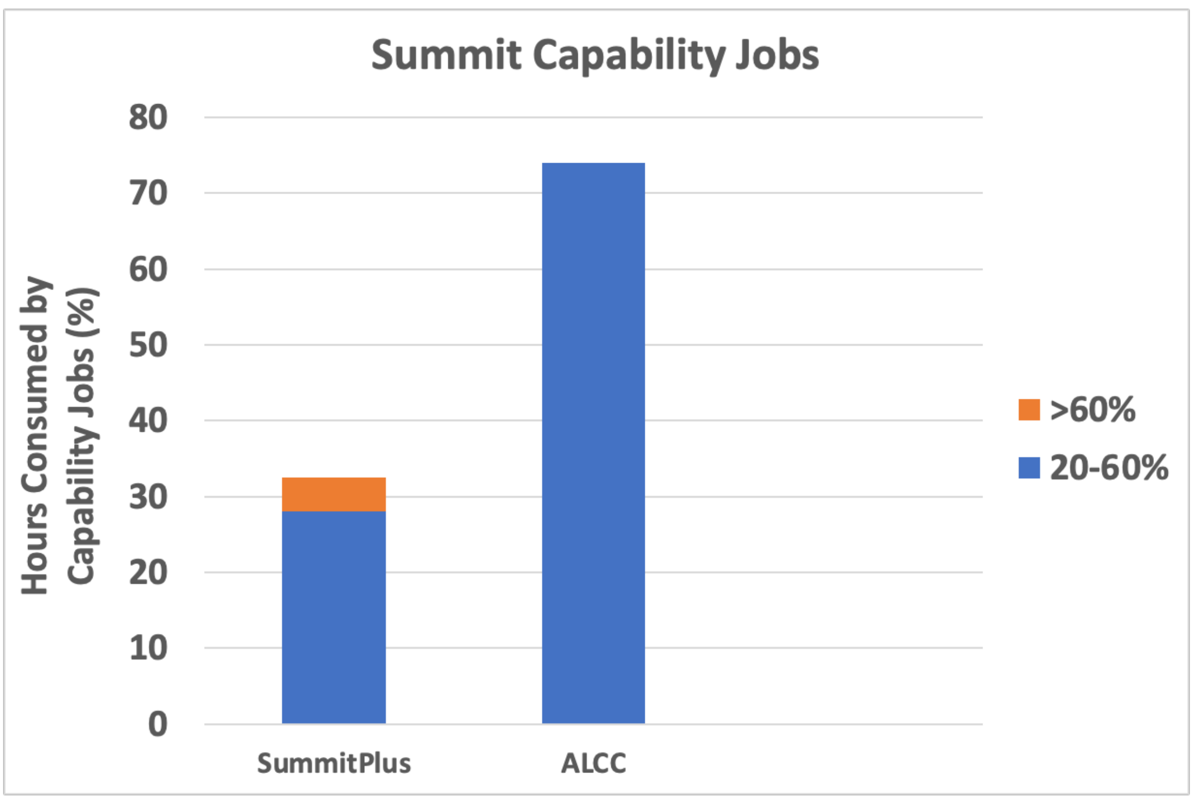


Figure 2-3. Summit capability usage by job size bins and project type.

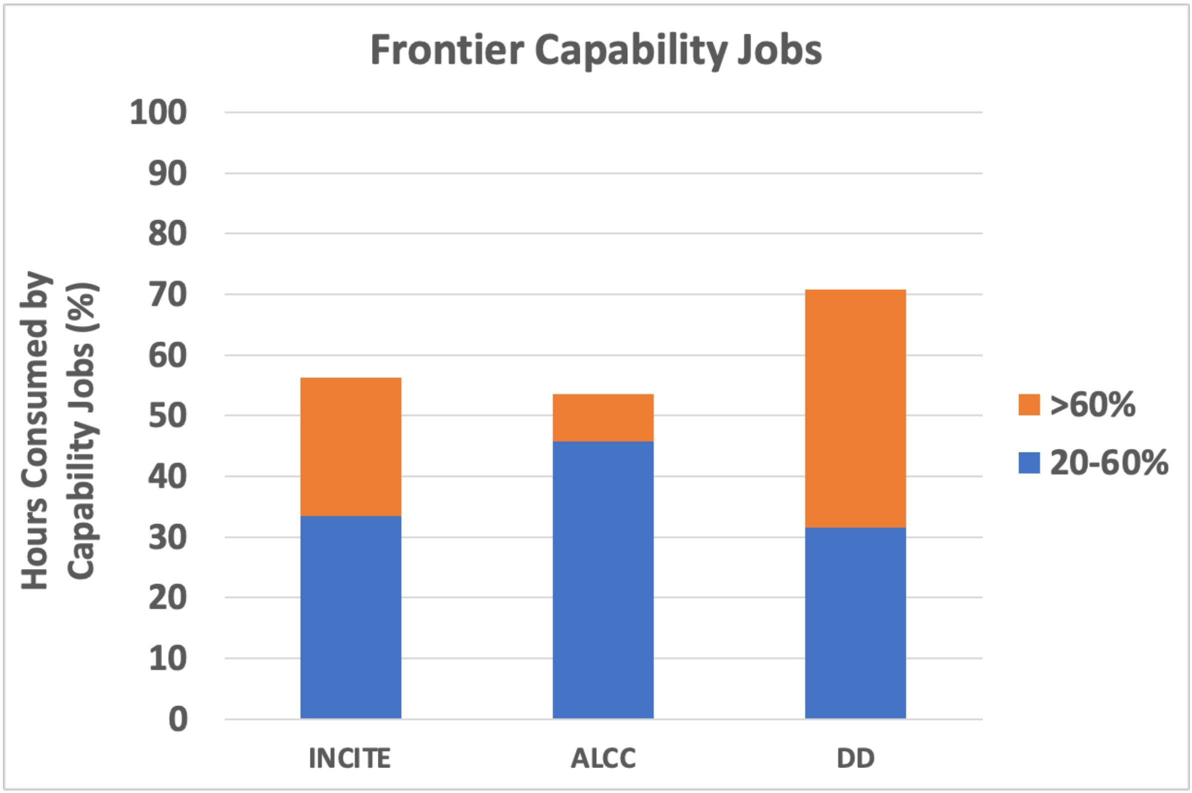


Figure 2-4. Frontier capability usage by job size bins and project type.

HIGH-PERFORMANCE COMPUTING FACILITY
2024 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY

April 2025

3. ALLOCATION OF RESOURCES

CHARGE QUESTION 3: (a) Did the allocation of computing resources conform with ASCR's published allocation policies (i.e., ratio of resources allocated between INCITE, ALCC, and DD)? (b) Was the allocation of DD computing resources reasonable and effective? (c) Did the Facility encounter issues with under- or over-utilization of user allocations? If so, was the Facility's management of these issues effective in maximizing productive use of resources while promoting equity across the user population?

OLCF RESPONSE: Yes. The OLCF continues to enable high-impact science results through access to its leadership-class systems and support resources. The allocation mechanisms are robust and effective. The OLCF enables compute and data projects through the DD user program. This program seeks to enable researchers through goals that are strategically aligned with ORNL and DOE, as described in Section 3.1. The allocation of resources among INCITE, ALCC, and DD programs matched the agreed-upon decomposition: 60% INCITE, 30% ALCC, and 10% DD program. As of December 31, 2024, 105.3% of the total allocated hours for INCITE 2023 had been consumed on Frontier. By extending a customary additional month of access to 2024 INCITE allocations that were not renewed for CY 2025, the OLCF ultimately delivered 112% of the total allocated hours for INCITE 2024 on Frontier by January 31, 2025.

3.1 SUMMARY OF ALLOCATION PROGRAMS

The primary allocation programs that provide access to OLCF resources are the INCITE program, the ALCC program, and the OLCF's DD program. The agreed-upon apportionment of resources for these programs is as follows: 60% of available resources were allocated to INCITE, 30% to ALCC, and 10% to DD. This decomposition marks a return to an allocation model that predates the ECP's use of Frontier for measurement and development. It also reflects the full production status of Frontier in 2024 (Table 3-1).

Table 3-1. Percentages of delivered time per allocation program for Frontier.

Resource	INCITE	ALCC	DD
Frontier	61.8%	28.1%	10.1%

Aside from these well-established allocation programs, 2024 saw an additional and unique OLCF allocation program launched and completed on a one-year timescale: SummitPLUS.

In late 2023, the OLCF determined that it would retain Summit for an additional year beyond the planned-upon retirement of the machine in 2023. As no allocation had been offered on Summit via INCITE or ALCC for 2024 in anticipation of this retirement, use of the machine would require a new, ad-hoc allocation program. The OLCF obtained concurrence from ASCR to launch such an allocation program in September 2023. The program, called SummitPLUS, provided allocations to computationally ready projects and ran from January to November of 2024. Applications for SummitPLUS opened on September 18, 2023, and were closed on October 30, 2023. Subject matter experts in computational and computer sciences reviewed the applications, and awardees were notified in December 2023.

The OLCF received 169 proposals for the SummitPLUS program and awarded time to 108 projects. The total amount of time awarded was 19.04 million Summit node-hours, which represented an over-allocation of approximately 150%. This was done to ensure high utilization of the resource throughout the award year. The average project size was 175,000 Summit node-hours, and project sizes ranged from 20,000 to 300,000 Summit node-hours. The program started in late January 2024 and continued until final Summit decommissioning on November 15, 2024. In total, 21.4 million node-hours were delivered, representing 111% of the original allocation.

3.2 FACILITY DIRECTOR'S DISCRETIONARY RESERVE TIME

The OLCF primarily allocates time on leadership resources through the INCITE program and through the facility's DD program. The OLCF seeks to enable scientific productivity via capability computing through both programs. Accordingly, a set of criteria is considered when making allocations, including the strategic impact of the expected scientific results and the degree to which awardees can effectively use leadership resources.

The goals of the DD program are threefold.

(1) To enable users to prepare for leadership computing competitions such as INCITE and ALCC (e.g., to improve and document application computational readiness)

(2) To broaden the community of researchers capable of using leadership computing by enabling new and nontraditional research topics

(3) To support R&D partnerships, both internal and external to ORNL, to advance the DOE and ORNL strategic agendas

These goals are aligned particularly well with three of the four OLCF missions.

(1) To enable high-impact, grand-challenge science and engineering that could not otherwise be performed without leadership-class computational and data resources

(2) To enable fundamentally new methods of scientific discovery by building stronger collaborations with researchers at experimental facilities as well as DOE offices that have large computing and data science challenges

(3) To educate and train the next-generation workforce in the application of leadership computing to solve the most challenging scientific and engineering problems

R&D partnerships are aligned with DOE and ORNL strategic agendas. These partnerships may be entirely new areas for HPC, or they may be areas that need nurturing. Examples include projects associated with the ORNL Laboratory Directed Research and Development (LDRD) program; programmatic science areas (e.g., fusion, materials, chemistry, climate, nuclear physics, nuclear engineering, bioenergy science and technology); and key academic partnerships (e.g., the University of Tennessee Oak Ridge Innovation Institute).

Also included in this broad category are projects that come to the OLCF through the Accelerating Competitiveness through Computational ExceLlence (ACCEL) Industrial HPC Partnerships outreach, which encourages opportunities for industrial researchers to access the leadership systems through the usual leadership computing user programs to conduct research that would not otherwise be possible.

The actual DD project lifetime is specified upon award: allocations are typically for 1 year or less. However, projects may request 3-month extensions or renewals up to an additional 12 months. The average size of a DD award on Frontier in CY 2024 was roughly 28,800 Frontier node-hours, but awards ranged from 5,000 to 500,000 node hours or more. In 2024, the OLCF DD program participants used approximately 10.1% of the total user resources on Frontier (Table 3-1). See APPENDIX E for a full list of DD projects for CY 2024.

3.3 ALLOCATION PROGRAM CHALLENGES

The INCITE program experienced minimal underutilization on Frontier throughout much of CY 2024. The OLCF modified job priorities for INCITE jobs relative to the usual policies to ensure that all allocated hours were delivered to the program in CY 2024. Specifically, starting in October 2024, the INCITE projects that had exceeded their allocation but had still used less than 125% of their allocation were exempted from usual over-allocation penalties. This strategy enabled INCITE projects to use (and exceed) their remaining time within the calendar year.

HIGH-PERFORMANCE COMPUTING FACILITY
2024 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY

April 2025

4. OPERATIONAL AND TECHNICAL INNOVATIONS

CHARGE QUESTION 4: (a) Have technical innovations been implemented that have improved the facility's operations? (b) Have management/workforce innovations been implemented that have improved the facility's operations? (c) Is the facility effectively utilizing their postdoctoral fellows?

OLCF RESPONSE:

(4a) Yes, the Facility has implemented technical innovations that improved the Facility's operations. Details can be found in Section 4.1.

(4b) Yes, the Facility has implemented management and workforce innovations that have improved the Facility's operations. Details can be found in Section 4.2.

(4c) Yes, the Facility has effectively utilized its postdoctoral fellows. Details can be found in Section 4.3.

4.1 TECHNICAL INNOVATIONS

OLCF has developed and implemented multiple technical innovations in 2024. These innovations and their operational impacts and improvements are described below.

4.1.1 High-Performance Linpack and High-Performance Linpack Mixed Precision Benchmarking at Leadership Scale

Achieving exascale results on both the High-Performance Linpack (HPL) and HPLMxP (Mixed-Precision) benchmarks marks a major milestone in HPC. This achievement highlights Frontier's position at the forefront of leadership computing by demonstrating a dramatic leap in computational power, efficiency, and scalability. The HPL benchmark, which focuses on solving dense linear systems, remains the gold standard for evaluating raw double-precision floating-point performance—the precision most commonly used in scientific applications. Reaching exascale on HPL signals that Frontier can handle massive, complex computations at an unprecedented speed, thereby offering transformative capabilities across a wide range of scientific, engineering, and industrial disciplines. Meanwhile, the HPLMxP benchmark demonstrates the system's ability to perform efficiently at lower precisions, showcasing the growing convergence of HPC and AI workloads. This capability enables exceptional performance while also improving energy efficiency—an essential feature for next-generation computing systems.

A team at the OLCF played a pivotal role in implementing an optimized HPL code, enabling Frontier to achieve 1.353 exaflops on 9,604 nodes for an increase of over 12%. This breakthrough was made possible through a collaboration with AMD, which introduced a new technique that delivered 65.8% efficiency—the highest ever recorded for an exascale system, exceeding the performance of any other exascale supercomputer. This success was driven by a combination of factors, including node scaling, code enhancements, and improvements to the system, such as the removal of lower-performing nodes. A paper detailing this work has been submitted to the ISC 2025 conference. Additionally, a slightly improved version of the HPLMxP code achieved 11.39 exaflops on an expanded number of nodes, marking a nearly 12% increase from the previous result of 10.2 exaflops. This reinforces Frontier's position as a leader in exascale computing, with remarkable advancements in both raw performance and energy efficiency.

4.1.2 Slingshot Compatible OpenMPI Deployment on Frontier

As part of a multiyear effort to provide multiple software options to users and to mitigate the risk associated with relying on a single vendor's software platform, the OLCF has been developing alternative communication library support to deploy on Frontier. This effort has centered around extending widely used open-source software projects with support for the network architecture in Frontier to fully leverage the available hardware communication channels provided by the architecture. Of particular focus was the implementation of intranode communication mechanisms that allowed application communication to bypass the network and move data using the high-speed AMD proprietary interconnect (XGMI) that connects all the GPUs and CPUs present in a node. As part of this, the OLCF has designed, implemented, and began to upstream features to both the libfabric and OpenMPI communication libraries to allow them to interface with the Slingshot network devices installed across the nodes of Frontier. In 2024, this effort culminated in the upstreaming of the initial functional implementations of these features to the libfabric project. This in turn allowed the deployment of a fully supported version of OpenMPI on Frontier, which is broadly available to the OLCF user community.

4.1.3 Data Management and Tiering Improvements

4.1.3.1 PoliMOR

Building upon the team's recent development efforts, in 2024, a team consisting of Rick Mohr, Christopher Brumgard, Anjus George, and Bradly Gipson deployed PoliMOR to Orion, the Lustre-based storage system serving Frontier, in limited-production capacity. PoliMOR is a state-lite, agent-based, service system that functions as a distributed, extensible, and automated data management policy engine. It offers services such as data migration, purging, data collection, and telemetry for file systems such as Lustre through a straightforward policy framework. For Lustre systems such as Orion, this capability is crucial for adhering to total capacity and hierarchical storage tier limitations. PoliMOR is an ongoing project and is being developed to address a wider range of use cases and meet the operational requirements. PoliMOR is open source and can be accessed at https://github.com/olcf/polimor.

4.1.3.2 HIS_XFER: Simplified Data Transfer Tool

A new data transfer tool created at the OLCF could be available to facilities nationwide after making its debut at the OLCF. The tool, called *hsi_xfer,* was created by HPC storage systems engineer Jake Wynne to help transfer large quantities of data from the facility's HPSS to the newly deployed Kronos nearline storage system. Kronos is a 134 PB, multiprogrammatic nearline storage system that also provides tape-based backups for all data as a disaster-recovery measure.

After decades of service, HPSS is set to be decommissioned in early 2025, and users are working to migrate their data before the January 31, 2025, deadline. As the HPC Storage Systems team prepared for the mass exodus earlier this year, the need for an efficient data transfer tool from HPSS to Kronos was obvious. Wynne developed an idea for the tool with his colleague, Gregg Gawinski, another HPC storage systems engineer at the OLCF. First, they disabled Globus, a command-line tool and web interface commonly used for seamless data transfer. This enabled Wynne to focus solely on *hsi,* the remaining tool available for facilitating data transfer from HPSS.

The *hsi_xfer* transfer tool is named after the *hsi* command-line interface for the HPSS. The *hsi_xfer* tool builds on the capabilities of *hsi* by offering a more efficient and streamlined way to transfer data. To distinguish this enhanced version from the original *hsi*, the team coined the name *hsi_xfer.* Wynne wrote the script to optimize the management of HPSS and data transfer node resources during this period of large-scale data retrieval. The goal was to prevent overloading the tape library's robotic tape-retrieval mechanisms and provide users with a simpler, more efficient way to access their data. The tool batches all requested files from a single tape and streams them together, thereby minimizing the robotic movement, reducing tape loading and seek times, and extending the lifespan of both the tapes and the hardware. The *hsi_xfer* tool is essentially a wrapper around *hsi*—it uses the existing *hsi* tool but adds features such as concurrent transfer threads, checksumming, and checkpointing while offering a more user-friendly interface. The enhanced functionality, proven efficiency, and ease of use have attracted interest from other computing facilities. The script has demonstrated superior transfer performance compared to other tools while also offering data-integrity features typically found in tools such as Globus—features that are not available in the standard *hsi*. Like Globus, the script includes a checkpointing feature that allows users to quickly recover from interrupted transfers, resuming exactly where they left off with minimal overhead. The script is being open-sourced.

4.1.4 OLCF-6 AI and Workflow Benchmarks

4.1.4.1 FORGE: AI Benchmark

As AI workloads become increasingly important on leadership-computing platforms, the OLCF incorporated an AI-specific benchmark, FORGE, into the OLCF-6 procurement process as part of the request for proposals. The FORGE benchmark, designed to stress-test system capabilities, emphasizes low-precision compute, communication bandwidth, and energy efficiency in training LLMs on scientific datasets. FORGE is based on mainstream frameworks that support distributed training across data, tensor, and pipeline parallelism. The benchmark assesses system performance by using computational metrics such as teraflops, scaling efficiency, and energy efficiency (teraflops/Watt). FORGE also provides loss tracking and time-to-solution metrics. The primary figure of merit combines time-to-solution and energy efficiency, and metrics such as scaling efficiency support a comprehensive system evaluation.

FORGE stresses compute systems with demanding low-precision calculations in float16 or bfloat16, requiring efficient communication patterns (e.g., AllReduce, AllGather, ReduceScatter) to manage message sizes 3× the model's parameters. OLCF-6 benchmark run rules include using specific tokenizers, optimizers, architectures, and training over 257 billion tokens. System performance is maximized through customized kernel optimizations and adjustable learning rates and batch sizes tailored to training scales. The benchmark ensures comprehensive evaluations of computational efficiency, scalability, and energy use in LLM training.

Related document/dataset: OLCF-6 FORGE RFP Benchmark, Junqi Yin, Sajal Dash, Feiyi Wang, and Mallikarjun (Arjun) Shankar, https://www.olcf.ornl.gov/wp-content/uploads/OLCF-6_FORGE_description-1.pdf; Tokenized Data for FORGE Foundation Models, https://doi.org/10.13139/OLCF/2004951.

4.1.4.2 OLCF-6 Workflow Benchmark

The OLCF-6 Workflow Benchmark evaluates HPC system capabilities for handling dynamic data stream workloads through ML4NSE (Machine Learning for Neutron Scattering Experiment) applications. Using a Temporal Fusion Transformer (TFT) model, ML4NSE analyzes neutron scattering data in real-time, processing a nuclear peak and six satellite peaks from magnetic ordering in single-crystal samples. The system architecture channels multiple data streams through a gateway node using open-source streaming transport (ZeroMQ, RabbitMQ), with a control stream managing data multiplexing and filtering. The Data Orchestrator connects to both fast internal networks (RDMA) and external networks, managing data flow between compute nodes and consumers.

The benchmark includes weak scaling experiments (training TFT models on varying voxel configurations, requiring 9–4,608 ranks) and strong scaling experiments (fixed 32 × 32 × 32 input with variable level-1 mapping). Performance is measured primarily through time-to-solution, with secondary metrics including gateway node exchange bandwidth, stream multiplexing scalability, and control operation latency. A reference benchmark run demonstrates scaling capabilities up to 4,608 GPUs on Frontier, with comprehensive performance tracking for both weak and strong scaling efficiency.

Related document/data: OLCF-6 Workflow RFP Benchmark, Rafael Ferreira da Silva, Patrick Widener, Junqi Yin, and Ketan Maheshwari, https://www.olcf.ornl.gov/wp-content/uploads/OLCF-6_Workflow_description-7.5.24.pdf.

4.1.4.3 NCCS Software Provisioning

The NCCS Software Provisioning (NSP) system is a framework for deploying software stacks on HPC systems. NSP leverages Ansible to automate the deployment of Spack environments using templates and to manage installation procedures for non-Spack software through custom roles. Additionally, NSP enhances vendor-managed Lmod installations using hooks, enabling dynamic and responsive software layouts that adapt seamlessly to changes in the programming environment. The NSP framework is currently used to deploy software for the OLCF's Frontier supercomputer and the Andes analysis cluster. NSP's Lmod hooks have also allowed the OLCF to easily tag module files for various purposes, including to indicate which modules are provided through Extreme-Scale Scientific Software Stack (E4S)-supported Spack packages, part of the User-Managed Software program, or provided by the facility. NSP provides a clear software structure that allows users to distinguish the type of support provided by each individual package (e.g., compiled with GPU support). Additionally, it is used for systems in both the Air Force Weather (AFW) and National Climate-computing Research Center (NCRC) program.

4.2 OPERATIONAL INNOVATION – MANAGEMENT/WORKFORCE

The OLCF understands the significance of maintaining a healthy management and workforce structure for ensuring successful and continuous operation. To this end, the OLCF is pursuing a multipronged pipeline approach. The first stage of the OLCF's comprehensive pipeline starts with student programs followed by on-the-job training programs. To complement these efforts, the OLCF is also collaborating with ORNL's HR and communications organizations to develop a robust recruitment campaign.

4.2.1 ORNL's Next Generation Pathways to Computing Summer School

ORNL held the second iteration of the NGP Computing Summer School, one of five Pathways Summer Schools funded by DOE's WDTS Reaching a New Energy Sciences Workforce initiative. In 2024, the summer school hosted 24 high school students from East Tennessee for 5 weeks. During the program, students were grouped into pairs and matched with an ORNL mentor. OLCF staff led the development of the curriculum for NGP, including introductory programming lessons, and provided an overview of HPC and AI topics. NGP participants were also invited to attend the HPC Crash Course offered by the OLCF and had in-person office hours to work through issues and explore topics more in-depth. In addition to learning about HPC and AI topics, the students utilized NVIDIA JetBot to gain hands-on experience and apply the concepts learned throughout the program.

Thanks to the success of the program in 2023 and 2024, NGP was renewed by WDTS for summer 2025, and funding was increased to expand the program to 30 students.

4.2.2 On-the-Job Training Opportunities

The other stage of the OLCF's pipeline focuses on on-the-job training. This stage has four distinct elements: job rotation, coaching, job instruction, and training through step-by-step assignments. The OLCF has instituted a staff development program for all new operators and currently has three early career systems admins. As staff members in the Operations Control Room (OCR) age out, the goal is to promote operators to an early career admin position and replace them with a technician with an AS or a new college graduate with a BS in an IT-related field. The goal for the new operators is to learn from the ground up: start with NCCS facility operations, develop systems administration skills, and obtain further education in IT as needed. Although it is difficult to estimate their time in the OCR prior to promotion to early career administrator, the OLCF has a 2–3-year goal for these candidates to advance. As the early career staff members are hired, the OLCF reviews their skill sets and offers placement opportunities in groups across the division. The goal is for each staff member to develop rapidly and to use their abilities with their embedded groups to transfer tasks that can be delegated to the operators. The desire is for individuals to stay in the role for 1–2 years. Through this progression, the OLCF can have a healthy mix of early career and seasoned OCR operators to increase the OCR abilities/skill sets and to provide homegrown talent for the other groups. The OLCF is currently seeing success with their efforts and current staff. Since establishing the program, the group has progressed 1 operator to a junior systems admin position, and 2 junior systems admins from OCR to different groups across the OLCF.

4.2.3 NCCS Virtual Career Fair 2024

Originally launched in 2021, the OLCF hosted the third iteration of the NCCS Virtual Career Fair on November 8, 2024. The event focused on recruiting for open positions in NCCS across all groups and was held via the GatherTown platform. The virtual nature of the event makes it easily accessible and enables the OLCF and NCCS to reach a broader audience. The event had 171 registrants from a wide range of universities and institutions. The candidates were interested in both internship and staff opportunities. More information about the NCCS Virtual Career Fair can be found at https://www.olcf.ornl.gov/calendar/virtual-career-fair-2024/.

4.3 POSTDOCTORAL FELLOWS

4.3.1 Program

The DOE recognizes the need to train and retain computational scientists in a broad range of disciplines that support DOE and the nation's critical missions to maintain the US competitive advantage in high-performance and data-intensive scientific computing. When considering the ever-increasing capability of high-end computer architectures, there is a continuing and increasing need to cultivate and maintain a well-trained computational science workforce in academia and industry and at the national laboratories. To address this need, DOE proposed that ASCR establish a postdoctoral training program at its user facilities, including the OLCF, the ALCF, and NERSC. This program, known as Computational Scientists for Energy, the Environment, and National Security (CSEEN) has the following objectives: (1) ensure an adequate supply of scientists and engineers who are appropriately trained to meet national workforce needs, including those of DOE, for high-end computational science and engineering with skills relevant to both exascale and data-intensive computing; (2) make ASCR facilities available through limited-term appointments for applied work on authentic problems with highly productive work teams and increasingly cross-disciplinary training; and (3) raise the visibility of careers in computational science and engineering to build the next generation of leaders in computational science. In CY 2019, the OLCF began to leverage additional funding from the ECP to augment the CSEEN program with additional postdoctoral fellows.

The OLCF CSEEN postdoctoral program seeks to provide opportunities to bridge the experience gap between the need to address domain science challenges and the need to cultivate high-performance software development expertise. One of the focus areas is to provide the skills required to port, develop, and use software suites on the leadership computing resources at the OLCF. The software development activities can occur in conjunction with a CAAR project (which will be OLCF-6 funded starting in 2026). This model offers the greatest potential for scientific breakthroughs through computing and provides ample opportunity to publish in domain science literature. This approach will ensure that the postdoctoral trainees continue to build their reputations in their chosen science communities. Participants in the CSEEN postdoctoral program are encouraged to attend tutorials, training workshops, and training courses on select computer science topics. One of the most important outcomes for the postdoctoral trainee is the opportunity to publish and present research accomplishments.

In CY 2024, a total of 15 postdocs were members of the OLCF workforce. Of those, 10.5 were fully supported by OLCF funds, and 4.5 of the postdocs were at least partially supported by sources outside the OLCF, including EERC, GreenSight, LUCID, Distinguished Staff Fellowship, SciDAC, and LDRD projects. The background and current work of these postdocs in the Science Engagement section is described below.

4.3.1.1 Samuel Fagbemi

Samuel Fagbemi joined ORNL in February 2024 and is mentored by Ramanan Sankaran from the Advanced Computing for Physical Sciences Section in the Computational Sciences and Engineering Division (CSED) and Philip Roth of the NCCS Science Engagement section. He studies non-equilibrium chemical processes for achieving electrified heterogeneous catalysis using programmable Joule heating (JH). Specifically, he models JH within non-woven carbon-graphite fiber geometries at the fiber-scale and couples the radiative effects with electrochemical phenomena for propane dehydrogenation. Samuel is an active contributor to Quilt—a software platform for level-set modeling of multiphysics processes within the ORNL-led Non-Equilibrium Energy Transfer for Efficient Reactions center. He carried out large-scale simulations and strong scaling studies of JH processes on the OLCF Summit system and has a publication describing the work under development. Samuel is the PI of an OLCF DD allocation project that supports energy efficiency and performance optimization of coupled JH simulations on the OLCF's Frontier system. Samuel served as mentor in a 2024 HPC-focused Faculty Hackathon organized by the National Science Foundation–funded Science Gateways Center of Excellence, in which he mentored faculty members on HPC best practices and ways to implement HPC in their curricula by using HPC programs such as the Frontier Pathways to Supercomputing Initiative. He participates in the Science Engagement section's Uncertainty Quantification working group.

4.3.1.2 Chao Lu

Chao Lu is a Postdoctoral Research Associate in the Advanced Computing for Life Sciences & Engineering group at the OLCF. Prior to joining ORNL, Chao earned his PhD in electrical engineering from the University of Texas at Dallas in 2024. At the OLCF, his research focuses on developing and testing QC algorithms for solving partial differential equations. These algorithms are developed to tackle canonical fluid dynamics problems, which involve accelerating the existing quantum algorithms to solve complex fluid dynamics problems. He collaborates with external ORNL researchers and internal ORNL teams focused on integrating HPC and QC.

4.3.1.3 Nasik Muhammad Nafi

Nasik Muhammad Nafi is a postdoctoral research associate in the Advanced Computing for Life Sciences & Engineering group at the OLCF. Prior to joining ORNL, Nafi earned his PhD in computer science from Kansas State University in 2024. At the OLCF, his research focuses on large-scale distributed training of cutting-edge AI models for image processing and segmentation tasks. These models are applied to areas such as medical imaging, manufacturing microstructure analysis, oceanographic turbulence modeling, and climate modeling. He collaborates with external ORNL researchers and internal ORNL teams funded by LDRDs and is part of a team competing for the 2025 Gordon Bell Prize.

4.3.1.4 Bai-Shan Hu

Bai-Shan Hu joined the OLCF in 2022. His activity is currently funded by the Physics Division through a SciDAC award, and he is mentored by Gustav Jansen. Baishan had three publications in CY 2023, two of which are in the prestigious *Physical Review Letters* journal.

Publications

Neupane, S., N. Kitamura, Z. Y. Xu, R. Grzywacz, S. J. Novario, J. Okołowicz, M. Płoszajczak, B.-S. Hu, J. M. Allmond, A. Chester, J. M. Christie, I. Cox, J. Farr, I. Fletcher, J. Heideman, D. Hoskins, T. T. King, A. Laminack, S. N. Liddick, M. Madurga, A. L. Richard, P. Shuai, K. Siegl, P. Wagenknecht, and R. Yokoyama, ""First β-delayed neutron spectroscopy of 24O," *Phys. Rev. C*, Volume: 110 Issue: 3, pg 034323 (2024): https://doi.org/10.1103/PhysRevC.110.034323.

Yuan, Q. and B.-S. Hu, "Ab initio calculations of anomalous seniority breaking in the πg9/2 shell for the N = 50 isotones," *Physics Letters B*, Volume 858, 2024, 139018, ISSN 0370-2693, https://doi.org/10.1016/j.physletb.2024.139018.

Hu, B.-S., Z. H. Sun, G. Hagen, G. R. Jansen, and T. Papenbrock, "Ab initio computations from 78Ni towards 70Ca along neutron number N = 50," *Physics Letters B*, Volume 858, 2024, 139010, ISSN 0370-2693, https://doi.org/10.1016/j.physletb.2024.139010.

Hu, B.-S., Z. H. Sun, G. Hagen, and T. Papenbrock, "Ab initio computations of strongly deformed nuclei near 80Zr," *Phys. Rev. C* 110 (1) L011302 (2024): https://doi.org/10.1103/PhysRevC.110.L011302.

Hu, B.-S., "How do mirror charge radii constrain density dependence of the symmetry energy?" *Physics Letters B*, Volume 857, 138969, ISSN 0370-2693 (2024): https://doi.org/10.1016/j.physletb.2024.138969.

Acharya, B., B.-S. Hu, S. Bacca, G. Hagen, P. Navrátil, and T. Papenbrock, "Magnetic Dipole Transition in 48Ca," *Phys. Rev. Lett.* 132 (23) 232504 (2024): https://doi.org/10.1103/PhysRevLett.132.232504.

4.3.1.5 Tor Djaery

Tor Djaerv joined the group as a postdoctoral associate in 2022. His research focuses on nuclear structure calculations using the Nuclear Tensor Contraction Library (NTCL), and he is mentored by Gustav Jansen. Tor has been making significant contributions to NCTL on the Frontier supercomputer to improve the capabilities of an existing deformed coupled clusters code. In CY 2024, Tor had two publications and anticipates publishing a technical paper on NTCL soon.

Publications

Jiang, W. G., C. Forssén, T. Djärs, and G. Hagen, "Emulating ab initio computations of infinite nucleonic matter," *Phys. Rev. C* 109 (6) 064314 (2024): https://doi.org/10.1103/PhysRevC.109.064314.

Jiang, W. G., C. Forssén, T. Djärs, and G. Hagen, "Nuclear-matter saturation and symmetry energy within Δ-full chiral effective field theory," *Phys. Rev. C* 109 (6) L061302 (2024): https://doi.org/10.1103/PhysRevC.109.L061302.

4.3.1.6 Steve Fromm

Steve Fromm joined ORNL as a postdoctoral research associate in late CY 2024. Steve's research focuses on developing and improving methods used in large-scale simulations of core-collapse supernovae. Steve is a US Army veteran from Michigan. He earned his PhD in physics and computational mathematics, science, and engineering from Michigan State University, where his research focused on improved methods for general relativistic radiation hydrodynamics in simulations of neutron star mergers and core-collapse supernovae.

4.3.1.7 Matthias Heinz

Matthias Heinz joined the group in the fall of CY 2024 via the ORNL's Distinguished Staff Fellowships Program. The program offers outstanding early career scientists and engineers a unique and prestigious entry to a career at ORNL. Matthias's research will focus on first-principles computations of heavy deformed atomic nuclei, aiming to predict their properties and gain deeper insights into fundamental physics. He earned his BS in physics and computer science from the Ohio State University and his MS and PhD in theoretical nuclear physics from the Technische Universität Darmstadt. In 2024, Matthias had a publication that he finalized at ORNL.

Publication

Tichai, A., P. Arthuis, K. Hebeler, M. Heinz, J. Hoppe, T. Miyagi, A. Schwenk, and L. Zurek, "Randomized low-rank decompositions of nuclear three-body interactions," *Phys. Rev. Res.* 6 (4) 043331 (2024): https://doi.org/10.1103/PhysRevResearch.6.043331.

4.3.1.8 Margot Fitz-Axen

Margot Fitz-Axen started with the group in the fall of 2024. Margot earned her PhD in astronomy from the University of Texas at Austin, where she conducted groundbreaking research on numerical simulations of molecular cloud collapse and star formation with cosmic ray transport. Her current work centers on the development and analysis of simulations for core-collapse supernovae.

4.3.1.9 Spancer Dong

Spancer Dong is a postdoctoral research associate in the Advanced Computing for Chemistry and Materials group in the OLCF. Prior to joining ORNL, Spancer earned her PhD in theoretical and computational chemistry from the University of Louisville, Kentucky, in 2024, under supervision of Prof. Lee Thompson. At ORNL, her research concerns the application of the non-orthogonal configuration interaction for fragments method to molecular assemblies involved in electron and exciton transfer processes, such as in photovoltaics and photo-electronics devices. She is a member of the international GronOR development and application team with researchers at the University of Barcelona, University of Rovira i Virgili, and the University of Groningen.

Sousa, C., X. López, X. Dong, R. Broer, T. Straatsma, and C. de Graaf, "Non-orthogonal Configuration Interaction for Singlet Fission: Beyond the Dimer," *Journal of Physical Chemistry C*, 129, no. 8 (2025): https://doi.org/10.1021/acs.jpcc.4c08656.

4.3.1.10 Bharath Raghavan

Bharath Raghavan is a postdoctoral research associate in the Advanced Computing for Chemistry and Materials group at ORNL's NCCS. He is mentored by David Rogers, Van Ngo, and Dilip Ashtagiri. His research interests at ORNL lie at the intersection of computational drug design, AI, quantum biophysics and HPC. His doctoral research with Prof. Paolo Carloni in the Computational Biomedicine institute at Forschungszentrum Jülich involved the problem of noninvasive detection of glioma through HPC and quantum mechanics (QM)/molecular mechanics (MM) simulations of the IDH1 protein. This included integrating the MiMiC QM/MM software package into an HPC-based virtual screening protocol, and the development of the MiMiCPy toolkit to streamline MiMiC simulation setup.

4.3.1.11 Noor Md Shahriar Khan

Noor Md Shahriar Khan joined the group in January 2025 as a postdoctoral research associate and is mentored by Dmytro Bykov. His research focuses on lanthanide-aqua complexes, for which he applies high-level electronic structure theory to investigate their unique electronic and magnetic properties. Specifically, he conducts multireference calculations on lanthanide-aqua systems, focusing on the impact of spin-orbit coupling and relativistic effects with an eye toward their potential applications as single-molecule magnets for quantum materials. Prior to his work at ORNL, Shahriar gained significant experience at the Georgia Institute of Technology, where he performed quantum chemistry and molecular dynamics simulations to study electrochemical reactions, particularly anodic coupling reaction mechanisms, using density functional theory–based molecular dynamics. His research during that time contributed to the publication of the paper, "Catalytic Role of Methanol in Anodic Coupling Reactions Involving Alcohol Trapping of Cation Radicals" (https://doi.org/10.1021/acs.joc.4c02227). At ORNL, he utilizes advanced computational resources at the OLCF to perform large-scale parallel calculations to further the understanding of lanthanide-aqua complexes in quantum chemistry and material design.

4.3.1.12 Tanvir Sohail

Tanvir Sohail joined the group in September 2023 and specializes in multiscale materials modeling using DOE leadership-class supercomputers. His research spans large-scale simulations of alloys, quantum materials, and polymer composites. In CY 2024, he published one paper and presented at two conferences. Under the mentorship of Swarnava Ghosh, he has explored the stability of magnetic defects in B20 materials, optimized distributed ML training, and advanced GPU code porting. He is also developing efficient workflow tools on Frontier to enhance ensemble simulations for materials research.

Publication

Georgiadou, A., H. Monge-Camacho, T. Sohail, S. Ghosh, A. V. Parambathu, D. N. Asthagiri, D. Bykov, T. Athawale, and T. L. Beck. "Ensemble Simulations on Leadership Computing Systems," *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, IEEE (2024): 394–401.

4.3.1.13 Khadiza Begam

Khadiza Begam joined the OLCF in 2024. As a postdoctoral research associate in ORNL's Advanced Computing for Chemistry and Materials Group at the NCCS, she is working toward virtual instrumentation of dynamic nuclear polarization to enable neutron crystallography. The research will advance the computational method for density mapping of hydrogens to locate their positions and is crucial in various applications (e.g., drug design, hydrogen fuel generation) to understand the interactions and functionality of biomolecular systems. Before joining ORNL, Khadiza completed her PhD (physics) from Kent State University. The graduate research focused on electronic structure calculations, condensed phase spectral calculations, and quantum mechanical modeling to study the intra- and intermolecular charge and energy transfer, transport mechanisms, solvated NMR method implementation, and applications.

Publication

B. Khadiza Begam et al., "Antioxidative Triplet Excitation Energy Transfer in Bacterial Reaction Center Using a Screened Range Separated Hybrid Functional," *The J. of Phys. Chem. B*, 128, no. 18 (2024): 4315–4324.

4.3.1.14 Maria Batool

Maria Batool**,** who is jointly appointed by NCCS and CSED, is actively engaged in projects related to drug discovery and collaborative research efforts. Mentored by Jens Glaser from Science Engagement in NCCS and Debsindhu Bhowmik from Advanced Computing Methods for Health Sciences in CSED, Maria's primary focus is the development of HPC-enabled AI methods for identifying small molecule inhibitors targeting disease-related enzymes. Specifically, she contributes to ongoing collaborations aimed at discovering inhibitors for the serine hydroxymethyltransferase, a critical enzyme in various aggressive cancers. Maria's expertise in computational chemistry and drug discovery enhances her contributions to user projects at NCCS, in which she utilizes AI tools to predict novel molecules for testing. Additionally, Maria's involvement in projects that explore the effects of low-dose radiation exposure underscores her commitment to advancing understanding and mitigation strategies through molecule generation methods and computational simulations. This research holds promise for the development of novel radioprotective agents with applications across industries and medical contexts.

4.3.1.15 Hector Hernandez Corzo

Hector Hernandez Corzo has been an integral part of the team since 2022. His research focuses on leveraging quantum mechanical methods to analyze the electronic structure of atoms and molecules. In CY 2024, Hector extended his scientific interests into the realm of AI, where he undertook projects aimed at advancing software development through AI applications. Hector's postdoctoral appointment ended in January 2025.

5. RISK MANAGEMENT

CHARGE QUESTION 5: Does the Facility demonstrate effective risk management practices?

OLCF RESPONSE: Yes. The OLCF has a history of successfully anticipating, analyzing, rating, and retiring both project- and operations-based risks. The OLCF risk management approach is modeled after the Project Management Institute's best practices. Risks are tracked and retired, reclassified, or mitigated as appropriate. A change history is maintained for historical reference.

The major operational risks for the OLCF in CY 2024 are listed and described in this section. Planned mitigations and implementations are included in the subsequent descriptions. As of this writing, the OLCF has one high-priority operational risk, which involves the maturity and performance of OpenMP (originally a risk from the OLCF-5 project but moved into Operations upon project completion). This number is subject to change based on continuous reevaluation of existing risks (and potential new risks).

5.1 RISK MANAGEMENT SUMMARY

The OLCF's Risk Management Plan describes a regular, rigorous, proactive, and highly successful review process that is reviewed at least annually and updated as necessary. Projects, such as OLCF-6, have their own project-specific risk management plans. Although plans are different, the approach in each is similar. Risks are tracked in a risk registry database application that can track project and operational risks separately.

Operations risks are continually assessed and monitored by the risk owners, the facility management team, OLCF group leaders and section heads, and other stakeholders. When assessing risks, the OLCF management team focuses its attention on the high and moderate risks as well as any low risks within the impact horizons associated with the risk. Trigger conditions and impact dates are recorded in the risk notes narrative section of the register. Risk owners are proactive in tracking trigger conditions and impact horizons for their risks and bringing appropriate management attention to those risks, regardless of the risk rating level.

The OLCF reports a change summary of affected operations risks to the DOE program office as part of its monthly operations report. At the time of this writing, 34 active entries are in the OLCF operations risk register. These range from facility-wide risks such as safety, funding, expenditures, and staffing to more focused risks dealing with specific components such as computing platforms, storage, and networks.

The costs of handling risks are integrated in the budgeting exercises for the entire facility. For operations, the costs of risk mitigation are accepted, and residual risk values are estimated by expert opinion and are accommodated, as much as possible, in management reserves. This reserve is continually reevaluated throughout the year.

5.2 MAJOR RISKS TRACKED IN 2024

Table 5-1 contains the major risks tracked for OLCF operations in 2024. The selected risks are all rated medium or high in impact.

Table 5-1. 2024 OLCF major risks.

Risk ID/description	Probability/impact	Action	Status
723: Safety – personal injury	Low/Medium	Mitigating	Reduce risk by monitoring worker compliance with existing safety requirements, having daily toolbox safety meetings, conducting periodic surveillance using independent safety professionals, having joint walk-downs by management and work supervisors, and by emphasizing the stop-work authority of all personnel. Observations from safety walk-downs are evaluated for unsatisfactory trends (e.g., recurring unsafe acts). Unsatisfactory performance will receive prompt management intervention commensurate with the severity of the safety deficiencies.
406: System cybersecurity failures	Low/High	Mitigating	The OLCF has developed a cybersecurity plan that implements protection to the MODERATE level. This includes such things as two-factor authentication and periodic formal tests and reviews.
1240: Failure to handle export-controlled information properly	Low/High	Mitigating	Staff with elevated privileges go through annual training on handling export-controlled information. Projects are reviewed by the ORNL Export Control Office, and project PIs are briefed based on that review. The OLCF has automated permission enforcement on resources/areas where export-controlled information can be stored.
1245: System unavailability due to mechanical/electrical system failure	Low/High	Mitigating	The system was designed with leak detection in mind. Mitigation involves performing all preventative maintenance, performing inspections, and monitoring where possible.

5.3 NEW OR RECHARACTERIZED RISKS SINCE LAST REVIEW

5.3.1 Recharacterized Risks

No risks in the OLCF Operations Risk Register were recharacterized during the review year.

5.3.2 New Risks in This Reporting Period

No risks in the OLCF Operations Risk Register were recharacterized during the review year.

5.4 RISKS RETIRED DURING THE CURRENT YEAR

No risks were retired from the OLCF Operations Risk Register during the review year.

5.5 MAJOR RISKS FOR NEXT YEAR

With the decommissioning of both Summit (at the end of the 2024) and HPSS (in early 2025), risks specific to those systems will be retired. For 2025, Frontier will be the primary OLCF compute system, Orion the primary storage system, and Kronos the primary archival system.

Day to day operations, routine maintenance, and various deliveries can involve working with large equipment or moving large equipment through hallways and affect areas used for ingress, egress, and normal traffic for staff and visitors. Additionally, CY 2025 begins with a major activity of the de-installation of Summit. For this reason, the OLCF's safety risk (723) remains a major risk factor to track for the year. Staff must be reminded that safety is a top priority and that they each have stop work authority if they observe a situation that may be unsafe.

Additionally, risks related to cybersecurity (406) and handling of export-controlled information (1240), both of which were major risks in CY 2024, remain major risks tracked in 2025. Cyberthreats are ever present and require constant vigilance not only by the cybersecurity team but by all staff members. Staff must also be vigilant for issues related to handling of export-controlled information, especially as operations continue under the Scalable Protected Infrastructure (SPI) program. Further discussion of the center's cybersecurity approach and the associated risk mitigation is discussed in Section 7 of this report.

5.6 RISKS THAT OCCURRED DURING THE CURRENT YEAR AND THE EFFECTIVENESS OF THEIR MITIGATION

No operations risks were triggered in CY 2024.

6. ENVIRONMENT, SAFETY, AND HEALTH

CHARGE QUESTION 6: Has the Facility demonstrated effective Environment, Safety, and Health (ES&H) practices to benefit staff, users, the public, and the environment?

OLCF RESPONSE: Yes. The OLCF is committed to providing a safe workplace and leveraging ES&H practices that benefit ORNL employees, subcontractors, visitors, and the environment. The OLCF is committed to the zero-accident philosophy and believes that all accidents are preventable. The OLCF utilizes DOE safety regulations as specified in 10 CFR 851, ORNL's Standards Based Management System (SBMS), the Battelle Safe Conduct of Research Principles, and many other systems/work practices to obtain zero-accident performance.

The core functions of the Integrated Safety Management (ISM) process require defining the work scope, analyzing the scope for hazards, developing controls, performing work safely, and continuous improvement. These functions are the basis for safe operations in the OLCF data centers. ORNL SBMS contains several procedures for ISM that are implemented in the OLCF. ORNL's SBMS Subject Area Work Control is broken into smaller procedures that provide the requirements for research staff, maintenance staff, space management, and non-employees. The procedures for ISM in Research and Development, Implement Work Control for Operations and Maintenance, Maintain ISM in Laboratory Space, and Implement ISM for Nonemployees are utilized in the OLCF's safety management program. Each program is tailored for the target audience and provides different types of work control documents, research safety summary (RSS) for research staff and lab spaces, a work plan and job hazard evaluation for maintenance staff, and a subcontract with ES&H requirements, including subcontractor/vendor submittal of a hazard analysis. Each type of work control requires the identification of the hazards and the development of controls. Additionally, each type of work control is reviewed and approved. The approval level is a graded approach that depends on the hazard set. Approvals can include ES&H (ORNL and vendor), supervisor, company/vendor representative, technical project officer, group leader, division director, subject matter experts, or a combination. In addition to utilizing the ORNL subject matter experts, the OLCF center manager is a certified safety professional and is fully involved in work planning and safe execution of all center work.

Daily safety management in the facility also utilizes multiple tools. These tools include inspections of job sites or walk throughs, pre-task briefings, briefings on applicable lessons learned/safety snapshots, safety talks, hazard analysis briefings, stop-work issuance, post-task reviews, and management assessments. Management assessments are formal reviews and are recorded in the Assessment and Commitment Tracking System. The tracking system documents the completion of the ORNL ISM process and provides a means for analysis. The DOE ORNL Site Office participates in field implementation and documentation of all operational safety reviews and partners with the ORNL Offices of Institutional Planning and Integrated Performance Management and the Safety Services Division on independent safety management system assessments. The culture of safety at ORNL and the OLCF is reflected in these processes, which are designed to reduce and prevent injuries to personnel and potential exposure to hazards associated with operation of the facility.

6.1 NORMAL DAY-TO-DAY OPERATIONS

Normal day-to-day operations of the OLCF include small system installations (same-rack installations), multiple-rack installations, un-racking and excessing equipment, infrastructure upgrades, routine facility maintenance, mechanical system maintenance, 24-7-365 operations support, and oversight of vendor/subcontractors performing on-site work.

6.1.1 Safety Performance

For CY 2024, the normal day-to-day operations met the zero accident performance criteria and remained safe, efficient, and effective. The ES&H performance for normal day-to-day operations included zero total recordable cases, zero Days Away Restricted or Transferred (DARTs), and zero first aid cases.

6.1.2 Normal Day-to Day ES&H Highlights

6.1.2.1 Center Support

CY 2024 was a busy year for the installation of cabinets and removal of outdated systems for the OLCF. The facility staff integrated several contractors into operations for accomplishing these tasks. Each contractor performed work in accordance with a hazard analysis that was developed and tailored to the work steps, hazards, and controls. In total, 34 additional cabinets of hardware/tape storage were added to the OLCF. Facilities also removed 50 cabinets of outdated equipment and shredded (with a vendor) 42,000 storage drives. The OLCF also de-energized and removed the water supply for the Summit machine and an additional large-scale machine. Finally, the OLCF and its construction partners constructed a new 24-cabinet pod, including power and piping upgrades. All these tasks were successfully completed without injury or any other upset condition.

6.1.2.2 Work Control/Monitoring

As in past years, the RSS was revised this year, and noise monitoring was conducted for the three data centers. The addition of new machines has impacted the noise levels in several areas. Noise monitoring results will be added to the RSS for staff members to review. Additional postings have also been added to mark the new areas that require hearing protection.

6.1.2.3 Safety/Insight Talks

Safety/insight talks are utilized to promote conversations about safety, safety culture, and the Safe Conduct of Research principles. The talks provide an opportunity to reinforce key safety culture principles by making them part of everyday conversations. Safety talk topics can be task-related safety items, facility safety items, or items related to safety away from work. Operations staff performed 134 safety/insight talks in CY 2024.

6.1.2.4 Center Hazard Analysis

All installations that require vendor support were conducted in accordance with the center’s hazard analysis. All individuals on site for site work received a pre-task hazard analysis briefing from the center manager or center staff.

6.2 LARGE-SCALE SUPERCOMPUTER INSTALLATION/ACTIVITIES

The CY 2024 large-scale workload included the operation of two large-scale supercomputers.

For CY 2024, the large-scale installation/activities also met the zero accident performance criteria. All work was accomplished safely with zero total recordable cases, zero DARTs, and zero first aid cases. In addition, the work was performed without adverse impact to the facility, co-located ORNL employees, or the environment.

6.2.1 Large-Scale Supercomputer Installation/Activities ES&H Highlights

6.2.1.1 Vendor Hazard Analysis

All vendor personnel were briefed on the new hazard analysis by the OLCF center manager. In addition to the hazard analysis, (tasks, hazards, and controls), the applicable safe conduct of research principles, human performance, applicable lessons learned, and stop/suspend work authorization are covered during the briefing.

6.2.1.2 Frontier's Cooling Loops and Fluid Replacement

Stabilization of Frontier's cooling loops, which contained the original vendor-supplied water, as well as the replacement of the water with new coolant began during CY 2023 and was completed in CY 2024. The effort included a refresh of each cooling distribution unit, which included removal of the cooling water, flushing, and replacement of the cooling water. Facility staff worked with the vendor to analyze the work, develop a schedule and controls, and to ensure safe execution of the work, including a variance in ORNL waste water discharges for disposal of the removed cooling water.

6.2.1.3 Subcontractor ES&H Engagement

The engagement of the subcontractor's ES&H support is critical to ensuring that the vendor is fully invested in staff member safety and health and is fulfilling the subcontract's ES&H requirements. In CY 2024, OLCF subcontractors added several on-site, hands-on support personnel. The subcontractor utilized a staff augmentation service to provide potential candidates. These candidates adopted ORNL's safety culture from the OLCF facility staff through briefings and were mentored by other subcontractor staff. The subcontractor site manager and the OLCF data center manger observed these employees closely during their mentoring period. The subcontractor and OLCF data center manager discussed each candidate's work performance in relation to safety, and the appropriate candidates were hired by the subcontractor.

7. SECURITY

CHARGE QUESTION 7: (a) Has the Facility demonstrated effective cyber security practices? (b) Does the Facility have a valid cyber security plan and Authority to Operate? (c) Does the Facility have effective processes for compliance with applicable national security policies related to Export Controls and foreign visitor access?

OLCF RESPONSE: Yes. The OLCF maintains a strong culture of continuous operational improvement, especially in cybersecurity. Updates to the ORNL cybersecurity program plan as well as the supercomputing security zone system security plan were made in CY 2024 in accordance with the latest NIST 800-53 Rev 5 policy guidance. The Authorizing Official at the DOE Oak Ridge Office recently retired, and the new Authorizing Official has granted interim Authority to Operate to ORNL and subsequently the OLCF while policy standards are assessed and vetted over the course of FY 2025 (see Figure 7-1). The OLCF expects to have a longer, permanent Authority to Operate by the end of CY 2025 following that assessment.

The technical staff members track and monitor existing threats and vulnerabilities to assess the risk profile of the OLCF operation. The OLCF is committed to innovating in this area by developing open-source tools and employing cutting-edge practices that enhance the operation without increasing the OLCF's risk profile.

The OLCF employs ORNL policies related to export control and foreign visitor access.

Department of Energy
Office of Science
Oak Ridge National Laboratory Site Office
P.O. Box 2008
Oak Ridge, Tennessee 37831-6269

February 12, 2025

Mr. K. Eddie Bishop
Information Systems Security Manager
Oak Ridge National Laboratory
Post Office Box 2008
Oak Ridge, Tennessee 37831-6045

Dear Mr. Bishop:

EXTENSION OF INTERIM AUTHORIZATION TO OPERATE (ATO) FOR THE SUPERCOMPUTER ENCLAVE

References: 1) Letter from K. Eddie Bishop to William W. Wheeler, subject, *Contract DE-AC05-00OR22725, Request for Full Authorization to Operate for the Supercomputing System*, dated November 7, 2024

2) Letter from William W. Wheeler to Stephen K. Streiffer, subject, *Reauthorization of Unclassified Information Systems at Oak Ridge National Laboratory (ORNL)*, dated November 27, 2024

The request in Reference 1 has been reviewed and is not approved. This decision was based primarily on the provided Risk Assessment (RA) not being aligned with NIST SP 800-30, *Guide for Conducting Risk Assessments*.

The existing Interim ATO for the Supercomputing Enclave is extended until November 30, 2025. Consistent with Reference 2, please ensure the Supercomputer Enclave is included among the early authorization packages provided for approval, and that all RA documents provided with the updated authorization packages are consistent with NIST SP 800-30 and Office of Science Cyber Security Program Plan expectations.

My point of contact for this action is Cody Carpenter at Cody.Carpenter@science.doe.gov.

Sincerely,

William W. Wheeler
Authorizing Official

Figure 7-1. OLCF Authority to Operate.

7.1 SUMMARY

All IT systems that operate for the federal government must have authority to operate. This involves developing and obtaining approval for a policy and implementing a continuous monitoring program to confirm that the policy is effectively implemented. The ORNL accreditation package currently uses the

NIST Special Publication 800-53, revision 5, Security and Privacy Controls for Federal Information Systems and Organizations, and the US Department of Commerce Joint Task Force Transformation Initiative (August 2009) as guidelines for security controls. The OLCF determined that the highest classification of data is moderate based on the guidelines for information classification in the Federal Information Processing Standards Publication 199, Standards for Security Categorization of Federal Information and Information Systems, Computer Security Division, Information Technology Laboratory, NIST. The OLCF is accredited at the moderate level for protecting the confidentiality and integrity of user and system information, and this accreditation authorizes the OLCF to process sensitive, proprietary, and export-controlled data.

Security at the OLCF is built upon a strong configuration management baseline. Puppet, among other tools, is used to enumerate and deploy both security and operational configurations required by policy and best practices. HPC system images delivered from the vendor are augmented with a Puppet configuration to bring all nodes of a system into compliance. Other important controls include enforcement of multifactor authentication in the OLCF moderate enclave, lightweight and well-adopted configuration management procedures, adoption of DevSecOps principles (e.g., tight intergroup coordination and use of continuous integration), and strong incident response and triage capabilities (e.g., operational and security dashboards and frequent practice).

Year after year, cybersecurity planning is becoming more complex as the center continues its mission to enable world-class science. The center is very proactive and views its cybersecurity plans as dynamic documentation that it will preemptively respond with and modify as needed to provide an appropriately secure environment. The OLCF abides by the Health Insurance Portability and Accountability Act’s Privacy and Security Rule to provide supercomputing resources to projects that contain PHI. The OLCF also abides by the International Traffic and Arms Regulations (ITAR) and DOE I 471.7, Controlled Unclassified Information for projects that contain these levels of sensitive information.

7.2 SECURITY OPERATIONS

The security team performs a wide range of activities, including security policy development and assessment, event monitoring, incident handling and reporting, vulnerability scanning and triage, and security system engineering. The team maintains a DevSecOps mindset to automate security when possible and to integrate with the software development and systems/operations teams in the center to help design and secure new capabilities as they are developed. The security team has developed and enforced procedures and policies to help formalize security-related processes and to improve security posture. This includes a security data review policy, which will help provide adequate information from data owners and allow the security team to properly analyze the data for any potential security issues. The OLCF has been involved in external documentation such as the NIST SP 800-223 *High-Performance Computing (HPC) Security: Architecture, Threat Analysis, and Security Posture* document that was finalized in 2024. The NIST SP 800-223 document outlines security architecture, threat analysis, and security posture for HPC.

To facilitate continuous operational improvement, the security team has an open “Security Questions” channel through which anyone can ask security-related questions about vulnerabilities, operational procedures, policy, and more. The team also participated in the HPC Security Technical Exchange conference hosted at LLNL, where they met with other security professionals from other labs and exchanged ideas and information to work toward continuous improvement. The security team was also involved in SC24, where a team member was on the organizing committee of the International Workshop on Cyber Security in High Performance Computing and participated in the Zero Trust talk. The security team also met with other organizations to share and discuss how to improve security posture during the conference. Additionally, the team purchased the enterprise version of Fleet/Osquery, which is used for vulnerability reporting, detection engineering, and device management. Ongoing efforts to maintain and upgrade the open-source version of Fleet/Osquery were more expensive than purchasing an enterprise license. This purchase has allowed the organization to save money through reduction of security engineering effort and allowed the security team to provide access to information system owners owing to additional capabilities that the enterprise license unlocks. Direct access to Fleet/Osquery empowers OLCF information system owners to resolve security issues as well as monitor their systems for troubleshooting purposes without requiring security engineer intervention.

The OLCF has developed Secure LLM Inspection and Malware Evaluation (SLIME) capabilities in tandem with ORNL enterprise security engineers and researchers from the ORNL Center for Artificial Intelligence Security Research. SLIME allows OLCF security engineers to spot-check parallel file systems and other high-performance storage areas for indicators of compromise and unwanted data that may exist in AI/ML datasets. Future work will enable the OLCF to scan and evaluate the safety and security of LLM models that OLCF users bring to the facility.

In addition to project work, the security team also tracks significant cybersecurity events that are above and beyond normal baseline threats reported in the OLCF ticket tracking system. Baseline threats such as background SSH (Secure Shell) scanning, firewall probing, and software vulnerabilities patched within quarterly patching windows are expected. Examples of events above this baseline include the public disclosure of serious software vulnerabilities, detected suspicious user behavior, and observed but unexpected patterns of system event logs and metrics. In CY 2024, the OLCF security team tracked 21 above-baseline threats, none of which resulted in a compromise of OLCF systems by malicious attackers. Each of these events is treated as a live incident and used to practice the OLCF incident response plan several times throughout the year.

Because HPC systems are scientific instruments, care must be taken to determine the appropriate response for each vulnerability and potential threat. Industry-standard responses such as emergency patching and application allow-listing are difficult to implement quickly because scientific software stacks provided by the facility and users may need to be rebuilt and then tested to ensure the correctness of scientific results.

7.3 OLCF USER VETTING

The OLCF follows a set of rigorous controls for vetting user access, as defined by ORNL and DOE policy, to ensure compliance with export-control regulations and foreign visitor access policies.

7.3.1 OLCF Projects

Users must be added to an approved OLCF project to obtain access to OLCF resources. An ORNL export control officer reviews the scope of work for all OLCF user projects to determine whether there are any export-control restrictions to which the OLCF must adhere and to place an internal designation of category 1, category 2, or category 3 on each project. These categories then drive the business processes that are followed for each applicant and defined in Table 7-1.

Table 7-1. Export control review categories for projects.

Category designation	Category description	PI actions before project activation
Category 1	The category 1 rating is applied if the project is open fundamental research that does not involve proprietary input and/or output, sensitive data, and/or export-control restrictions above EAR99.	Sign OLCF PI agreement
Category 2	The category 2 rating is applied if the project involves proprietary input and/or output, sensitive subject areas, and/or export-control restrictions above EAR99 but below ITAR.	Sign OLCF PI agreement Participate in mandatory security call to review risks/restrictions associated with category 2 projects
Category 3	The category 3 rating is applied if the project involves proprietary input and/or output, sensitive subject areas, PHI, PII, and/or export-control restrictions above EAR99 to include ITAR.	Sign OLCF PI agreement Participate in mandatory security call to review risks/restrictions associated with category 3 projects

Sensitive information, including proprietary and export-controlled information, is segregated and protected in the specific project area to protect it from unauthorized access, and specific storage rules and requirements are relayed to the PI and individual project users to further prevent information mishandling. If a project is rated category 2 or above, then the project PI must participate in a mandatory security call with the Information Systems Security Officer (or designee) to review the risks and restrictions before the project is enabled.

Once the security call is complete and all other project requirements are met, the project is enabled in the OLCF RATS and labeled with the appropriate category, and any export control restrictions are added to the project.

7.3.2 OLCF Users

All users requesting access to OLCF resources are required to fill out the OLCF account application form and provide the project identification and PI for the project they are requesting to join. Based on the category of the project designated in RATS, the following requirements (Table 7-2) must be met before the user can be added to the project and provided access to OLCF resources.

Table 7-2. OLCF project category requirements.

Project category	PI approval	ORNL Personnel Access System (PAS) [1] or Restricted Party Screening (RPS) [2]	User Agreement [3]	Sensitive data rules [4]	Level 2 identity proofing [5]
Category 1	PIs must approve all user account requests to access to their project.	a. Approved PAS is required for all applicants that are not US citizens or lawful permanent residents. b. All US citizens or lawful permanent residents go through Restricted Party Screening (RPS).	Must have valid user agreement	N/A	Required
Category 2	PIs must approve all user account requests to access to their project.	c. Approved PAS is required for all applicants that are not US citizens or lawful permanent residents. d. All US citizens or lawful permanent residents go through RPS.	Must have valid user agreement	Must return signed sensitive data rules	Required
Category 3	PIs must approve all user account requests to access to their project.	e. Approved PAS is required for all applicants that are not US citizens or lawful permanent residents. f. All US citizens or lawful permanent residents go through RPS.	Must have valid user agreement	Must return signed sensitive data rules	Required

[1] PAS: The system that ORNL uses to process on-site and/or remote access for foreign nationals and non-employees.
[2] RPS: ORNL maintains a subscription to the Descartes Visual Compliance tool, which is used to look up applicants and their institutions that do not require PAS approval. If any hits are found on the user or the user's institution, then the information is turned over to the export control officer. The officer then works with the Counterintelligence Office to look at the applicant or institution in more detail and informs the OLCF whether it is acceptable to proceed.
[3] User Agreement: Serves as the master agreement that establishes the general terms and conditions, including the disposition of intellectual property, for work at any of ORNL's user facilities. A User Agreement must be executed with each user's institution.
[4] Sensitive data rules: This form contains a user acknowledgment that documents that users on a category 2 project are aware of the risks and rules for accessing the sensitive project.
[5] Level 2 identity proofing: The OLCF uses RSA SecurID tokens for authenticating to OLCF moderate resources. Level 2 identity proofing of all applicants is required as part of the OLCF moderate Certification and Accreditation. To achieve Level 2 identity proofing, applicants must have their identity and RSA SecurID token verified by a notary or an OLCF-designated registration authority. The token is not enabled until all the above steps are completed, including the return of the original notarized form.

8. STRATEGIC RESULTS

CHARGE QUESTION 8: (a) Are the methods and processes for monitoring scientific accomplishments effective? (b) Has the facility demonstrated effective engagements with technology vendors and /or engaged in effective research that will impact next generation technology relevant to the facility's mission? (c) Has the Facility demonstrated effective engagements with critical stakeholders (such as the SC Science Programs, DOE Programs, DOE National Laboratories, SC User Facilities, and/or other critical U.S. Government stakeholders [if applicable]) to both enable mission priorities and gain insight into future user requirements?

OLCF RESPONSE: Yes. OLCF projects and user programs are advancing DOE's mission to ensure US security and prosperity by addressing its energy, environmental, and nuclear challenges through transformative science and technology solutions. The selected accomplishments described in this section highlight how the OLCF is advancing two strategic objectives of DOE's Strategic Plan Goal 1, "Science and Energy: Advance foundational science, innovate energy technologies, and inform data driven policies that enhance economic growth and job creation, energy security, and environmental quality...," as stated in the DOE Strategic Plan: 2014–2018 (March 2014):

- Strategic Objective 2: Support a more economically competitive, environmentally responsible, secure, and resilient US energy infrastructure
- Strategic Objective 3: Deliver the scientific discoveries and major scientific tools that transform the understanding of nature and strengthen the connection between advances in fundamental science and technology innovation

8.1 SCIENCE HIGHLIGHTS AND ACCOMPLISHMENTS

The OLCF collects and reports annually the number of refereed publications that result (at least in part) from the use of OLCF resources. For the Leadership Computing Facilities (LCFs), tracking is done for a period of 5 years following the project's use of the facility. This number may include publications in press or accepted but not those submitted or in preparation. This is a reported number, not a metric. Additionally, the facility may report other publications when appropriate.

The facility also regularly searches for and actively solicits information about possible scientific highlights produced by projects that use OLCF resources. Specific questions about the possible availability of such highlights are asked as part of each quarterly and annual report solicitation for projects. In addition, scientific liaisons to INCITE projects regularly report potential scientific highlights to the facility's management. These active avenues for discovering potential scientific highlights are augmented by a regular (i.e., roughly monthly) examination of the ongoing publication tracking described above.

Publications in high-impact journals or papers that garner a significant number of early citations (among other criteria) are deemed good candidates for features.

8.1.1 OLCF Publications Report

In 2024, 587 publications resulting from the use of OLCF resources were published based on a data collection completed on April 10, 2025. In this document, "year" refers to the calendar year unless it carries the prefix FY. In the 2023 OLCF OAR, 495 publications were reported (this number has been significantly revised upward; see Table 8-1 below). Guidance allows accepted and in-press publications to be reported, but the OLCF reports only publications appearing in print in the year under review. However, the OLCF continues to search for publications after the OAR is submitted to DOE each year, and the number of publications shown in previous OARs is updated in the current report. Table 8-1 provides the updated, verified, and validated publications count for the 2012–2024 period, showing overall consistent growth in both the total publications count and the number of publications in journals with high-impact factors

Table 8-1. Summary of unique OLCF publications for 2012–2024.

Year	Unique, confirmed OLCF publications	High-impact publications with JIF* >10
2024	581	20
2023	560	28
2022	575	22
2021	565	27
2020	524	37
2019	459	31
2018	496	20
2017	477	27
2016	467	33
2015	366	21
2014	315	16
2013	364	9
2012	342	20

*JIF = Journal impact factor

8.1.1.1 OLCF Publication Methodology

The facility collects and reports annually the number of refereed publications resulting (at least in part) from use of OLCF resources. This number may include publications in press or accepted but not those submitted or in preparation. This is a reported number, not a metric. In addition, the OLCF may report other publications where appropriate. The OLCF employs a multipronged approach to discovering, validating, and reporting publications, and this approach is explained in more detail below.

8.1.1.2 Discovery Methods for OLCF Publications

There are three primary methods of identifying OLCF publications: COBRA, Resolution, and self-reports.

COBRA is a system developed by ORNL to automate the discovery and management of publications related to scientific facilities and organizations. It uses various automatic methods to search for publications across multiple sources such as Web of Science, Google Scholar, Scopus, and more. It also verifies the publications by looking at the metadata, full text, acknowledgments, and funding agencies, among other information.

Resolution is ORNL's publication clearance and tracking system for publications with an ORNL author. To capture publications from Resolution, the COBRA administrator manually queries all publications in Resolution to collect any that contain OLCF keywords. If any are found, then the publications are imported into COBRA.

The final method of discovery is self-reporting from users. These reports are generally collected through quarterly and closeout reports. Upon receipt of reports, all publications listed in the report are manually verified first by querying COBRA to see if the publication has already been captured. If the publication is not already in COBRA, then OLCF staff will search online for the publication. Once found, a review of the acknowledgment and funding sections is conducted to locate the keywords and/or OLCF acknowledgment statement. If they meet the criteria, then they are added to COBRA.

8.1.1.3 Publication Validation

Once the publications are in COBRA, the next step is to validate each publication to ensure the work utilized OLCF resources. This is done by ORNL's library staff and/or OLCF staff. This is generally done by reviewing the publication and looking for the acknowledgment statement, associated keywords, and/or directly contacting the authors of the publication. Once validated, the publication is marked "confirmed" in COBRA.

8.1.1.4 How COBRA Results are Disseminated

Once the discovery process has been completed, COBRA has many features for publication management. COBRA provides the user with an interactive web-based GUI that allows for searching, filtering, and exporting of the data. The publication information data also comes from multiple sources and sources that currently have no openly available API (self-reports, internal systems, Google alerts). To accommodate these sources, COBRA also allows for publications to be entered manually. To increase efficiency in manual entry, COBRA can pull most of the metadata associated with a publication from its DOI (digital object identifier). COBRA uses the Web of Science and Cross-Ref APIs to automatically populate many fields, including title, authors, and publication date. COBRA also calculates metrics based on the publication metadata, such as high impact, citation count, and highly cited publications, while keeping track of where publications are in their publication process. If greater detail is required, then a query can be executed on the data, and those records can be downloaded into a spreadsheet for further review.

The OLCF considers COBRA to be an invaluable addition to the publication workflow, and it greatly improves publication discovery and overall management beyond reliance on self-reports and ad-hoc tracking of publications (e.g., an Excel spreadsheet).

8.1.1.5 Scientific Accomplishments

The OLCF advances DOE's science and engineering enterprise by fostering robust scientific engagement with its users through the INCITE liaison program, the user assistance program, and the OLCF DD program. The following subsections provide brief summaries of select scientific and engineering accomplishments as well as resources for obtaining additional information. Although they cannot capture the full scope and scale of achievements enabled by the OLCF in 2024, these accomplishments advance the state of the art in science and engineering R&D across diverse disciplines and advance DOE's science programs toward mission goals. OLCF users produced many breakthrough publications in high-impact journals in 2024, which is an additional indication of the breadth of these achievements (Table 8-2).

Table 8-2. Publications in high-impact journals in 2024.

Journal	Number of publications
ACS Applied Materials & Interfaces	1
ACS Nano	4
Advanced Functional Materials	1
Advanced Materials	1
Angewandte Chemie-International Edition	1
Bioinformatics	1
Bulletin of the American Meteorological Society	1
Chemical Science	1
Chemsuschem	1
Green Chemistry	1
Journal of Physical Chemistry Letters	1
Journal of the American Chemical Society	3
Nano Energy	1
Nature	3
Nature Nanotechnology	1
Nature Physics	1
New Phytologist	1
Physical Review Letters	18
Physical Review X	1
Proceedings of the National Academy of Sciences of the United States	3
Reports on Progress in Physics	2
Reviews of Modern Physics	1
Science	1

Altogether in 2024, OLCF users published 50 papers in journals with a JIF of greater than 7 and published 20 papers in journals with a JIF greater than 10.

8.1.2 Science Highlights

8.1.2.1 Game-Changing Quantum Chemistry Calculations Push New Boundaries of Exascale Frontier

Researchers conduct largest and most accurate molecular dynamics simulations to date of 2 million correlated electrons by using the world's fastest supercomputer

PI: Giuseppe Barca, University of Melbourne

Allocation Program: DD

The Science

Historically, researchers have hit a wall trying to simulate the physics of molecular systems with highly accurate models because there just was not enough computing power. So, they have been confined to only simulating small molecules, but many of the interesting problems they want to solve involve large models. For drug design, for example, large proteins are often responsible for causing disease. Those proteins can be as large as 10,000 atoms—too many to simulate without using highly approximate models running at a reduced precision. Reduced-precision models are more computationally efficient but produce less accurate results. Modern drug design requires the ability to simultaneously model large proteins and match them with libraries of small molecules intended to bind to the large protein and block it from functioning. Most molecular simulations use simplified models to represent the forces between atoms. However, these force-field approximations do not account for certain essential quantum mechanical phenomena such as bond breaking and formation—both important to chemical reactions—and other complicated interactions. To overcome these limitations, the University of Melbourne–led team developed the Extreme-scale Electronic Structure System (EXESS) code specifically designed for exascale systems with hybrid architectures that utilize a combination of CPUs and GPU accelerators.

The Impact

EXESS enabled the first time-resolved quantum chemistry simulation to exceed an exaflop—over a quintillion, or a billion-billion calculations per second—using double-precision arithmetic. In addition to setting a new benchmark, the achievement also provides a blueprint for enhancing algorithms to tackle larger, more complex scientific problems using leadership-class exascale supercomputers.

Summary

A University of Melbourne-led team (see Figure 8-1) won the Association for Computing Machinery's (ACM's) 2024 Gordon Bell Prize on November 14, 2024, for their study that used the Frontier supercomputer to calculate the dynamics of more than 2 million electrons, calculating the number of atoms in a molecular dynamics simulation 1,000× greater in size and speed than any previous simulation of its kind. This is the first time-resolved quantum chemistry simulation to exceed an exaflop when using double-precision arithmetic. In addition to setting a new benchmark, the achievement also provides a blueprint for enhancing algorithms to tackle larger, more complex scientific problems using leadership-class exascale supercomputers. This work has applications in drug design.

Funding

This research was supported by the DOE Office of Science's ASCR program.

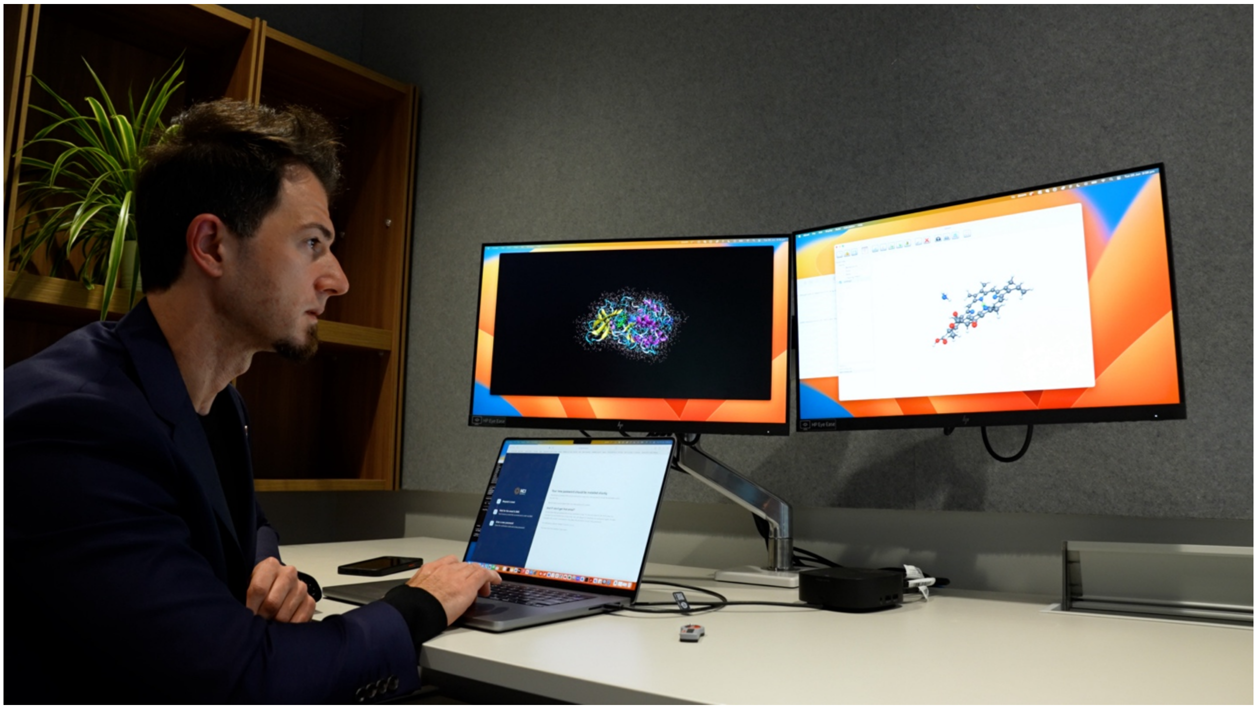


Figure 8-1. Giuseppe Barca, an associate professor at the University of Melbourne, led a team of researchers in breaking new ground in the exascale computing era. Using Frontier, the team conducted a molecular dynamics simulation 1,000× greater in size and speed than any previous state-of-the-art simulation of its kind (Credit: University of Melbourne).

Publication: R. Stocks, et al., *SC24: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis.* November 2024, Article No. 9, 1-12.

Related Link: "Game-Changing Quantum Chemistry Calculations Push New Boundaries of Exascale Frontier," OLCF News (July 17, 2024)

8.1.2.2 Frontier Users' Exascale Climate Emulator Nominated for Gordon Bell Climate Prize

The highly scalable climate emulator offers faster, radically enhanced high-resolution simulations without the need for massive data storage

PI: David Keyes, King Abdullah University of Science and Technology (KAUST)

Allocation Program: DD

The Science

A research team from KAUST used the Frontier supercomputer to develop an exascale climate emulator with radically enhanced resolution but without the computational expense and data storage requirements of state-of-the-art climate models. Climate models are incredibly complex and can take weeks or months to run, even on the fastest supercomputers. They generate massive amounts of data that become nearly impossible to store, and it is becoming a bigger and bigger problem as climate scientists are constantly pushing for higher resolution.

The Impact

The KAUST team's climate emulator solves two problems: speeding up computations and reducing storage needs. It is designed to mimic model outputs on demand without storing PBs of data. Instead of saving every result, all they have to store are the emulator code and necessary parameters. This, in principle, allows the team to generate an infinite number of emulations whenever they need them. By leveraging the latest advances in GPUs and mixed-precision arithmetic, the team's climate emulator offers a remarkable resolution of 3.5 km (approximately 2.2 miles) and can replicate local conditions on a timescale from days to hours (see Figure 8-2). This work was awarded the ACM's 2024 Gordon Bell Prize for Climate Modelling.

Summary

A research team led by KAUST won the 2024 Gordon Bell Prize for Climate Modelling. The KAUST team used the Frontier and Summit supercomputers to help them develop a climate emulator that offers radically enhanced resolution without the need to store massive amounts of data.

Funding

This research was supported by the DOE Office of Science's ASCR program.

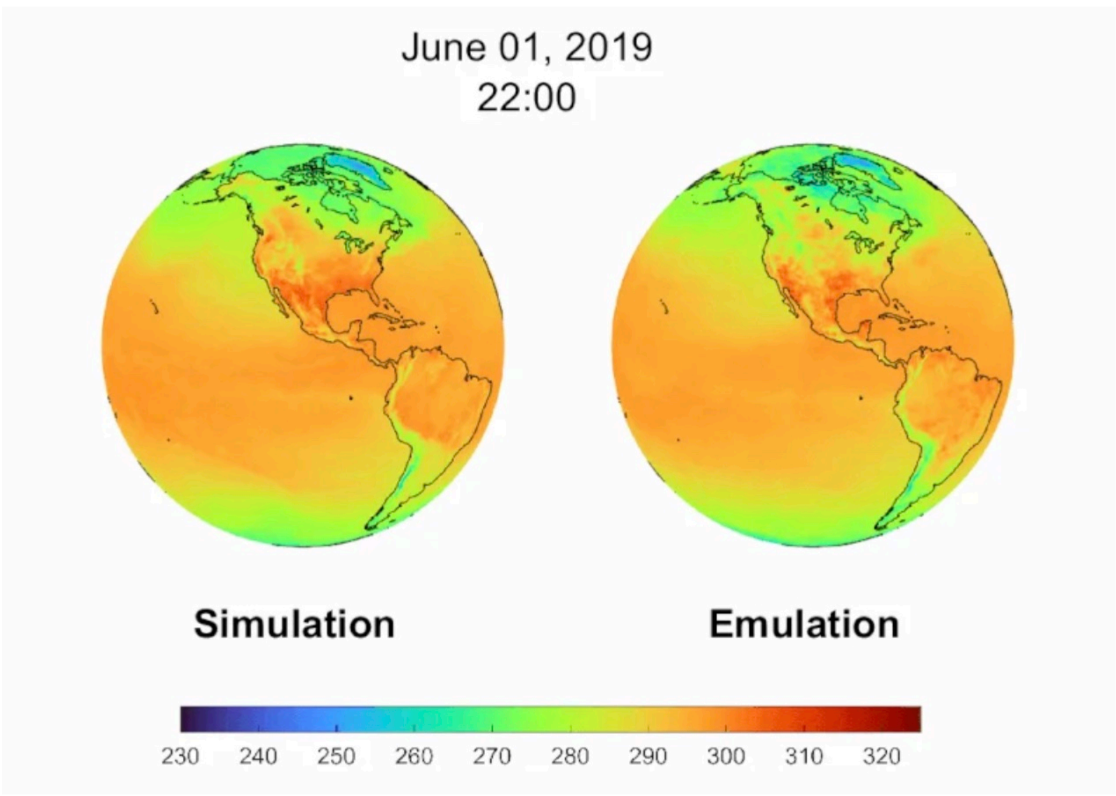


Figure 8-2. A research team led by KAUST used the Frontier and Summit supercomputers to help them develop a climate emulator that offers radically enhanced resolution without the need to store massive amounts of data (Credit: KAUST). Publication: S. Abdulah, et al., *SC24: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*. November 2024, Article No. 2, 1-12.

Related Link: "Frontier Users' Exascale Climate Emulator Nominated for Gordon Bell Climate Prize," OLCF News (October 30, 2024)

8.1.2.3 Protein Design on Demand

OLCF's Frontier trains AI to devise new foundations for life

PI: Arvind Ramanathan, Argonne

Allocation Program: ECP, DD

The Science

An Argonne-led research team used the Frontier supercomputer to train an AI model to draw up blueprints for the building blocks of life. LLMs can learn how to order and construct new patterns based on vast amounts of training data and sophisticated numerical techniques. The models can improve on that learning over time with additional training. The larger the model and its training datasets, the better its performance—but also the higher its demand for computational power. Proteins, the building blocks of life, perform vital functions that can range from transporting molecules, to replicating cells, to kicking off metabolic reactions. A single, average-sized protein can consist of as many as 300 amino acids strung together in unique sequences. Scientists have spent decades classifying those sequences in search of a means to predict a protein's function based on its structure (see Figure 8-3). Arvind Ramanathan and his team wanted to find out whether an LLM could identify ideal protein sequences for a given biochemical function.

The Impact

The team used Frontier to train an LLM on proteins as part of a high-speed digital workflow with four other supercomputing systems. Besides Frontier, the workflow included Argonne's exascale machine Aurora, the Swiss National Supercomputing Centre's Alps, the CINECA data center's Leonardo, and NVIDIA corporation's PDX. The workflow's simulations ran at exascale with at a peak speed of roughly 5.57 exaflops at mixed precision. This work earned the team a finalist nomination for the ACM's 2024 Gordon Bell Prize for outstanding achievement in HPC.

Summary

An Argonne-led research team was named a finalist for the ACM's 2024 Gordon Bell Prize. The team used the Frontier supercomputer to train an AI model to draw up blueprints for the building blocks of life.

Funding

This research was supported by the National Institutes of Health (NIH) and by the DOE Office of Science's ASCR program.

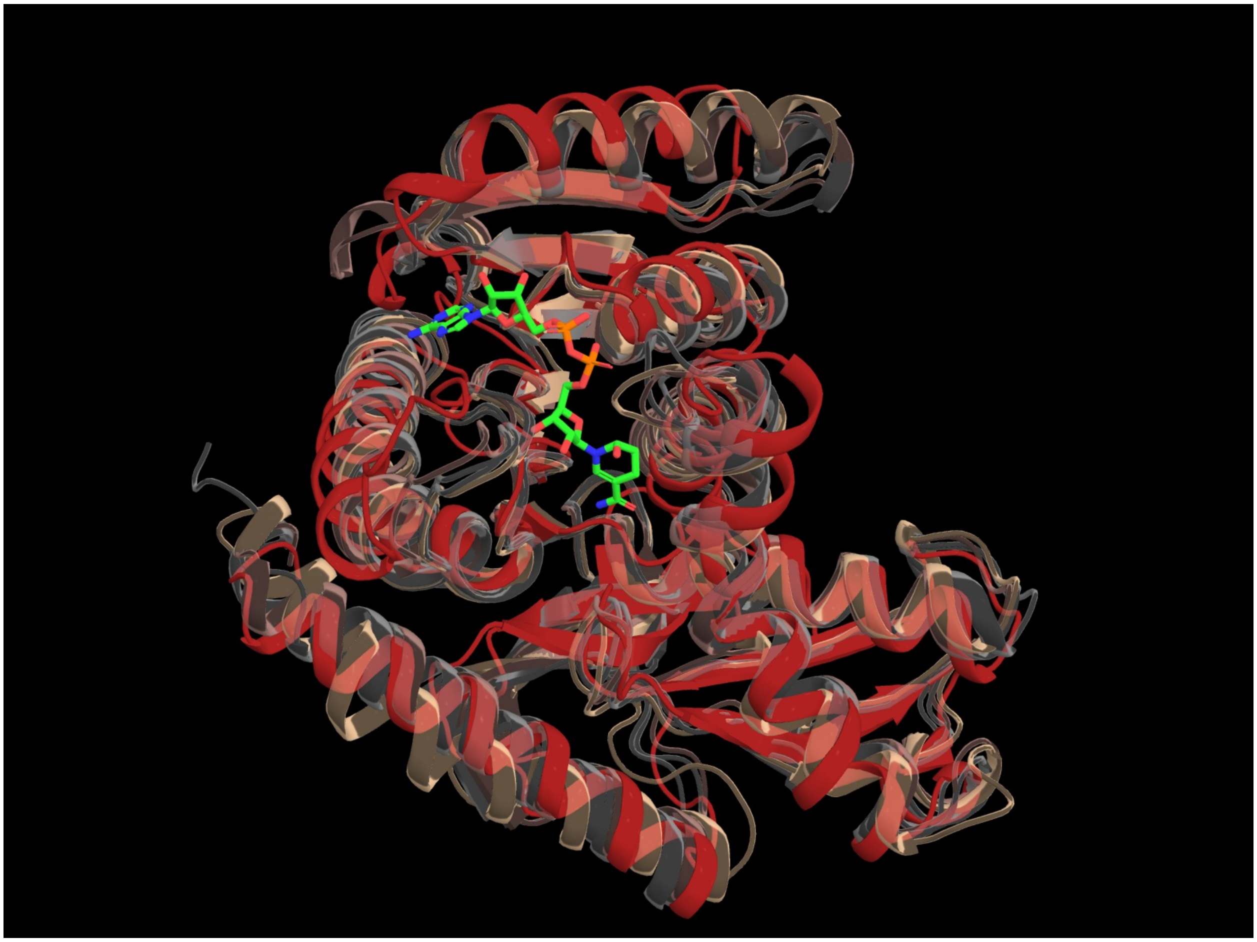


Figure 8-3. Scientists used Frontier to train an AI model that identified synthetic versions of malate dehydrogenase that preserve the protein's critical structure and key binding areas (Credit: Arvind Ramanathan, Argonne). Publication: G. Dharuman, et al., *SC24: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*. November 2024, Article No. 7, 1-13.

Related Link: "Protein Design on Demand," OLCF News (November 14, 2024)

8.1.2.4 Record-Breaking Run on Frontier Sets New Bar for Simulating the Universe in the Exascale Era

World's largest simulation of the cosmos lays new computational foundation for simultaneous extreme-scale dark matter and astrophysical investigations

PI: Salman Habib, Argonne

Allocation Program: INCITE (HEP142)

The Science

In November 2024, researchers used the fastest supercomputer on the planet at the time to run the largest astrophysical simulation of the universe ever conducted. A true understanding of structure formation in the Universe requires cosmological hydrodynamics simulations. That is, simulations of both dark matter—which only interacts through gravity—and atomic matter—which involves complex physics such as hot gas dynamics, star formation, and black holes. Until now, simulating a universe-scale volume with that level of detail was unthinkable.

The Impact

A research team from the DOE’s Argonne used the Frontier supercomputer at ORNL to set a new benchmark for cosmological hydrodynamics simulations and provide a new foundation for simulating the physics of atomic matter and dark matter simultaneously (Figure 8-4). The simulation size corresponds to surveys undertaken by modern, large synoptic telescope surveys, a feat that until now has not been possible at this scale. In addition, cosmological hydrodynamics simulations are vastly more computationally demanding than those focusing solely on gravity. The supercomputer code used in the simulation is called the Hardware/Hybrid Accelerated Cosmology Code, which was significantly upgraded through the ECP’s ExaSky effort.

Summary

In early November 2024, researchers at the DOE’s Argonne used Frontier, the fastest supercomputer on the planet at the time, to run the largest astrophysical simulation of the universe ever conducted. The team set a new benchmark for cosmological hydrodynamics simulations and provide a new foundation for simulating the physics of atomic matter and dark matter simultaneously. The simulation size corresponds to surveys undertaken by modern, large synoptic telescope surveys, a feat that until now has not been possible at this scale.

Funding

This research was supported by the ECP, ASCR, and the High Energy Physics programs under the DOE’s Office of Science.

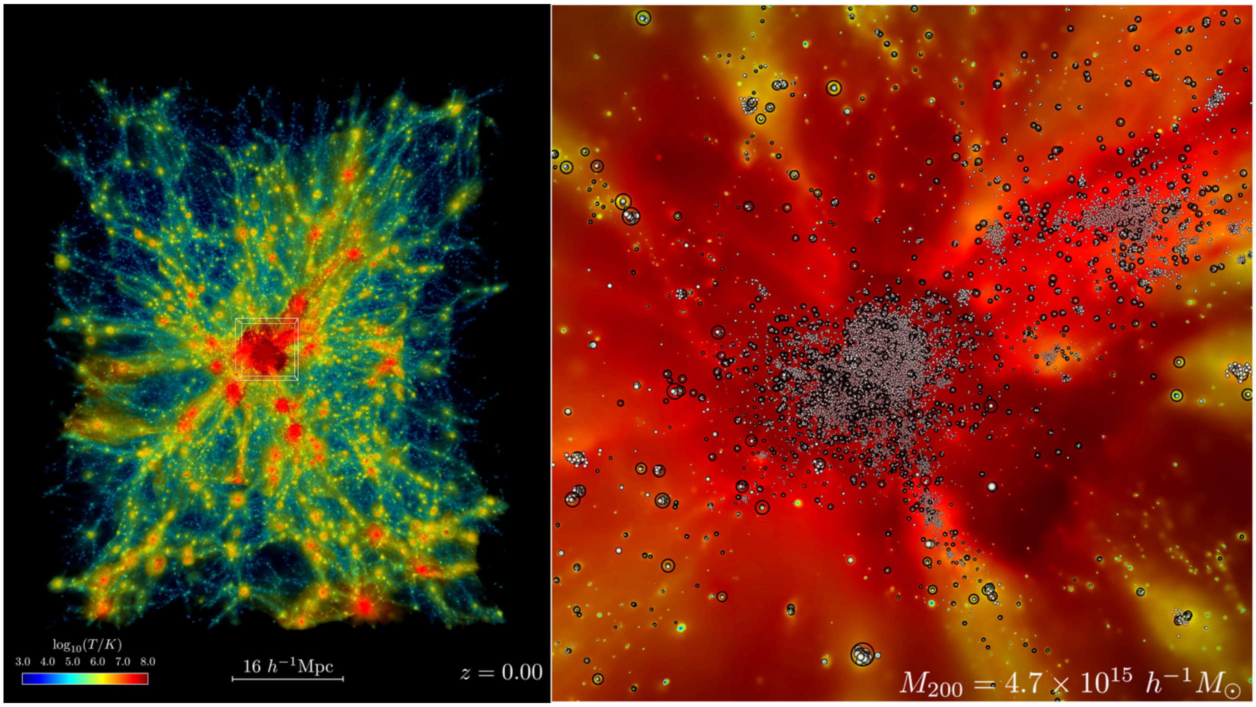


Figure 8-4. A small sample from the Frontier simulations reveals the evolution of the expanding universe in a region that contains a massive cluster of galaxies from billions of years ago to present day (left). Red areas show hotter gasses, with temperatures reaching 100 million Kelvin or more. Zooming in (right), star tracer particles track the formation of galaxies and their movement over time (Credit: Argonne, DOE).Publication: in development.

Related Link: "Record-Breaking Run on Frontier Sets New Bar for Simulating the Universe in the Exascale Era," OLCF News (November 19, 2024)

8.1.2.5 Frontier Simulations Provide New Insights into Calcium-48's Controversial Nuclear Magnetic Excitation

The world's most powerful supercomputer is helping resolve conflicting research results that have puzzled scientists for more than a decade and could shine new light inside collapsing stars

PI: Gaute Hagen, Oak Ridge National Laboratory

Allocation Program: INCITE

The Science

Calcium-48 is an important isotope in a variety of settings. Its nucleus is composed of 20 protons and 28 neutrons—a combination that scientists call "doubly magic." Magic numbers—such as 20 and 28—are specific numbers of protons or neutrons that provide stability by forming a closed shell within the nucleus. The strong binding and simple structure of calcium-48 also make it an interesting test subject for studying the strong and weak nuclear forces that hold nuclei together or break them apart.

The Impact

An Oak Ridge-led team used a model called chiral effective field theory to connect nuclear forces to the fundamental theory of the strong nuclear force—the theory of quantum chromodynamics. They used a powerful numerical method called the coupled-cluster method to compute the properties of the calcium-48 nucleus (see Figure-8-5). The approach provides a compromise between high precision and detail and computational cost and is a perfect task for Frontier. The simulations showed that calcium-48's magnetic transition strength ranges from 7 to 10 nuclear magnetons squared, which is consistent with the results of the gamma ray experiments.

Summary

Nuclear physicists at ORNL used Frontier to calculate the magnetic properties of calcium-48's atomic nucleus. Their findings will not only provide a better understanding of how magnetism manifests inside other nuclei but will also help to resolve a decade-old disagreement between experiments that drew different conclusions about calcium-48's magnetic behavior. Additionally, the research could provide new insights into the subatomic interactions that happen inside supernovae.

Funding

The DOE Office of Science (Office of Nuclear Physics and Office of High Energy Physics) and DOE's SciDAC-5 NUCLEI Collaboration supported this research.

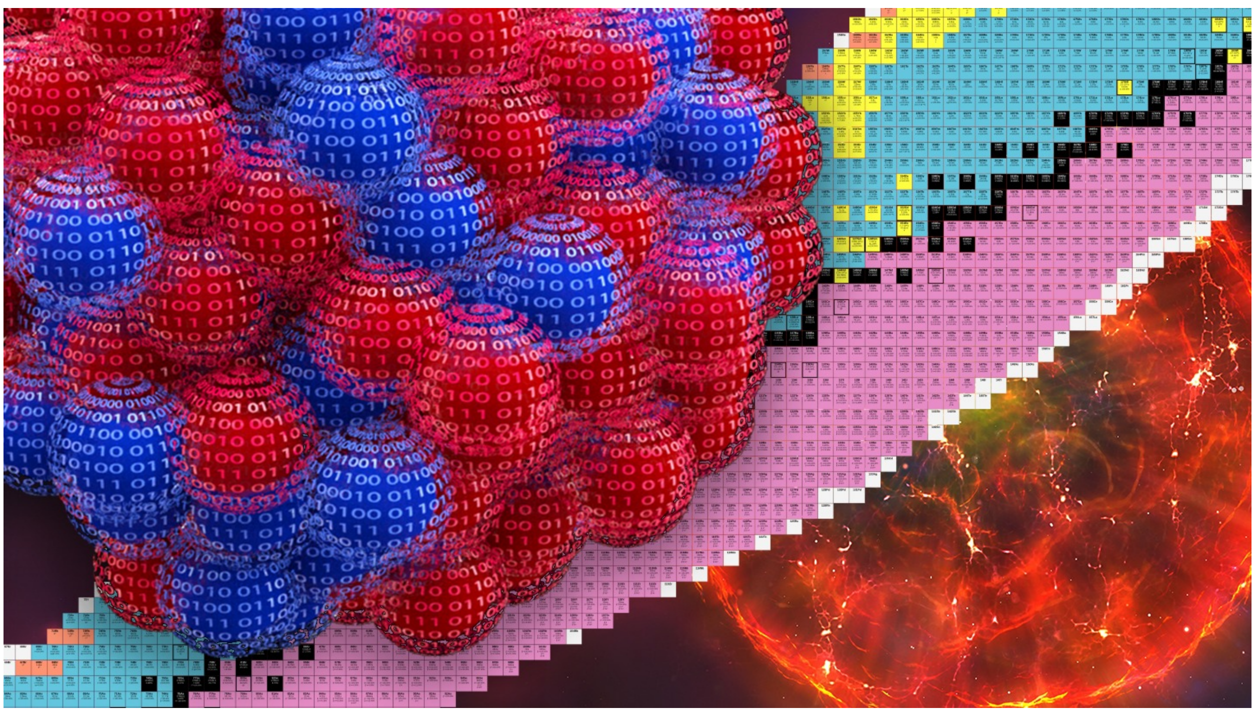


Figure-8-5. The Frontier supercomputer simulated magnetic responses inside calcium-48, depicted by red and blue spheres. Insights into the nucleus's fundamental forces could shed light on supernova dynamics (Credit: ORNL, DOE).

Publication: B. Acharya et al, "Magnetic Dipole Transition in Ca48," *Physical Review Letters* (2024): https://doi.org/10.1103/PhysRevLett.132.232504.

Related Link: "Frontier Simulations Provide New Insights Into Calcium-48's Controversial Nuclear Magnetic Excitation," OLCF News (September 3, 2024)

8.1.2.6 Flying Quieter and Cleaner

Startup aerospace company Whisper Aero uses Summit to test its novel concepts for an ultraquiet electric airplane

PI: Vineet Ahuja, Whisper Aero

Allocation Program: SummitPLUS

The Science

Startup aerospace firm Whisper Aero has designed an advanced electric ducted fan intended to replace the traditional fossil fuel–burning engines used on most aircraft. According to the company's tests, its electric ducted fans are at least 100× quieter and 20% more efficient overall than those of its competitors. Whisper Aero is also refining its vision for a nine-passenger electric airplane called the Whisper Jet. To predict how their electric ducted fans would perform in the Whisper Jet design, Whisper Aero's engineering team turned to the OLCF's Summit.

The Impact

Access to Summit and its GPU-accelerated architecture gave the company's researchers a 10× speedup in compute times compared to their in-house, CPU-only systems. The team put Summit's aero predictions to the test in Virginia Tech's wind tunnel with a ¼-scale drone demonstrator. "Summit's pretest predictions were not just close—they were dead on," said Vineet Ahuja, Whisper Aero's head of flight sciences.

Summary

Tennessee-based aerospace startup Whisper Aero is working to develop quieter and more fuel-efficient aircraft engines. Their engineering team turned to the OLCF for its robust computing power. Access to Summit and its GPU-accelerated architecture gave the company's researchers a 10× speedup in compute times compared to their in-house, CPU-only systems and enabled advanced simulations via a unified approach that was not previously possible (see Figure 8-6). The team plans to continue their work with the OLCF to investigate vertical takeoff and landing and short takeoff and landing vehicle designs that leverage Whisper's propulsors.

Funding

This research was supported by the DOE Office of Science's ASCR program.

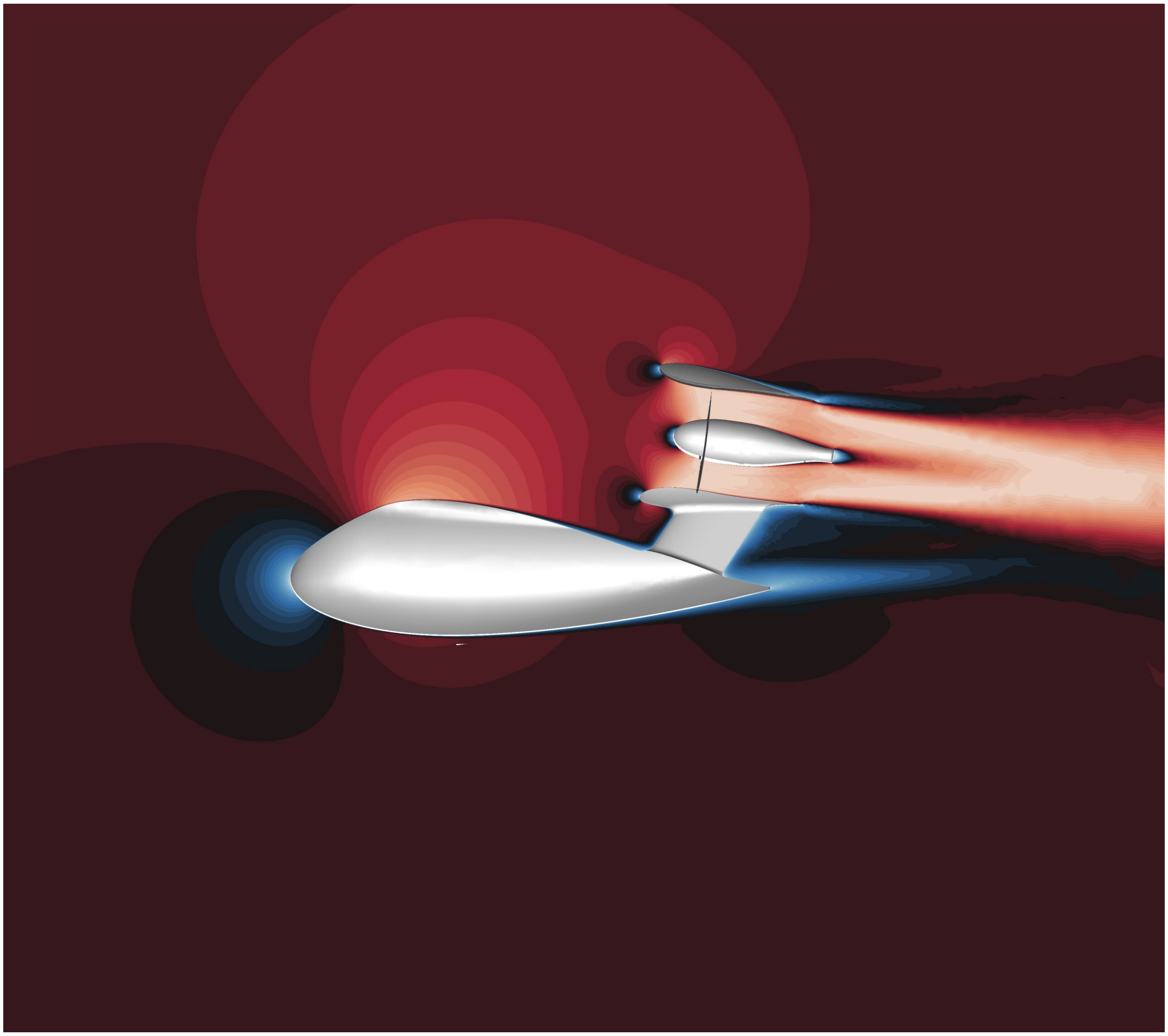


Figure 8-6. Whisper Aero conducted computational fluid dynamics simulations on Summit of their electric drone design, aeropropulsive integration, and airframe design and analysis (Image Credit: Whisper Aero). Publication: Jedamski, D., et al. "Distributed Electric Propulsion and Vehicle Integration with Ducted Fans," AIAA AVIATION Forum (2023): https://doi.org/10.2514/6.2023-3455.

Related Link: "Flying Quieter and Cleaner," OLCF News (August 12, 2024)

8.1.2.7 Frontier Simulations Could Help Build a Better Diamond

Supercomputer study guides effort to create otherworldly BC8 on Earth

PI: Ivan Oleynik, University of South Florida

Allocation Program: ALCC

The Science

Diamonds form when extreme heat and pressure pack carbon atoms together to create the hardest material on this planet. The precious stones not only decorate fine jewelry but perform service in the most demanding industrial jobs worldwide. Mining crews pierce bedrock with diamond-tipped drill bits. Engravers use diamond knives and saws to make incisions in quartz and granite. Scientists have theorized an even harder substance exists on other planets—a super diamond known as BC8, made up of eight carbon atoms for a diamond's every four. Extreme pressures and temperatures at the core of an exoplanet twice Earth's size or larger could create the necessary conditions to produce such materials. Synthesizing BC8 under laboratory conditions could open a new vista of industrial possibilities, courtesy of a material harder than the hardest thing in nature. The effort would require 10 million times the pressure of Earth's atmosphere and temperatures equal to those on the sun's surface. Various experiments have tried to create BC8 and failed, all at great expense. To predict what it might take to synthesize this substance, researchers would need a highly accurate way to simulate these complex interactions between carbon atoms in a billion-atom sample under a variety of conditions.

The Impact

A team from the University of South Florida used Frontier to simulate synthesizing a material harder and tougher than diamond—or any other substance on Earth (see Figure 8-7). The study used Frontier to predict the likeliest strategy to synthesize such a material, thought to exist only within the interiors of giant exoplanets, or planets beyond the solar system—at least so far. Frontier's exascale speeds of more than 1 quintillion calculations per second brought the goal within reach for the first time. None of the traditional classical interatomic or quantum models could offer that kind of detail at scale. To achieve the required level of accuracy, the team trained a novel ML interatomic model by using an extensive cache of quantum-mechanical data on various states of carbon, including BC8.

Summary

A University of South Florida team used Frontier to predict the likeliest strategy to synthesize a super-hard, diamond-like material thought to exist only within the interiors of giant exoplanets or planets beyond the solar system. Diamonds are used in some of the most demanding industrial tasks in the world, and significant demand exists for this class of materials. None of the traditional classical interatomic or quantum models could offer the needed detail at scale, and Frontier's exascale speeds brought the goal within reach for the first time.

Funding

This research was supported by the DOE Office of Science's ASCR program.

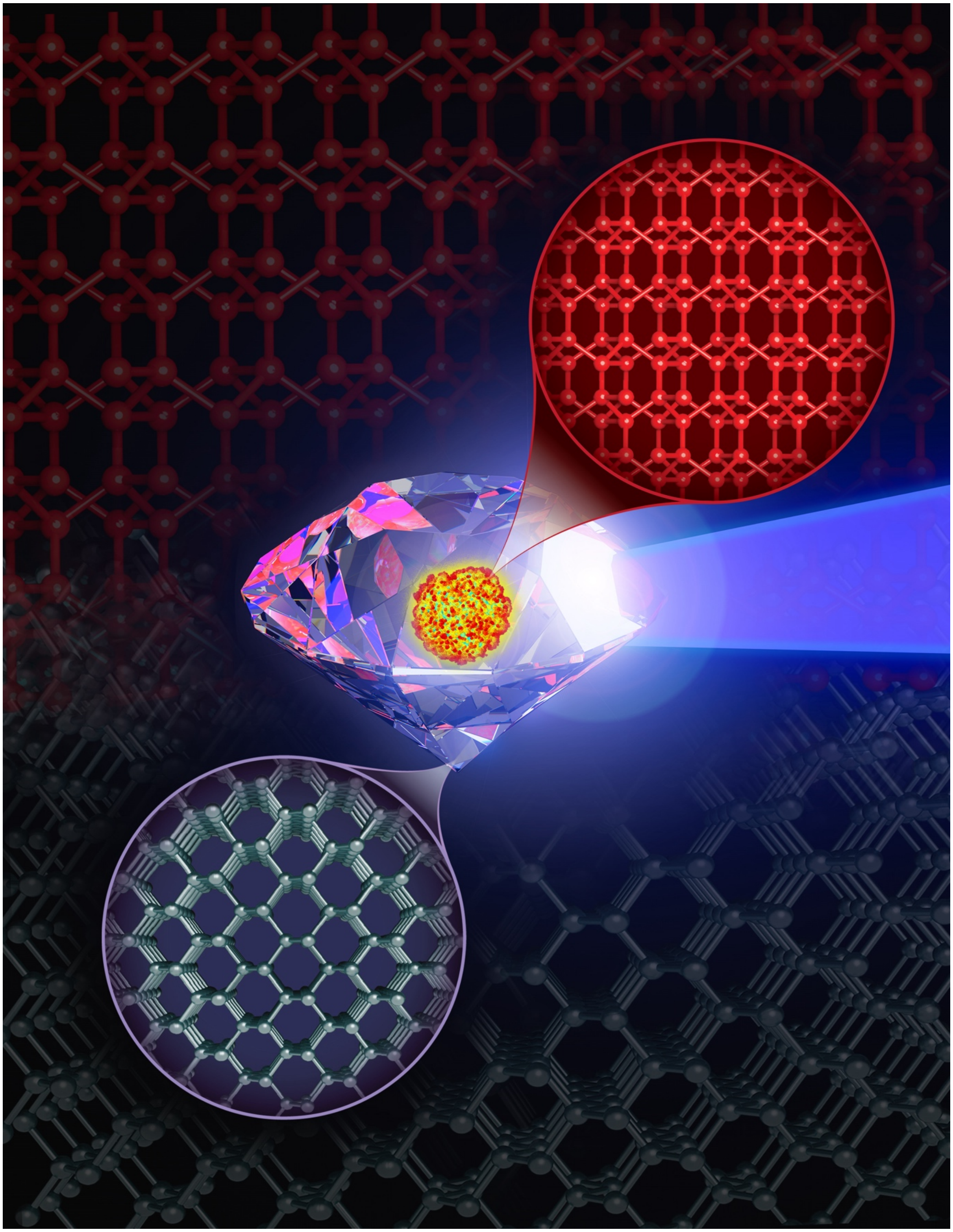


Figure 8-7. Researchers used ORNL's Frontier supercomputer to simulate synthesizing BC8, a material harder and tougher than diamond. BC8 could offer a wide range of potential uses for industry and other applications (Credit: Mark Meamber, LLNL).

Publication: Kien Nguyen-Cong et al., "Extreme Metastability of Diamond and its Transformation to the BC8 Post-Diamond Phase of Carbon," *Journal of Physical Chemistry Letters* 15 (2024): https://doi.org/10.1021/acs.jpclett.3c03044.

Related Link: "Frontier Simulations Could Help Build a Better Diamond," OLCF News (July 26, 2024)

8.1.2.8 New Clues to Improving Fusion Confinement

Using ORNL's Frontier exascale supercomputer, researchers discover an unexpected feature of multiscale turbulence in the edge of a fusion reactor's plasma

PI: Emily Belli, General Atomics

Allocation Program: INCITE

The Science

Scientists have developed sophisticated techniques to improve the confinement of superhot plasma within tokamak fusion reactors. Confinement in tokamaks is limited by the small but steady heat loss caused by turbulence. One important method to improve tokamak confinement is a byproduct of the commonly used neutral beam injection system, which uses an intense particle beam to heat the plasma to 150 million degrees Celsius to induce fusion. In addition to heating the plasma, this beam also causes the plasma to rotate around the tokamak's chamber. Previous studies had indicated that this rotation would reduce the turbulence caused by ions. It was also theorized that the rotation would not affect the electrons because they are so light and fast. The results of a new study on Frontier contradict that theory.

The Impact

The researchers found that with the interaction between ions and electrons, the rotation can increase the overall level of turbulence and subsequently reduce the quality of confinement (see Figure 8-8). Although the ions' turbulence is favorably reduced, the total turbulence can increase, which is unfavorable for a reactor. The team's findings could help to optimize the design of future tokamaks, including ITER, the international collaboration in France to build the world's largest tokamak.

Summary

A research team from General Atomics used Frontier to study the plasma edge region close to the tokamak's outer walls in a fusion reactor. This region is important because it sets the global confinement of energy and heat in the plasma. But calculating the turbulence in this edge region is very difficult. To do so, researchers must capture the subtle interaction between very heavy ions and very light electrons—an interaction that is frequently ignored. Previous simulations in this edge region were limited computationally and could only have the ion physics or only have the electron physics. They could not describe the complex interplay between the two. This became possible for with Frontier. The team's findings could help to optimize the design of future tokamaks, including ITER.

Funding

This research was supported by the DOE Office of Science's ASCR program.

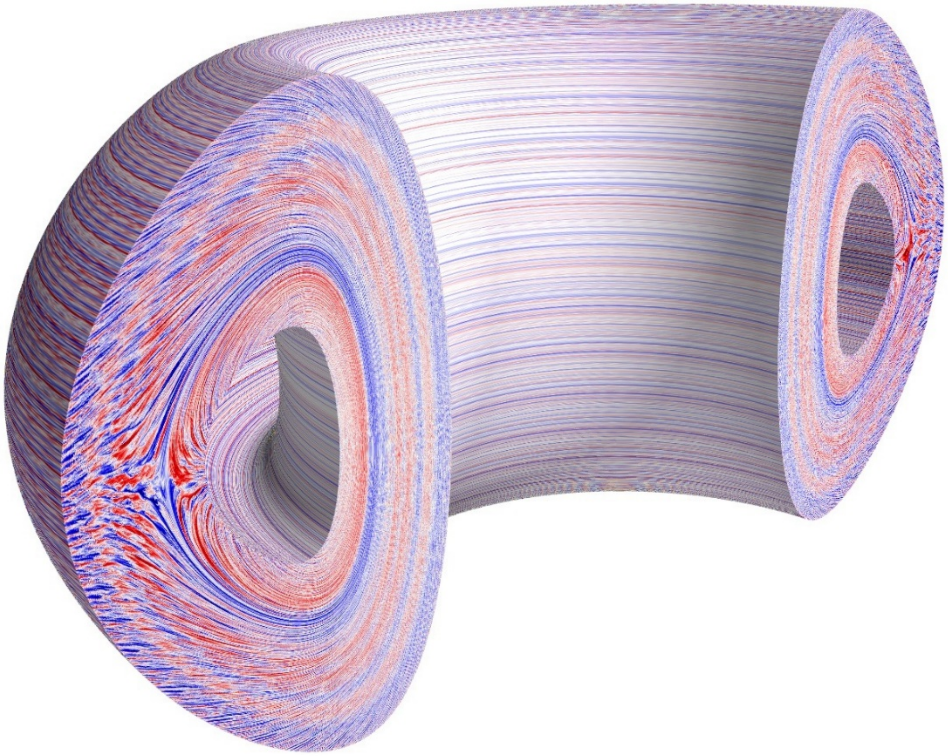


Figure 8-8. Researchers used the CGYRO gyrokinetic code to create a multiscale simulation of plasma temperature fluctuations driven by turbulence. The model was based on data from DOE's DIII-D National Fusion Facility for an "ITER baseline" tokamak scenario (Image: Emily Belli, General Atomics).

Publication: E. A. Belli, J. Candy, and I. Sfiligoi. "Flow-shear destabilization of multiscale electron turbulence," *Plasma Physics and Controlled Fusion* (2024): https://doi.org/10.1088/1361-6587/ad2c28.

Related Link: "New Clues to Improving Fusion Confinement," OLCF News (June 24, 2024)

8.1.2.9 Steering Toward Quantum Simulation at Scale

Study models key quantum state at unprecedented size

PI: Andrew Potter, University of British Columbia

Allocation Program: QCUP

The Science

A challenge in QC is building capabilities on a quantum computer to solve problems that are difficult to calculate on conventional computers. Researchers are working to steer a quantum system toward a particular state while competing with the quantum fluctuations away from this state. There is a transition point at which these competing effects exactly balance. That point separates a phase where the steering succeeds and where it fails. The farther the system moves out of equilibrium, the more likely classical versions of the model will break down because of the size and complexity of the equations. A University of British Columbia–led research team sought to use QC to model those dynamics.

The Impact

A University of British Columbia–led team used the Quantinuum H1-1 quantum computer to model a quantum version of a classical mathematical model that tracks how a disease spreads. The model used qubits to simulate the transition between active states, such as infection, and inactive states, such as death or recovery. The team used the quantum processor to simulate a system in which active qubits can activate neighboring qubits or become inactive. By monitoring the system in real time at each step and testing as they went, the team could detect the likelihood that performing a quantum gate on a qubit could affect the state of a qubit and, if not, remove it from the calculation. This lowers the likelihood that errors creep in.

Summary

A University of British Columbia–led team used the Quantinuum H1-1 quantum computer to model a quantum version of a classical mathematical model that tracks how a disease spreads (see Figure 8-9). The goal of this study was to work toward building capabilities on a quantum computer to solve this problem and others like it that are difficult to calculate on classical computers.

Funding

This research was supported by the DOE Office of Science's ASCR program.

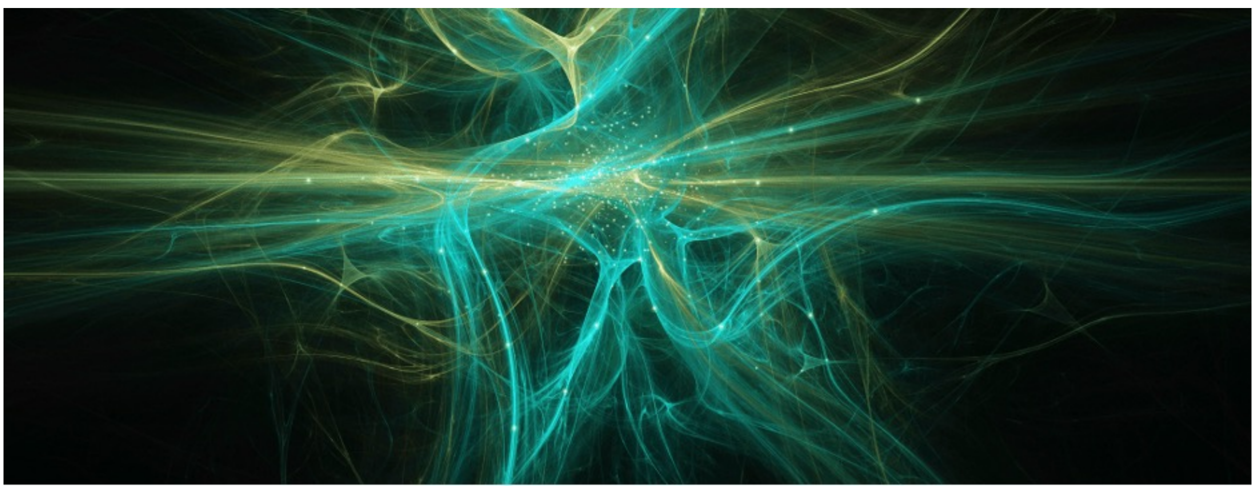


Figure 8-9. Researchers relied on support from ORNL's QCUP to simulate a key quantum state at one of the largest scales reported. The findings could mark a step toward improving quantum simulations. (Credit: Getty Images).

Publication: Chertkov, E., Z. Cheng, and A. C. Potter et al., "Characterizing a non-equilibrium phase transition on a quantum computer," *Nat. Phys.* 19 (2023): 1799–1804, https://doi.org/10.1038/s41567-023-02199-w.

Related Link: "Steering Toward Quantum Simulation at Scale," OLCF News (April 22, 2024)

8.1.2.10 Planning for a Smooth Landing on Mars

Work toward a potential crewed NASA mission to Mars was launched on the OLCF's Summit supercomputer in 2019 and flew to new heights on Frontier in 2023

PI: Eric Nielsen, NASA Langley

Allocation Program: INCITE

The Science

Since 2019, a team of NASA scientists and their partners has been using NASA's FUN3D software on OLCF supercomputers to conduct computational fluid dynamics (CFD) simulations of a human-scale Mars lander. Unlike in recent Mars missions, parachutes are not considered feasible. Mars's thin atmosphere—about 100× less dense than Earth's—will not support a parachute landing for such a large craft. Instead, the leading candidate for landing humans on Mars is using retropropulsion to decelerate—firing forward-facing rockets built into the craft's heat shield. Because the physics of such a landing cannot be replicated on Earth, high-fidelity CFD simulations are required to predict the controllability of the landing craft.

The Impact

With guidance from NASA mission planners, the team formulated a multiyear plan consisting of increasingly sophisticated simulations. In 2019, on Summit, the team was able to characterize static vehicle aerodynamics at anticipated throttle settings and flight speeds ranging from Mach 2.5 down to Mach 0.8. After optimizing its code for GPUs in 2020, the team's 2021 campaign addressed the complex interactions of the liquid oxygen/methane rocket engines with the Martian atmosphere. In 2023, Frontier enabled the team to successfully simulate 35 seconds of controlled flight, descending from 8 km altitude to about 1 km (see Figure 8-10).

Summary

A NASA-led team is using Frontier, following years of working on Summit, to conduct CFD simulations of a human-scale Mars lander. Unlike recent Mars missions, parachutes are not considered feasible. Mars's thin atmosphere—about 100× less dense than Earth's—won't support a parachute landing for such a large craft. Instead, the leading candidate for landing humans on Mars is using retropropulsion to decelerate—firing forward-facing rockets built into the craft's heat shield. Frontier enabled the team to successfully simulate 35 seconds of controlled flight, descending from 8 km of altitude to about 1 km.

Funding

This research was supported by the DOE Office of Science's ASCR program.

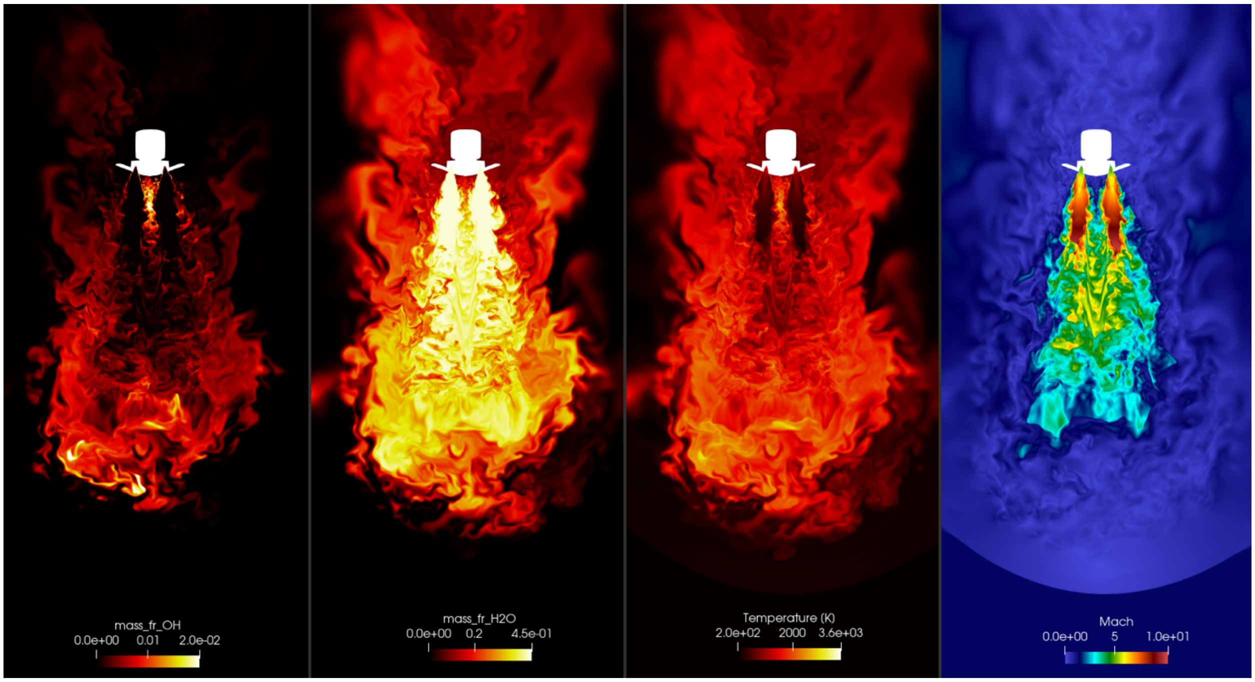


Figure 8-10. Instantaneous solution quantities shown for a static Mach 1.4 solution on a mesh consisting of 33 billion elements using 33,880 GPUs or 90% of Frontier. From left to right: contours show the mass fractions of the hydroxyl radical and H2O, the temperature in Kelvin, and the local Mach number (Image: Gabriel Nastac/NASA).

Publication: Carlson, J., et al. "High-Fidelity Simulations of Human-Scale Mars Lander Descent Trajectories," AIAA AVIATION Forum (2023): https://arc.aiaa.org/doi/10.2514/6.2023-3693.

Related Link: "Planning for a Smooth Landing on Mars," OLCF News (February 29, 2024)

8.2 RESEARCH ACTIVITIES/VENDOR ENGAGEMENT FOR FUTURE OPERATIONS

8.2.1 Advanced Computing Ecosystem

Computing systems and the associated architectures required to support emerging applications and interconnected/integrated workflows are becoming more complex. To better understand these new technologies, evaluate their pros and cons, assess their operational impacts and capabilities, and develop tools to mitigate missing capabilities, the OLCF has aggregated related efforts under the ACE umbrella. ACE is designed to help the OLCF with the upcoming OLCF-6 project and with DOE's IRI program. ACE consists of four fronts: test beds, technology evaluations, R&D of foundational technologies, and IRI science pilots and workflows.

8.2.2 Quantum HPC Integration

During 2024, the OLCF made significant progress in advancing the integration of QC technologies into classical HPC infrastructure. At ORNL and centered in the Computing and Computational Sciences Directorate (CCSD), staff members are coordinating activities involving 3 organizations: the Quantum Science Center, the Quantum Information Sciences section of CSED, and the NCCS/OLCF. At the OLCF, staff members have focused efforts on steps to integrate quantum computers into the leadership HPC resources. The goal is to be a technology hub for the hardware development, integration software, and alternatives analysis that will prepare the OLCF for future procurements that optimally address mission needs. The work is closely coupled with the development of efficient computational methods and software for tackling domain science problems in chemistry, materials, optimization, AI/ML, logistics, CFD, and national security.

Building on efforts begun in a working group with a goal to develop a roadmap for quantum integration (initiated in 2023 and led by Tom Beck, section head of Science Engagement and Sarp Oral, section head of Advanced Technologies in NCCS), a team of roughly 20 researchers across CCSD published a paper outlining a strategy that addresses hardware, software, workflows, user programs, benchmarking, applications, and a general software framework for the integration. This publication (see below) was solicited as part of a special issue on QC in *Future Generation Computer Systems*.

Building on the publication, several members of the team successfully competed for ORNL LDRD funding that is supporting three themes: (1) the quantum integration framework, (2) application development that applies QC to computational chemistry and materials problems (molten salt clusters), and (3) quantum networking. In addition, NCCS discussions were initiated to form a new Quantum-HPC group within the Science Engagement section, and that group was formed at the beginning of January 2025. The group focuses on the intersection of quantum hardware/software integration and science applications, where the most intense action in the rapidly developing QC field will likely occur. The team is currently developing the integration software that will create, to the extent possible, a hardware-agnostic framework that will facilitate a test bed environment for analyzing hardware options. An installation of prototype Quantum Brilliance hardware (nitrogen vacancy qubits in diamond) is planned for late spring or early summer. This hardware will enable testing of the developing framework and some initial quantum-driven simulations.

Publication

Thomas Beck, Alessandro Baroni, Ryan Bennink, Gilles Buchs, Eduardo Antonio Coello Pérez, Markus Eisenbach, Rafael Ferreira da Silva, Muralikrishnan Gopalakrishnan Meena, Kalyan Gottiparthi, Peter Groszkowski, Travis S. Humble, Ryan Landfield, Ketan Maheshwari, Sarp Oral, Michael A. Sandoval, Amir Shehata, In-Saeng Suh, and Christopher Zimmer. “Integrating quantum computing resources into scientific HPC ecosystems,” *Future Generation Computer Systems* 161 (2024): 11–25.

ORNL LDRD

Software Framework for Integrated High Performance + Quantum Computing (PI Daniel Chaves Claudino and co-PI Tom Beck).

8.2.3 Integrated Research Infrastructure–Related Activities

8.2.3.1 Test Beds

The ACE test bed is a unique OLCF capability that provides a centralized, sandboxed area for deploying heterogeneous computing and data resources and facilitating the evaluation of diverse workloads across a broad spectrum of system architectures. ACE is designed to fuel productization of new HPC technologies as applicable to the OLCF and DOE missions. ACE is an open-access environment consisting of HPC production-capable resources, and it allows researchers and HPC system architects to assess existing and emerging technologies more freely without the limitations of a production environment. Topics of interest include the following:

- IRI workflows and patterns (i.e., time-sensitive, data-intensive)
- Emerging compute architectures and techniques for HPC (e.g., Arm, AI appliances, reconfigurable)

- Emerging storage architectures and techniques for HPC (e.g., object storage)
- Emerging network architectures and techniques (e.g., data processing units)
- Cloudification of traditional HPC architectures (e.g., multitenancy, preemptible queues)

The ACE test bed currently has the following computer resources:

- Defiant
  - 36 nodes: AMD Epyc CPU, 4 AMD MI100 GPUs, Slingshot 10 networking
  - Former Frontier early access system
- Holly
  - Single Supermicro server with 8 NVIDIA H100 GPUs

The ACE test bed currently has the following storage resources deployed with close engagement with vendors:

(1) Polis – Lustre
  - ~1.6 PB
  - Primarily spinning disk with some flash, connected to Defiant

(2) VastData – NFS
  - ~600 TB
  - NFS-over-RDMA storage appliance
  - Flash, connected to the IB fabric

(3) DAOS – Object Storage
  - 8 servers with ~30 TB flash each and IB connections (currently being upgraded from HDR to NDR)

The ACE test bed also contains the OLCF's Wombat cluster, which was established in 2018 to evaluate the maturity and compatibility of Arm's AArch64 architecture for scientific HPC applications. Over time, the test bed has been updated as needed. Latest update included 8 Hopper nodes.

Additional details on ACE, supported foundational technology developments, and collaboration projects are available in the OLCF technical report summarizing FY 2024 ACE activities supporting IRI (https://info.ornl.gov/sites/publications/Files/Pub223426.pdf).

8.2.3.2 NCCS Advanced Computing Ecosystem Report

An end-to-end streaming demo has been designed and prototyped using a series of software and hardware components. The data source is managed by https://github.com/frobnitzem/lclstream, an API for starting, stopping, and listing stream-push operations on the SLAC side. A flexible network topology for streaming is provided by a stackable message cache. The cache can operate at SLAC, the OLCF, or both. It has been containerized and is ready to be run within the ACE streaming infrastructure. On the compute side, an AI dataset storage mechanism has been implemented to store the data for AI inference with https://code.ornl.gov/99R/local_sampler. This sampler fits within the PyTorch ecosystem and is compute-efficient because it only samples from node-local data. An API provides the job control functionality necessary for launching stream-processing associated compute jobs on Summit and Defiant. Finally, the full workflow has been tied together with launch scripts (https://code.ornl.gov/99R/streamrun) that document the install, start, and run phases of the project.

The OLCF center highlight (linked to an initial collaborative paper) is available at https://www.olcf.ornl.gov/2024/05/06/ornl-and-slac-team-up-for-breakthrough-biology-projects/, (https://www.sciencedirect.com/science/article/pii/S0959440X24000356).

Quick list of software output:

- https://github.com/lcls-users/lclstream
- https://gitlab.com/frobnitzem/nng_stream
- https://github.com/frobnitzem/stream.py
- https://github.com/ORNL/certified
- https://frobnitzem.github.io/psik_api/

8.2.3.3 Data Streaming to HPC

Scientific workflows are being developed to concurrently couple experimental science facilities with HPC facilities to enable analysis of observational data while the experiment is ongoing and where analysis results are potentially fed back to the experiment in a time-sensitive manner for use in control decisions such as instrument tuning or to steer the experiment. These scientific workflows require foundational technologies that enable bidirectional streaming of data between the experimental science facilities and the HPC systems located within the computational facilities.

In early FY 2024, the OLCF formed a working group that incorporated both researchers and operational staff to design a new OLCF capability for Data Streaming to HPC (DS2HPC). The product of this working group was an architectural approach to supporting DS2HPC in a flexible way while supporting the deployment of many technological solutions. This approach was informed by several science use cases with different data streaming workload characteristics. The motivating use cases, architecture, and requirements are documented in an OLCF technical report published in March 2024 (https://www.osti.gov/biblio/2338264).

In the latter half of FY 2024, the working group focused on prototyping the DS2HPC capability within the ACE test bed. A full demonstration was realized by the end of FY 2024. This demo utilized a commodity data streaming framework (i.e., RabbitMQ) deployed on allocatable data streaming nodes via the Data Streaming Orchestration API and associated service of the OLCF Secure Scientific Service Mesh (S3M). Additionally, a new DS2HPC benchmark was developed to exercise and measure the performance of deployed streaming services for both latency-sensitive and throughput-oriented DS2HPC workflows.

8.2.3.4 LCLStream Collaboration (OLCF and SLAC)

The OLCF and SLAC are closely collaborating on the development, testing, and deployment of new technologies that will be critical for expanding LCLS-II (Linac Coherent Light Source II) analysis capabilities and fully leveraging its high resolution, high-rate data. Initial work focused on performance improvements for data loading and GPU-accelerated inference training for image data, resulting in multiple code improvements to LCLS's Masked X-Ray Image Encoder. Led by David Rogers in the Science Engagement Section, a data streaming pipeline was built and demonstrated at 24 Gbps data streaming from SLAC's S3DF to the OLCF's ACE test bed (https://gitlab.com/frobnitzem/nng_stream) with workflow control secured by mutual public-key authentication (https://github.com/ORNL/certified). A survey of useful technology building blocks supporting this work was published as a guide to the Integrated Research Infrastructure (IRI) community (https://www.osti.gov/biblio/2498436). Efforts are currently underway to deploy elements from the LCLStream pipeline into the standing infrastructure of both S3DF and NCCS, including potential pathways toward seamless data streaming to compute. Progress from this collaboration will increase the scientific user community's access to high-speed data analysis at ASCR compute facilities.

8.2.3.5 Gamma-Ray Energy Tracking Array/Deleria

Collaborating with a team led by Dr. Mario Cromaz from Lawrence Berkeley National Laboratory (LBNL) and ESnet, Dr. Jansen in the Science Engagement Section has developed the software that powers the data pipeline of GRETA (Gamma-Ray Energy Tracking Array), a cutting-edge gamma-ray detector nearing completion for the DOE Office of Science, Nuclear Physics. GRETA is the flagship instrument at the Facility for Rare Ion Beams at Michigan State University, where its mission is to advance the understanding of nuclear structure and reactions, making this development crucial for furthering the understanding of matter in the universe. The Deleria (Distributed Event-Level Experiment Readout and Integrated Analysis) software is unique because it can operate in a distributed environment, thereby allowing the detector and computing resources to be located at different facilities. It is precisely this capability that the Deleria team, with the help of a dedicated team at the OLCF led by Ross Miller, demonstrated last year when they successfully ran a distributed experimental workflow between the ESnet test bed at LBNL and the ACE test bed at the OLCF. The demonstration ran for several minutes and saturated the assigned available bandwidth between ESnet and the OLCF at 40 Gbps, which is more than the rate at which GRETA will produce data at full capacity. The GRETA detector (see Figure 8-11) can support a much higher event rate than the current data pipeline is designed for. A potential future upgrade calls for a completely new computing cluster and online storage and backup to be installed and maintained next to the detector. Deleria offers an alternative to this model in that computing and storage can be available on-demand from one or several of DOE's HPC and HPD facilities, providing flexibility and scalability for future needs.

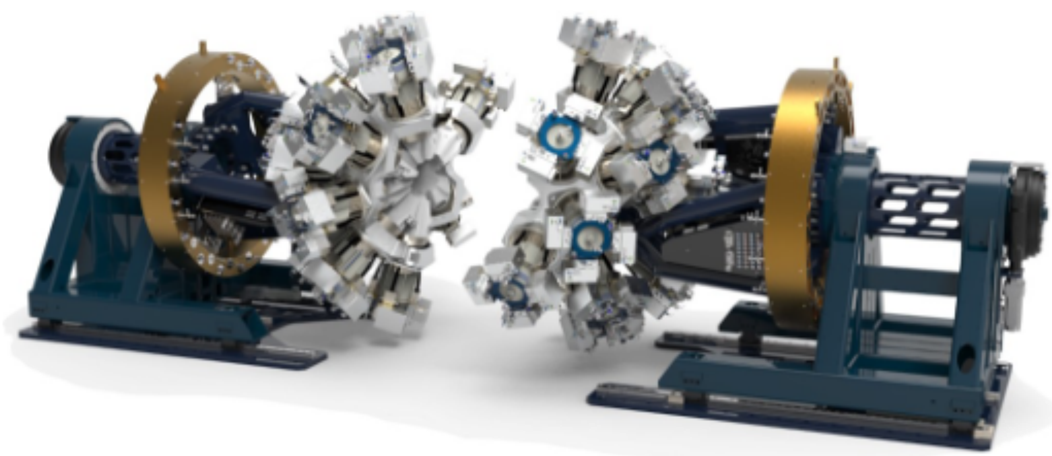


Figure 8-11. The GRETA/Deleria detector.

8.2.3.6 Integrating ORNL's HPC and Neutron Facilities with a Performance-Portable CPU/GPU Ecosystem: Experimental Validation of DOE's Integrated Research Infrastructure Program Using the OLCF's Advanced Computing Ecosystem Test Bed

The "Integrating ORNL's HPC and Neutron Facilities with a Performance-Portable CPU/GPU Ecosystem" (http://dx.doi.org/10.1109/SCW63240.2024.00264) research paper, which won the best paper award at the 6th Annual Workshop on Extreme-Scale Experiment-in-the-Loop Computing (XLOOP24 held in conjunction with SC24), demonstrates a groundbreaking approach to integrating two of DOE's premier scientific facilities: the OLCF and the SNS. The paper presents a performance-portable CPU/GPU ecosystem that uses the Julia programming language and the JACC.jl package to modernize neutron scattering data reduction workflows. The proposed implementation achieved significant speedups—up to 299× on GPUs compared to current production code—while maintaining portability across different hardware architectures, including NVIDIA and AMD GPUs.

The work directly advances DOE's IRI program goals by creating seamless integration between experimental and computational facilities. By enabling efficient processing of neutron scattering data from SNS instruments on the OLCF's supercomputing resources, this research provides a blueprint for future facility integration efforts. The paper's approach to performance portability through Julia and LLVM demonstrates how modern scientific software can bridge the gap between experimental data acquisition and HPC resources. This success story showcases how DOE facilities can work together to accelerate scientific discovery through improved computational workflows.

8.2.3.7 Enabling Low-Overhead HT-HPC Workflows at Extreme Scale using GNU Parallel

GNU Parallel enables highly efficient high-throughput computing workflows at extreme scale on supercomputers (Frontier and Perlmutter), maintains remarkably low overhead even when managing over a million concurrent tasks, and significantly simplifies workflow management compared to traditional approaches.

The "Enabling Low-Overhead HT-HPC Workflows at Extreme Scale using GNU Parallel" research paper (http://dx.doi.org/10.1109/SCW63240.2024.00257) demonstrates GNU Parallel's effectiveness as a workflow management tool for high-throughput computing at extreme scale on the OLCF's Frontier supercomputer. In this work, researchers conducted comprehensive benchmarking that showed GNU Parallel's ability to scale efficiently up to 9,000 nodes (96% of Frontier), managing over 1.15 million concurrent tasks with significantly lower overhead compared to traditional workflow systems. Key achievements include optimization of CPU/GPU utilization, efficient use of node-local NVMe storage, and innovative data motion strategies that achieved 200× speedup over sequential transfers.

This work exemplifies strong cross-laboratory collaboration, with researchers from the OLCF and NERSC validating GNU Parallel's capabilities across multiple DOE computing facilities, including the OLCF's Frontier and NERSC's Perlmutter. The team demonstrated GNU Parallel's versatility through several real-world applications, including the FORGE LLM development, Celeritas GPU-accelerated particle transport simulations, and massive Darshan log processing workflows.

8.2.3.8 Smoky Mountain Conference

At the 2024 Smoky Mountains Computational Sciences and Engineering Conference hosted by ORNL in September 2024, the Advanced Technologies Section of the NCCS organized a session of introductory talks and demonstrations to highlight OLCF efforts to develop and deploy capabilities useful for enabling cross-facility scientific user facilities, as envisioned by the DOE's IRI program. This session consisted of two topic areas—one focused on technologies that support IRI goals and another focused on IRI application use cases. OLCF personnel gave all but one of the technology-focused presentations, with the other presentation covering IRI technologies emerging from the ORNL Interconnected Science Ecosystem LDRD initiative, which also includes contributions from several OLCF staff.

In particular, three key OLCF ACE technologies were highlighted: (1) the ACE test bed, which provides computing and storage resources and supports collaborative development of IRI science workflows; (2) the S3M, which provides user-facing APIs and integrated secure services for utilizing ACE test bed resources; and (3) the DS2HPC capability (https://www.osti.gov/biblio/2338264), which leverages both S3M and the ACE test bed to enable bidirectional memory-to-memory data streaming between experiments running at external science facilities and applications running within OLCF's HPC systems.

Two of the four application-oriented presentations focused on OLCF collaborations with external science facilities to develop new workflows that use the ACE test bed. The GRETA/Deleria project is a collaboration with LBNL and ESnet to stream detector event data from the GRETA instrument at the

Facility for Rare Isotope Beams to the OLCF's Defiant HPC system for online analysis. The LCLStream project is a collaboration with the SLAC National Accelerator Laboratory and Stanford University to utilize the OLCF's Defiant for training an ML model using data streamed from the Linac Coherent Light Source.

As part of the Smoky Mountain Conference in 2024, Feiyi Wang and Christopher Zimmer from the OLCF chaired a session called "Exascale Accomplished! What's Next? The Challenges of Deploying HPC Systems in a Radically Diverging Post-Exascale World." This session hosted the following speakers:

- Glenn Lockwood (Microsoft) presented on large-scale cloud deployments that target large AI foundation models.
- Kate Clark (NVIDIA) presented on the use of mixed precision in QCD applications to improve performance and reduce energy consumption.
- Hans-Christian Hoppe (Jülich) presented on harnessing heterogeneity in the exascale era.
- Brad Beckman (AMD) presented on the use of gem5 for defining advancements in next-generation processors.

8.2.3.9 Workflows Community Summit 2024

The 2024 Workflows Community Summit was an international gathering that brought together 111 experts from 18 countries to discuss future trends and challenges in scientific workflows. The summit focused on six key areas: (1) time-sensitive workflows, (2) convergence of AI and HPC workflows, (3) multifacility workflows, (4) heterogeneous HPC environments, (5) user experience and interfaces, and (6) FAIR (findability, accessibility, interoperability, and reusability) computational workflows.

The summit produced critical recommendations (http://dx.doi.org/10.5281/zenodo.13844759), including developing standardized cross-facility authentication protocols, investing in AI-driven workflow optimization, implementing distributed-by-design workflow models, creating comprehensive workflow benchmarks, formulating workflow-specific UX principles, and establishing a FAIR workflow maturity model.

The summit highlighted significant shifts in scientific computing, particularly the integration of AI with exascale computing systems and the growing need for workflows that span multiple facilities. Key findings emphasized the importance of managing heterogeneous environments, addressing data movement challenges, and ensuring workflow reproducibility across diverse computing infrastructures. The discussions led to proposals for developing frameworks for cloud-HPC integration, establishing multifacility authentication protocols, and accelerating AI integration in workflow management.

As the executive director of the Workflows Community and summit convener, Rafael Ferreira da Silva's leadership role at the OLCF positions the laboratory at the forefront of these developments. The summit's outcomes directly align with the OLCF's mission, particularly in areas such as AI-HPC integration and managing heterogeneous computing environments. The recommendations regarding standardized protocols and workflow benchmarks provide the OLCF with strategic guidance for enhancing its workflow capabilities and supporting next-generation scientific applications. This alignment between community-driven initiatives and the OLCF's strategic direction strengthens the laboratory's role in shaping the future of scientific workflows.

8.2.4 Large-Scale Network Scaling and Performance Evaluation

In the spring of 2024, the OLCF undertook a comparative study that evaluated the design differences between two generations of supercomputer network architectures. The purpose of this evaluation was to measure the impact of cost optimization strategies when designing a network topology. Broadly, operators of capability-class scientific and AI supercomputers have assumed that non-tapered fat-tree topologies provide superior performance but come with significant cost. Owing to this substantial cost, many computing centers with budgetary constraints have adopted less costly network topologies, such as the tapered dragonfly, that trade cost for reduced global network performance. In 2024, the OLCF was in a unique position of operating two supercomputers (Summit and Frontier), with one using the non-tapered fat-tree network topology (Summit) and the other using the tapered dragonfly topology (Frontier). The OLCF took this opportunity to conduct a comparative performance evaluation of both topologies, at scale, using actual workloads relevant to the OLCF's mission. The results of the study showed that the selection of a cost-optimized dragonfly network topology for Frontier imposed negligible impact on large-scale application performance while at the same time resulting in significant cost savings. The results of the evaluation were published at SC24 (Awais et al., "An Evaluation of the Effect of Network Cost Optimization for Leadership Class Supercomputers," SC24.), where it was a finalist for the best paper award.

In 2024, the OLCF also worked with HPE to develop simulation studies of various alternative network architectures based on the dragonfly topologies. These simulations were used to evaluate the ability of dragonfly-based networks to support leadership-class supercomputers that target applications capable of running across the entire machine. These simulations provided input for the design process and larger considerations used as part of the evaluation effort for the OLCF-6 procurement. The findings of this effort led to a better understanding of dragonfly topologies and the effect that tapered global bandwidth has on their performance at scale.

8.2.5 Building Conversational AI Assistant on Frontier

ChatHPC is an innovative pipeline designed to create conversational AI assistants optimized for deployment on Frontier (Figure 8-12). By leveraging Frontier's unparalleled computational power, chatHPC enables the efficient pre-training, fine-tuning, deployment, and evaluation of LLMs tailored for domain-specific applications. The pipeline incorporates cutting-edge technologies such as distributed fine-tuning with DeepSpeed ZeRO, self-improved instruction generation, and retrieval-augmented generation. These features facilitate rapid alignment of LLMs to specific scientific domains, ensuring they are both accurate and context aware.

For Frontier users, chatHPC represents a transformative tool for scientific exploration. By enabling conversational interaction with their domain-specific data, scientists can quickly query, interpret, and extract insights from complex datasets or simulation results. This reduces the time and expertise required for data analysis, allowing researchers to focus on discovery and innovation. Whether troubleshooting computational workflows, generating optimized scripts, or answering nuanced scientific questions, chatHPC acts as an intelligent assistant that accelerates research and fosters collaboration across disciplines.

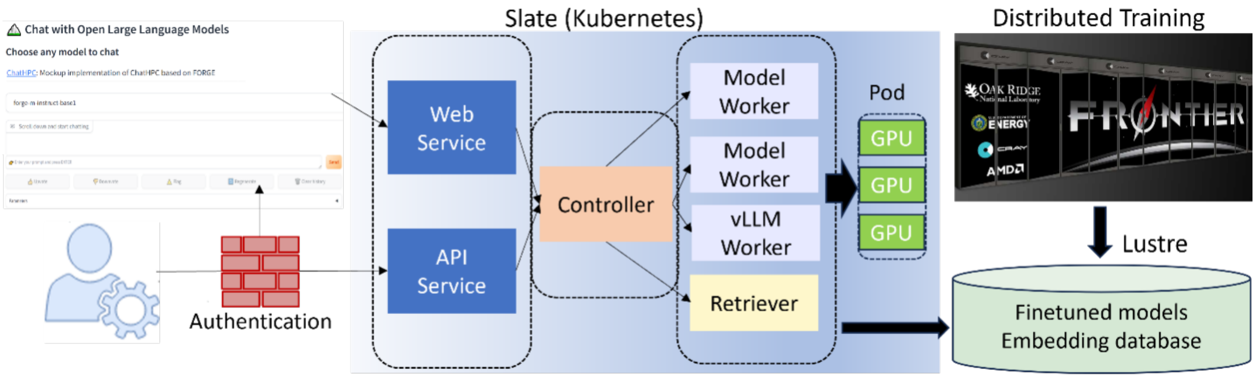


Figure 8-12. chatHPC architecture.

Related publication: Junqi Yin, Jesse Hines, Emily Herron, Tirthankar Ghosal, Hong Liu, Suzanne Prentice, Vanessa Lama, and Feiyi Wang, "chatHPC: Empowering HPC Users with Large Language Models," *Journal of Supercomputing* 81, no. 1 (2025): 194.

8.2.6 OLCF-6 AI and Workflow Benchmarks

8.2.6.1 OLCF-6 AI Benchmarks

Because AI workloads are becoming an increasingly common task for leadership-computing platforms, the OLCF incorporated an AI-specific benchmark, FORGE, into the OLCF-6 procurement process as part of the OLCF-6 request for proposals. The FORGE benchmark is designed to stress test system capabilities and emphasizes low-precision compute, communication bandwidth, and energy efficiency in training LLMs on scientific datasets. FORGE is based on mainstream frameworks that support distributed training across data, tensor, and pipeline parallelisms. The benchmark measures system performance with metrics such as teraflops, scaling efficiency, and energy efficiency (teraflops per Watt), ensuring high energy efficiency and computational performance. FORGE also provides loss tracking and time-to-solution metrics. The primary figure of merit combines time-to-solution and energy efficiency, while metrics such as scaling efficiency support comprehensive system evaluation.

FORGE stresses compute systems with demanding low-precision calculations in float16 or bfloat16 data types, requiring efficient communication patterns (e.g., AllReduce, AllGather, ReduceScatter) to manage message sizes 3× the model's parameters. Compliance with OLCF-6 benchmark run rules includes using specific tokenizers, optimizers, architectures, and training over 257 billion tokens. System performance is maximized through customized kernel optimizations and adjustable learning rates and batch sizes tailored to training scales. The benchmark ensures comprehensive evaluations of computational efficiency, scalability, and energy use in LLM training.

Related Document/Data

Yin, Junqi, Sajal Dash, Feiyi Wang, and Mallikarjun (Arjun) Shankar, *OLCF-6 FORGE RFP Benchmark*, Oak Ridge National Laboratory, Oak Ridge, TN (2024): https://www.olcf.ornl.gov/wp-content/uploads/OLCF-6_FORGE_description-1.pdf.

Yin, Junqi, Sajal Dash, Feiyi Wang, and Mallikarjun (Arjun) Shankar, *Tokenized Data for FORGE Foundation Models*, Oak Ridge National Laboratory, Oak Ridge, TN (2023): https://doi.org/10.13139/OLCF/2004951.

8.2.6.2 OLCF-6 RFP Workflow Benchmark

The OLCF-6 Workflow Benchmark evaluates HPC system capabilities in handling dynamic data stream workloads through ML4NSE applications. Using a TFT model, it analyzes neutron scattering data in real time, processing a nuclear peak and six satellite peaks from magnetic ordering in single-crystal samples. The system architecture channels multiple data streams through a gateway node using open-source streaming transport (ZeroMQ, RabbitMQ), with a control stream managing data multiplexing and filtering. The Data Orchestrator connects to both fast internal networks (RDMA) and external networks, managing data flow between compute nodes and consumers.

The benchmark includes weak scaling experiments (training TFT models on varying voxel configurations, requiring 9–4,608 ranks) and strong scaling experiments (fixed 32 × 32 × 32 input with variable level-1 mapping). Performance is measured primarily through time-to-solution, with secondary metrics including gateway node exchange bandwidth, stream multiplexing scalability, and control operation latency. A reference benchmark run demonstrated scaling capabilities up to 4,608 GPUs on Frontier, with comprehensive performance tracking for both weak and strong scaling efficiency.

Related Document/Data

Ferreira da Silva, Rafael, Patrick Widener, Junqi Yin, and Ketan Maheshwari, *OLCF-6 Workflow RFP Benchmark*, Oak Ridge National Laboratory, Oak Ridge, TN (2024): https://www.olcf.ornl.gov/wp-content/uploads/OLCF-6_Workflow_description-7.5.24.pdf.

8.2.7 ExaDigiT

ExaDigiT is a framework developed by a team at the OLCF for modeling digital twins of liquid-cooled supercomputers. The framework consists of three core modules: (1) the Resource Allocator and Power Simulator (RAPS); (2) a cooling system model that simulates the entire central energy plant, from cooling towers to distribution racks; and (3) a visual analytics module featuring augmented reality and a dashboard interface. ExaDigiT significantly lowers the barrier to developing digital twins for data centers.

The OLCF has used ExaDigiT to create a digital twin of Frontier, the first exascale supercomputer. The RAPS module allows for the replay or rescheduling of system telemetry data and for virtual prototyping. This capability has enabled the OLCF to explore potential improvements in energy efficiency, such as virtually evaluating cost benefits of running the system on medium voltage direct current or investigating smart-load sharing rectifiers that dynamically engage based on blade-level loads.

Additionally, ExaDigiT was employed for capacity planning in the OLCF-6 effort to evaluate whether the current central energy plant can support two exascale supercomputers. The results were published in the *Proceedings of the SC24 International Conference on High Performance Computing, Networking, Storage, and Analysis* (https://dl.acm.org/doi/10.1109/SC41406.2024.00029). Beyond supporting the OLCF operations, ExaDigiT has been open-sourced, and an international community has been formed to explore new use cases. Over the past year, the framework has been adopted by 10 supercomputing centers worldwide and continues to grow. Additionally, the OLCF continues to collaborate closely with vendors such as HPE and NVIDIA to further its development. A worldwide consortium is being assembled to steward further development, support, and deployment of ExaDigiT with participating supercomputing organizations.

8.2.8 National AI Research Resource Secures Pilot Collaborations with NIH/NCATS

As part of the National Secure Data Service (NSDS) Demonstration Project, this collaborative effort between the NIH and the OLCF is designed to guide both the NSDS and the National Artificial Intelligence Research Resource (NAIRR) pilot by developing tools and techniques that enable tiered access via synthetic data generation. The project seeks to deepen the understanding of how synthetic data generators interact with large real-world datasets within Frontier's secure supercomputer environment, CITADEL. The data generators are planned to be used as privacy-enhancing technologies within this context, and the primary source is the National Covid Cohort Collaborative data. These synthetic data are produced through advanced statistical and ML techniques, which create new records that retain the aggregate statistical characteristics of the original dataset while safeguarding privacy and confidentiality. By leveraging Frontier, the project will explore open-source, AI-driven methods for generating synthetic data. Moreover, employing synthetic data facilitates a tiered access approach to restricted datasets that might otherwise be difficult to access. This collaboration will contribute to the creation of a robust synthetic data generator toolkit that includes methods for assessing privacy risks, data utility, and the application of advanced AI techniques for synthetic data creation. Ultimately, the outcomes of this project will extend beyond these methods, offering further innovations and insights into the potential of synthetic data generation for handling confidential, large-scale, real-world data within Frontier's CITADEL environment.

Additionally, as part of the NAIRR secure pilot effort, the OLCF is working to deploy a robust and stable JupyterHub environment that is tightly integrated with the Frontier supercomputer. This will enable users to run Jupyter notebooks easily on Frontier and other OLCF compute resources. Once in production, a similar methodology will be used in the OLCF's moderate enhanced enclave to enable use of Jupyter notebooks with protected data within the OLCF's Scalable Protected Infrastructure. As part of this collaboration, the OLCF is also working with NIH/NCATS to deploy a retrieval augmented generation pipeline on Frontier as part of a query-response user workflow. This query-response agent was deployed on Frontier, and a live demonstration was given at SC24 in November 2024.

8.2.9 Science Engagement Research Activities

8.2.9.1 Biomolecular Simulations in the Exascale Era

With the birth of exascale computing, researchers are at the threshold of a new era of biomolecular simulations—an era that opens new avenues to tackle challenges of vital national interest in drug discovery, personalized medicine, and molecular biotechnology. The simultaneous growth in the science of big data and AI built upon ML significantly enhances the ability to discover informative patterns in both simulation and experimental datasets, allowing the synergistic use of two different streams of information that has hitherto been only loosely connected. A CECAM (Centre Européen de Calcul Atomique et Moléculaire) workshop helped crystallize these ideas into a publication [1] that presents a vision for biomolecular simulations in the exascale era (Figure 8-13). The eventual goal within the coming decade is to study a simple cell in all its molecular detail.

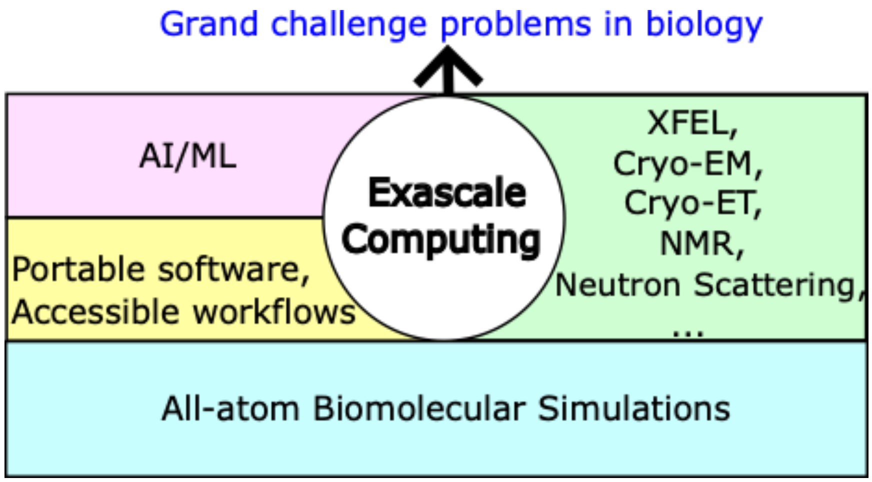


Figure 8-13. Vision for biomolecular simulations in the exascale era.

In the pre-exascale era, simulation studies were feasible for complexes such as the nuclear pore complex or the ribosome. However, these are still only parts of a cell. A successful system simulation on Frontier contained approximately 155 × 109 molecules (about 0.465 × 1012 atoms) of a realistic model of liquid water in a cube of size 1.67 microns, slightly larger than the average size of a common bacterium. With these results, researchers can credibly envision creating an in-silico twin of a minimal biological cell.

A game-changer on the experimental side is the arrival of serial femtosecond x-ray crystallography for investigating biomolecular structure and dynamics. With the new generation of x-ray free-electron lasers, which generate ultrabright x-ray pulses at megahertz repetition rates, ultrafast conformational changes and charge movement in biomolecules can be rapidly probed. Synergizing this development with exascale computing has the potential to accelerate discovery in biomolecular sciences and usher in a new era of dynamic structural biology. The vision and status for this IRI was communicated by members of SLAC and ORNL [3, 4] (below).

As large-scale simulations of biological systems are pursued, care must be taken in how those simulations are performed. For example, a key finding [5,6] showed that the common choices of time-step for integrating the equations of motion for simulating liquid water led to the incorrect partitioning of thermal energy between translational and rotational motion of the water molecule, a key requirement for correct sampling of all aqueous phase processes. This work cautions against a 47-year-old assumption that has held sway over the field of biomolecular simulations but also suggests simple steps that can ensure correct sampling.

Highlights and Publications:

(1) Beck, T. L., P. Carloni, and D. N. Asthagiri, "All-Atom Biomolecular Simulation in the Exascale Era," *J Chem Theory Comput* (2024).

(2) Lakin, M. "Breaking Benchmarks: Frontier Supercomputer Sets New Standard in Molecular Simulation," https://www.ornl.gov/news/breaking-benchmarks-frontier-supercomputer-sets-new-standard-molecular-simulation.

(3) Mous, S., F. Poitevin, M. S. Hunter, D. N. Asthagiri, and T. L. Beck, "Structural Biology in the Age of X-Ray Free-Electron Lasers and Exascale Computing," *Curr. Opin. Struct. Biol.* 86, 102808 (2024): https://doi.org/10.1016/j.sbi.2024.102808.

(4) Turczyn, C. “ORNL, SLAC Team up for Breakthrough Biology Projects,” https://www.ornl.gov/news/ornl-slac-team-breakthrough-biology-projects.

(5) Asthagiri, D. N. and T. L. Beck, “MD Simulation of Water Using a Rigid Body Description Requires a Small Time Step to Ensure Equipartition,” *J Chem Theory Comput* 20 (2024): 368–374, https://doi.org/10.1021/acs.jctc.3c01153.

(6) Turczyn, C. “Something in the Water Does Not Compute,” https://www.olcf.ornl.gov/2024/05/06/something-in-the-water-does-not-compute/.

8.2.9.2 Quantum Simulations of Fluid Dynamics

This study evaluates the Harrow-Hassidim-Lloyd quantum linear solver algorithm for the idealized Hele–Shaw flow, a theoretical model with applications in microfluidics, groundwater flow, and bioengineering. The team from the OLCF’s Science Engagement section benchmarked the algorithm on simulators and real quantum devices through the OLCF’s QCUP. Results show that accuracy and computational cost depend on the problem’s condition number, and error suppression and mitigation techniques enhance performance, making such fluid flow problems valuable benchmarks for noise mitigation in QC.

Publication:

Gopalakrishnan Meena, M., K. C. Gottiparthi, J. G. Lietz, A. Georgiadou, and E. A. Coello Pérez, “Solving the Hele-Shaw flow using the Harrow-Hassidim-Lloyd algorithm on superconducting devices: A study of efficiency and challenges,” *Physics of Fluids* 36, no. 10 (2024): 101705, (https://doi.org/10.1063/5.0231929).

8.2.9.3 The Association for Computing Machinery’s 2024 Gordon Bell Prize

The Gordon Bell Prize in supercomputing went to researchers from the University of Melbourne, AMD, QDX, and DOE’s ORNL (including Dmytro Bykov, Science Engagement section). Using the entire Frontier system at the OLCF, the team calculated a system containing more than 2 million correlated electrons. This represents a quantum molecular dynamics simulation 1,000× greater in size and speed than any previous simulation of its kind. It is also the first real-world scientific simulation that exceeded an exaflop when using sustained double-precision arithmetic.

8.2.9.4 Nature Publication on the Element Promethium

Promethium, or element 61, was discovered at ORNL nearly 80 years ago. In all that time, it remained the only lanthanide the chemical properties of which had not been characterized at the lab. Radioactive promethium is particularly elusive in nature owing to its short half-life of only 2.5 years. In 2024, a team of researchers from ORNL’s High Flux Isotope Reactor, Radiochemical Engineering Development Center, the OLCF (team member Dmytro Bykov from the Science Engagement Section), and Brookhaven National Laboratory’s National Synchrotron Light Source II were finally able to purify and study the element at the lab, both spectroscopically and theoretically. The high-fidelity theory simulations were conducted at the OLCF and shed light on electronic structure of the element for the first time. The results of their work were recently published in *Nature*.

8.2.10 Industry Engagement

ACCEL, the OLCF’s industrial partnership program, had a very strong year in 2024 with 38 projects underway across two systems. For 2024, 27 new projects launched, and 11 prior projects continued throughout the year. Industrial projects ran on both Summit and Frontier: 10 projects had Summit allocations, and 27 had Frontier allocations. In addition, 1 project continuing from a previous year had an allocation on Crusher.

These 38 industrial projects represented approximately 7% of the Frontier projects, 9% of the Summit projects, and 8% of the total Summit and Frontier projects combined for INCITE, ALCC, DD, and SummitPlus (including the ECP). These projects used 3,832,688 Summit node hours (approximately 18% of the total Summit hours) and 8,381,081 Frontier node hours (approximately 13% of the total Frontier hours). Collectively, these hours represented approximately 14% of the total hours across Summit and Frontier combined.

8.2.10.1 Industrial Project Allocations Across External User Programs

In 2024, the total allocated industrial project hours on Frontier and Summit were allocated as follows:

- 54% through INCITE,
- 25% through ALCC,
- 4% through the OLCF DD program, and
- 17% through the SummitPlus program.

In 2024, the total allocated industrial project hours on Frontier were allocated as follows:

- 65% through INCITE,
- 30% through ALCC, and
- 5% through the OLCF DD program.

In 2024, the total allocated industrial project hours on Summit were allocated as follows:

- 100% through SummitPlus.

8.2.10.2 Observations about the Industrial Projects

This year was notable for new projects from prior users Pratt & Whitney, Boeing, and insurance giant FM (formerly FM Global). All three had paused their use of the OLCF's leadership computing systems as they took a measured approach to adopting GPUs. Pratt & Whitney and FM returned this year with Frontier DD awards. Boeing returned last year as part of an aerospace industry team that had a Summit DD allocation, and this year they received an INCITE award on Frontier—their first since a Jaguar INCITE award in 2012.

The 27 new projects in 2024 received their awards via INCITE (6 projects), ALCC (2 projects), DD (9 projects), and SummitPlus (10).

INCITE

Industry projects had a larger influx of six new INCITE awards, with industry partnering with universities or labs in three of the awards, ensuring broader community engagement.

ALCC

Industry had two new ALCC awards from long-time user GE Aerospace (formerly General Electric) and from small business Microsurgeonbot. This was Microsurgeonbot's first ALCC Award.

OLCF DD

SummitPLUS Awards

Industry users launched 10 SummitPlus awards in 2024. Although most of the awards went to companies that had previously used Summit, GE Vernova received its first OLCF allocations. GE Vernova was spun off from GE's power, wind, and electrification businesses as a new legal entity in April 2024. Their SummitPlus projects were related to the development of wind turbines and high-efficiency gas turbines.

Frontier Awards

In addition to welcoming back Pratt & Whitney and FM, several other firms were awarded their first projects at the OLCF on Frontier through the DD program.

- AI startup Atomic Canyon is using Frontier to develop a foundational AI model to enable smart search of millions of documents from the Nuclear Regulatory Commission's database to accelerate regulatory and compliance processes for nuclear power plants.
- Lockheed Martin is leading a public-private sector aerospace team to assess CFD capabilities to predict dynamic flow distortion for a civil, high bypass ratio aero-engine at crosswind conditions. Results will improve the understanding of unsteady fan-intake interactions pertinent to novel engine architectures. They will be shared through the American Institute of Aeronautics and Astronautics, ensuring broad community awareness and impact.
- XTI Aerospace is using Frontier to accelerate development of its TriFan 600 aircraft, the company's vertical takeoff and landing business aircraft that combines the operational flexibility of helicopters with the speed and range of fixed-wing airplanes, creating an entirely new aircraft category—the vertical lift crossover airplane.
- QC small business IonQ is using Frontier to develop a basic tool chain on the AMD/AMD hybrid architecture on Frontier, where the QPU would be an emulator of a quantum computer.
- GE Vernova received a Frontier award related to predicting the impact of manufacturing techniques for heavy-duty gas turbine fan–shaped diffuser holes.

8.2.10.3 Industrial Project Highlight

Summit Helps Forge Stronger Flights

Companies such as RTX (formerly Raytheon Technologies) rely on sturdy titanium alloys as a cornerstone material to build jet-engine turbine blades, landing gear, and exhaust ducts. Titanium is stronger and lighter than steel, resistant to rust and corrosion, and resilient past the melting points of most other metals, making it ideal for aircraft components.

However, forging those alloys requires extreme heat and precision, which make the process vastly more difficult than typical manufacturing. Under the intense heat and pressure, components can crack and break during production.

The RTX research team, under the leadership of RTX senior technical fellow for advanced manufacturing Tahany El-Wardany, theorized that adding copper to raw titanium could yield an alternate alloy, and that 3D-printing techniques could be used to knit the alloy's grains into an equiaxed, or latticed, microstructure with grains running horizontally and vertically to hold up under greater stress. To tackle this problem, the RTX team partnered with ORNL distinguished scientist Balasubramaniam Radhakrishnan through DOE's High Performance Computing for Manufacturing program.

Radhakrishnan used Summit to simulate this equiaxed microstructure to see what causes it to form and how stable it is. He employed phase-field modeling, a computationally demanding and time-consuming mathematical technique that rigorously captured the physics of the process. Microstructures were simulated down to the nanometer under a wide range of extreme conditions (see Figure 8-14). The results suggested higher ratios of copper and niobium made the equiaxed microstructure more likely to form, and cooling rates could be tailored to produce the stronger alloy.

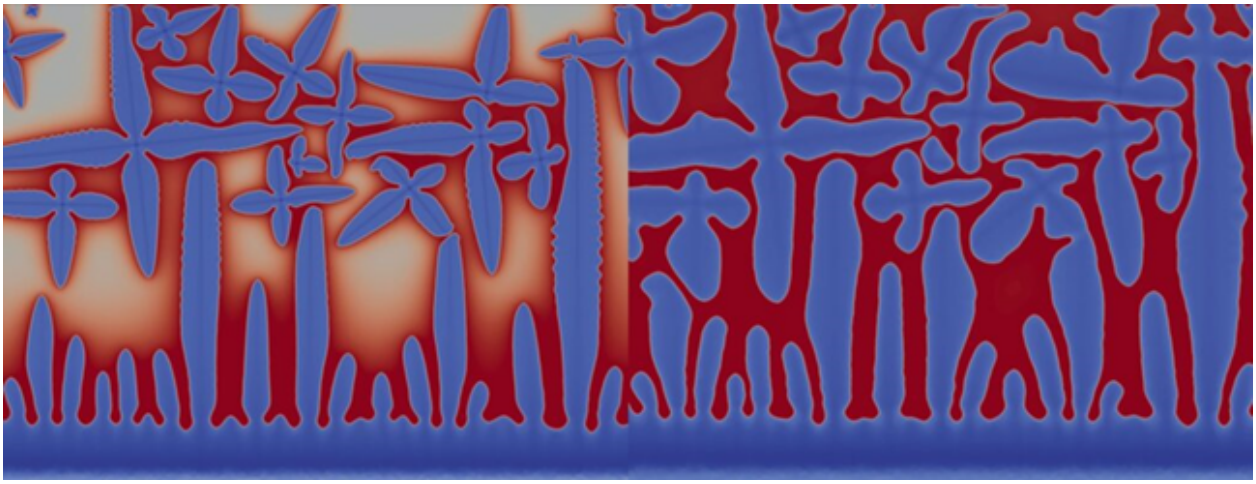


Figure 8-14. Evolution of the microstructures that make up a titanium alloy designed via simulation on Summit (Credit: ORNL).

Related Link: https://www.olcf.ornl.gov/2024/04/30/summit-helps-forge-stronger-flights/, OLCF News (June 17, 2023).

The Impact

The Summit simulations captured so much detail that RTX found their production process needed tighter standards to reproduce the study's findings. El-Wardany explained, "We trust the computational results over the production results in this case, and we're fine-tuning our production to meet the computational standards. These simulations have already saved us years of work, and we can continue to apply this approach as we seek better, more advanced materials and designs."

The results hold promise for improvements across the aerospace field and beyond. The new alloy could cut in half the annual $273 million production costs of machining titanium components and save the company as much as 2.5 quadrillion Btus in energy costs by 2050, according to the research team's estimates.

"Thanks to Summit, we have a candidate for an improved titanium alloy. Summit's predictive simulations shrank a decade of physical testing into what we hope will be 2 or 3 years," El-Wardany said.

8.3 DOE PROGRAM ENGAGEMENTS/REQUIREMENTS GATHERING

Through diverse partnerships, the OLCF cultivates strategic engagements with other DOE Office of Science programs, DOE Applied Programs, the ECP, and other federal agencies to broaden the OLCF user community, better understand current and future DOE mission needs related to the facility, and leverage opportunities for additional hardware, software, application, and operational innovations. Below are examples of some of the collaborations and outcomes that benefitted the OLCF and the broader DOE community.

8.3.1 Ongoing Engagements with Office of Science Observational and Experimental Facilities Investigators

Although PIs funded by DOE Office of Science program offices commonly use the OLCF, in recent years many PIs have explored discretionary allocations to establish workflows from their observational facilities to the OLCF resources. Using institutional resources, Office of Science Biological and Environmental Research–sponsored investigators in the ARM program use operational data workflows that originate from their sensors in their data management facility and have set up exploratory efforts with OLCF data and analytics (JupyterHub) portals. BES-sponsored scientists who use facilities such as ORNL's Center for Nanophase Materials Science and the SNS have successfully won discretionary time on Summit to perform AI-driven electron microscopy and design-of-experiment campaigns (porting their codes to GPUs), respectively. The movement of data from the facilities is supported by traditional data transfers and new tools to simplify the user experience (such as the DataFlow tool discussed in Section 4.1.3). Additionally, scientists from the LCLS femtosecond x-ray laser facility at SLAC have obtained three SummitPLUS allocations that will facilitate new workflows for rapid processing of data and training of AI models on massive datasets. This work will help analyze and compress emergent raw data and steer ongoing experiments. The partnership will enhance the coupling of a major BES experimental facility with OLCF exascale resources. The lessons learned from workflows for DOE's High Energy Physics programs (e.g., ATLAS-supporting BigPANDA workflows), Nuclear Physics programs (e.g., ALICE), and FES's deep learning for real-time prevention of disruption have informed workflows deployed on Slate for Summit and the design of the Frontier login nodes.

8.3.1.1 Ongoing Engagements with Atmospheric Radiation Measurement

In 2023, NCCS partnered with the ARM program to enhance their data transformation workflows. NCCS deployed over 5 PB of disk storage and 28 PB of tape storage into the NCCS open nearline system (Themis) to provide ARM with a highly durable storage resource. The OLCF offers data management support and expertise in these resources, thereby ensuring safe, long-term storage that also performs well enough to meet researcher demands.

The hosting of ARM's HPC resources and data at the OLCF offers several benefits to both the OLCF and ARM. In 2024, the OLCF leveraged the Multi-Category Security feature of Slurm to provide a limited node sharing capability to ARM. This improved cluster throughput and reduced the time from job submission to end data result, better enabling these workflows in an HPC space. Supporting a project such as ARM showcases the unique service capabilities and expertise that the OLCF provides.

8.3.2 Engagement with the National Institutes of Health and the National Cancer Institute

The collaboration between ORNL and the NIH's National Cancer Institute (NCI) began as a strategic partnership in 2016 with the aim of advancing specific areas of cancer research and HPC development by applying advanced computing, predictive ML/deep learning models, and large-scale computational simulations.

Collaborative research is performed under the project known as MOSSAIC (Modeling Outcomes using Surveillance data and Scalable Artificial Intelligence for Cancer) and consists of (1) developing large-scale, state-of-the-art transformer language models for clinical information extraction; (2) building new capabilities for biomarker and recurrence detection; (3) pushing novel research in abstention and uncertainty quantification so that models can be effectively deployed in clinical practice; (4) conducting lab studies to evaluate the performance of the models in real-world cancer registry settings; and (5) enabling large-scale transformer training on Leadership Computing Facility systems. Collectively, these efforts aim to modernize the national cancer surveillance program and enable near real–time cancer surveillance in the United States.

This research poses several significant computational challenges. Transformer models require pretraining on very large datasets with billions of elements, which itself requires access to large HPC resources. For example, these models require hundreds of Summit nodes for training by using data parallelism. Additionally, transformer models must be trained on the raw text of cancer pathology reports, and this text includes PHI.

To better support these efforts, the OLCF launched a security framework called CITADEL, which allows researchers to use supercomputers for research that contains PHI. ORNL's unmatched combination of a secure PHI enclave and a protected computing environment—through CITADEL—allows the secure use of Summit for MOSSAIC research and is crucial for the application of large-scale transformer models on the cancer pathology report corpus. CITADEL is a good example of an innovation achieved in one program that can benefit multiple programs, such as the US Department of Veterans Affairs (VA), and other OLCF user communities.

NIH/NCI's collaboration with the OLCF has led to several notable outcomes, both for the broader HPC community and for the MOSSAIC research team. The computing capabilities of the OLCF enable efficient iterative development of deep learning models, driven by the challenges of extracting information from the NCI data. Deploying in this way will allow integration with the new CITADEL capability and support running the transformer training pipeline with sensitive PHI data on the OLCF systems. The transformer training modules will also be made available to the OLCF user community, allowing domain scientists in other fields to train their own domain-specific transformer models on these systems.

Other major accomplishments of this collaboration have been the development of a new scalable and more efficient way to develop and deploy APIs to the Surveillance, Epidemiology, and End-Results (SEER) registries and other relevant stakeholders. The MOSSAIC team developed a new, modular PyTorch API designed to enable quick and easy swapping between (1) deep learning models; (2) different data preprocessing techniques, uncertainty quantification methods, and other pre- and post-processing methods; and (3) different cancer surveillance tasks, including but not limited to path-coding, identifying reportable path reports, identifying report type, detection of biomarkers, and identifying cases of recurrent metastatic disease.

In 2024, the MOSSAIC team continued to grow the Case-Level Multi-Task Hierarchical Self Attention Network, which is a natural language processing algorithm that autocodes cancer pathology reports at participating SEER registries and does so 18× faster than traditional methods. The algorithm saves 46,000 person hours per year and is a significant step toward near real–time population cancer surveillance. As the algorithm is adopted by more registries, the team is working on privacy-preserving methods that would enable data- and model-centric approaches to train the algorithm in a distributed environment. These advances demonstrate the power of AI for improving population health surveillance.

Building on this foundation, the team recently used the MIMIC III dataset to evaluate various attention mechanisms used to train AI models to focus on specific information in clinical documents. They then used these findings to improve the models that extract cancer data from electronic cancer pathology reports.

The MOSSAIC team also used Summit and CITADEL to develop a specialized transformer model called Path-BigBird, which they used to process 2.7 million cancer pathology reports from six SEER registries.

Path-BigBird has the potential to streamline the extraction of pathology data and could outperform traditional deep learning approaches for gathering cancer data such as sites, histology, and incidences.

8.3.3 Engagement with the US Department of Veterans Affairs

In 2016, the VA partnered with DOE and ORNL to revolutionize the health care of veterans and, by extension, all Americans via advanced data analytics and HPC. The DOE-VA collaboration advances the missions of both the VA and DOE and leverages each agency's unique resources to support efforts that would not otherwise be possible. The VA's Office of Research and Development uses DOE's supercomputing facilities and expertise in big data and HPC to advance veterans' treatment and the VA's medical and genomic research.

The Million Veteran Program used the Summit supercomputer to identify genetic markers for early-onset prostate cancer (before the age of 55). Supported by the security framework CITADEL, which protects sensitive personal health information (PHI) so it can be studied directly on the supercomputer, the project leveraged computationally efficient transformer models to capture complex interactions within and between genes, a capability not viable without large-scale HPC.

With Summit now retired, this research continues on Frontier. CITADEL remains essential to this work, enabling secure, scalable, and privacy-preserving access to sensitive biomedical data. These partnerships have resulted in unprecedented advancements in secure computing, empowering the VA and NIH to conduct groundbreaking research while continuing to strengthen the HPC ecosystem at ORNL.

8.3.4 Engagement with Air Force Weather and National Oceanic and Atmospheric Administration

ORNL and AFW launched two HPE Cray EX supercomputers, named Miller and Fawbush, to support the AFW's numerical weather modeling at much higher speed and fidelity. Thanks to the powerful systems, which reached Certification of Operational Readiness in February 2021, AFW is developing new specialized models such as full physics cloud forecasting and a global hydrology model. ORNL is also providing a unique system of safeguards, including separate, dedicated power sources for each machine, failsafe features through the Slurm resource scheduler, and dynamic load balancing. In 2024, ORNL worked with AFW to provide access to ORNL's newly deployed nearline storage file system, Kronos.

ORNL and NOAA established the NCRC through a strategic partnership in 2009 with a goal to leverage leading HPC and information technologies to develop, test, and apply state-of-the-science, computer-based global climate simulation models based on a strong scientific foundation. In 2021, based on success achieved over the program's lifespan, the program was extended another 5 years. The NCCS hosts the program's primary computational resource, named Gaea, which consists of an HPE-Cray EX 3000 HPC system providing over 10 petaflops of peak computation. In 2023, work began to procure and install a new climate modeling system with a peak performance of over 10 petaflops. Combined with the program's existing HPC resource, the new system will increase the program's peak performance to over 20 petaflops. The NCRC supports and complements NOAA's climate mission to understand climate variability and change and improve society's ability to plan and respond. The partnership enhances the capacity of NOAA's existing climate research centers by supporting the development of next-generation models, building computational capacity, cultivating a highly trained computational workforce, and engaging the global user community.

The partnership allows not only the opportunity for collaborative R&D but also the opportunity to develop and harden tools, methods, and best practices that improve operation and efficient use of HPC resources within the NCCS. For example, a project that began in 2021 and continued in 2024 has been developing the ability to run portable workloads within containers on NOAA's HPC resources. The NCCS project members are also undertaking similar efforts to develop containers that will run on Summit and Frontier. Lessons learned from each effort benefit both NOAA and the OLCF, and the broadened perspective helps to harden and improve the results for all programs in the NCCS.

APPENDIX A. 2024

There were no recommendations received regarding the 2023 Operational Assessment.

APPENDIX B. 2024

Table B-1. Science highlights.

Date	Science Highlights Submitted to DOE Title	Writer	URL
01/05/24	Exascale's New Frontier: Combustion-Pele	Matt Lakin	https://www.olcf.ornl.gov/2024/01/05/exascales-new-frontier-combustion-pele/
01/10/24	AI director featured at White House roundtable, Capitol Hill briefings	Elizabeth Rosenthal	https://www.ornl.gov/news/ai-director-featured-white-house-roundtable-capitol-hill-briefings
01/17/24	Fungal 'Bouncers' Patrol Plant-Microbe Relationship	OLCF Staff Writer	https://www.olcf.ornl.gov/2024/01/17/fungal-bouncers-patrol-plant-microbe-relationship/
01/19/24	Exascale's New Frontier: E3SM-MMF	Matt Lakin	https://www.olcf.ornl.gov/2024/01/19/exascales-new-frontier-e3sm-mmf/
01/19/24	ORNL's Tourassi named IEEE Fellow	Scott Jones	https://www.ornl.gov/news/ornls-tourassi-named-ieee-fellow
01/30/24	Exascale's New Frontier: SLATE	Coury Turczyn	https://www.olcf.ornl.gov/2024/01/30/exascales-new-frontier-slate/
02/05/24	CyberShake Study Uses Summit Supercomputer to Investigate Earthquake Hazards	Quinn Burkhart	https://www.olcf.ornl.gov/2024/02/05/cybershake-study-uses-summit-supercomputer-to-investigate-earthquake-hazards/
02/22/24	Exascale's New Frontier: LatticeQCD	Coury Turczyn	https://www.olcf.ornl.gov/2024/02/22/exascales-new-frontier-latticeqcd/
02/29/24	Planning for a Smooth Landing on Mars	Coury Turczyn	https://www.olcf.ornl.gov/2024/02/29/planning-for-a-smooth-landing-on-mars/
03/15/24	Scientists Use Summit Supercomputer to Explore Exotic Stellar Phenomena	Quinn Burkhart	https://www.olcf.ornl.gov/2024/03/15/scientists-use-summit-supercomputer-to-explore-exotic-stellar-phenomena-2/
04/10/24	INCITE 2025 Call for Proposals	Katie Bethea	https://www.olcf.ornl.gov/2024/04/10/__trashed/
04/15/24	New Data Processing Automation Grows Plant Science at ORNL	Betsy Sonewald	https://www.olcf.ornl.gov/2024/04/15/new-data-processing-automation-grows-plant-science-at-ornl/
04/15/24	ORNL collaboration helps secure 2023 Gordon Bell Prizes	Betsy Sonewald	https://www.olcf.ornl.gov/2024/04/15/ornl-collaboration-helps-secure-2023-gordon-bell-prizes/

Date	Science Highlights Submitted to DOE Title	Writer	URL
04/22/24	Steering Toward Quantum Simulation at Scale	Matt Lakin	https://www.olcf.ornl.gov/2024/04/22/steering-toward-quantum-simulation-at-scale/
04/29/24	Adaptable IO System Delivers the Data	Coury Turczyn	https://www.olcf.ornl.gov/2024/04/29/adaptable-io-system-delivers-the-data/
04/30/24	Summit Helps Forge Stronger Flights	Matt Lakin	https://www.olcf.ornl.gov/2024/04/30/summit-helps-forge-stronger-flights/
05/06/24	Something in the Water Does Not Compute	Coury Turczyn	https://www.olcf.ornl.gov/2024/05/06/something-in-the-water-does-not-compute/
05/06/24	ORNL and SLAC Team Up for Breakthrough Biology Projects	Coury Turczyn	https://www.olcf.ornl.gov/2024/05/06/ornl-and-slac-team-up-for-breakthrough-biology-projects/
05/13/24	Breaking Benchmarks: Frontier Supercomputer Sets New Standard in Molecular Simulation	Matt Lakin	https://www.olcf.ornl.gov/2024/05/13/breaking-benchmarks-frontier-supercomputer-sets-new-standard-in-molecular-simulation/
05/14/24	Meet the NCCS and OLCF Director: Arjun Shankar	Betsy Sonewald	https://www.olcf.ornl.gov/2024/05/14/meet-the-nccs-and-olcf-director-arjun-shankar/
05/14/24	Going Big: World's Fastest Computer Takes On Large Language Modeling	Matt Lakin	https://www.olcf.ornl.gov/2024/05/14/going-big-worlds-fastest-computer-takes-on-large-language-modeling/
05/20/24	Exascale's New Frontier: ADIOS	Coury Turczyn	https://www.olcf.ornl.gov/2024/05/20/exascales-new-frontier-adios/
06/07/24	Quantum Chemistry and Simulation Help Characterize Coordination Complex of Elusive Element 61	Betsy Sonewald	https://www.olcf.ornl.gov/2024/06/07/quantum-chemistry-and-simulation-help-characterize-coordination-complex-of-elusive-element-61/
06/21/24	Untangling the Entangled	Matt Lakin	https://www.olcf.ornl.gov/2024/06/21/untangling-the-entangled/
06/24/24	New Clues to Improving Fusion Confinement	Coury Turczyn	https://www.olcf.ornl.gov/2024/06/24/new-clues-to-improving-fusion-confinement/
07/09/24	Exascale's New Frontier: SuperLU/STRUMPACK	Coury Turczyn	https://www.olcf.ornl.gov/2024/07/09/exascales-new-frontier-superlu-strumpack/
07/17/24	Game-Changing Quantum Chemistry Calculations Push New Boundaries of Exascale Frontier	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/07/17/game-changing-quantum-chemistry-calculations-push-new-boundaries-of-exascale-frontier/

Date	Science Highlights Submitted to DOE Title	Writer	URL
07/19/24	DOE Opens Bids to Construct ORNL's Next-Generation Supercomputer	Katie Bethea	https://www.olcf.ornl.gov/2024/07/19/doe-opens-bids-to-construct-ornls-next-generation-supercomputer/
07/26/24	Frontier Simulations Could Help Build a Better Diamond	Matt Lakin	https://www.olcf.ornl.gov/2024/07/26/frontier-simulations-could-help-build-a-better-diamond/
07/29/24	Exascale's New Frontier: ExaWind	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/07/29/exascales-new-frontier-exawind/
08/12/24	Flying Quieter and Cleaner	Coury Turczyn	https://www.olcf.ornl.gov/2024/08/12/flying-quieter-and-cleaner/
08/22/24	2024 Notable System Changes: Summit and HPSS	Katie Bethea	https://www.olcf.ornl.gov/2024/08/22/2024-notable-system-changes-summit-and-hpss/
08/27/24	Quantum Computing Experts Gather for Fifth Annual User Forum at ORNL	Angela Gosnell	https://www.olcf.ornl.gov/2024/08/27/quantum-computing-experts-gather-for-fifth-annual-user-forum-at-ornl/
08/28/24	Study Seeks to Unite High-Performance Computing, Quantum Computing for Science	Matt Lakin	https://www.olcf.ornl.gov/2024/08/28/study-seeks-to-unite-high-performance-computing-quantum-computing-for-science/
09/03/24	Frontier Simulations Provide New Insights Into Calcium-48's Controversial Nuclear Magnetic Excitation	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/09/03/frontier-simulations-provide-new-insights-into-calcium-48s-controversial-nuclear-magnetic-excitation/
09/04/24	Exascale's New Frontier: CANDLE	Coury Turczyn	https://www.olcf.ornl.gov/2024/09/04/exascales-new-frontier-candle/
09/06/24	DOE, ORNL Announce Opportunity to Define Future of High-Performance Computing	Katie Bethea	https://www.olcf.ornl.gov/2024/09/06/newfrontiers/
09/06/24	Molecular simulations, supercomputing lead to energy-saving biomaterials breakthrough	Stephanie Seay	https://www.ornl.gov/news/molecular-simulations-supercomputing-lead-energy-saving-biomaterials-breakthrough
09/10/24	OLCF Pioneers Approaches to Energy Efficient Supercomputing	Coury Turczyn	https://www.olcf.ornl.gov/2024/09/10/olcf-pioneers-approaches-to-energy-efficient-supercomputing/
09/17/24	Retiring — and Shredding — the Alpine Storage System	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/09/17/retiring-and-shredding-the-alpine-storage-system/
09/18/24	Users of Oak Ridge Leadership Computing Facility gather for 20th annual meeting	Angela Gosnell	https://www.olcf.ornl.gov/2024/09/18/users-of-oak-ridge-leadership-computing-facility-gather-for-20th-annual-meeting/
09/23/24	Exascale's New Frontier: ExaAM	Matt Lakin	https://www.olcf.ornl.gov/2024/09/23/exascales-new-frontier-exaam/

Date	Science Highlights Submitted to DOE Title	Writer	URL
09/23/24	Exascale's New Frontier: ExaBiome	Matt Lakin	https://www.olcf.ornl.gov/2024/09/23/exascales-new-frontier-exabiome/
09/30/24	Exascale's New Frontier: GAMESS	Matt Lakin	https://www.olcf.ornl.gov/2024/09/30/exascales-new-frontier-gamess/
10/14/24	Standing up the Nation's Supercomputing Pipeline for Streaming Big Data in Real Time	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/10/14/standing-up-the-nations-supercomputing-pipeline-for-streaming-big-data-in-real-time/
10/17/24	Frontier Quantum Chemistry Calculations Nominated to Win HPC's Highest Prize	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/10/17/frontier-quantum-chemistry-calculations-nominated-to-win-hpcs-highest-prize/
10/22/24	Goodbye HPSS, Hello Kronos	Angela Gosnell	https://www.olcf.ornl.gov/2024/10/22/goodbye-hpss-hello-kronos/
10/23/24	Surprising Details in a More Precise Description of Fission	Coury Turczyn	https://www.olcf.ornl.gov/2024/10/23/surprising-details-in-a-more-precise-description-of-fission/
10/23/24	Two ORNL quantum research papers selected for editor's choice list	Mark Alewine	https://www.ornl.gov/news/two-ornl-quantum-research-papers-selected-editors-choice-list
10/30/24	Frontier Users' Exascale Climate Emulator Nominated for Gordon Bell Climate Prize	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/10/30/frontier-users-exascale-climate-emulator-nominated-for-gordon-bell-climate-prize/
11/01/24	Bigger, Faster, Smarter Genetics Research	Coury Turczyn	https://www.olcf.ornl.gov/2024/11/01/bigger-faster-smarter-genetics-research/
11/04/24	Two ORNL staff receive 2024 Great Minds in STEM Luminary award	Angela Gosnell	https://www.ornl.gov/news/two-ornl-staff-receive-2024-great-minds-stem-luminary-award
11/12/24	Gordon Bell Prize Nomination Recognizes Efforts to Train Extreme Scale Large Language Models Using Frontier	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/11/12/gordon-bell-prize-nomination-recognizes-efforts-to-train-extreme-scale-large-language-models-using-frontier/
11/13/24	Fine-Tuning Forecasts: ORBIT Brings Long-Range Weather Prediction Within Reach	Matt Lakin	https://www.olcf.ornl.gov/2024/11/13/fine-tuning-forecasts-orbit-brings-long-range-weather-prediction-within-reach/
11/14/24	Protein Design on Demand	Matt Lakin	https://www.olcf.ornl.gov/2024/11/14/protein-design-on-demand/
11/15/24	DOE's Quantum Computing User Program releases Request for Information to Gather Input on Quantum Computing Access	Katie Bethea	https://www.olcf.ornl.gov/2024/11/15/qcup-rfi/

Date	Science Highlights Submitted to DOE Title	Writer	URL
11/18/24	Frontier Supercomputer Hits New Highs in Third Year of Exascale	Matt Lakin	https://www.olcf.ornl.gov/2024/11/18/frontier-supercomputer-hits-new-highs-in-third-year-of-exascale/
11/18/24	INCITE Program Awards Supercomputing Time to 81 High-Impact Projects	Katie Bethea	https://www.olcf.ornl.gov/2024/11/18/incite-program-awards-supercomputing-time-to-81-high-impact-projects/
11/19/24	Record-Breaking Run on Frontier Sets New Bar for Simulating the Universe in the Exascale Era	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/11/19/record-breaking-run-on-frontier-sets-new-bar-for-simulating-the-universe-in-the-exascale-era/
11/19/24	Oak Ridge National Laboratory receives honors in 2024 HPCwire Editors' Choice award	Scott Jones	https://www.ornl.gov/news/oak-ridge-national-laboratory-receives-honors-2024-hpcwire-editors-choice-award
11/22/24	Game-Changing Quantum Chemistry Calculations on Frontier Earn Gordon Bell Prize	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/11/22/game-changing-quantum-chemistry-calculations-on-frontier-earn-gordon-bell-prize/
11/22/24	Gordon Bell Climate Prize Goes to KAUST Frontier Users' Exascale Climate Emulator	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/11/22/gordon-bell-climate-prize-goes-to-kaust-frontier-users-exascale-climate-emulator/
12/03/24	Summit's Bonus Year of Scientific Achievement	Coury Turczyn	https://www.olcf.ornl.gov/2024/12/03/summits-bonus-year-of-scientific-achievement/
12/04/24	New Data Transfer Tool Developed at ORNL to Be Made Available for Public Use	Angela Gosnell	https://www.olcf.ornl.gov/2024/12/04/new-data-transfer-tool-developed-at-ornl-to-be-made-available-for-public-use/
12/10/24	Calculations on Frontier earns Gordon Bell Prize, ORNL major presence at SC24	Mark Alewine	https://www.ornl.gov/news/calculations-frontier-earns-gordon-bell-prize-ornl-major-presence-sc24
12/12/24	ORNL Researchers Honored with Award for Best Event Report	Angela Gosnell	https://www.olcf.ornl.gov/2024/12/12/ornl-researchers-honored-with-award-for-best-event-report/
12/13/24	OLCF Conference Outreach Teaches HPC Fundamentals, Prepares Young Professionals	Jeremy Rumsey	https://www.olcf.ornl.gov/olcf-news/olcf-conference-outreach-teaches-hpc-fundamentals-prepares-young-professionals/
12/18/24	Summit Helps Veterans Affairs Connect Genetic Dots	Matt Lakin	https://www.olcf.ornl.gov/2024/12/13/olcf-conference-outreach-teaches-hpc-fundamentals-prepares-young-professionals/
12/19/24	The OLCF's 2024 in Review	Angela Gosnell	https://www.olcf.ornl.gov/2024/12/19/the-olcfs-2024-in-review/
12/23/24	Decoding Atmospheric Effects of Gravity Waves	Jeremy Rumsey	https://www.olcf.ornl.gov/2024/12/23/decoding-atmospheric-effects-of-gravity-waves/

APPENDIX C. 2024

The Oak Ridge Leadership Computing Facility (OLCF) provided the Frontier computational resource (Table C-1) and the Alpine and High-Performance Storage System (HPSS) data resources for production use in 2024. Supporting systems such as EVEREST, Andes, and data transfer nodes were also offered. Metrics for these supporting systems are not provided. The following systems were operational during this reporting period.

HPE CRAY EX (FRONTIER) RESOURCE SUMMARY

Debuting in 2022 at 1.102 exaflops, Oak Ridge National Laboratory's (ORNL's) Frontier supercomputer is the world's first exascale system. ORNL's long history of supercomputing excellence enables scientists to expand the scale and scope of their research, solve complex problems in less time, and fill critical gaps in knowledge. The OLCF team tuned the system and improved the Linpack score in May 2024, adding 0.012 exaflops to the result for a new score of 1.206 exaflops. They improved the score again in November 2024, adding 0.147 exaflops for a new score of 1.353 exaflops. Frontier is an HPE Cray EX supercomputer comprising 77 Olympus rack HPE cabinets, each with 128 AMD compute nodes, for a total of 9,856 AMD compute nodes. Each Frontier compute node consists of one 64-core AMD Optimized 3rd Gen EPYC CPU (with two hardware threads per physical core) with access to 512 GB of DDR4 memory. Each node also contains four AMD MI250X GPUs, each with two Graphics Compute Dies (GCDs) for a total of eight GCDs per node. The programmer can think of the eight GCDs as eight separate GPUs, each having 64 GB of high-bandwidth memory. The CPU is connected to each GCD via an Infinity Fabric CPU-GPU link, allowing a peak host-to-device and device-to-host bandwidth of 36 + 36 GB/s. The two GCDs on the same MI250X are connected with an Infinity Fabric GPU-GPU link with a peak bandwidth of 200 GB/s. The GCDs on different MI250X GPUs are also connected with an Infinity Fabric GPU-GPU link, and the peak bandwidth ranges from 50 to 100 GB/s based on the number of Infinity Fabric connections between individual GCDs.

IBM AC922 (SUMMIT) RESOURCE SUMMARY

The OLCF installed and deployed an IBM AC922 system, Summit, which became available for full production on January 1, 2019. In 2023, despite plans to shut down the IBM Power System AC922 Summit supercomputer after 5 years of service, with support from ASCR, the OLCF was able to extend Summit's life for an additional year, operating until November 2024. At the time, Summit remained the fourth-fastest supercomputer in the United States and ranked among the top 10 worldwide. This extension led to the launch of the SummitPLUS allocation program, which provided more than 19 million hours of compute time to 108 research projects across academia, government and industry to support critical scientific studies from January through October 2024. Summit's extended operation allowed researchers to continue benefiting from its powerful computing capabilities, marking a productive final year before its shutdown on November 15, 2024. Summit comprises 4,608 high-density compute nodes, each equipped with two IBM POWER9 CPUs and six NVIDIA Volta GPUs (Table C-2). In total, the Summit system is capable of 200 petaflops of peak computational performance and was recognized as the most powerful system in the world for its performance on both the High-Performance Linpack and the Conjugate Gradient benchmark applications from June 2018 until June 2020. Three new cabinets with a higher memory footprint were added to the Summit system in July 2020 to support COVID-19 research. Additionally, three more cabinets were added to the system in March 2021. These cabinets provided spare parts for Summit in its 6th year of operation in 2024.

Table C-1. Frontier 2024.

System	Access	Type	CPU	GPU	Computational description			Interconnect
					Nodes	Node configuration	Memory configuration	
Frontier	Full production	HPE Cray EX	AMD Optimized 3rd Gen EPYC CPU (64 core)	AMD Instinct MI250X GPUs, each feature 2 Graphics Compute Dies (GCDs) for a total of 8 GCDs per node	9,856	64-core AMD Optimized 3rd Gen EPYC CPU (1/node) + AMD Instinct MI250X GPUs (4/node)	700 PB HDD + 11 PB Flash Performance Tier 9.4 TB/s and 10 PB Metadata Flash Lustre 3.8 TB node-local NVMe per Frontier node	4-port HPE Slingshot 200 Gbps (25 GB/s) NICs providing a node-injection bandwidth of 800 Gbps (100 GB/s)

Table C-2. Summit 2024.

System	Access	Type	CPU	GPU	Computational description			Interconnect
					Nodes	Node configuration	Memory configuration	
Summit	Full production	IBM AC922	3.45 GHz IBM POWER9 (22 core)	1,530 MHz NVIDIA V100 (Volta)	4,608	IBM POWER9 CPUs (2/node) + NVIDIA Volta V100 GPUs (6/node)	(4,608) 512 GB DDR4 and 96 GB HBM2 per node; (54) 2TB DDR4 and 192 GB HBM2 per node; >10 PB DDR4 + HBM + Non-volatile aggregate	Mellanox EDR 100 Gbps InfiniBand (Non-blocking Fat Tree)

GENERAL PARALLEL FILE SYSTEMS (ALPINE AND WOLF) RESOURCE SUMMARY

In January 2019, the OLCF deployed Alpine as its next-generation global file system to support the computational resources in the OLCF. Alpine is a single GPFS (General Parallel File System) namespace with a usable capacity of 250 PB and a file system–level performance of 2.5 TB/s. The Alpine file system is the default high-performance parallel file system for all of the OLCF's moderate compute resources. In January 2024, Alpine was decommissioned and replaced with Alpine 2.

In March 2017, the OLCF procured, installed, and deployed the Wolf GPFS, which serves as the center-wide file system for the computational resources in the Open Production enclave. Wolf provides a total storage capacity of 8 PB and up to 120 GB/s of performance.

LUSTRE FILE SYSTEM (ORION) RESOURCE SUMMARY

In April 2023, the OLCF deployed Orion as its next-generation global file system to support the computational resources in the OLCF. Orion is a multitiered (flash and HDD), single Lustre namespace with a usable capacity of 700 PB. The demonstrated file system–level performance from Fronter is 11 TB/s write and 14 TB/s read. The Orion file system is the default high-performance parallel file system for all the OLCF's moderate compute resources.

DATA ANALYSIS AND VISUALIZATION CLUSTER (ANDES) RESOURCE SUMMARY

A new data analytics cluster named Andes went into production on December 9, 2020. The primary purpose of the data analytics cluster is to provide a conduit for large-scale scientific discovery through pre- and post-processing of simulation data generated on Summit. Users with accounts on Innovative and Novel Computational Impact on Theory and Experiment (INCITE) or ASCR Leadership Computing Challenge (ALCC) supported projects are automatically given accounts on the data analytics cluster. Director's Discretionary (DD) projects may also request access to this cluster. Andes is a 704-node cluster, and each node contains two 16-core 3.0 GHz AMD EPYC processors and 256 GB of main memory. Andes offers nine additional heterogeneous nodes, each of which provides 1 TB of main memory and two NVIDIA Tesla K80 (Kepler GK210) GPUs.

HIGH-PERFORMANCE STORAGE SYSTEM RESOURCE SUMMARY

The OLCF provides a long-term storage archive system based on the HPSS software product co-developed by IBM, Los Alamos National Laboratory, Sandia National Laboratories, Lawrence Livermore National Laboratory, Lawrence Berkeley National Laboratory, and ORNL. The ORNL HPSS instance is currently over 136 PB in size and provides ingestion rates of up to 30 GB/s via a 22 PB front-end disk cache backed by a 21-frame Spectra Logic TFinity tape library that houses 81 IBM TS1155 tape drives and over 17,000 tape media slots, thereby providing ORNL a current capacity of 180 PB, which is expandable well into hundreds of PBs. The archive's average ingestion rate ranges between 100 and 150 TB/day. The archive environment consists of hardware from Dell, Brocade, DataDirect Networks, and Spectra Logic. In 2024, in preparation for the decommissioning of HPSS, the OLCF brought online Kronos, a 134 PB, multiprogrammatic nearline storage system that also provides tape-based backups for all data as a disaster-recovery measure. The system uses IBM's Storage Scale—known to many as the General Parallel File System—and the disaster recovery component leverages a magnetic tape format called the Linear Tape File System. These two systems are linked by IBM's Storage Archive hierarchical storage manager.

VISUALIZATION RESOURCE SUMMARY

The EVEREST facility has three computing systems and two separate state-of-the-art visualization display walls. The primary display wall spans 30.5 × 8.5 ft. and consists of eighteen 1920 × 1080 Barco projection displays arranged in a 6 × 3 configuration for a 32:9 aspect ratio at 11,520 × 3,240. The secondary display wall is being upgraded to eighteen 1920 × 1080 Barco LCD displays arranged in a 6 × 3 configuration to provide a secondary 32:9 aspect ratio. There are several 86 in. mobile interactive touch displays for easy and fast collaboration. Multiple augmented reality systems provide an interactive scalable space equipped for mixed-reality data exploration and analysis. These combined technologies, along with OLCF staff expertise, allow scientists to analyze complex scientific datasets in an immersive environment and communicate abstract concepts in an intuitive visual format.

NEARLINE STORAGE (THEMIS) SUMMARY

Themis entered production in late CY 2022 to bridge the retention requirement gap between scratch (immediate term) and archival (permanent) data storage use cases while streamlining resource access requirements. In response to user needs, the system is also intended to facilitate gateways and other data-sharing mechanisms within the National Center for Computational Sciences, ORNL, and the broader research community. Themis consists of a 14 PB Spectrum Scale file system backed by a 37 PB IBM TS4500 tape library. The Spectrum Scale layer will be capable of an aggregate bandwidth of 45 GB/s, moving to 65 GB/s after network upgrades. The tape component offers dual-copy to provide resiliency with an aggregate bandwidth of 7.68 GB/s. The system was expanded by 11 PB of Spectrum Scale and 7 PB of tape in late Q1 of CY 2023, with additional expansions to both disk and tape components later in the year.

PROTECTED DATA INFRASTRUCTURE SUMMARY

The OLCF now provides a production enclave to support the processing of data subject to the Health Insurance Portability and Accountability Act and International Traffic and Arms Regulations. Using the infrastructure in this enclave, users can submit and run Protected Data workloads on Summit. The enclave provides a dedicated login node for submission of Protected Data jobs and access to the Protected Data GPFS, Arx. Arx provides a total storage capacity of ~3.3 PB and up to 30 GB/s of performance.

OLCF HPC RESOURCE PRODUCTION SCHEDULE

The OLCF computational systems entered production according to the schedule listed in Table C-3. This list includes historical data associated with the Cray XT5, the very small overlap in December 2011 beginning with the introduction of the Cray XK6, and the series of Cray XK systems first available in 2012 and 2013.

Table C-3. OLCF HPC system production dates, 2008–present.

System	Type	Production date [a]	Performance end date [b]	Notes
Frontier	HPE Cray EX	April 5, 2023	Present	
Summit	IBM AC922	March 19, 2021	November 15, 2024	Production with 4,608 hybrid CPU-GPU nodes: IBM POWER9 CPUs (2/node) + NVIDIA Volta V100 GPUs (6/node). Three cabinets added as a spare parts cache for an optional 6th year.

System	Type	Production date [a]	Performance end date [b]	Notes
Summit	IBM AC922	July 1, 2020	November 15, 2024	Production with 4,608 hybrid CPU-GPU nodes: IBM POWER9 CPUs (2/node) + NVIDIA Volta V100 GPUs (6/node). Three cabinets added for COVID-19 research.
Summit	IBM AC922	January 1, 2019	November 15, 2024	Production with 4,608 hybrid CPU-GPU nodes: IBM POWER9 CPUs (2/node) + NVIDIA Volta V100 GPUs (6/node).
Spider III (Alpine)	GPFS	January 1, 2019	Present	250 PB GPFS single namespace file system.
Spider II	Lustre parallel file system	October 3, 2013	August 2, 2019	Delivered as two separate file systems: /atlas1 and /atlas2, 30+ PB capacity.
Orion	HPE ClusterStor E1000 v.2	April 5, 2023	Present	700 PB Lustre single namespace file system.
Eos	Cray XC30	October 3, 2013	August 2, 2019	Production with 736 Intel E5, 2,670 nodes.
Titan	Cray XK7	May 31, 2013	August 2, 2019	Production with 18,688 hybrid CPU-GPU nodes: AMD Opteron 6274/NVIDIA K20X.
JaguarPF	Cray XK6	September 18, 2012	October 7, 2012	Production at 240,000 cores until September 18, when partition size was reduced to 120,000 AMD Opteron cores. Additional Kepler installation. TitanDev access terminated.
JaguarPF	Cray XK6	February 13, 2012	September 12, 2012	Full production until September 12, when partition size was reduced to 240,000 AMD Opteron cores. Beginning of Kepler installation.
JaguarPF	Cray XK6	February 2, 2012	February 13,2012	Stability test. Restricted user access. 299,008 AMD Opteron 6274 cores. Includes 960-node Fermi-equipped partition.
JaguarPF	Cray XK6	January 5, 2012	February 1, 2012	Acceptance. No general access. 299,008 AMD Opteron cores.
JaguarPF	Cray XK6	December 12, 2011	January 4, 2012	142,848 AMD Opteron cores.
JaguarPF	Cray XT5	October 17, 2011	December 11, 2011	117,120 AMD Opteron cores.
JaguarPF	Cray XT5	October 10, 2011	October 16, 2011	162,240 AMD Opteron cores.
JaguarPF	Cray XT5	September 25, 2009	October 9, 2011	224,256 AMD Opteron cores.
JaguarPF	Cray XT5	August 19, 2008	July 28, 2009	151,000 AMD Opteron cores.

[a] The production date used for computing statistics is either the initial production date or the production date of the last substantive upgrade to the computational resource.
[b] The performance end date is the last calendar day that user jobs were allowed to execute on that partition.

APPENDIX D. 2024

2023 OPERATIONAL ASSESSMENT GUIDANCE

Scheduled Availability

Scheduled availability (SA) (Eq. [D.1]) in high-performance computing facilities is the percentage of time that a designated level of resource is available to users, excluding scheduled downtime for maintenance and upgrades. The user community must be notified of the need for a maintenance event window no less than 24 hours in advance of the outage (emergency fixes) for it to be considered scheduled downtime. Users will be notified of regularly scheduled maintenance in advance on a schedule that provides sufficient notification, no less than 72 hours before the event and preferably as many as 7 calendar days prior. If that regularly scheduled maintenance is not needed, then users will be informed of the maintenance event cancellation in a timely manner. Any service interruption that does not meet the minimum notification window is categorized as an unscheduled outage.

$$SA = \left(\frac{\text{time in period} - \text{time unavailable due to outages in period}}{\text{time in period} - \text{time unavailable due to scheduled outages in period}}\right) * 100 \qquad \text{(D.1)}$$

A significant event that delays a return to scheduled production by more than 4 hours will be counted as an adjacent unscheduled outage, unscheduled availability, and additional interrupt.

Overall Availability

Overall availability (OA) (Eq. [D.2]) is the percentage of time that a system is available to users. Outage time reflects both scheduled and unscheduled outages.

$$OA = \left(\frac{\text{time in period} - \text{time unavailable due to outages in period}}{\text{time in period}}\right) * 100 \qquad \text{(D.2)}$$

Mean Time to Interrupt

Mean time to interrupt (MTTI) is the time, on average, to any outage of the full system, whether unscheduled or scheduled (Eq. [D.3]).

$$MTTI = \left(\frac{\text{time in period} - (\text{ duration of scheduled outages} + \text{duration of unscheduled outages })}{\text{number of scheduled outages} + \text{number of unscheduled outages} + 1}\right) \qquad \text{(D.3)}$$

Mean Time to Failure

Mean time to failure (MTTF) is the time, on average, to an unscheduled outage of the full system (Eq. [D.4]).

$$MTTF = \frac{\text{time in period} - (\text{ duration of unscheduled outages })}{\text{number of unscheduled outages} + 1} \qquad \text{(D.4)}$$

System Utilization

System utilization (SU) is the percentage of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user group, including staff and vendors (Eq. [D.5]).

$$SU = \left(\frac{\text{core hours used in period}}{\text{core hours available in period}}\right) * 100 \qquad \text{(D.5)}$$

APPENDIX E. 2024

DIRECTOR'S DISCRETIONARY PROJECTS UTILIZING FRONTIER (ENHANCED ENCLAVE) ENABLED (AT ANY POINT) IN CY 2024

Table E-1. Director's Discretionary projects utilizing Frontier (Enhanced Enclave) in CY 2024.

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
LRN032_MDE	Jessica Inman	Georgia Tech	12,500	42,438	Large Scale Formal Verification of Neural Networks

DIRECTOR'S DISCRETIONARY PROJECTS UTILIZING FRONTIER (MODERATE ENCLAVE) ENABLED (AT ANY POINT) IN CY 2024

Table E-2. Director's Discretionary projects utilizing Frontier (Moderate Enclave) in CY 2024.

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
CSC380	Bronson Messer	ORNL	355,000	95,384	CAAR for Frontier
AST181	Evan Schneider	University of Pittsburgh	15,000	571,379	Cholla
CSC524	Van Ngo	ORNL	20,000	559	Benchmarking and Development of the MiMiC Framework for Highly Efficient QM/MM Simulations on Leadership Supercomputers
AST163	Christian Cardall	ORNL	20,000	220	INCITE Preparation for 3D+1D core-collapse supernova simulations with GenASiS
CHM191	Andre Severo Pereira Gomes	CNRS	15,000	18	PRECISE
CSC533	William Severa	Sandia	20,000	41,813	COINFLIPS
CPH005	Dario Alfe	University College London	20,000	265	New Frontiers for Material Modeling via Machine Learning Techniques with Quantum
CMB150	Olivier Desjardins	Cornell	20,000	0	High Fidelity Modeling of Spray Formation and Dispersion
AST182	Jay Kalinani	RIT	20,000	19,668	Benchmarking for the new GPU-accelerated open-source GRMHD code AsterX
CSC538	Irina Rish	University of Montreal	20,000	0	Large-Scale AI Models
CHM181	Kurt Mikkelsen	University of Copenhagen	15,000	3,011	Massively parallel GPU-enabled cluster perturbation methods

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
ATM136	Wei Zhang	ORNL	20,000	283	Quantifying Feedbacks of Climate Intervention Under Climate Change
CFD175	Peter Vincent	Imperial College London	16,550	3,179	Application of PyFR to Methane Super-Emitter Detection using Frontier
BIP235	Ron Dror	Stanford	30,000	62,382	Advancing the Rational Design of Functionally Selective GPCR-targeted Drugs
CFD173	Yuuichi Asahi	JAERI	30,000	41	Optimization of City wind flow simulation code on Frontier
FUS152	Colin McNally	General Fusion	20,000	19,558	Beyond Neoclassical Closures via Kinetic Monte-Carlo Calculations
NFI125	Aurel Bulgac	U. of Washington	50,000	38,516	Microscopic Simulations of Real-Time Nuclear Dynamics in the Exascale
CSC536	Chathika Gunaratne	ORNL	20,000	8,523	SAGESim - Scalable Agent-based GPU Enabled Simulator
CHM193	Taisung Lee	Rutgers	20,000	0	A next-generation extendable simulation environment for affordable, accurate, and efficient free energy simulations
GEO151	Kim Olsen	San Diego State University	35,000	0	Preparing AWP-ODC for Exascale Earthquake Ground Motion Research on Frontier
MAT263	Wenzhe Yu	ANL	8,000	3	GPU acceleration of large-scale many-body perturbation theory using OpenMP 5
ENG123	Ramesh Balakrishnan	ANL	15,000	4,435	Direct Numerical Simulation of Separated Flow over a Speed Bump at Higher Reynolds Numbers
LRN032	Jessica Inman	Georgia Tech	12,500	0	Large Scale Formal Verification of Neural Networks
LGT114	Amy Nicholson	University of North Carolina at Chapel Hill	15,000	0	Electromagnetic corrections to the nucleon axial charge
CMB147	Joseph Oefelein	Georgia Tech	15,000	13	Analysis of Combustion and Wave Dynamics in Rotating Detonation Engines

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
CFD154	Spencer Bryngelson	Georgia Tech	25,000	13,679	Accelerated Sub-grid Multi-component Flow Physics
CHM194	Kedan He	Eastern Connecticut State University	1,000	597	Computational Chemistry and Drug Design: Molecular Recognition and Binding Prediction Powered by Machine Learning and Deep Learning Approaches
MAT226	Jan Michael Carrillo	ORNL	20,000	11,824	Molecular Dynamic Simulations of Amphiphilic Oligomer Membranes: Design Rules Towards Stable Membranes Capable of Learning and Memory
AST031	Pierre Ocvirk	Strasbourg Astronomical Observatory	5,000	6,290	Reionization and Its Impact on The Local Universe: Witnessing Our Own Cosmic Dawn
ATM112	Wei Zhang	ORNL	60,000	71,894	Aim High: Air Force R&D Collaboration
AST192	David Vartanyan	Carnegie Institution for Science	10,000	8,820	Core-Collapse Supernovae to Shock Breakout
MAT265	Victor Fung	Georgia Tech	20,000	466	Foundational graph neural network models for materials chemistry
CSC547	Abhinav Bhatele	U. of Maryland	24,000	3,511	Performance Analysis and Tuning of HPC and AI Applications
CSC550	Ramakrishnan Kannan	ORNL	15,000	10	gnn1e12
PHY129	Alexander Tchekhovskoy	Northwestern University	20,000	0	Simulating Neutron Star Binary Merger Remnant Disks and Tilted Thin Disk
LRN031	Stephan Irle	ORNL	20,000	4,890	Inverse design of UV-vis property of molecules using a workflow with machine learning methods
CSC549	Dhabaleswar Panda	Ohio State University	20,000	17,587	MVAPICH
GEO152	Dalton Lunga	ORNL	20,000	18,569	Large Vision Models for Geospatial Applications

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
LRN033	Mayur Mudigonda	Vayuh Inc.	20,000	17,174	toward foundation models for weather forecasting using physics and deep learning
CLI138	Moetasim Ashfaq	ORNL	20,000	6,208	Analytical Frameworks for Sub-Seasonal to Multi-Decadal Climate Predictions and Impact Assessments
TUR141	Niclas Jansson	Kungliga Tekniska Högskolan (KTH Royal Institute of Technology)	20,000	5,519	Extreme-scale high-fidelity turbulence simulations of convection
CSC546	Guojing Cong	ORNL	20,000	0	Generative pretraining with graph neural networks for materials for a diverse set of property predictions
LRN036	Aristeidis Tsaris	ORNL	20,000	1,900,161	AI Foundational Model for Weather and Climate Project
BIP167	Philip Kurian	Howard University	20,000	0	Computing Many-body Van Der Waals Dispersion Effects in Biomacromolecules
CSC528	Ang Li	PNNL	18,000	734	nwqsim
PHY180	Mark van Schilfgaarde	NREL	20,000	442	Green's function description of the electron phonon interaction
CPH154	Alberto Nocera	University of British Columbia	20,000	6,433	Quantum Annealing simulation of spin glasses using large scale Tensor Network Methods
AST193	Ka Ho Yuen	LANL	15,000	6,987	Exascale multiphase magnetized turbulence simulations with cosmic ray support and chemistry network
FUS147	Noah Reddell	Zap Energy Inc.	40,000	24,014	Study of High Energy Density Z-Pinch Plasma Stability by Kinetic Model on GPU
CLI137	Forrest Hoffman	ORNL	10,000	6	Earth System Grid Federation 2 (ESGF2)
MAT268	Patxi Fernandez-Zelaia	ORNL	10,000	0	Physics informed generative AI for materials design
CHM174	Santanu Roy	ORNL	5,000	7,281	Multi-scale/multi-physics molecular simulations at the Chemical Sciences Division

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
LRN039	Allan Grosvenor	Microsurgeonbot	20,000	64,869	Autonomy for DOE Simulations
CLI181	Samuel Hatfield	European Centre for Medium-Range Weather Forecasts	20,000	3,033	Preparing for INCITE 2024
CLI900	Valentine Anantharaj	ORNL	10,000	1,687	Provisioning of Climate Data
LRN037	Pei Zhang	ORNL	20,000	49,593	Multiscale foundation models for physical systems
PHY184	Nicholas Sauter	LBNL	20,000	29,656	ExaFEL
BIP251	Joshua Vermaas	Michigan State University	20,000	20,304	Bacterial Microcompartment Permeability Assays in the Computational Microscope
CFD165	Federico Municchi	NREL	20,000	1,392	CFD simulations of heat and mass transport in solar powered membrane distillation systems
BIF147	Kathy Yelick	LBNL	20,000	20,155	Exabiome
BIP232	Pin-Kuang Lai	Stevens Institute of Technology	20,000	2	Integrating molecular dynamics simulations and machine learning to accelerate antibody development
BIP250	Micholas Smith	UTK	18,000	1,312	BPGbio
GEO158	Kaushik Bhattacharya	Caltech	20,000	44,280	A learning-based multiscale model for reactive flow in porous media
GEO157	Kohei Fujita	University of Tokyo	40,000	7,205	Finite-element seismic simulation with model optimization on Frontier
TUR143	Mahendra Verma	IIT Kanpur	15,000	9,273	Anisotropic Magnetohydrodyamic Turbulence with Applications to Solar Physics
CSC143	Norbert Podhorszki	ORNL	100,000	21,525	ADIOS - The Adaptable IO System
CFD183	Vittorio Michelassi	Nuovo Pignone	20,000	7,640	Fluent GPU performance assessment for high-efficiency waste heat recovery axial expanders.
CSC581	Mahantesh Halappanavar	PNNL	10,000	76,525	ExaGraph

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
MAT187	Danny Perez	LANL	1,000,000	1,038,718	2.2.1.04 ADSE04-EXAALT: Molecular Dynamics at the Exascale - Materials Science
AST153	Francis Alexander	BNL	500,000	248,182	2.2.6.08 ADCD08-ExaLearn: CoDesign Center for Exascale Machine Learning Technologies
ENG145	Slaven Peles	ORNL	10,000	46	ExaSGD: Stochastic Grid Dynamics at Exascale (CSC359)
CSC570	Hong-Jun Yoon	ORNL	15,000	988	Automatic information extraction from environmental mitigation plans
BIP247	Davide Mandelli	Forschungszentrum Juelich GmbH	80,000	66,694	Unraveling bacterial prodrug resistance by extremely parallel QM/MM simulations
NRO108	Xiao Wang	ORNL	20,000	3,748	Advanced Deep Learning for High-Resolution MR Mouse Brain Image Segmentation
CSC582	Susan Mniszewski	LANL	20,000	1,663	Optimizing PROGRESS/BML Libraries for Biosystems
CSC571	Arthur Lorenzon	UFRGS	20,000	8,124	Optimizing Exascale Computing for Power and Energy Sustainability
UMS002	Sameer Shende	University of Oregon	5,000	3	E4S
PHY181	Travis Humble	ORNL	80,000	107,275	Quantum Supremacy
BIP249	Laxmikant Kale	UIUC	20,000	219	Charm++ and NAMD Scaling with MPI and OFI on Frontier
CSC574	Joseph Schuchart	SUNY Stony Brook	15,000	6,925	ICL DisCo
CSC577	Ramakrishnan Kannan	ORNL	10,000	0	elements
AST208	David Radice	Pennsylvania State University	20,000	11,198	Exascale Simulations of Compact Binary Mergers
CSC588	Bogdan Nicolae	ANL	20,000	27	VELOC
CSC583	Peter McCorquodale	LBNL	10,000	392	FFTX high-performance FFT library
CSC586	Howard Pritchard	LANL	10,000	0	Open MPI

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
LRN035	Prasanna Balaprakash	ORNL	20,000	0	trillion parameter model training
MAT201	Panchapakesan Ganesh	ORNL	80,000	35,776	Center for Nanophase Materials Sciences
BIP252	Juan Perilla	U. of Delaware	20,000	697	Molecular mechanisms of HIV-1 infection
ENG146	Marcos Vanella	NIST	20,000	8,816	Physics Based Modeling for Forest Fuel Management
CSC568	Jeffrey Vetter	ORNL	10,000	80	cosmic dawn
CSC519	Maciej Cytowski	Pawsey Supercomputing Centre	5,000	0	Pawsey SC access for early testing purposes
CSC579	Hemanth Kolla	Sandia	15,000	2,426	Scalable Tensor Decompositions
CSC590	Irina Rish	University of Montreal	19,000	11,058	Large-Scale AI Models
CSC575	Michael Breitenfeld	The HDF Group	15,000	7,381	HDF5
CSC332	Robert Latham	ANL	20,000	2	2.3.4.10 STDM12-DataLib: Data Libraries and Services Enabling Exascale Science
CFD196	John Bell	LBNL	20,000	14,291	Performance Optimization of AMReX
CLI192	Gautam Bisht	PNNL	15,000	1,854	Development of River Dynamical Core for E3SM
CSC587	Jonathan Hu	Sandia	5,000	1,219	Trilinos Performance Optimization and Scaling
MAT284	John Coleman	ORNL	20,000	15,234	Process-structure predictions for metal additive manufacturing
CSC589	Chao Yang	LBNL	500	43	Spare Matrix-matrix Multiplication Kernel for Many-nucleon Calculations
CFD195	Xiaoyi Lu	FM Global	20,000	157	GPU-acceleration for large-scale fire simulations: code development and performance profiling
AST136	Bronson Messer	ORNL	20,000	866	ExaStar: Exascale Models of Stellar Explosions: Quintessential Multi-physics Simulation

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
CSC580	Stephen Hudson	ANL	15,000	6,240	Plasma Accelerator optimization using libEnsemble
NFI129	Steven Hamilton	ORNL	20,000	476	Demonstration of Monte Carlo neutron transport for analysis of nuclear systems
ENG144	Jaime Maldonado	GE Vernova	20,000	54,612	Use of Large Eddy Simulation to Predict Impact of Manufacturing Techniques on Film Cooling Effectiveness of Heavy Duty Gas Turbine Fan Shaped Diffuser Holes
LRN051	Pradeep Ramuhalli	ORNL	20,000	48	Foundations of Artificial Intelligence for Robust Engineering and Science
CSC593	Mark Gates	UTK	20,000	1,167	CLOVER
CLI193	Jerry Watkins	Sandia	10,000	238	Framework for Antarctic System Science in E3SM
MAT286	Haixuan Xu	UTK	10,000	25,358	Modeling Refractory High Entropy Alloys
MAT285	Sumit Gupta	ORNL	20,000	24,351	In silico atomistic design of hierarchically nanostructured interphases for high-performance composites
GEO160	Mark Carroll	NASA Goddard	25,000	11,725	Foundation Model for Earth Observing Satellite Data
HEP143	Seth Johnson	ORNL	15,000	107	Developing scalable simulation workflows for Celeritas
CSC567	Jeffrey Vetter	ORNL	10,000	3	brisbane
CMB162	Andrew Smith	Pratt & Whitney	20,000	40,629	Validating Multiphysics Models for Simulating Aviation Combustors
CSC591	Pieter Ghysels	LBNL	15,000	2	STRUMPACK Sparse Solver and Rank-Structured Preconditioner
CPH166	Joachim Brand	Massey University of New Zealand	20,000	15,019	Rimu.jl: massively parallelized quantum many-body problem solver in Julia
CSC597	Richard Mills	ANL	15,000	0	Developing, Benchmarking, and Sustaining PETSc/TAO Composable Solvers

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
CSC594	Jeffrey Vetter	ORNL	10,000	109	Bluestone
SYB111	Erica Teixeira Prates	ORNL	100,000	142,698	The Discovery of Longitudinal and Temporal Climatype Patterns
CLI194	David Keyes	KAUST	20,000	168,955	Saving PetaBytes in Earth System Model Outputs using Stochastic Approximations
CSC605	Irina Rish	University of Montreal	30,000	27,378	Continual Training of Foundation Models
CSC607	Oscar Hernandez Mendoza	ORNL	8,000	792	OpenSHMEM
MED130	Daoud Meerzaman	National Cancer Institute	15,000	0	Multimodal machine learning for ovarian cancer
CHM213	Giuseppe Barca	University of Melbourne	20,000	806,701	Enabling high accuracy ab initio molecular dynamics for drug discovery
LRN052	Kristian Kielhofner	Atomic Canyon, Inc.	20,000	4,633	AI Model for Search and Retrieval Trained on NRC ADAMS Dataset
ENG147	Srikanth Allu	ORNL	15,000	1	Rapid Operational Validation Initiative
CSC611	Sudip Seal	ORNL	20,000	144	A High-Speed At-Scale Framework for What-If Analysis of Disruptive Phenomena
BIP253	Monte Pettitt	University of Texas Medical Branch at Galveston	20,000	-	Understanding liquid-liquid phase transitions coupling to protein disorder
AST209	Yuan-Sen Ting	The Australian National University	10,000	21,638	Continual Pretraining of Large Language Models for Astronomy
CSC615	Hartmut Kaiser	Louisiana State University	10,000	3,537	Testing HPX-Kokkos using Octo-Tiger
BIF148	Erik Garrison	UT Health Science Center	20,000	22	Embedding Life in Multimodal DNA/Language Models
BIF152	Arvind Ramanathan	ANL	20,000	116,059	Foundation Models for Molecular Epidemiology
CSC612	Qinglei Cao	SLU	20,000	130,334	Optimizing PaRSEC on Frontier Supercomputer through Matrix Computations in Climate and Weather Prediction

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
FUS170	Rogerio Jorge	U. of Wisconsin	15,000	0	Design of Stellarator Devices using Surrogate Models
MAT291	Elie Huerta	ANL	20,000	1	MOFA: Generative AI-driven MOF discovery for carbon capture at exascale
CFD199	Kenji Miki	NASA Glenn	20,000	57	GlennHT-GPU
MAT287	Felipe Jimenez Angeles	Northwestern University	20,000	0	Computational study of ion transport in dissociable, polarizable, and electro-responsive materials
CSC596	Kangil Kim	Gwangju Institute of Science and Technology	20,000	0	Theoretical Stagnation of Training Dynamics in Neural Network Architectures for Foundation Models
CLI197	Brandi Gamelin	ANL	2,000	12	SCREAM_GPU multiple RRM for mid-century extreme events
BIP255	Debsindhu Bhowmik	ORNL	20,000	3,976	Large-scale Simulations of Functional Lipidomic Asymmetry and Membrane Permeability
NEL107	Mathieu Luisier	ETH Zurich	20,000	0	Towards exascale simulations of nanoelectronic devices in the GW approximation
CSC617	Terry Jones	ORNL	20,000	2	STEP: Software Tools Ecosystem Project
AST212	Mordecai-Mark Mac Low	American Museum of Natural History	20,000	4	Global protoplanetary disk models: gas and dust instabilities in the inner regions
MAT258	Corey Oses	Johns Hopkins University	15,000	22	Startup
NPH164	Keh-Fei Liu	University of Kentucky	20,000	5,243	Neutron electric dipole moment from lattice QCD with chiral fermions
CHM215	Konstantinos Vogiatzis	UTK	5,000	0	Molecular Engineering for Drug Innovation using Computational Intelligence
MAT289	Monica Olvera de la Cruz	Northwestern University	20,000	0	Computational investigation of confined ion transport for iontronics and neuromorphic computing

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
BIP256	Mahmoud Moradi	University of Arkansas, Fayetteville	20,000	12,470	Respiratory Syncytial Virus Fusion Glycoprotein Conformational Free Energy Landscape
LRN053	Noah Smith	Allen Institute for Artificial Intelligence	20,000	0	OLMo v.2.0
MED131	Aldo Badano	US FDA	20,000	0	Large-scale Digital Twin Simulations
NPH165	Andre Walker-Loud	LBNL	15,000	11,381	Nuclear Interactions from QCD
CSC621	Richard Pausch	Helmholtz-Zentrum Dresden-Rossendorf	60,000	1,280	Accelerating Plasma Physics on large scale systems
CSC608	Jiajia Li	NC State	10,000	0	Sparse tensor methods development for hypergraphs and more
CHP131	Sarom Leang	EP Analytics	16,000	2,867	GAMESS Performance Analysis and Debugging
CFD200	Qiong Liu	New Mexico State University	20,000	0	Extreme-scale data assimilation and resolvent analysis for urban air-shed control
MAT293	Sara Ferry	MIT	15,000	0	Primary Damage Modeling of Vanadium Multicomponent Alloys for Fusion Power Plants
AST211	Oliver Benedikt Zier	Harvard-Smithsonian Center for Astrophysics	20,000	4,064	Scaling Tests for THESAN-XL
NTI115	Phani Motamarri	IISc Bangalore	20,000	51,295	Large-scale density functional theory calculations using finite-element basis
ARD175	Luigi Martinelli	Princeton	15,700	4,669	Direct numerical simulation of the Boltzmann equation for hypersonic transitional boundary layers
CSC629	Oscar Hernandez Mendoza	ORNL	7,000	19,129	OpenSHMEM
LSC120	Marcos Sotomayor	University of Chicago	20,000	0	Large-scale all-atom molecular dynamics simulations of mechanotransduction complexes

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
CSC613	Scott Parker	ANL	20,000	25,567	Evaluation of Performance, Portability, and Programming Models on GPU Architectures
CFD201	Eric Johnsen	U. of Michigan	15,000	507	Shocks, droplets, and bubbles
BIP257	Jerry Parks	ORNL	10,000	14	Targeting group I introns in pathogenic fungi
HEP144	Charles Leggett	LBNL	5,000	2	HEP-CCE
CSC609	Seung-Hwan Lim	ORNL	15,000	24	Split-Parallel Scalable Graph Neural Network Training
HEP146	David Shih	Rutgers	10,000	0	Foundation Models for HEP
AST213	Thomas Quinn	U. of Washington	10,000	1	Modeling Galaxy Clusters
CSC624	Mariam Kiran	ORNL	10,000	37	Scalable Quantum Network Simulation and AI
GEO161	Johansell Villalobos	CeNAT	20,000	35	Advancing flood modeling with the SERGHEI code
AST215	Sebastian Keller	CSCS	30,000	0	Exascale SPH and N-body benchmarking with SPH-EXA
CHM219	Edward Valeev	VA Tech	20,000	39	mpqc
CSC623	Jong Youl Choi	ORNL	20,000	873	AI for Scientific Discovery in in Computational Fluid Dynamics
AST214	Jens Mahlmann	Columbia University	20,000	11	Next-Generation Magnetized Plasma Turbulence Modeling in Astrophysics with the GPU-PIC Code Entity
LRN065	Josh Arnold	ORNL	20,000	69	Scalable AI Model Vulnerability Evaluation
NFI131	Benjamin Collins	ORNL,	20,000	217	Coupled Microreactor Simulations
CPH172	Xiao Liang	William & Mary	20,000	19,858	Tensor representation of backflow corrections on solving strongly correlated quantum many-body systems
LGT130	Henry Monge Camacho	ORNL	5,000	0	Computing four-quark matrix elements using Feynman-Hellmann theorem
FUS172	Eugenio Schuster	Lehigh U.	20,000	4	Gyrokinetic Simulations of Plasma Instabilities in NSTX(U)

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
CLI199	Siddhartha Bishnu	MIT	20,000	0	A Verification Suite of Ocean Dynamical Core
BIE119	Abhishek Singharoy	Arizona State University	10,800	8,210	Membrane models of biological energy transfer
MED133	Avinash Sahu	University of New Mexico	20,000	19	ProtChemPilot-: AI-copilot for Translating Protein-Sequences and Chemical-Structures into Functional Insights
AST216	Yuan-Sen Ting	Ohio State University	20,000	0	Decoding the Stars: Spectral Foundational Models for Next-Generation Astrophysics
AST218	Mike Guidry	UTK	5,000	0	FERN-FENN Compute Node Development
CFD203	Yue Ling	Baylor, USC	10,000	759	2PhaseExaComp
CHP132	Rajni Chahal	ORNL	18,000	0	Development and Benchmarking of Uncertainty-Guided Automated Workflow for Trustworthy and Energy-Efficient Machine Learning Interatomic Potentials
MAT721	Elsa Olivetti	MIT	20,000	0	Iterative Construction of an Automated Synthesis Knowledge Graph
FUS173	Noah Mandell	Type One Energy Group	20,000	56	Enabling high fidelity turbulence-based predictions of stellarator reactor performance on Frontier
CMB163	Muhsin Mohammed Ameen	ANL	15,000	1	NekGT
CFD205	Marc Olivier Delchini	ORNL	20,000	5	VERTEX-CFD: Advanced Multiphysics Solver for Core Applications
MPH121	David Rogers	ORNL	20,000	2	Fast Data Reduction for Experimental X-Ray Light Source Facilities
CHP134	Peter Frazier	Cornell	20,000	0	von

Project ID	PI	Institution	Most recent frontier allocation	Frontier usage	Project name
ENG154	Samuel Fagbemi	ORNL	20,000	3	Modeling and Simulation of Efficient Non-equilibrium Thermochemical Processes for Industrial Heating Applications Using Pulsed Heating and Quenching
AST200	Evan Schneider	University of Pittsburgh	15,000	0	From Supernovae to Galactic Winds: The ISM-Halo Connection
ARD180	John Anastos	XTI Aircraft	12,000	0	Extended VTOL Aircraft and Ducted Fan Design in Transition and Flight
ARD178	Neal Domel	Lockheed Martin Corporation	50,000	0	American Institute of Aeronautics and Astronautics Propulsion Aerodynamics Workshop
LRN069	Masako Yamada	IonQ	5,000	0	Quantum + AI
CSC651	Prasanna Date	ORNL	20,000	0	SuperNeuro-HPC: An Exascale Simulator for Neuromorphic Algorithms
NPH168	Aurel Bulgac	U. of Washington	10,000	0	Nucleon transfer in heavy-ion collisions
BIF153	Yingfeng Wang	UT Chattanooga	20,000	0	Large Language Models for Breast Cancer Detection
CFD208	Siva Chaluvadi	GE Vernova	24,000	0	Massively Parallel Large Eddy Simulations for High-Efficient Gas Turbines Operating with Hydrogen and High Aerodynamic Loading

PROJECTS UTILIZING SUMMIT (ENHANCED ENCLAVE) ENABLED (AT ANY POINT) IN CY 2024

Table E-3. Director's Discretionary projects utilizing Summit (Enhanced Enclave) in CY 2024.

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
BIF144_MDE	Mayanka Chandra Shekar	ORNL	35,000	0	Cincinnati Children's Hospital Medical Center (CCHMC) Mental Health Trajectories
LRN060_MDE	Madhumita Sushil	University of California - San Francisco	17,250	906	Pre-training a generative selective state space model, the Mamba model, on UCSF-specific deidentified clinical notes and time-series structured data
MED116_MDE	Heidi Hanson	ORNL	1,000	0	Developing bio preparedness models informed by heterogeneous data streams - Partner to MED117

PROJECTS UTILIZING SUMMIT (MODERATE ENCLAVE) ENABLED (AT ANY POINT) IN CY 2024

Table E-4. Director's Discretionary projects utilizing Summit (Moderate Enclave) in CY 2024.

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
COS001	Katrin Heitmann	ANL	100,000	958	LSST DESC HOS
CLI188	David Keyes	KAUST	206,780	166,426	Saving PetaBytes in Earth System Model Outputs using Stochastic Approximations
BIP245	Rommie Amaro	UC San Diego	200,000	209,606	Interrogating Virus Aerostability in High pH Conditions
FUS164	William Fox	Princeton	250,000	152,264	Magnetic Field Generation and Reconnection in High Energy Density Plasmas
CLI187	Lawrence Cheung	Sandia	250,000	258,113	Measurement-based high fidelity wind farm simulations for realistic, complex atmospheric conditions
LRN048	Karl Pazdernik	PNNL	150,000	26,069	Scaling Multimodal Foundation Models for Scientific Document Reasoning

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
ARD172	Akanksha Baranwal	Sandia	250,000	79,326	Nonequilibrium effects in hypersonic boundary layers: DNS and data-driven RANS modeling
CPH159	Philip Kurian	Howard University	30,000	0	Computing many-body dispersion and super-radiance effects in bio-macromolecular dynamics in aqueous environments
BIP244	Abraham Lenhoff	U. of Delaware	100,000	71,648	Molecular simulation of monoclonal antibody binding to protein A
LRN047	Robert Patton	ORNL	100,000	8,282	Scalable Swarm Intelligence
MPH120	Stephan Irle	ORNL	250,000	163,847	Inverse Design of Near Infrared Fluorophores for Quantum Network Repeaters
MPH119	Vardhan Satalkar	Georgia Tech	200,000	61,850	Developing a Generative AI Model for Protein Disorder
FUS157	Jacob Merson	Rensselaer Polytechnic Institute	100,000	11,283	High-fidelity Coupled SOL Impurity Transport Simulations in 3D Complex Geometry Fusion Devices
CHM205	Giuseppe Barca	University of Melbourne	218,000	194,701	Enabling high-accuracy exascale ab initio molecular dynamics
BIP240	Jaan Mannik	UTK	110,000	56,748	Modeling key cell cycle processes in bacteria
ENG142	Slaven Peles	ORNL	100,000	49,157	High Fidelity Operational Reliability Modeling
CHM202	Vyacheslav Bryantsev	ORNL	150,000	1,587,470	Reactive, Generalizable Machine Learning Potentials for Molten Salts Modeling at Scale
BIP237	George Khelashvili	Cornell, Weill Cornell Medicine	200,000	207,579	Molecular mechanisms of GPCR-mediated phospholipid scrambling
CMB153	Joseph Oefelein	Georgia Tech	250,000	141,925	Benchmark Simulations of Turbulent Multiphysics Processes in a Laboratory-Scale Supersonic Combustor
CHM196	Walter Chapman	Rice U.	100,000	16,413	Permeability of Gases in Polymers at Cryogenic Conditions from Molecular Simulation
NFI127	Dillon Shaver	ANL	100,000	590,829	CFD for Advanced Nuclear Reactors

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
NRO109	Thomas Uram	ANL	100,000	16,797	PeakBrain: Generalizable segmentation models for connectomics
LGT127	Hai-Tao Shu	BNL	297,000	277,788	Searching for the critical end point using lattice QCD
BIP242	James Gumbart	Georgia Tech	173,600	4,825	Dynamics and energetics of bacterial pili extension and retraction
BIF143	Jens Glaser	ORNL	211,344	721,680	Contrastive Learning for Drug Discovery
ARD163	Zhi Wang	University of Kansas	250,000	171,155	High Order Wall-Modeled Large Eddy Simulation of High-Lift Configurations
CHM207	Alex Ivanov	ORNL	100,000	4,279	Revealing Promethium Aqua Ion Chemistry Using Relativistic Calculations and Machine Learning Approaches
AST198	Bart Ripperda	University of Toronto	250,000	446,640	Radiative MHD of bright transients from neutron stars
CSC559	Vicente Leyton Ortega	ORNL	120,000	33	Exploring the Frontiers of Simulated Quantum Computing: Performance and Limitations of Tensor Network Simulators in Solving MaxCut and Many-Body Problems
AST203	Austin Harris	ORNL	250,000	141,430	Simulations of core-collapse supernovae in 3D with rotation
CSC562	Peter Groszkowski	ORNL	20,000	526	Exploring and benchmarking prospects for HPC-Quantum integration on a leadership-scale computing platform
AST204	Yuran Chen	WashU, WUSTL	100,000	97,053	PIC Simulations of the Polarized X-ray Emission from Magnetars
ATM145	Ngoc-Cuong Nguyen	MIT	20,000	0	High-Fidelity Physics-Based Modeling of the Ionosphere-Thermosphere
ATM146	Walter Hannah	LLNL	183,000	392,095	Energy Exascale Earth System Model Project
LGT128	Henry Monge Camacho	ORNL	280,000	430,383	Electromagnetic and strong isospin breaking corrections to strong dynamics

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
BIP243	Ron Dror	Stanford	235,620	268,927	Revealing the Structural Basis of Functional Selectivity to Create Safe, Effective Drugs
MAT272	Ayana Ghosh	ORNL	250,000	257,188	Domain dynamics of ferroelectric heterostructures at large scale using causal-informed scientific machine learning and atomistic simulations
CFD185	Parisa Mirbod	UIC	170,000	1,770	Interface-resolved simulations of scalar transport in turbulent bubbly flows
CHM198	Tjerk Straatsma	ORNL	250,000	260,365	Comparative Performance Analysis of Programming Models Used in GronOR
CFD189	Wesley Harris	MIT	250,000	519,649	Direct numerical simulations of hypersonic boundary layer receptivity and transition
CHP125	Ang Li	PNNL	150,000	10,665	Advanced Numerical Simulation for Measurement based Computing of Quantum Chemistry
CHM206	Thanh Do	UTK	200,000	63,687	Developing a workflow for the automation of large-scale parallel tempering MD simulations for advancing drug discovery
CMB156	Jorge Salinas	Sandia	240,000	114,359	Thermal and chemical nonequilibrium effects in detonation waves revealed by high-fidelity simulations
CHM208	Vanda Glezakou	ORNL	100,000	81,096	Structure and reactivity at complex interfaces
CMB157	Bruno Souza Soriano	Sandia	250,000	196,297	Fundamental study of soot formation and flame dynamics of sustainable aviation fuel using DNS
CHP129	Jonathan Nickels	University of Cincinnati	130,000	3,101	Excess entropy strategy for constraining AI parameterized force fields from ab initio simulation.
CLI180	Peter Thornton	ORNL	128,000	89,824	High-Resolution E3SM Land Model on GPUs
CSC564	Richard Mills	ANL	30,000	0	Scalable Simulation and Data Analytics with PETSc
FUS160	Abdourahmane Diaw	ORNL	250,000	12,681	Tungsten Erosion Modeling in WEST through Synthetic Diagnostic

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
FUS161	Randy Churchill	Princeton Plasma Physics Laboratory	250,000	57,810	Frameworks for Multiscale Transport Modeling in Fusion Plasmas
GEO153	Philip Maechling	USC	244,775	384,894	SCEC Earthquake Ground Motion Modeling Research
GEO155	David McCallen	LBNL	260,000	259,543	EQSIM regional earthquake simulations for San Francisco Bay Area
LRN045	Ryan Coffee	SLAC National Accelerator Laboratory	50,000	0	Dynamic Information Flow for Secure and Real-Time Integration of Edge and HPC
LRN046	Anima Anandkumar	Caltech	100,000	2,067	Neural Operators for Learning Multi-Scale Multi-Physics Processes
CMB159	Shrey Trivedi	Sandia	200,000	90,948	Molecular level simulations of reacting flows under thermal and chemical non-equilibrium
MAT267	Sophie Blondel	UTK	100,000	14,909	Modeling plasma facing and structural materials for fusion applications
MAT273	Victor Fung	Georgia Tech	180,000	475	Foundational graph neural network models for chemistry and materials science
CSC556	Juri Papay	UKRI-STFC	100,000	3,225	SciMLBench
MAT276	Swarnava Ghosh	ORNL	250,000	818,991	First-principles understanding of the electronic and magnetic properties of doped-magnetic quantum topological materials
MAT277	Stephen DeWitt	ORNL	250,000	117,578	Interplay between length scales as spot melts solidify at varied power profiles
NPH160	Martha Constantinou	Temple University	250,000	711,214	pion and kaon twist-3 GPDs
MPH118	Travis Wheeler	University of Arizona	288,000	21,476	6,000 kinase simulations for a new molecular dynamics repository
NPH161	Keh-Fei Liu	University of Kentucky	250,000	290,692	Gravitational Form Factors from Lattice QCD
AST196	Mark Krumholz	The Australian National University	115,000	706,187	Metal Loading of Galactic Winds
SYB112	Daniel Jacobson	ORNL	200,000	549,347	Building Ensembles of Single Cell Predictive Expression Networks

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
MED123	Dario Alfe	University College London	150,000	48,397	Modelling new materials for hydrogen storage applications
FUS158	Derek Schaeffer	UCLA	150,000	57,632	Kinetic Simulations of Quasi-Parallel Collisionless Shocks in Laboratory Plasmas
CFD186	Myoungkyu Lee	University of Houston	200,000	190,557	Particle-Resolved Direct Numerical Simulation of Particulate Buoyancy-Driven Turbulent Convection
BIP236	Emad Tajkhorshid	UIUC	200,000	189,497	Microscopic Characterization of the Full Transport Cycle of a Major Neurotransmitter Transporter in Human Brain
BIP246	Aleksei Aksimentiev	UIUC	117,500	164,753	Decoding Sequence-Dependent Dynamics of Holliday Junctions in DNA Self-Assembly and Beyond
CSC565	Irina Rish	University of Montreal	250,000	1,089,738	Large-Scale Multimodal AI Foundation Models
CPH156	Marina Filip	Oxford	250,000	142,245	Photophysics of Excitons in Low-Dimensional Organic-Inorganic Semiconductors
NFI126	Aurel Bulgac	U. of Washington	250,000	249,500	Microscopic Framework for Fission Dynamics of Odd-Mass Nuclei
ARD166	Sanjiva Lele	Stanford	235,000	114,758	Shock Unsteadiness in Transonic Flow over Supercritical Laminar Flow Control Airfoil
ARD169	Jonathan MacArt	Notre Dame	221,000	101,738	Deep Learning Closure for Large Eddy Simulation of Transitional Hypersonic Shockwave-Boundary Layer Interactions
AST199	Christian Cardall	ORNL	100,000	15,206	Ensemble Surveys of Core-collapse Supernovae
AST200	Evan Schneider	University of Pittsburgh	200,000	155,647	From Supernovae to Galactic Winds: The ISM-Halo Connection
ATM148	Rahul Ramachandran	NASA Marshall Space Flight Center	128,476	64,442	Understanding Extreme Weather Events with AI Forecast Emulators
CHM203	Ramanan Sankaran	ORNL	250,000	49,586	Modeling and Simulation of the Non-Equilibrium Energy Transfer for Efficient Reactions

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
CHM209	Kenneth Jordan	University of Pittsburgh	250,000	8	Multi-determinant diffusion Monte Carlo calculations of isomers of C20 fullerene
CHP126	Roberto Car	Princeton	250,000	241,423	Deep potential molecular dynamics of electrochemical and atmospherically-relevant aqueous interfaces
FUS159	Kevin Woller	MIT	150,000	146,871	Computational Design of Multi-principle element alloys for fusion energy
MAT270	Lang Yuan	USC	250,000	642	Multiscale Modeling of Subgrain Cellular Structure across Melt Pools in Additive Manufacturing
NFI128	April Novak	UIUC	173,000	243,248	Advanced Computing for Scientific Discovery of Molten Salt Reactor Dynamics
NPH162	Andre Walker-Loud	LBNL	250,000	81,073	Nuclear interactions from QCD
PHY182	Travis Humble	ORNL	110,000	201,641	Quantum Supremacy
PHY185	Revathi Jambunathan	LBNL	211,000	44,721	First-principles QED-PIC Simulations of High Energy Emission from Pulsars
TUR145	Suhas Suresh	Georgia Tech	140,000	56,070	High-fidelity simulations of particle-laden turbulent separating flows
CPH160	Alberto Nocera	The University of British Columbia	50,000	0	Probing dynamical correlations and information scrambling in Quantum Annealing devices using GPU optimized Tensor Network Methods
FUS155	Noah Reddell	HPE	300,000	315,843	Study of Z Pinch Plasma by 3D Kinetic Model on slimmed-memory GPU
MAT275	Zongtang Fang	Nissan Technical Center North America	100,000	23,304	First-Principles Study of NMC-Carbon Interfaces
TUR144	Ali Uzun	Analytical Mechanics Associates	250,000	637,350	Direct Numerical Simulation of Smooth-Body Flow Separation at a High Reynolds Number
ARD168	Michal Osusky	GE	150,000	280,858	Development of LES-informed AI/ML models for vortical flows in gas turbines

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
CSC555	Jana Thayer	SLAC National Accelerator Laboratory	100,000	256	ILLUMINE
LRN044	FREDERIC POITEVIN	SLAC National Accelerator Laboratory	250,000	53,594	Building Foundational and Surrogate Models for Experiment Steering at LCLS
MAT274	Kwangnam Kim	LLNL	250,000	13,901	Rational Design of High-Performing electrodes in Energy Storage Devices
MED125	Diego Rossinelli	Stanford	100,000	49,434	Large-Scale Investigation of CSF Dynamics in the Human Optic Nerve
CFD191	Aditya Nair	University of Nevada - Reno	240,672	148,515	High-fidelity computation of high-Reynolds number multiphysics problems
CFD180	Ishan Srivastava	LBNL	30,000	20,746	HPC4EI-CapraBiosciences
ARD167	Vineet Ahuja	Whisper Aero Inc.	100,000	28,800	Aerodynamic and Aeroacoustic Simulations of a Regional Air Mobility Aircraft with Distributed Electric Propulsion
CFD184	Marcus Blohm	GE Vernova	247,400	2,652,052	Massively Parallel Large Eddy Simulations for High-Efficient Gas Turbines Operating with Hydrogen and High Aerodynamic Loading
LRN038	Allan Grosvenor	Microsurgeonbot	75,000	131,936	Autonomously Driven Software – SummitPLUS Pursuit of Level 4 Autonomy
FUS156	Walter Guttenfelder	Princeton Plasma Physics Laboratory	250,000	233,178	Stellarator performance predictions
AST136	Bronson Messer	ORNL	5,000	0	ExaStar: Exascale Models of Stellar Explosions: Quintessential Multi-physics Simulation
UMS018	Damien Lebrun-Grandie	ORNL	500	3	Kokkos
CFD188	Muhsin Ameen	ANL	100,000	17,694	High-Fidelity Simulations of Sustainable Propulsion and Power Generation Systems
LSC119	Aditya Balu	Iowa State	100,000	6,114	Development of foundational AI models for Agriculture
CSC452	Abhinav Bhatele	LLNL, U. of Maryland	1,000	405	Performance Analysis and Tuning of HPC and AI Applications

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
CFD190	Marc Olivier Delchini	ORNL	100,000	56,591	VERTEX—Advanced Multiphysics Simulations for Core Applications
CPH167	Feliciano Giustino	UT Austin	30,000	4,493	Computational design of novel semiconductors for power and energy applications
ARD171	Russell Powers	Department of Defense	200,000	227,100	Coupled Fan Motor Assembly and Wind Tunnel Circuit Predictions using WMLES
TUR146	Balaji Jayaraman	GE	249,600	145,972	Virtual Characterization of Atlantic Coast Windfarms using High-fidelity Large-eddy Simulations
MAT290	Addis Fuhr	ORNL	50,000	287,197	Digital Twins to guide Autonomous Synthesis of Hybrid Quantum Heterostructures
LRN059	David Gagne	UCAR	500,000	0	Community Runnable Earth Digital Intelligence Twin Training and Reforecasts
UMS014	Paul Hargrove	LBNL	1,000	2	Pagoda: UPC++/GASNet for Lightweight Communication and Global Address Space Support
LRN061	John Winder	Johns Hopkins University	175,000	2	Exploring Iterative Reasoning in LLMs and the Transferability to Multimodal Models
LRN055	Abhinav Shrivastava	U. of Maryland	100,000	2,100	Unified Representation Learning
LRN056	Anuj Karpatne	VA Tech	750,000	0	MRA: Advancing process understanding of lake water quality to macrosystem scales with knowledge-guided machine learning
LRN063	Anju Gupta	University of Toledo	500	0	AI-Driven Unbiased Career-Life Assessment for the ADVANCEment Marginalized STEM Faculty
LRN060	Madhumita Sushil	University of California - San Francisco	17,250	0	Pre-training a generative selective state space model, the Mamba model, on UCSF-specific deidentified clinical notes and time-series structured data
LRN062	Yuankai Huo	Vanderbilt	100,000	1,925	Securing Healthcare Privacy: Rendering Large-Scale Unlearnable Medical Imaging Data to Prevent Data Leaks

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
CHM161	Sara Isbill	ORNL	100,000	178,290	Informing Forensics Investigations of Nuclear Materials
CMB148	Aditya Konduri	IISc Bangalore	1,000	0	Scalable Mathematically Asynchronous Algorithms For Flow Solvers
MED112	Sumitra Muralidhar	VA	1,000	746	Genome-Wide Phenome-Wide Association Study in the Million Veteran Program
CSC380	Bronson Messer	ORNL	50,000	41,228	CAAR for Frontier
MED132	Anuj Kapadia	ORNL	150,000	13,790	Transforming CT Imaging: Optimizing Protocols with Virtual Trials on Summit Supercomputer