ORNL/SPR-2023/2901

US Department of Energy, Office of Science High Performance Computing Facility Operational Assessment 2022 Oak Ridge Leadership Computing Facility

April 2023

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

DOCUMENT AVAILABILITY

Reports produced after January 1, 1996, are generally available free via OSTI.GOV.

Website www.osti.gov

Reports produced before January 1, 1996, may be purchased by members of the public from the following source:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone 703-605-6000 (1-800-553-6847)
TDD 703-487-4639
Fax 703-605-6900
E-mail info@ntis.gov
Website http://classic.ntis.gov/

Reports are available to DOE employees, DOE contractors, Energy Technology Data Exchange representatives, and International Nuclear Information System representatives from the following source:

Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831
Telephone 865-576-8401
Fax 865-576-5728
E-mail reports@osti.gov
Website https://www.osti.gov/

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL/SPR-2023/2901

Oak Ridge Leadership Computing Facility

US DEPARTMENT OF ENERGY, OFFICE OF SCIENCE HIGH PERFORMANCE COMPUTING FACILITY OPERATIONAL ASSESSMENT 2022 OAK RIDGE LEADERSHIP COMPUTING FACILITY

Subil Abraham
J. Paul Abston
Ryan M. Adamson
Scott Atchley
Aaron Barlow
Ashley Barker
Tom Beck
Katie L. Bethea
Michael Brim
Christopher Fuson
Olga Kuchar
Ryan Landfield
Jack Lange
Hong Liu
Don Maxwell
Veronica Melesse Vergara
Bronson Messer
Ross Miller
Sheila Moore
James Morgan
Sarp Oral
Thomas Papatheodore
Suzanne Parete-Koon
Ryan Prout
William Renaud
James Rogers
Michael Sandoval
Mallikarjun Shankar
Scott Simmerman
Betsy Sonewald
In-Saeng Suh
Kevin G. Thach
Suzy Tichenor
Georgia Tourassi
Feiyi Wang
Spencer Ward
Justin L. Whitt
Christopher Zimmer

April 2023

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

ACCEL	Accelerating Competitiveness through Computational ExceLlence
AD	Application Development
ADAC	Accelerated Data Analytics and Computing
AFW	Air Force Weather
ALCC	ASCR Leadership Computing Challenge
ALCF	Argonne Leadership Computing Facility
APPL	Advanced Plant Phenotyping Laboratory
Argonne	Argonne National Laboratory
ASCR	Advanced Scientific Computing Research
BNL	Brookhaven National Laboratory
BSD	Biosciences Division
CAAR	Center for Accelerated Application Readiness
CBI	Center for Bioenergy Innovation
CCSD	Computing and Computational Sciences Directorate
CNMS	Center for Nanophase Materials Science
CNN	convolutional neural network
CSEEN	Computational Scientists for Energy, the Environment, and National Security
DARTs	Days Away Restricted or Transferred
DD	Director's Discretionary projects
DMP	data management plan
DOE	US Department of Energy
DVA	data visualization and analytics
ECP	Exascale Computing Project
ES&H	environment, safety, and health
FAIR	Findability, Accessibility, Interoperability, and Reuse
Fermilab	Fermi National Accelerator Laboratory
FL	federated learning
GPFS	General Parallel File System
HBCU	Historically Black Colleges and Universities
HIPAA	Health Insurance Portability and Accountability Act
HPC	high-performance computing
HPL	High Performance Linpack
HPL-MxP	High Performance Linpack-Mixed Precision
HPSS	High-Performance Storage System
HVAC	High Velocity AI Cache
IMDB	Internet Movie Database
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
INCITS	International Committee for Information Technology Standards
ISM	Integrated Safety Management
ITAR	International Traffic and Arms Regulations
JIF	Journal Impact Factor
LANL	Los Alamos National Laboratory
LASSO	LES ARM Symbiotic Simulation and Observability
LBNL	Lawrence Berkeley National Laboratory
LCF	Leadership Computing Facility
LES	large-eddy simulation
LLNL	Lawrence Livermore National Laboratory
LSMS	Locally Self-Consistent Multiple Scattering code

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
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
NIH	National Institutes of Health
NISQ	Noisy Intermediate Scale Quantum
NLP	natural language processing
NVFlare	NVIDIA Federated Learning Application Runtime Environment
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
PHI	personal health information
PI	principal investigator
PNNL	Pacific Northwest National Laboratory
QC	quantum computing
QCUP	Quantum Computing User Program
QNLP	quantum natural language processing
QRUC	Quantum Resource Utilization Council
RATS	Resource and Allocation Tracking System
RENEW	Reaching a New Energy Sciences Workforce
REST	representational state transfer
RSS	research safety summary
SA	scheduled availability
Sandia	Sandia National Laboratories
SBMS	Standards Based Management System
SC	International Conference for High Performance Computing, Networking, Storage, and Analysis
SDK	software development kit
SEER	Surveillance, Epidemiology, and End Results program
SGCI	Science Gateways Community Initiative
SME	subject matter expert
SPOKE	Scalable Precision Medicine Open Knowledge Engine
SSH	Secure Shell
ST	Science and Technology
SU	system utilization
UCSF	University of California, San Francisco
VA	US Department of Veterans Affairs
VC	venture capital
WDTS	Workforce Development for Teachers and Scientists

2022 0 – Executive Summary

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

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. As a US Department of Energy (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 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) 2022, the OLCF supported 1,681 users and 570 projects and exceeded all targets for user satisfaction. The facility received an average satisfaction score of 4.6 out of 5 on the annual user survey, and 96% of respondents reported a high satisfaction rate with the OLCF overall. Of the 3,212 user tickets submitted in CY 2022, OLCF staff resolved 97% within 3 business days. The facility also introduced several new services for users this year, including weekly virtual office hours with subject matter experts from ORNL and vendor partners; new views in MyOLCF that allow users to analyze allocation and compute usage for a project; the ability to build and run containers on Summit; and improved data visualization support and training resources.

Throughout the year, OLCF staff ensured that users had reliable access to the following HPC resources: the IBM AC922 Summit, the General Parallel File System (Alpine), and the archival storage system (High-Performance Storage System). Overall availability for both the compute resources and file systems exceeded 97% for CY 2022. The facility also successfully delivered on the allocation split of roughly 60%, 20%, and 20% of core-hours offered for the Innovative and Novel Computational Impact on Theory and Experiment (INCITE), the Advanced Scientific Computing Research Leadership Computing Challenge (ALCC), and Director’s Discretionary (DD) programs, respectively (Section 2). Over 800,000 Summit hours were dedicated to the COVID-19 High Performance Computing Consortium in CY 2022 through both the ALCC and DD programs. Table ES.1 summarizes the 2022 OLCF metric targets and the associated results. More information can be found in Section 2 for each OLCF resource.

Table ES.1. 2022 OLCF metric summary.

Metric description	CY 2022 Target	CY 2022 Actual
Overall OLCF score on the user survey will be 3.5 based on a statistically meaningful sample.	3.5	4.6
Time between Receipt of User Query (RT 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, Summit	35%	56.51%
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 SA to users, measured as a percentage of maximum possible scheduled.	95%	99.52%
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 OA to users measured as a percentage of maximum possible.	90%	97.87%
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 Alpine: Sustain availability to users measured as a percentage of maximum possible.	90%	97.57%
OA, archive storage: Sustain availability to users measured as a percentage of maximum possible.	90%	98.58%
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, Alpine: Sustain SA to users measured as a percentage of maximum possible scheduled.	95%	99.47%
SA, High Performance Storage System: Sustain SA to users measured as a percentage of maximum possible scheduled.	95%	99.88%

CY 2022 was a year of significant technological development and innovation for the OLCF. In May, the HPE Cray EX Frontier debuted as No. 1 on the TOP500 list as the first supercomputer to exceed 1 exaflops on the High Performance Linpack (HPL) benchmark. Beyond this achievement, the OLCF implemented several technical innovations that improved operations, including Advanced Computing Ecosystem Evaluations and test beds; leadership-scale capability benchmarks; scalable data hierarchy management; operational data workflow development; classical-quantum hybrid workflows; improved energy efficiency in Frontier; and launched efforts toward Federated Learning.

To maintain a healthy management and workforce structure, the OLCF pursued a multipronged pipeline approach, starting with student programs and followed by on-the-job training programs. These efforts were supported by a focused recruitment campaign developed in collaboration with HR and ORNL Communications.

The OLCF supported 13 postdocs in CY 2022 who produced excellent research and contributed to projects across the division.

Throughout the year, the OLCF maintained a strong culture of operational excellence, including risk management, workplace safety, and cybersecurity. The OLCF's rigorous risk management strategy anticipated and mitigated risks. Similarly, ORNL and the OLCF were committed to operating under the DOE's safety regulations that ensure a safe workplace. Technical staff tracked and monitored existing threats and vulnerabilities within the OLCF while continually developing tools and practices to enhance operations without increasing the facility's risk.

The OLCF carried a strong safety culture through its day-to-day operations, and the facility met the zero accident performance criteria during large-scale workloads and multiple supercomputer installations.

The scientific accomplishments of OLCF users are a strong indication of long-term operational success and broad scientific impact, and this year's user projects and programs have advanced strategic objectives from DOE. Users published 49 papers in high-impact journals such as *Science*, *Nature Physics*, *Nature Climate Change*, *Journal of the American Chemical Society*, and *Advanced Materials*. Crucial domain-specific discoveries facilitated by resources at the OLCF are described in the *High Performance Computing Facility Operational Assessment 2022: Oak Ridge Leadership Computing Facility* (OAR), Section 0.

The OLCF supported scientific accomplishments for a broad community of researchers in 2022, from traditional modeling and simulation projects to studies that exploit AI, ML, and big data analytics techniques. Several finalist teams, including one of the winning teams, used OLCF resources to conduct research in the competition for the prestigious Gordon Bell Prize. 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. Another finalist team used Summit for research that was nominated for the Gordon Bell Prize for HPC-based COVID-19 research.

This year, OLCF systems supported multiple projects that tackled fundamental scientific questions. For example, a team from Princeton University used Summit to produce the largest, most accurate simulation of nucleation—how water turns into ice. Data from their study could be used to improve the accuracy of weather and climate modeling or to advance the process of flash-freezing food. This study is one of the most significant applications of DeePMD, a simulation that won the 2020 Gordon Bell Prize. A team from Arizona State University modeled the molecular structures of proteins on Summit on the atomic level. A better understanding of the molecular structures of proteins can help the search for drug candidates to treat diseases.

Other studies used OLCF's Quantum Computing User Program (QCUP) to access the power of an IBM Q quantum computer to capture part of a calculation of two protons colliding. The team developed a quantum algorithm that runs part of the calculation on a quantum computer while allowing other calculations to run on classical computers.

The OLCF continued their strong engagement within the community to build the next generation HPC workforce by leading hands-on training sessions and participating in competitions and hackathons to reach audiences that are often underrepresented in the HPC community. OLCF staff were prominent participants at several conferences and events, such as the International Conference for High Performance Computing, Networking, Storage and Analysis, the PEARC Student volunteer program, and the Tapia Diversity in Computing Conference, where they led workshops and introductory training for students new to HPC.

In CY 2022, the OLCF continued to support and strengthen several strategic partnerships with other DOE Office of Science programs and federal agencies. These partnerships, which include ongoing collaboration and system support for Atmospheric Radiation Measurement, the National Cancer Institute, the US Department of Veterans Affairs, Air Force Weather, and NOAA, increase community engagement, contribute to the development of new software and capabilities, and support overall operational innovation.

The successful operation of the OLCF is the result of the extraordinary work by staff members in supporting the most capable HPC user facility in the world. The OLCF's staff is instrumental in identifying, developing, and deploying the innovative processes and technologies that support the advancement of science by the OLCF users and benefit the global HPC 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 OAR, an annual "Budget Deep Dive," an annual independent project review, and the OLCF annual report. 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 OCLF'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 2022 annual user survey, OLCF communications received a mean rating for users' overall satisfaction of 4.6. The OLCF uses a variety of methods to communicate with users, including the following:

- Weekly email message
- Welcome packet
- 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 networking
- Annual OLCF User Meeting
- Targeted training events (e.g., GPU hackathons or tutorials)

ES.2 SUMMARY OF 2022 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 2022. The 2022 metrics, target values, and actual results as of December 31, 2022, are noted throughout this report. The OLCF exceeded all agreed-upon metric targets.

ES.3 RESPONSES TO RECOMMENDATIONS FROM THE 2021 OPERATIONAL ASSESSMENT REVIEW

In April 2022 the operational activities of the OLCF for CY 2021 were presented to the DOE sponsor. The review committee of that report identified no recommendations.

2022 1 – User Support Results

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

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) at the US Department of Energy's (DOE's) Oak Ridge National Laboratory (ORNL) supported 1,681 users and 570 projects (including Quantum) in CY 2022. The OLCF continued to leverage an established user support model for effectively supporting users based on continuous improvement, regular assessment, and a strong customer focus. One key element of internal assessment is the annual user survey. As part of the survey, users are asked to rate their overall satisfaction with the OLCF on a scale of 1 to 5, with a rating of 5 indicating "very satisfied." The mean rating for overall satisfaction with the OLCF in 2022 was 4.6. Overall ratings for the OLCF were positive: 96% of all survey respondents reported being "satisfied" or "very satisfied" with the OLCF.

The OLCF measures its performance by using a series of quantifiable metrics. The metric targets are structured to ensure that users are provided prompt and effective support and that the organization responds quickly and effectively to improve its support process for any item that does not meet a minimum satisfactory score. The OLCF exceeded all metric targets for user satisfaction in 2022: 97% of tickets were resolved within 3 business days. The OLCF continued to enhance its technical support, collaboration, training, outreach, and communication and engaged in activities that promoted high-performance computing (HPC) to the next generation of researchers.

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 analysts, and scientific and data liaisons; multiple channels for stakeholder communication, including the OLCF User Council; 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 2022 survey was launched on October 12, 2022 and remained open for participation through November 28, 2022. The survey was sent to 1,396 users of the Innovative and Novel Computational Impact on Theory and Experiment (INCITE), Advanced Scientific Computing Research (ASCR) Leadership Computing Challenge (ALCC), and Director's Discretionary (DD) projects. These users also include members of the Exascale Computing Project (ECP) projects who logged into an OLCF system between January 01, 2022, and September 30, 2022. Although the ECP allocations come from the OLCF DD program, we once again tracked those responses separately from the DD responses. OLCF staff members were excluded from participation. A total of 645 users completed the survey for an overall response rate of 46%, as shown in Table 1-1. The results of the 2022 survey can be found on the OLCF website at https://www.olcf.ornl.gov/wp-content/uploads/2022-olcf-user-survey-report.pdf.

Table 1-1. Annual survey response rate for 2022.

Survey response	2021 target	2021 actual	2022 target	2022 actual
Number of users the survey was sent to	N/A	1,443	N/A	1,396
Total number of respondents	N/A	706	N/A	645
Percent responding	N/A	49%	N/A	46%

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 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 in part measured by the key survey responses for User Support in Table 1-2.

Table 1-2. Key survey OLCF user responses for 2022.

Survey question	2021 target	2021 actual	2022 target	2022 actual
Overall OLCF satisfaction score on the user survey	3.5/5.0	4.6/5.0	3.5/5.0	4.6/5.0
Overall Satisfaction with OLCF Support	3.5/5.0	4.6/5.0	3.5/5.0	4.6/5.0
Overall Satisfaction with OLCF Services	3.5/5.0	4.4/5.0	3.5/5.0	4.5/5.0
Overall Satisfaction with the Website	3.5/5.0	4.6/5.0	3.5/5.0	4.5/5.0
Overall Satisfaction with Communications	3.5/5.0	4.4/5.0	3.5/5.0	4.6/5.0
Overall Satisfaction with Problem Resolution	3.5/5.0	4.5/5.0	3.5/5.0	4.7/5.0
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 2021 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 2022 survey.

1.2 USER SUPPORT METRICS

The OLCF exceeded all user support metrics for 2022. The OLCF metric targets and actual results by CY for user support are shown in Table 1-3.

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

Survey area	CY 2021 target	CY 2021 actual	CY 2022 target	CY 2022 actual
Overall OLCF satisfaction rating	3.5/5.0	4.6/5.0	3.5/5.0	4.6/5.0

Average of all user support services ratings	3.5/5.0	4.5/5.0	3.5/5.0	4.5/5.0

1.2.1 Average Rating across All User Support Questions

The calculated mean of answers to the user support services–specific questions on the 2022 survey was 4.5/5.0, thereby indicating that the OLCF exceeded the 2022 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 (97%), communications (96%), documentation (96%), the projects and accounts team (96%), user assistance (95%), and the INCITE liaisons (95%). Included below are select open-ended responses to "What are the best qualities of the OLCF?" that highlight various aspects of user support:

In my experience the best qualities of OLCF are:

- "A leading-edge HPC user facility with comprehensive hardware/software ecosystem and robust support services. OLCF does really cater to a very broad set of users with diverse needs."

- "Apart from the awesome computational resources, for me the best quality of OLCF is the top-notch staff, in particular that liaising with users."

- "Highest compute power, good system availability, excellent system documentation and outstanding (in my experience) availability of user support."

- "OLCF has the best machines and the most competent technical team. You are the center I go to for my biggest challenge runs. Thank you!"

- "OLCF is the only place where I can do the kind of computations that I want to do. If you want to push the boundaries of science with a big computer, OLCF is the place to make it happen."

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. All questions scored above 3.5 on both the 2021 and 2022 surveys. Although we had no results that scored below satisfactory on the 2022 survey, we did a thorough review of the survey and identified areas in which we could improve or add new services to enhance the user experience at the OLCF. In some cases, the issue has already been addressed, or a solution is in the works and forthcoming.

- Ability for quick video chat – office hours
    - In May 2022, we began offering office hours for users of the Frontier Test and Development System, Crusher. The Crusher Office Hours are held weekly on Mondays and include staff from ORNL, HPE, and AMD. Users must sign up in advance by Thursday to attend the following Monday's session. Each office hour is limited to 5 different topics, and each one is assigned its own breakout room in Zoom. In the sign-up form, the project team must include information about the topic and details on the issues they would like to discuss. This helps us coordinate with vendors to ensure someone knowledgeable in the subject area can be available. From May to December 2022, representatives from 29 projects attended Crusher Office Hours to discuss issues and ask questions.

- Improve user ability to view allocation usage
  - We have continued to add new capabilities within MyOLCF. The initial capabilities were for analyzing system use on a user and/or project basis. When users log on to the MyOLCF GUI, they now have two new views: *Allocations* and *Analytics*. These new views enable users to analyze historical and current allocation usage and compute usage broken down by project and project members. We plan to continue growing the analytics available to users from MyOLCF, and these two capabilities are the first steps of that effort.

- Containers on Summit
  - The ability to build and run containers on Summit's compute resources was requested by users in the 2021 user survey. The OLCF provided the ability for Summit to run containers with Singularity, but users had to bring container images built on resources outside the OLCF. For Summit, building a container outside of Summit was not a straightforward endeavor. Users could not generally build their containers elsewhere because they would need access (with extra privileges) to a system with similar hardware as Summit, which is an extremely rare commodity. Building on Summit required extra privileges that could not be granted to the user community due to security concerns. The OLCF worked with Red Hat to update their Podman container framework to provide the ability to build containers on Summit's login nodes without the need for the extra privileges. This allowed users to build their own containers without needing separate hardware, thereby fulfilling the requests from the survey. The combined use of Podman and Singularity was tested, as was providing the appropriate separate storage locations on the login nodes needed by Podman for proper container building. Significant documentation with examples was also added to explain the process of building and running containers for various purposes (https://docs.olcf.ornl.gov/software/containers_on_summit.html), and a tutorial was presented during an OLCF user conference call.

- Improved data visualization support
  - From the 2021 user survey, users found the *getting started* process for visualization tools to be confusing, were frustrated with the current versions of ParaView and VisIt on both Summit and Andes and requested more data visualization training. To help streamline the getting started process, we updated the documentation (https://docs.olcf.ornl.gov/software/viz_tools/index.html) to include beginner-friendly quick start tutorials for ParaView and VisIt. In response to user feedback on older ParaView versions, the tool was regularly updated throughout the year on Andes and was similarly installed and newly maintained on Summit. VisIt, which did not reliably work in parallel jobs on Andes, was fixed and rebuilt by the User Assistance team, thus allowing the tool to be properly used again on Andes.
  - Building on the December 2021 user call and 2021 user survey feedback, the OLCF created the Data Visualization and Analytics (DVA) training series in 2022. The series is described in detail in Section 1.4.4.3. The hands-on events covered how to efficiently use the two main visualization tools (VisIt and ParaView) on our systems, along with how to analyze data by using Jupyter and NVIDIA Rapids on both Summit and Andes.

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.

- Average satisfaction ratings for questions on the user survey related to problem resolution are satisfactory or better.

1.3.1 Problem Resolution Metric Summary

In most instances, 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 users 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 2022, the OLCF issued 3,212 tickets in response to user inquiries. The OLCF resolved 97% of issues within 3 business days, as shown in Table 1-4.

More than 3/4 (86%) of the survey respondents submitted between one and five queries to the OLCF (via phone or email) in 2022, and 95% 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 (95%) and the quality of technical advice given (96%).

Table 1-4. Problem resolution metric summary.

Survey area	CY 2021		CY 2022	
	Target	Actual	Target	Actual
Percentage of problems addressed in 3 business days	80%	96%	80%	97%
Average of problem resolution ratings	3.5/5.0	4.6	3.5/5.0	4.7

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:

- a user support staff made up of account management liaisons, User Assistance analysts, Scientific Computing Group (SciComp) liaisons, data liaisons, and visualization liaisons;
- multiple pathways to communicate with users, sponsors, and vendors;
- developing and delivering training to 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 2022 and the specific OLCF staff resources available to assist users.

1.4.2 User Support Roles

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 2022 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) that offers PIs and project members timely, accurate data to empower decision-making for OLCF projects and self-service tools for project administration. To accomplish this, 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 center-mandated 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.

Twenty-six versions of myOLCF were deployed to production in 2022 and added many new features and improvements, including, but not limited to the following:

- Improved application responsiveness for users by increasing application runtime efficiency
- Added ability to view and filter Jira tickets
- Enabled DD applications
- Added Open Enclave support
- Added review step to account application forms
- Enable users to view and modify their preferred shell on the HPC systems

1.4.2.2 RATS CRM

The center's customer relationship management software, called RATS CRM, is under continuous development. Twenty-one versions of the software were deployed in 2022 and added new features and improvements:

- GUI redesign for an improved look-and-feel
- Support for managing new filesystems in the enhanced and open enclaves
- OLCF office-hour scheduling system

- New API for the automated setup of the NVIDIA Federated Learning Application Runtime Environment (NVFlare) server on behalf of users
- Major internal software stack upgrades

1.4.2.3 APPL

In 2022, the OLCF began working with data engineers from the Center for Bioenergy Innovation (CBI) in the Biosciences Division (BSD) at ORNL. The BSD recently opened the Advanced Plant Phenotyping Laboratory (APPL) at ORNL to offer one of the most diverse suites of imaging capabilities for plant phenotyping worldwide. The CBI data engineers are working to create an open data portal service that can serve their data to users and researchers around the world from their numerous data collection sites and analysis laboratories. They are also interested in data-adjacent compute for data analysis (plotting, filtering, transformation) that would create a seamless experience for their users. A team from the OLCF is working with the CBI data engineers to help them utilize the OLCF's OpenShift Kubernetes cluster, named *Slate*, and near-line storage system, named *Themis*, to provide a reliable and scalable platform for their data storage, distribution, and analysis needs. The OLCF is uniquely positioned to provide the infrastructure and platforms required to support the suite of integrated research services envisioned by the CBI.

1.4.2.4 OLCF Quantum Computing User Program

In 2022, we continued to develop the OLCF's strategic capability to provide OLCF users with access to state-of-the-art quantum computing (QC) resources for discovery and innovation in scientific computing applications, given the ever-increasing demand. The program enables users to become familiar with the unique aspects and challenges of QC and to implement and test quantum algorithms on the available systems. Users can explore prospective computational research applications and potentially accelerate existing scientific applications by using quantum processors and architectures. Research projects supported include work from every program of DOE's Office of Science—including Advanced Scientific Computing, Basic Energy Science, Biological Environmental Research, High-Energy Physics, Fusion Energy Science, and Nuclear Physics, among others. This continuing effort to support and expand the OLCF's Quantum Computing User Program (QCUP) required a team effort consisting of significant contributions from staff in many different areas inside and outside of the OLCF.

The team renewed contracts with three QC hardware vendors (IBM, Rigetti Computing, and Honeywell/Quantinuum), negotiated a new contract with a new hardware vendor (IonQ), and is continuing to monitor the quantum hardware landscape to provide the QC with the resources necessary to meet the needs of the user program. Each contract was a significant effort: the negotiations addressed user agreements, resource availability, pricing, and policies. In addition, the team worked with each vendor to maintain and establish the multistep processes required for issuing user and project accounts, monitoring allocations, enforcing user agreements, providing user support, tracking, and reporting, among other tasks. Importantly, the team also worked to integrate the QCUP proposal and merit review processes into the standard OLCF procedures by which researchers submit proposals for access to computing resources. The team also worked with the vendors to establish and maintain the user support model, whereby QCUP users request help and report issues with the resources. Because each vendor has very specific requirements for both access and support; there was no easy, one-size-fits-all solution when establishing these processes. The team maintained and expanded communication mechanisms with users to inform them of system downtimes, changes, training events, and more.

To provide the most effective user experience possible, the QCUP program expanded the operations teams by onboarding new User Assistance members, by adding subject-matter experts, and by incorporating a Software Services engineer to strengthen our QC user support capability, improve management software integration, and to initiate a quantum science engagement objective. The QCUP website, (https://www.olcf.ornl.gov/olcf-resources/compute-systems/quantum-computing-user-program/) was maintained and updated throughout the year and contains background information on the priorities of the QCUP program, FAQs, QCUP event info, and highlighted publications.

Additionally, the QCUP website now links to a new User Assistance documentation site (https://docs.olcf.ornl.gov/quantum/index.html) to provide user documentation and training materials to help approved users access and utilize the QC resources supplied by our current vendors. Full-fledged QC user guides for each vendor have been added, and an OLCF user call and QC crash-course training resources are in development along with quantum science engagement content—all of which reflect the unique aspects and services of each hardware vendor. Additionally, the website was updated regularly to reflect the changing capabilities and offerings of our vendor partners. Through the partnership between QCUP and the vendors, users were invited to participate in several quantum training events. In August, OLCF hosted the 2022 OLCF User Forum, which featured a tabletop session "Quantum at the OLCF." Also in 2022, users were invited to attend IBM Quantum's Qiskit Global Summer School in July, IBM Quantum's Fall Challenge in November, and numerous IBM Quantum webinars and workshops throughout the year. A new element of internal assessment added in 2022 was the first annual QCUP user survey, the results of which showed that 100% of QCUP users who participated in the survey and submitted QC-related tickets to help@olcf.ornl.gov were satisfied with how their tickets were resolved.

Because of the increased demand for quantum resource access in 2022, new changes were needed to project proposal merit evaluation and the resource allocation process for existing projects. The Quantum Resource Utilization Council (QRUC) was formed to evaluate new project proposals and discuss new allocation requests by existing projects at a biweekly QRUC meeting. Additionally, monthly QCUP operations meetings were held to address vendor capability updates, evolving user support needs, and vendor resource usage monitoring. Staff in the Science Engagement Section have been assigned as liaisons to all the ongoing QCUP projects to assist the users in scientific and computing topics, as well as technical access to the quantum hardware.

A new ongoing effort is using OLCF HPC systems for pre- and post-processing tasks before sending further computations off to a QCUP quantum machine, thereby establishing a hybrid computational workflow. Hybrid HPC with QC is a rapidly expanding area of interest at the OLCF, so we helped bridge the gap between our HPC users and our QC users by developing documentation for how to approach the hybrid workflow. This includes instructions on how to install relevant software, how to submit jobs to quantum machines from our HPC systems, and best practices when running quantum jobs on our machines. Each QCUP vendor has expressed interest in the availability of this workflow, so we have ensured that each vendor's relevant quantum software can be installed on our machines. This has been successfully tested on Andes, Crusher, and Frontier.

The OLCF supported 79 total QCUP projects and 244 QCUP users in 2022 and tracked 69 publications that resulted from the use of these resources.

1.4.2.5 Jupyter at OLCF

The OLCF's Jupyter provides a critical function that allows OLCF users to perform low-to-mid-level data analysis and visualization tasks, which do not require Summit's compute resources. Jupyter allows a user to simply browse to a URL and, by leveraging 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 with footprints on the OLCF's Open and Moderate enclaves, thereby allowing users to access the file systems and computing resources of both enclaves.

Jupyter launched in 2021 and saw significant use. The system's utilization increased substantially in 2022 with

- 2,000 more lab instances in 2022 vs. 2021,
- 2,100 more GPU lab instances in 2022 vs. 2021, and
- more unique users in 2022 vs. 2021.

Internal changes to Jupyter have also made it a more robust platform and allowed the OLCF to continue providing it as a flagship service. The entire service was stripped and rebuilt in early 2022 to allow updates to the service and underlying cluster to be pushed without interrupting ongoing user sessions. The updates also added multiple AI/ML packages in response to user requests. The Open and Moderate instances also saw a combined 172 training labs to provide quick access within the OLCF enclaves to support multiple training series.

1.4.2.6 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:

- We are fortunate to have OLCF's Matt Norman as both our INCITE project liaison, and a team member under our ECP project. With Matt Norman's assistance, combined with the many resources of the ECP project, we have been able to explore a wide variety of techniques for getting maximal performance out of the Summit GPUs. We have been able to resolve all our performance issues and out codes make very good use of the NVIDIA V100s. This work also prepared us very well for the A100s at NERSC. We are currently working on some new developments needed for our 2022 milestones. These are internal software development issues related to interactions between various physics model subcomponents.

- Our liaison, Mark Berrill, has been very helpful in answering questions, providing advice, and quickly diagnosing issues. In Q1, we reported some issues related to node failure to Mark, who provided some useful insight for diagnoses. In Q2, Mark also provided very helpful advice regarding preparing for future machines like Frontier and programming models for enhanced portability across different architectures. When scaling up to full Summit, Mark had also provided some insight regarding internal buffers in MPI overflowing that helped us to resolve the issue quickly.

- The user assistance has been very responsive and very helpful. Sheila Moore has kindly been providing me with monthly usage analytics for our project, which has been very useful for accounting. The user support also helped me with issues setting up Constellation DOI. In addition, Chris Fuson helped us to obtain a priority boost for a few weeks at the June, when the queue wait was very lengthy, which helped us to get our jobs through for our conference papers. Also, Reuben Budiadja also helped us with both providing advice regarding readiness for Frontier and with discussions about large data workflows and storage, which led to a paper for the Smokey Mountains Conference.

- Our OLCF liaison (Stephen Nichols) remains crucial to the effort and has played a central role in testing Tusas on Crusher and learning about the upcoming changes in Trilinos. Visualization on Andes with VisIt has remained vital as the standard approach for visualization in the project.

- We very much appreciate the support we received from our INCITE project liaison, Van Ngo. He has been very responsive to our requests, specifically with troubleshooting new software installations on Summit and Andes.

- Our OLCF liaison (Stephen Nichols) has been crucial to the effort so far. One hurdle for the project coming into the INCITE award was maintaining a stable Trilinos build with the MueLu package, due to its sensitivity to changes in the programming environment. After determining a Trilinos version and working set of modules, he is maintaining a shared Trilinos build in our project home space. He also has been a critical communication point between our team and the center, informing us of known issues, upcoming outages, and other important news from the OLCF.

- Murali has been fantastic and is ramping up his ability to help us as he understands our project better by attending our weekly meetings.

- We have multiple interactions with our INCITE Liaison in different contexts, and they have been uniformly positive. Having someone at OLCF who can help us understand the current state of affairs directly has been extremely valuable, and Dr. Budiardja is a collaborator on our INCITE award and has been active in our code development efforts.

- Our liaison, Vassili Mewes, has been very helpful, both for casual questions regarding system information, policy, software, and profiling, and for providing a perspective on stability improvements and performance optimizations in magnetohydrodynamics.

- Summit and Andes have been integral resources contributing to the success of this research. Factors that make this possible include the size and scale of Summit, staff support and training, and storage resources. Our INCITE liaison, Wayne Joubert, has been a key resource to estimate campaign node hours, optimize runs, and augment CoMet to include features that were needed to process new types of data. OLCF staff, and especially Veronica G. Meless Vergara, have been prompt and informative during any communications to resolve questions and assist with boosting of jobs, when necessary, which has greatly facilitated this research.

1.4.3 OLCF User Group and Executive Board

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 questions they may have about OLCF resources with OLCF staff, and it also offers training on timely topics to start the discussion. The training topics for the 10 events in 2022 spanned overviews of OLCF for newer users, such as this year's user calls on OLCF Best Practices, the Andes Analysis Cluster, and the High Perforce Storage System, to webinars on newer tools for HPC, such as this year's user calls about Containers on Summit, Automating Science Workflows at OLCF, and the Julia Programming Language. We received great user feedback from the Julia and HPCToolkit Overview calls; the users have requested a future Julia training series to further cover the language in-depth and hackathons involving HPCToolkit. The User Call also addresses topics that have been highly requested by users, such as this year's user call Machine Learning for HPC Simulations user call, which featured a talk by HPE representatives about how to use the SmartSim architecture to help integrate ML into HPC simulations and included a Git repository of example codes and use cases for users. In 2022, the calls were well attended with an average of 107 participants per call. The User Call that had the highest attendance (211) was the June announcement of the new Frontier system given by Director of Science Bronson Messer. The Frontier Announcement and Overview informed users of what to expect out of the new machine and what to expect when transitioning from our older machines. The 10 User Conference Calls in 2022 had a total of 1,043 attendees vs. 2021's 945 attendees.

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. OUG Executive Board terms are 3 years and are staggered so that three 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 3rd year on the board. In 2022, André Walker-Loud of Lawrence Berkeley National Laboratory (LBNL) replaced Eric Nielsen of NASA as chair. Steven Gottlieb is vice chair and will become chair for the 2023–2024 term. Ana Cunha, Eric Nielsen, and Mark Coletti began new 3-year terms, which will conclude in 2025. The current Executive Board is listed at https://www.olcf.ornl.gov/about-olcf/oug/.

This year, the OLCF user meeting was once again held virtually due to higher rates of COVID-19 occurring during its scheduled time. The meeting took place on October 18–19, 2022, and was hosted on Zoom. The first day began with overviews of the coming Frontier exascale system and its supporting ecosystem of software and hardware and concluded with user-led talks on early Frontier hardware. The second day started with overviews of software and systems designed to help users develop workflows for the data and simulation campaigns. The day concluded with deep-dive breakout sessions on pressing topics for OLCF users, such as HPC workflows, storage, training, tools, and math libraries used for HPC. The meeting drew 146 participants, including users and OLCF staff.

See https://www.olcf.ornl.gov/calendar/2022-olcf-user-meeting/ for more information.

1.4.4 Training

The OLCF training program provides our user community with general HPC training as well as special topics needed to fully leverage the facility's cutting-edge HPC resources. The OLCF offers training in the form of workshops, webinars, tutorials, seminars, and hackathons. In most cases, the training events are recorded, and the slides, recordings, and hands-on materials are made available to our users through the OLCF training archive: https://docs.olcf.ornl.gov/training/training_archive.html.

Survey respondents rated the training at 4.6 out of 5.0, and 97% were highly satisfied with training overall. Usefulness of the online training archive and quality of the training content received a high rating (96%), whereas the lowest rated aspect was the breadth of training events offered, but it still rated highly (94%).

In 2022, the COVID-19 pandemic again required the OLCF to adapt their training program to a fully virtual format. Message board tools such as Slack were used for each event to maintain communication between users and trainers throughout the event. One of the advantages of a purely virtual training is that it was logistically easier for multiple laboratories and vendors to co-organize and contribute to training events. Following the trend from last year, we continued to increase collaboration with staff from other DOE laboratories and our vendors to organize and present training events.

In 2022, the OLCF facilitated or collaborated in several hackathons, including 9 multiday virtual NVIDIA GPU hackathons and one virtual OpenMP ECP SOLLVE hackathon. The facility also hosted or contributed to several training series, including a five-part data visualization series, nine OLCF user conference calls/webinars, an eight-part Preparing for Frontier series, and eight Hands-On HPC Crash Course events. In addition, the OLCF delivered several individual virtual training sessions and workshops, including an NVIDIA-led training for software development kits (SDKs), Using GPUs with Standard C and Fortran, an Nsight tools profiler workshop, an OLCF-led 2-day R programming workshop in collaboration with NERSC and the Argonne Leadership Computing Facility (ALCF), a Score-P and Vampir training session, a Kokkos release briefing, and collaboration on NERSC-led Code and E4S training.

1.4.4.1 Using R on High Performance Computers Workshop

In August of 2022, OLCF led a workshop that focused on how to use the R programming language for statistical computing on multinode, multicore heterogeneous architectures. The workshop was conducted in collaboration with NERSC and ALCF, with all three facilities working to set up instructions and R so the hands-on examples could be done on their center's compute resources. The workshop drew 93 participants and covered possible workflows for developing R code for HPC and how to use multicore and distributed parallel concepts in R. The slides, recordings, and exercise for the workshop can be found on the event page: https://www.olcf.ornl.gov/calendar/using-r-on-hpc-clusters-webinar/.

1.4.4.2 Preparing for Frontier Series

In 2022, the OLCF launched the Preparing for Frontier series, which aimed to train users how to utilize Summit and other computing resources to prepare applications for Frontier. The series started in July and had eight events that covered overviews of Frontier, the primary programming models for offloading work to Frontier's GPUs, HIP and OpenMP Offload, and user experience talks from the Frontier Center for Accelerated Application Readiness (CAAR) and ECP teams. NERSC partnered with us in three events by helping to advertise the events and by contributing to the development of the OpenMP offload content and information about how to use it on their computing resources. Recordings and slides of the eight events were made available on the OLCF Training Archive. Links to each of the event pages may be found in the table below.

Topic	Date	Participants
Introduction to the Frontier Supercomputer	July 12, 2022	89
Introduction to HIP Programming	July 14, 2022	86
HIP for CUDA Programmers	July 21, 2022	75
OpenMP Offloading – Part 1: Offloading Basics	August 11, 2022	69
OpenMP Offloading – Part 2: Data Movement and Optimization	September 1, 2022	54
Lessons and Tips from OLCF's Crusher Hackathons	December 1, 2022	59
Crusher User Experience Talks	December 9, 2022	56
Using HIP and GPU Libraries with OpenMP	December 14, 2022	65

1.4.4.3 Data Visualization and Analytics (DVA) Training Series

In 2022, the OLCF launched a new DVA training series. Based on specific feedback from the 2021 user survey, the DVA training series covered how to use specific visualization tools that are installed on OLCF systems as well as best practices for how to analyze data generated on Summit and Andes. The series is ongoing (with tentative events planned for 2023), but the 2022 portion of the series spanned from July through October with five events. Although the focus was on OLCF systems, the DVA training series was also open to non-OLCF users, thereby allowing NERSC users to participate with us for the ParaView and

VisIt training events. Recordings and slides of the five events were made available on the OLCF Training Archive. Links to each of the event pages are presented in the table below.

Topic	Date	Participants
Jupyter at OLCF	July 14, 2022	41
Remote Visualization with VNC	July 27, 2022	51
Andes Overview	August 31, 2022	60
ParaView at OLCF	September 15, 2022	50
VisIt at OLCF	October 13, 2022	23

1.4.4.4 CUDA Training Series

The OLCF, NERSC, and NVIDIA again co-organized and delivered a CUDA training series as a follow-on to the 2020 series. Each module consisted of a 1-hour presentation followed by a hands-on session in which participants worked through example exercises meant to reinforce the material covered in the presentation. The following topics were covered:

- CUDA Multithreading with Streams: https://www.olcf.ornl.gov/calendar/cuda-multithreading/
- CUDA Multi Process Service: https://www.olcf.ornl.gov/calendar/cuda-multi-process-service/
- CUDA Graph: https://www.olcf.ornl.gov/calendar/cuda-graphs/

The series was attended by 185 users.

1.4.5 Community Engagement

OLCF staff design seminars, workshops, and tutorials to inform current and future scientists and engineers about the methods and techniques used for HPC and how HPC can be used to advance scientific research. Aspects of these activities are also targeted at the general public to convey the benefits and advances rendered by HPC-enabled science and engineering.

In 2022 these efforts also intentionally focused on broadening participation in computing.

1.4.5.1 Hands-On HPC Crash Course

During 2022, the Virtual Hands-On Summit course was expanded from an afternoon of training to a Hands-On HPC Crash Course that spanned 1–4 sessions. The course is designed for students who have little to no background in HPC. This year’s extension included foundational skills introductions to Unix, Vim, C, and Python, with accompanying hands-on exercises to expand the course to students who do not have coding experience.

The course’s coding challenges are made accessible to students by using a combination of lectures and self-paced coding workbooks that guide students to make a few changes or additions to existing codes, thereby allowing them to focus on learning a specific concept. The exercises build their knowledge cumulatively through increasingly complex tasks designed to teach them the basics of using a supercomputer. Students taking the course in 2022 were given two to four weeks access to the Ascent HPC cluster to do the coding exercise and hands-on challenges. Students were given access to a Slack Workspace to ask questions for the duration of their compute access. The course was taught both virtually via zoom and by in-person lectures in 2022 depending on the needs of the students.

Several noteworthy hands-on HPC Challenge Exercises were developed for the course through 2022. One popular new challenge, called PyTorch Basics, featured ML exercises aimed at introducing students to PyTorch and the process of training a convolutional neural network (CNN), which is a deep learning application commonly used for image classification. Students were motivated to tackle this challenge by the leaderboard, which tracked the top 10 students who trained CNNs that achieved >60% accuracy within an hour of computer wall time. A parallel data challenge was added that guided students to use the mpi4py and h5py Python tools to speed up and visualize a simplified galaxy formation simulation. A challenge was added to allow students to explore the basics of how to use profiling software to study, visualize, and improve the performance of a parallel application code and its subroutines. This challenge was developed in collaboration with an ORNL student intern who was studying profiling tools for her ORNL summer internship.

We also enhanced our course lectures by adding talks that covered scientific use cases for HPC and possible career-paths in HPC.

Midyear, we added a certificate for the course for students who completed seven or more of its hands-on HPC challenges. The motive was to encourage more students to complete the challenge exercises. We developed an opt-out contact list of the certificate earners so that we could communicate further HPC learning opportunities to them and survey them next year to understand if our course influenced them to do more computing. We ran a four-session, month-long version of the course called The Summer Virtual Hands-On HPC Crash Course that focused on ORNL interns. This was our first attempt at using the course certificate to motivate participation. We did see robust participation from the students in the form of student questions, both during the lectures and in the Slack workspace for the course. Of the 71 students enrolled, 34 earned certificates.

We used the Hands-On HPC Crash course as part of our outreach to underrepresented groups in computing. As with the previous year's Hands-on Summit Course, we offered the HPC-Crash Course to students at the Richard Tapia Celebration of Diversity in Computing and the Students Program at the 2022 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC22). The course was also offered as training for the 2022 Winter Classic Student Cluster Competition, organized by Intersect360 Research, and the Faculty Hackathon as part of Gateways 2022. When it was offered as part of the SIAM-MDS Broader Participation program, one of the students from the Summer virtual Crash Course served as the course's teaching assistant and gave a talk about why HPC was useful for her summer internship project.

Professors from Morehouse College, Florida A&M University, and the University of Tennessee Space Institute who were participating in the OLCF Pathways to Supercomputing Initiative requested the course, and we ran a three-session version of it called The Winter HPC Crash Course over their winter academic breaks. Because of our advertising, the course attendance grew to also include students and professors from San José State University, Bluffton University, Lehman College, Norfolk State University, Northeastern University, Oglala Lakota College, and the University of Texas at Austin. Students had 4 weeks of Ascent access to complete the course challenges. Of the 86 students who attended the lectures, 31 of them earned the certificate.

GitHub Repositories: https://github.com/olcf/hands-on-with-summit and https://github.com/olcf/foundational_hpc_skill

	Date	**Venue**	**Students**
Hands-On HPC Crash Course Pilot	January 18–21, 2022	Virtual	28

	Date	Venue	Students
Winter Classic Cluster Competition HPC Crash Course	March 15, 2022	Virtual	49
Summer Virtual Hands-On HPC Crash Course	June 20–July11, 2022	Virtual	71
HPC Crash Course at Tapia	September 8, 2022	In-person at Tapia	30
SIAM MDS22 Hands-On HPC Crash Course	September 22, 29, 2022	One virtual session, two in-person sessions at SIAM MDS	40
Faculty Hackathon Gateways 2022	October, 6, 2022	Virtual	4
SC22 Students HPC Crash Course	November, 10 and 13, 2022	One virtual session, two in-person sessions at SC22	67
Winter Virtual HPC Crash Course	December 13–16, 2022	Virtual	83

1.4.5.2 ECP HPC Workforce Development and Retention Group

OLCF played a key role in launching and leading the ECP HPC Workforce Development and Retention Group. The group's mission is to form a community within the national laboratories to share best practices for developing and retaining an inclusive and robust computing workforce. The HPC Workforce Development and Retention Action Group, led by OLCF HPC Engineer Suzanne Parete-Koon, is composed of researchers and engineers from universities and Argonne National Laboratory (Argonne), Brookhaven National Laboratory (BNL), Fermi National Accelerator Laboratory (Fermilab), LBNL, Lawrence Livermore National Laboratory (LLNL), Los Alamos National Laboratory (LANL), ORNL, Pacific Northwest National Laboratory (PNNL), Sandia National Laboratories (Sandia), and SLAC (including ALCF, NERSC, and OLCF). It has monthly community meetings in which it plans a workforce development webinar series and is developing a web page to share best practices for building a diverse, equitable, and inclusive workforce culture in computing.

Webinar series: https://www.exascaleproject.org/workforce-development-seminar-series/

1.4.5.3 Winter Invitational Classic Student Cluster Competition

OLCF staff members helped organize and support the 2022 Winter Classic Invitational Cluster Competition. The competition was a team-based event that consisted of students from Historically Black Colleges and Universities (HBCUs) and Hispanic Serving Institutions. Participating teams competed in scientific challenges created by the supporting mentoring institutions. The students attended a Hands-on with Summit introductory training workshop to provide a basis for the rest of their time with the OLCF. OLCF staff created a set of challenges around ML for these students to learn about and compete over. These challenges processed cosmological data, including supernovae. The teams were presented with a webinar to provide background on the science behind the code and how to run and optimize the code on the Ascent training system. The teams then had a week to complete the set of exercises and submit a brief report with their findings. In the end, 9 of the 12 teams completed a submission, with the majority of those receiving a perfect or near-perfect score. The students showed great interest in learning about the topic and demonstrated a clear advancement of their knowledge and skills by the end of their time working with the OLCF.

1.4.5.4 Faculty Hackathon

The Faculty Hackathon was a Science Gateways Community Initiative (SGCI) event that ORNL helped plan and organize. The purpose of the faculty hackathon was to help introduce faculty to HPC concepts, tools, and techniques that they could initiate in their own courses. This will expose students to HPC and prepare them for further studies or careers in the area. It is important because it promotes and increases the number of people with HPC skillsets and expertise as well as minority groups in HPC. It also creates the environment for collaboration between HPC educators from different institutions and industries. Furthermore, it enables researchers at various institutions of higher learning to gain access to HPC resources for research and instruction at their institutions.

ORNL helped plan the number of faculty teams that would participate, the number of mentors needed, and the target areas to focus on. ORNL also provided three mentors that helped the teams instill HPC concepts, exercises, and techniques within their courses. In addition, ORNL delivered a crash course to the faculty on HPC concepts and techniques. The motive behind supporting the faculty hackathon was to promote the OLCF's Pathways to the Supercomputing Initiative to this faculty, the majority of whom are from HBCUs and Minority Serving Institutions. A total of five faculty teams, including 10 faculty members from Winston-Salem State University, Jackson State University, Elizabeth City State University, North Carolina A&T State University, Morehouse College, Norfolk State University, and Austin Community College, participated. The faculty teams presented posters of their work at the 2022 SGCI conference in San Diego, CA, where TeamRams, one of the faculty teams, won the best poster for the entire conference.

See: https://hackhpc.github.io/FacultyHack-Gateways22/

1.4.5.5 SC22 Students@SC Program

OLCF staff members participated in the SC22 Students@SC program by serving as the lead organizers for the Resume Doctor event, the PitchIt! Perfecting your Elevator Speech workshop, and the Students HPC Crash Course. For the resume workshop, the organizing committee of 25 professionals from the national labs, industry, and universities reviewed student resumes and provided best practices and tips for how to target their resumes for different job fields. The event was well attended by students. 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. For the HPC Crash Course, students used OLCF resources to do hands-on exercises to learn the fundamentals of supercomputing from OLCF staff.

Resume Doctor Event: https://sc22.supercomputing.org/session/?sess=sess398

Students HPC Crash Course: https://sc22.supercomputing.org/presentation/?id=miscc101&sess=sess302

PitchIt! Perfecting your Elevator Speech:
https://sc22.supercomputing.org/presentation/?id=misc104&sess=sess303#038;sess=sess303

1.4.5.6 Smoky Mountains Computational Sciences and Engineering Conference, Data Challenge Session

This year, OLCF staff members organized the annual Smoky Mountains Computational Sciences and Engineering Conference, Data Challenge Session (SMCDC2022). SMCDC provides an opportunity to tackle scientific data challenges that come from datasets contributed by ORNL, industry, and academia. These datasets come from scientific simulations and instruments in physical and chemical sciences, electron microscopy, bioinformatics, engineering, materials science, neutron sources, urban development, and other areas and provide open-research questions and tasks for participants to solve. These challenges are intended to draw students, scientists, and researchers who may be at the beginning stages of incorporating data analytics into their workflow, but they may also appeal to data analytics experts interested in applying novel techniques to datasets of national importance. This year, 35 teams registered, 19 teams submitted solution papers before the deadline, and 7 papers were selected for publication by Springer. We also received positive feedback from students who appreciated the opportunity to work on real-world scientific data.

"My favorite part of the internship was having the chance to work on the SMCDC with a group of other interns and start from nothing and end up with a finished analysis and a paper—it was fun and rewarding. Despite it being challenging and even frustrating at times, I bonded with my fellow interns and ended up with a product I could be proud of," said Kevin Wang, a 2022 Pathways to Pathways to Computing intern.

See: https://www.olcf.ornl.gov/2022/12/04/supercomputers-lead-to-super-learning-experiences/

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 2022 reporting period includes full CY production periods for the following HPC resources: the IBM AC922 Summit, the General Parallel File System (GPFS) (Alpine), and the archival storage system (High-Performance Storage System [HPSS]). In 2022, 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 2022 for the Summit computational resources (Table 2-1), HPSS data resources (Table 2-2), and the Alpine resources (Table 2-3). Supporting systems such as EVEREST, Andes, data transfer nodes, and Themis (nearline 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 resource 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 **Error! Reference source not found.**.

Operational performance metrics are provided for OLCF's Summit, the HPSS archive system, and the Alpine external GPFS file system (Table 2-3–Table 2-5).

Table 2-1. OLCF Operational performance summary for Summit.

	Measurement	2021 target	2021 actual	2022 target	2022 actual
IBM AC922 Summit	SA	90%	99.71%	95%	98.52%
	OA	90%	97.88%	90%	97.87%
	MTTI (hours)	NAM[a]	612	NAM	857
	MTTF (hours)	NAM	2,184	NAM	1,726

[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 HPSS.

	Measurement	2021 target	2021 actual	2022 target	2022 actual
HPSS	SA	95%	99.96%	95%	99.88%

	OA	90%	97.56%	90%	98.58%
	MTTI (hours)	NAM	342	NAM	480
	MTTF (hours)	NAM	1,459	NAM	1,750

NAM data is provided as reference only.

Table 2-3. OLCF Operational performance summary for the Alpine external GPFS file system.

	Measurement	2021 target	2021 actual	2022 target	2022 actual
Alpine	SA	95%	99.88%	95%	98.47%
	OA	90%	99.07%	90%	97.57%
	MTTI (hours)	NAM	789	NAM	777
	MTTF (hours)	NAM	2,187	NAM	1,438

NAM data is provided as reference only.

Summit and Alpine both saw significant decreases in MTTF in 2022. With the addition of Frontier clients mounting Alpine, we experienced some network instability and scalability issues that we continued to work through over the CY. The frequency of outages increased (hence the lower MTTF), but they were normally short, as indicated by the relatively static rates of scheduled and OA.

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 2, the SA target for an HPC compute resource increases to at least 90%, and the OA target increases to at least 85%. For year 3 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 2 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; on a schedule that provides sufficient notification: no less than 72 hours prior to the event and preferably as much 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 2021 and 2022 (Table 2.4).

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

Table 2-4. OLCF operational performance summary: SA.

	System	2021 target	2021 actual	2022 target	2022 actual
SA	IBM AC922	95%	99.71%	95%	99.52%
	HPSS	95%	99.96%	95%	99.88%
	Alpine	95%	99.88%	95%	99.47%

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).

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

As shown in Table 2-5, the OLCF exceeded the OA targets of the facility’s resources for 2021 and 2022.

Table 2-5. OLCF operational performance summary: OA.

	System	2021 target	2021 actual	2022 target	2022 actual
OA	IBM AC922	90%	97.88%	90%	97.87%
	HPSS	90%	97.56%	90%	98.58%
	Alpine	90%	99.07%	90%	97.57%

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 2-6.

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

Table 2-6. OLCF operational performance summary: MTTI

	System	2021 actual	2022 actual
MTTI[a] (hours)	IBM AC922	612	857
	HPSS	342	480
	Alpine	789	777

[a] MTTI is not a 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 2-7.

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

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

	System	2021 actual	2022 actual
MTTF[a] (hours)	IBM AC922	2,184	1,726
	HPSS	1,459	1,750
	Alpine	2,187	1,438

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

2.2 TOTAL SYSTEM UTILIZATION IN 2022

2.2.1 Resource Utilization Snapshot

During the operational assessment period (January 1, 2022–December 31, 2022), 36,649,091 Summit node hours were used outside of outage periods from an available 40,585,844 node hours. The total system utilization for the IBM AC922 Summit was 90.3%.

2.2.2 Total System Utilization

2.2.2.1 2022 Operational Assessment Guidance

System utilization (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 2022, 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 node hours by job. This method is extended to track the consumption of Summit node hours by program, project, user, and system with high fidelity. Figure 2-1 summarizes the IBM AC922 Summit utilization by month and by program for CY 2022. 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 fourth production year of Summit, utilization was 90.3%, which is in line with year 3 at 89.68%.

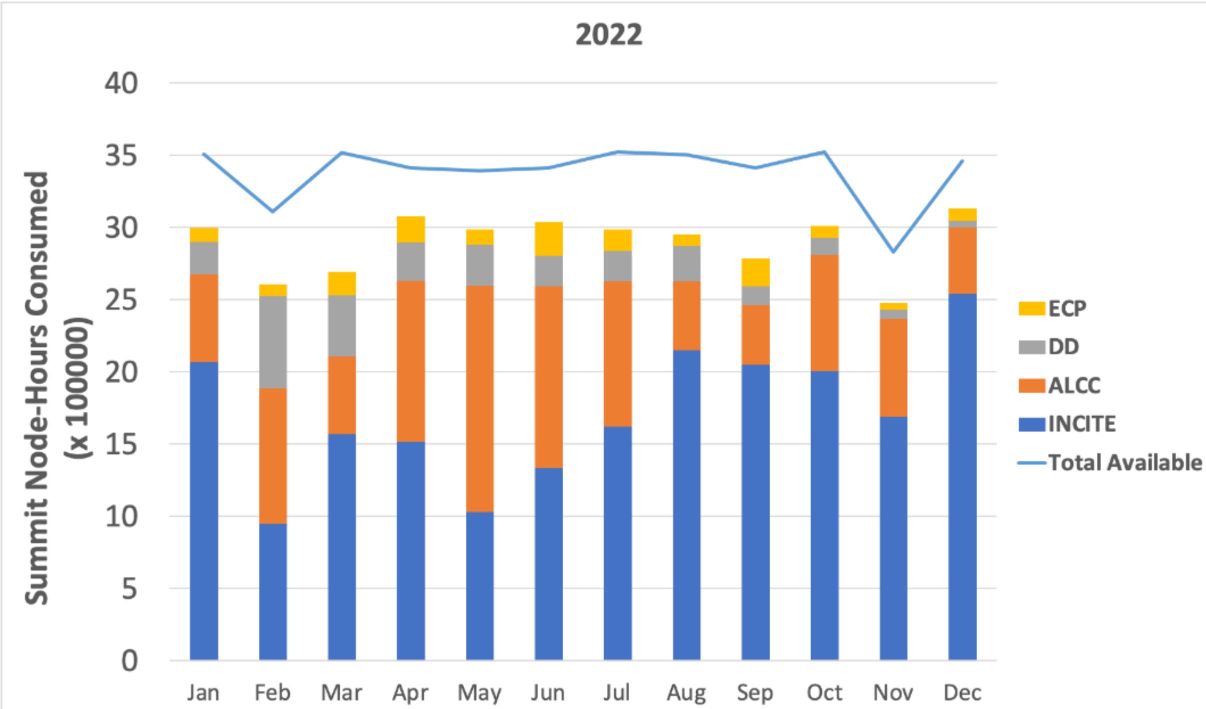

Figure 2-1. IBM AC922 resource utilization: Summit node hours by program for 2022.

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 2022 INCITE allocation program was the largest program in 2022, with a commitment of 18.8 million Summit node hours. The consumption of these allocation programs is shown in Table 2-8. As shown, all commitments were exceeded for each allocation program on Summit for 2022. This programmatic overachievement is due in part to the high uptime and diligent work of the OLCF staff.

Table 2-8. The 2022 allocated program performance on Summit.

Program	Allocation	Hours consumed	Percent of total
INCITE[a]	18,800,000	19,743,952	59%
ALCC[b]	Allocation spans multiple CY	10,714,034	32%
DD[c]	–	2,955,384	9%
Total[d]		33,413,370	100%

[a] Includes all 2022 INCITE program usage (including the 13th bonus month usage in January 2023).
[b] Includes all ALCC program usage for CY 2022.
[c] Includes ECP.
[d] Does not include usage outside of the three primary allocation programs.

Non-renewed INCITE projects from 2022 continued running through January 2023 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 2022 allocation period (ending June 30, 2022) were also granted extensions as appropriate.

CY 2022 saw a continuation of Summit allocations to the COVID-19 High Performance Computing Consortium. Over 800,000 Summit hours were provided to the consortium and span both the ALCC and DD programs. The allocations declined as the year came to an end.

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 2022 was 35% for Summit based on years of service, as described above. The OLCF continues to exceed expectations for capability usage of its HPC resources (Table 2-9 and Figure 2-2).

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.

Table 2-9. OLCF capability usage on the IBM AC922 Summit system.

Leadership usage	CY 2021 target	CY 2021 actual	CY 2022 target	CY 2022 actual
INCITE	NAM[a]	53.52%	NAM	57.38%
ALCC	NAM	36.76%	NAM	53.25%
All projects	35%	50.68%	35%	56.51%

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

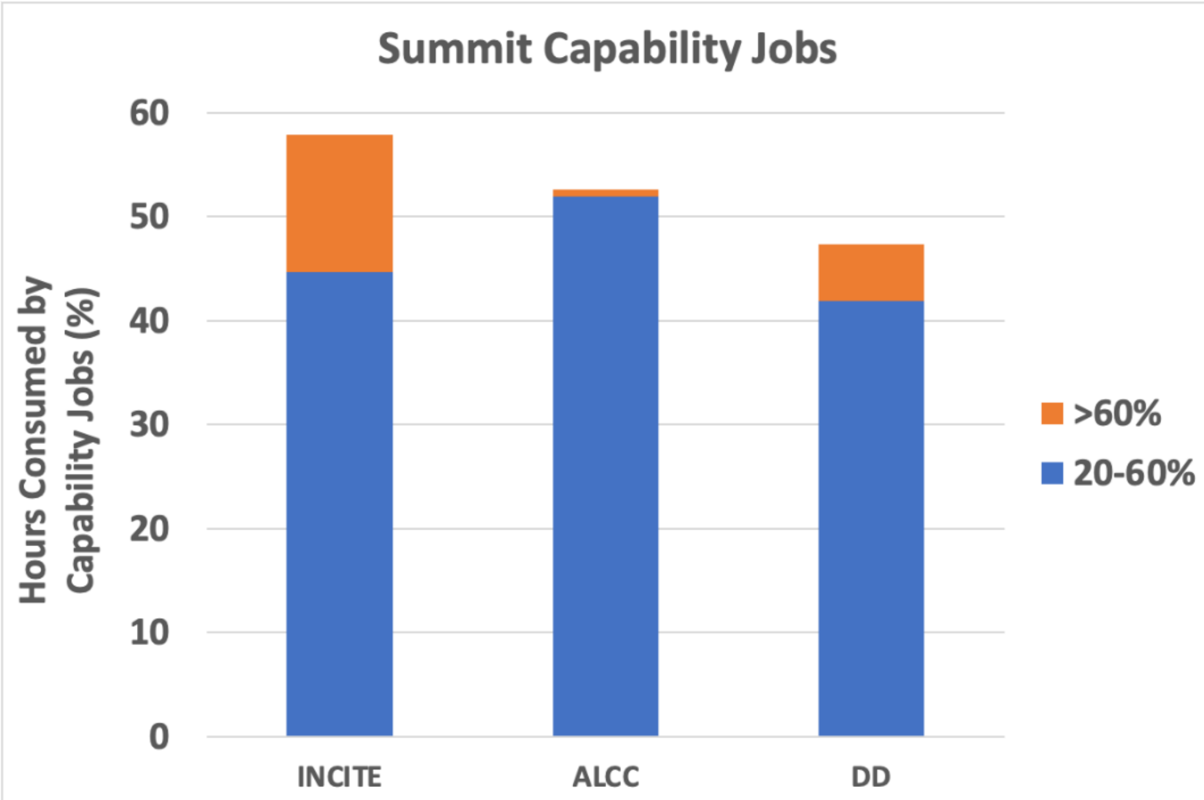

Figure 2-2. Summit capability usage by job size bins and project type.

2022 3 – Allocation of Resources

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

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, DD, and ECP)? (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. The OLCF successfully modified priorities for INCITE jobs to ensure that all allocated hours were delivered to the program for CY 2022. As of December 31, 2022, 99.8% of the total allocated hours for INCITE 2022 had been consumed on Summit. By extending our customary additional month of access to 2022 INCITE allocations that were not renewed for CY 2023, OLCF ultimately delivered 105% of the total allocated hours for INCITE 2022 by January 31, 2023.

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, 20% to ALCC, and 20% to DD, with up to half of the DD allocation (i.e., 10%) dedicated to ECP Application Development (AD) and Software Technology (ST) projects during each quarter. We report these ECP hours as a separate category in the following discussion. In CY 2022, the percentage of available compute time delivered to each of these programs was as follows: 59.1% INCITE, 28.4% ALCC, 8.2% DD, and 4.3% ECP. The ECP resources were generally under-allocated in each quarter of CY 2022, and the resulting time available in each quarter was often used by other projects from the other allocation programs.

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 DOE and ORNL strategic agendas

These goals are aligned particularly well with three of the OLCF's four 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 experimental facilities as well as DOE offices that have large compute 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 program; programmatic science areas (e.g., fusion, materials, chemistry, climate, nuclear physics, nuclear engineering, and 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 carry out research that would not otherwise be possible. See Section 8.2.9 for more information about ACCEL.

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 in CY 2022 was roughly 22,300 Summit node hours, but awards can range from 1,000s to 200,000 node hours or more. In 2022, the OLCF DD program participants (including ECP users) used approximately 13% of the total user resources on Summit, thereby consuming more than 3.5 million Summit node hours. Of that total, 34% of the hours were consumed by ECP AD and ST projects. See APPENDIX E for a full list of DD projects for CY 2022.

3.3 ALLOCATION PROGRAM CHALLENGES

The INCITE program suffered from under-utilization throughout much of CY 2022. The OLCF modified job priorities for INCITE jobs relative to our usual policies to ensure that all allocated hours were delivered to the program in CY 2022. Specifically, starting in November 2022, we increased the relative priority of INCITE jobs by increasing (by 5 days) the effective number of days they appeared to have been queued. Normally, INCITE and ALCC projects have a priority boost relative to DD and other jobs of a single day of aging. This increase in aging by 4 days enabled INCITE projects to use their remaining time within the calendar year, thereby falling short by less than 32,000 node-hours (from a total allocation of 18.8 million node-hours) on the final day of CY 2022. By continuing our usual practice of extending the availability of Summit to non-continuing INCITE projects, the 2022 INCITE program consumed 953,952 node hours over the allocated time by the end of January 2023. Importantly, this priority boost was essential but not sufficient to achieve this result. The projects also significantly ramped up their usage during this time to effectively drive up the demand for this increase in supply.

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

4. OPERATIONAL INNOVATION

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 2022. Key innovations and their operational impacts and improvements are discussed below.

4.1.1 Advanced Computing Ecosystem Evaluations and Test Beds

Computing systems and the associated architectures needed to support emerging applications and interconnected and integrated workflows are becoming more complex. To better understand these new technologies, evaluate their pros and cons, and assess their operational impacts and capabilities, the OLCF has established multiple test beds and continued their evaluation on multiple fronts.

4.1.1.1 Wombat Test Bed

The Wombat test bed was established in 2018 to evaluate the maturity and compatibility of the ARM AArch64 architecture for scientific HPC applications. Over time, the test bed has been updated as needed. Between February 2022 and July 2022, we ran a large-scale evaluation of HPC software on the Wombat cluster. Eleven teams composed of people from 10 different institutions worldwide evaluated traditional HPC modelling and simulation software. The purpose was two-fold. First, we wanted to see how well the software performed on current AArch64+GPU hardware. The second purpose was to evaluate the overall AArch64 software ecosystem to see how much it had improved since the 2019 evaluation and to note any shortcomings that could be problematic if the OLCF were to deploy this architecture at scale. Results of the evaluation were presented at the International Workshop on Arm-based HPC: Practice and Experience (IWAHPCE-2023), which is part of the HPC Asia 2023 conference. Comprehensive findings and results were published on Arxiv: Application Experiences on a GPU-Accelerated Arm-based HPC Testbed.

4.1.1.2 Edge-to-Exascale Test Bed

The edge-to-exascale test bed is a first attempt to establish and evaluate an edge computing workflow, and it was assembled with a particular use case in mind. The use case comes from researchers at the Center for Nanophase Materials Science (CNMS) who are using an atomic force microscope to probe the electrical and magnetic properties of different materials. The computing hardware consists of an NVIDIA DGX along with two other nodes with different NVIDIA GPUs in them. Additionally, the test bed has real-time data acquisition and FPGA hardware from National Instruments. Software for the CNMS use case, which is currently in an advanced prototype stage, is a replacement for existing analysis software written in Matlab. This old software was single threaded, did not use any accelerators, and required all the raw data to be written to a file before it could be processed. The new software is written in C++ and uses CUDA and OpenMP. It reads the raw data from the National Instruments hardware and processes it in real time. The advantage of the new software is that results are available immediately rather than a few hours after the microscope scan is finished. Additionally, because results are available as the atomic force microscope scans, it is possible to see if the scan is progressing properly and adjust or restart the scan if necessary. With the previous software, researchers were unaware of problems until the scan had completed and the data was processed, which would take several hours. This effort allowed OLCF staff to gain a deeper understanding for establishing and optimizing an edge computing workflow. This experience will be useful in future OLCF operations which will require integrated research infrastructures with interconnected workflows that involve edge and exascale systems.

4.1.1.3 Cloud Evaluation for Scientific HPC

In the fall of 2022, OLCF organized a rapid evaluation effort for three of the leading cloud vendors to assess their capabilities to support large-scale HPC computing in their environments. This was a preliminary evaluation effort to establish baseline capabilities of the vendors to support an environment relevant to decision-making and planning for the OLCF-6 procurement. Additionally, the study was meant to develop an initial understanding of cloud infrastructure so that the OLCF could evaluate its ability to service other DOE workloads that fall outside of the Leadership Class category. The study indicated that although current public cloud infrastructures are incapable of supporting a true leadership-class environment, they do present opportunities and features that are attractive for small-to-moderate-scale workloads. The full results of this study were published in an ORNL technical report (ORNL/TM-2022/2760) and will be submitted for publication at SC23. In addition, presentations that cover the study’s results will be presented at the 2023 SOS 25 Workshop and the 2023 Salishan Conference on High Speed Computing.

4.1.1.4 Advanced Computing Ecosystems Test Bed for Integrated and Interconnected Workflows

In 2022, the OLCF started to evaluate ecosystem design architectures that can operationally support integrated and interconnected scientific workflows. As part of this exercise, the OLCF began planning for the deployment of an integrated research infrastructure test bed. This test bed is intended to explore the necessary system requirements (i.e., software, configuration) to enable the integrated facilities workflow and provide other facilities access cycles to the test bed to develop and grow use cases that need more powerful clusters. The test bed will incorporate novel compute and storage technologies to identify the most suitable technologies for the use cases needed at ORNL. This test bed will also be used to identify the capability of multitenant scheduling to meet the real-time scheduling constraints of facility coupling. This test bed should be deployed in FY 2023.

4.1.2 Capability Benchmarking and Development at the Leadership Scale

In 2022, OLCF made great strides in developing and executing leadership-scale capability benchmarks, including the High Performance Linpack (HPL) benchmark and the High Performance Linpack - Mixed Precision (HPL-MxP) benchmark.

4.1.2.1 HPL

Frontier became the first supercomputer to exceed 1 exaflops on the HPL benchmark, which is used to rank the TOP500 supercomputers. Frontier scored 1.102 exaflops, which placed it on top of the June 2022 list. This effort was successful in part due to crucial contributions of OLCF staff. For the HPL effort, HPE relied on AMD to port the single-node HPL to AMD's CPU and GPUs for the Bard Peak node used in Frontier. This node consists of one AMD EPYC 7A53 Trento CPU and four AMD Instinct MI250X GPUs connected by AMD's Infinity Fabric coherent links. HPE's own benchmarking team then added proprietary optimizations to improve the scalability when using multiple nodes in parallel. As HPE worked to bring the system up, they used HPL to measure performance but also to identify under-performing hardware (e.g., CPUs, GPUs, network components). When HPE achieved good performance on all 128 nodes in a cabinet, their team would scale up HPL to the full cabinet. Once multiple cabinets were running well, they would then start running HPL across multiple cabinets.

When HPE started using about 20% of Frontier, they noticed odd performance that neither they nor the OLCF staff had ever seen on any previous system. They observed the performance peak, as it should, but then immediately start to fall off, which it should not. In addition to looking at the performance output, OLCF staff looked at the power profiles of the runs. A typical HPL profile shows maximum performance for about 60%–70% of the run and then a gradual fall off to idle as the amount of work decreases. What OLCF saw was an oscillating pattern in the power output that indicated swings of multiple megawatts every minute or two. This pattern frustrated HPE and AMD as well as OLCF. At OLCF's insistence, HPE added OLCF staff to both the AMD-led single-node team and to the HPE-led scaling team. OLCF staff were able to quickly provide some optimizations that increased single-node performance by 5%–7%. However, the performance oscillations continued.

As the deadline to submit a final HPL result loomed, HPE combined the teams, and we had daily calls to share possible solutions and test them out. OLCF staff also asked for more performance profiles. AMD staff showed the HPE staff how to collect profiles from a subset of nodes. HPE then captured another set of profiles and shared it the next day. One of the OLCF staff members took the profiles and plotted an interesting relationship between a process's MPI rank (i.e., ID) and its performance in the critical area. He shared that information during the next day's call. This was the *ah-ha* moment that triggered a lively discussion about how the plot looked like the expected performance when performing a linear lookup. HPE confirmed during the call that their MPI did not use any linear searches, but they were able to confirm that the low-level communication library did have one by default. Because MPI did not need that functionality, HPE was able to recompile MPI without it. During that evening's HPL runs, we achieved over 900 petaflops by using 8,192 nodes, and the oscillation was gone. The next morning, the HPE installation performed maintenance and prepared the system for another go. That night's first six runs were frustrating because they crashed with node failures or bugs in HPE's HPL. For the seventh run, the HPE staff member running the jobs identified an optimization that was causing HPL to crash and altered the parameters to avoid the suspect code path. He then launched the seventh run, which was the 1.1 exaflops result that placed Frontier on top of the TOP500 list and second on the Green500 list (behind one of the Frontier Test and Development systems). That result was the combined efforts of HPE, AMD, and the OLCF.

4.1.2.2 HPL-MxP

A team of mostly OLCF staff researchers has designed, developed, and deployed an implementation of the HPL-MxP benchmark. This benchmark, formerly known as HPL-AI, seeks to highlight the ability of applications to achieve superior performance by utilizing modern hardware such as graphics processing units (GPU) by utilizing lower precision computation such as 16-bit (half precision) along with iterative refinement to bring solutions back to full 64-bit (double precision) accuracy. These algorithms can demonstrate both high accuracy and employs hardware that is primarily used for accelerating AI workloads. This strategy is an important area of investigation for supercomputing applications to improve performance and more efficiently use available hardware.

The OLCF implementation addressed both performance at scale and performance portability due to having two top supercomputers (Frontier and Summit) in the facility. This was demonstrated by the rankings of the HPL-MXP which is similar to the rankings of the traditional HPL Top500 listing. The OLCF code was used to achieve a #1 ranking for Frontier at 7.942 exaflops, the #2 placement for the EuroHPC LUMI at 2.168 exaflops, and the #5 with Summit at 1.411 exaflops. While the application was used for other systems as well, these three systems are unique in that they delivered exascale performance at scale with HPL-MXP. The knowledge and expertise in developing this code was utilized in aiding the achievement of Frontier's 1.102 exaflops on the Top500 HPL rankings.

These results show the capability of National Center for Computational Sciences (NCCS)/OLCF staff to combine their deep understanding of system architectural details and use it to optimization performance. Working with these and other benchmarks also gives our staff the ability to understand the needs of various algorithms and optimization opportunities which will help guide future procurements.

OLCF has also made progress on the development of other Leadership benchmarks and efforts.

4.1.2.3 MLCommons

From the outset, OLCF staff have been active participants in the MLCommons scientific working group. Our contribution to the group is exemplified through ORNL's innovative AI application, STEMDL, which specializes in scanning transmission electron microscopic image reconstruction. We are proud to have STEMDL selected as one of the four benchmark models released by the working group during ISC 2022. This benchmark release demonstrates our commitment to providing cutting-edge solutions and advancing the field of AI for science. Further details about the benchmark release are included in this publication: https://link.springer.com/chapter/10.1007/978-3-031-23220-6_4.

4.1.2.4 WfBench

In 2022, staff from OLCF Data Lifecycles and Scalable Workflows group collaborated with researchers from the University of Southern California, University of Hawaii, and Argonne to develop workflow benchmarks (WfBench) that can be used to evaluate the performance of workflow systems on current and future software stacks and hardware platforms. WfBench automates the generation of realistic workflow benchmark specifications that are easily translated into benchmark code that is executable with current workflow systems. It generates workflow tasks with arbitrary performance characteristics regarding CPU, memory, and I/O usage, and generates realistic task dependency structures based on that seen in real-world scientific workflow applications. The benchmarks have been used at OLCF for evaluating the ability to run HPC workloads in Clouds (Google Cloud, Microsoft Azure, and Amazon Web Services), and they have also gained much interested from the community (specifically NERSC and NVIDIA) to execute them during the procurement phase of new machines.

4.1.3 Scalable Data Hierarchy Management

In 2022, the OLCF worked on several efforts to operationally better manage the vast amounts of data generated or stored on OLCF systems.

4.1.3.1 Data Management Plans (DMPs)

In 2022, staff from the OLCF, ORNL legal, and cybersecurity teams began working on a DMP template in accordance with the DOE's *Suggested Elements for a Data Management Plan*, effective January 1, 2022. From the beginning, the DMP was conceived as providing insight and guidance into the data being created at OLCF but also connecting Principal Investigators (PIs) with the appropriate data, legal, and cybersecurity teams. The DMP was also designed to make it easier for PIs to fill out information by creating a flow of checkboxes and prompts. The plan was further refined through interviews with members of the ORNL Computing and Computational Sciences Directorate's Interconnected Science Ecosystem (INTERSECT) initiative, ORNL APPL, ORNL Business Division, OLCF user proposals, and other research projects. In 2023, we will continue refining the DMP with an emphasis on transforming it into a Machine Actionable Data Management Plan (MaDMP) that brings curators, legal, and cyber closer to the research as it happens. Our work received support from the Data Curation Network Big Data Interest group as a creative way to shift curation from a fundamentally reactive stance to a proactive and collaborative one. The DMP directly impacts enhanced data operations and storage management at OLCF, enhances user experience, and provides rigor for a key element of research.

4.1.3.2 Constellation

Constellation, OLCF's Open Data Portal, continues to grow with a 43% increase in datasets for CY 2022. Constellation now holds approximately 3 PB of data and over a million files (some compressed). The largest dataset is approximately 2 PB, and the largest single file is 17 TB. Staff in the OLCF Data Lifecycle and Scalable Workflows group have been meeting the challenges of large, unstructured datasets the management of which often requires creative and collaborative solutions. Constellation directly improves the sharing of scientific results and its quality. For example, this year a PI deposited 10 million directories that corresponded to molecules. Curation discovered that one of the directories was empty, and upon further discussion with the PI, it was determined that this was a mistake, and that the experiment would need to be rerun. The result was that 41 unprocessed molecules were reduced to 13. Thus, curation actively contributed to scientific achievement at the OLCF. Other curation activities uncovered incomplete datasets for another depositor, and the existence of usernames in yet another dataset. A process enhancement for depositing datasets was implemented this year that requires PIs to deposit READMEs to ensure that datasets are understandable, now and in the future. At the end of 2022, the Data Lifecycle and Scalable Workflows group led a national laboratory working group on big-data challenges with participants from PNNL, LLNL, SNL, and LANL. In 2023, this national laboratory working group will start collating big-data tools developed at the labs into a common (and curated) GitHub repository for easy access. Also, Constellation will be redesigned to share the repository code with others so that we can all work from a common platform. The OLCF is uniquely positioned to take a leadership role in managing and curating large data.

4.1.3.3 QuickSilver

QuickSilver is a policy engine for data and storage management developed by OLCF staff. With newer systems beginning to utilize multiple storage technologies with different performance profiles, some level of automation for migrating data between the storage tiers is desired. There is also a need to periodically purge old data from the storage system, and this is often handled by a set of simple single-purpose scripts. QuickSilver is designed to provide a single framework for handling both tasks while being flexible enough to incorporate new storage policies in the future. This software also serves as risk mitigation for (or an alternative to) vendor supplied solutions for the Orion file system.

4.1.3.4 HVAC

The High Velocity AI Cache (HVAC) is an automated caching tier for Summit and Frontier designed to alleviate the bottleneck of the I/O subsystem on AI and ML applications. HVAC works by using a copy-on-read mechanism to cache requested files into the node-local NVMe devices used in the scheduled job. HVAC uses algorithmic hashing in place of metadata storage and retrieval to locate the requested data in the allocation. These techniques offer significant improvements over reading multiple epochs of data from the parallel file system. The HVAC prototype was successfully developed and tested at a scale of 1,024 nodes on Summit against several ML training applications and benchmarks. The results of this work were published at Cluster in September 2022. The HVAC prototype is now in the preproduction phase and will be used by applications on OLCF supercomputers.

4.1.3.5 UnifyFS/ECP

UnifyFS (https://unifyfs.readthedocs.io/en/dev/) is an ECP software technology product developed by OLCF and ORNL staff in collaboration with LLNL that enables HPC applications to utilize node-local storage devices under a unified shared namespace, similar to how applications already use shared parallel file systems. Currently, due to a lack of system-provided software that enables shared file data access to the node-local storage on the OLCF Summit and Crusher systems, applications must treat these storage devices independently, and that means node-local data from remote nodes in a job is not accessible. UnifyFS overcomes this limitation and is easy to integrate with current production applications because it does not require code changes (in most applications) due to its seamless I/O interception method and because it has customizable semantics to align with the most common parallel I/O use cases. The OLCF has made recent versions of UnifyFS available to users on both Summit and Crusher.

4.1.4 Data Workflows

OLCF is actively pursuing enhancement of its operational data workflow capabilities. In 2022, OLCF has worked on several projects to this end. OLCF also organized science data hackathons to facilitate efficient access to large volumes of datasets that the OLCF has published.

4.1.4.1 Open Science Data Hackathons

Researchers affiliated with many institutions lack the resources to acquire and analyze multi-PB datasets. DOE HPC user facilities can marshal the expertise and capacity to support large-scale data hackathons that bring together the science community from multiple international institutions to focus on science inquiries with minimal data management overhead. Open science data hackathons have the potential to accelerate discovery science for societal benefits.

One of the OLCF's staff members was a liaison for an INCITE project that generated 4 PB of data from climate simulations, the Experimental Nature Run at 1-km (XNR1K). The data were made publicly available via the OLCF. However, interested researchers were faced with logistical challenges of downloading and managing large volumes of data. In 2022, we facilitated direct access to the XNR1K data and provisioned access to OLCF computational resources via an Open Science Data Hackathon. The data hackathon was launched on July 1, 2022, and scheduled to last for a year to provide sufficient time for the teams to synthesize their results. The participating teams have access to Andes for data analytics, Ascent for ML and AI workloads, and Jupyter Notebook instances on Slate for interactive analytics. Mentors have been providing help and support for the systems and the data collection. Also, in the summer of 2022, a subset of the data was used by a team of students in the EU to develop innovative visualization techniques. Another team of students from a graduate summer school used the data to study atmospheric processes in the lower stratosphere and presented their results during the 2022 Fall Meeting of the American Geophysical Union. The participants of the hackathon could bring their own analytics tools and apply them to the XNR1K data. Currently, we have 11 international teams from different institutions pursuing various science interests including (1) observing system simulation experiments for developing instruments for future satellites and for mission planning; (2) nowcasting techniques for severe weather using AI techniques; (3) data-driven subgrid scale parameterization methods for earth system models; (4) improved understanding of topography-induced circulations at the meso and local scales; (5) innovative visualization techniques; and (6) mentoring and training the next generation of climate scientists via a summer school.

4.1.4.2 Zambeze

In 2022, staff from the OLCF Data Lifecycle and Scalable Workflows group embarked on researching and developing a distributed orchestration prototype system called Zambeze. With ASCR's vision of having an integrated research infrastructure, experiment campaigns can span multiple facilities, and that requires processing, storage, and advanced network capabilities for large data movement. On the other hand, defining and executing these campaigns should not diminish the user experience due to increased system complexity. Zambeze addresses this need and seeks to integrate services (e.g., workflow systems, data management tools) via a distributed orchestration architecture in which agents (deployed in each system) translate user requests into actions to be performed on the computational platform. User requests are formatted as messages and distributed by using a message queueing service. These messages are then received by distributed agents that can satisfy the user request. A single agent will then process the message by enacting actions to plugins that represent integrated services. This tool will enable OLCF to participate in a larger DOE ecosystem and enhance user experiences across the DOE complex, thus increasing operational efficiency. An initial prototype of Zambeze (version 0.3) will be demonstrated in CY 2023.

4.1.4.3 DataFlow

OLCF staff worked to further develop and deploy the DataFlow software. DataFlow is a web-based data ingestion solution designed to provide researchers with a seamless method of easily moving their captured instrument data to storage locations where they can use compute resources to analyze the data. Based on the feedback of researchers, additional data adapters were added to DataFlow in 2022, thereby bringing its support to three primary storage adapters: Globus, DropBox, and Microsoft OneDrive. This addition has vastly reduced the time and difficulty for researchers to move their instrument data to analysis platforms. Over the course of 2022, DataFlow was publicized to numerous groups throughout ORNL and external facilities with an article in ORNL Today along with several presentations, including a featured presentation at DOE Data Days 2022. This has garnered significant attention and interest in deploying DataFlow to various instrument labs in several directorates and divisions. The deployment process has been simplified largely due to collaborating with ORNL's Information Technology Services Division in making DataFlow a lab-wide offering. As more researchers can quickly and easily move their data off their instruments and closer to compute resources, they will be well-positioned to participate in integrated research infrastructures and leverage OLCF resources.

4.1.4.4 FlowCept

In 2022, OLCF started a new research project called FlowCept to address the need to integrate data in cross facility scientific experiments. Because these experiments deal with a myriad of facilities, tools, models, and data, efficiently integrating the data to enable holistic data analyses is critical to support scientific discoveries. Holistic integration helps facilitate the understanding and explanation of an experiment's results, FAIR (Findability, Accessibility, Interoperability, and Reusability) and trustworthy AI verification, and runtime monitoring and steering. Resource management algorithms can also benefit by using the data for decisions on when and how to scale their workflows and identify reasons or risks for performance bottlenecks. FlowCept connects to any existing system that generates data and seamlessly intercepts their dataflows to build an integrated data view on the fly. In 2023, we will prioritize development to deal with heterogeneous tools and data and also facilitate user adoption. We have been building a plugin- and event-oriented architecture to address heterogeneous facilities, systems, and data. We will develop plugins for Zambeze, MLFlow, Dask, and TensorFlow and run on at least one real use case using an OLCF machine.

4.1.5 Bridging Classical and Quantum Systems – Hybrid Workflows

QCUP is an OLCF program that provides access to commercial QC resources for testing and evaluation of scientific computing. QCUP fosters the development of scientific applications for quantum computers by enabling competitive access to these computing resources and curating a portfolio of projects that address scientific research priorities. QCUP also tracks the status of existing capabilities as measured through characterization and benchmarking with in-house testing that assesses the feasibility and readiness of a commercial QC system as well as monitoring of external evaluations distributed through published research. QCUP engages with users in the development of scientific applications for quantum computers and develops new infrastructure for these innovations. Thus, QCUP addresses the above goals by operating a merit-based access program for using commercial QC resources. Management of the program is overseen by the QCUP PI, who leads a team that spans three functional areas: (1) operations, (2) science engagement, and (3) technical integration. Staff from all functional areas contribute to a QRUC that prioritizes and selects proposals for QC resource allocation based on technical merit, scientific impact, and mission alignment.

As a QC and HPC integration component on a system integration level, QCUP provides the remote-server model as a loose HPC and QC integration method with the OLCF HPC systems and the quantum service providers (i.e., currently IBM, Rigetti, and Quantinuum). Although lightweight code can be run on the front-end nodes, more compute-intensive code should be submitted to the compute nodes by using a batch job. Then, the compute nodes can communicate with external URLs (e.g., IBM Quantum backends) through HTTP proxy settings in a batch script. However, there are many constraints to resolve in this model, such as conflicts of policies between the HPC systems and external quantum service providers, network latency, and security. Currently, this integration model is working with the IBM Quantum and Quantinuum back ends but does not yet work with Rigetti. To maximize the HPC and QC system integration, it would be best for the quantum device to be co-located with the OLCF HPC center.

At the software integration level, emulating QPUs on the HPC systems can be a very powerful tool in QC. About 45% of jobs and 58% of executions in the total IBM quantum back-end usage of QCUP users run on the IBM quantum simulators to test their quantum code before running on the real QPU. Therefore, it is important for QCUP users to use local HPC resources for their quantum simulations. Currently, QCUP provides quantum platforms such as Qiskit (IBM, Quantinuum), PyQuil/Forest SDK (Rigetti), and PennyLane (Xanadu) on the HPC systems. We are also working to integrate the TN-QVM (developed at ORNL), NWQ-Sim (PNNL), and QTensor (Argonne) quantum simulators on the OLCF's HPC systems.

Within the QCUP and under an internally submitted proposal, the OLCF has established a hybrid workflow capability to bridge the gap between HPC and QC platforms. We successfully built a prototype and developed three fully automated examples for demonstration: the Traveling Salesperson Problem, Grover's Search Algorithm, and Shor's Factoring Algorithm. We employ workflows to generate inputs from the OLCF's HPC systems and transfer them to IBM Q systems in the cloud, where the quantum calculations produce results that then return to OLCF resources. This workflow-based approach provides additional benefits, including (1) end-to-end programmatic automation of the entire process, (2) an out-of-the-box tool for interfacing with HPC schedulers and quantum middleware, and (3) concurrency of independent tasks (e.g., running the same algorithm over a simulator and a real quantum device simultaneously). The work was presented at ParslFest 2022 at the University of Chicago, and a paper is being submitted to the 2023 International Conference on Computational Science. This positions the OLCF for next-generation research problems that will utilize QC resources.

Further expanding OLCF's capabilities in bridging HPC and QC, the Facility has also been working on an ECP-funded project in 2022 that focuses on Quantum Natural Language Processing (QNLP). QNLP is a cutting-edge application of QC that aims to develop natural language processing (NLP) models that will be executed on quantum computers. In this project, we assess the feasibility and accuracy of QNLP by using numerical simulators on large-scale HPC systems and actual QC hardware. Based on related QNLP work, we perform crucial groundwork necessary to develop a Quantum Transformer deep learning model and to simulate QNLP on the exascale Frontier supercomputer at ORNL. To this end, we first use HPC-Quantum simulators and deploy the HPC and Quantum hybrid workflow to use QC devices to implement quantum pipelines and neural networks with default datasets. Second, we apply this workflow to CANDLE/MOSSAIC (Modeling Outcomes using Surveillance data and Scalable Artificial Intelligence for Cancer) datasets to demonstrate a real-world application on HPC and Noisy Intermediate Scale Quantum (NISQ) devices. Thus, performing Quantum Transformer simulations by using the datasets on Frontier with the hybrid model, we can conduct parameter tuning and QNLP model development faster, with improved accuracy, and at a lower energy cost.

The transformer architecture has revolutionized the world of NLP. Based on the self-attention mechanism, transformers have outperformed other deep learning methods (e.g., recurrent neural networks and CNNs) for NLP tasks such as machine translation, generation of text, question answering, and text classification. However, the main issue in neural networks is that the huge parameter sizes usually require a huge training dataset and a massive computing network to conduct the training. This issue is particularly clear for large transformer models, which can require days to train at exascale. Finding a way to fundamentally reduce the size and training cost of these neural networks would be transformative for computational science. Supercomputers have been perfected over the years to solve many problems, but they cannot scale with the same potential as quantum computers when enough variables are added. Quantum computers, likewise, are unable to solve all computational problems, but are very efficient for many that classical computers cannot solve. Furthermore, combining these two resources in workflows can allow for significant speedups in many applications and offer relief for the end of endless, exponentially increasing computing power promised by Moore's Law. As shown in Figure 4-1, we have successfully evaluated and tested this HPC and QC hybrid workflow model with the quantum pipeline by using the quantum trainer in QNLP between the supercomputer at ORNL and the IBM Quantum simulators and back ends. Thus, this project aimed to extend QNLP and Quantum Transformer deep learning applications to biomedical datasets and integrate the most successful quantum approaches with Frontier at ORNL.

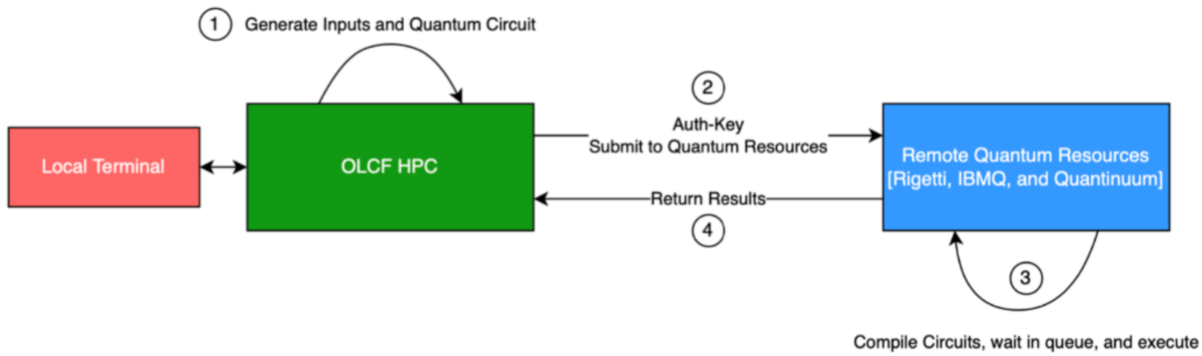

Figure 4-1. OLCF HPC-QC hybrid workflow model.

On the other hand, and within the existing DOE-NCI (National Cancer Institute) partnership, the CANDLE ECP project works with the MOSSAIC project to develop innovative deep learning for population-level cancer surveillance on DOE supercomputing resources. This involves studying the NCI's Surveillance, Epidemiology, and End Results (SEER) program corpus of cancer pathology reports. The reports in this corpus present challenges for traditional NLP approaches due to their unstructured text, highly specialized vocabulary, and irregular data quality. Current state-of-the-art performance for the MOSSAIC cancer pathology corpus is provided by a transformer deep learning model. Because these reports contain personal health information (PHI), preliminary studies are typically conducted on non-PHI datasets. For this project, we consider two open datasets for testing and the MOSSAIC dataset itself:

1. The Internet Movie Database (IMDB) is a widely studied NLP dataset for sentiment classification. The IMDB dataset contains 50,000 movie reviews for NLP or text analytics. For binary sentiment classification, this dataset contains substantially more data than previous benchmark datasets. We provide a set of 25,000 highly polar movie reviews (i.e., 1-star vs. 5-star ratings) for training and 25,000 polar movie reviews for testing. So, the quantum classifiers predict the number of positive and negative reviews by using either classification or deep learning algorithms.

2. The CANDLE P3B3 dataset is a corpus of synthetically generated cancer pathology reports. It is computationally similar to the MOSSAIC dataset.

3. The MOSSAIC dataset consists of several million unstructured text–based cancer pathology reports. This dataset is well-studied with a variety of deep learning NLP approaches, and current state-of-the-art performance is provided by a transformer model. Because the dataset contains PHI, it resides in the secure PHI enclave at ORNL, and a tokenization protocol has been developed to allow it to be used outside of the enclave.

We utilize the Frontier supercomputer at ORNL to simulate quantum transformer for binary classification. This experiment demonstrates the HPC and QC hybrid model in real-world QNLP applications. As preliminary work, we used open-source codes (e.g., lambeq, torchquantum, qtransformer) and adopted the straightforward inner-product, self-attention network.

However, for AI, performing QNLP with NISQ devices is currently still in its infancy. Specifically, scalability is lacking when using a large-scale dataset that requires a lot of computational processing. Thus, when performing QNLP simulations using the large-scale CANDLE/MOSSAIC biomedical datasets on Frontier with the HPC-QC hybrid model, parameter tuning and QNLP model development should improve with more accurate, faster, and more efficient operation.

4.1.6 Energy Efficiency

OLCF staff members have built a system that enables continuous and long-term power profiling of jobs that run on OLCF systems to optimize per-job and overall energy consumption to support more efficient operations. This system indexes power measurement time-series data by using the job allocation history and produces extensively pre-processed time–series based aggregates and statistics. Unlike prior efforts, which depended on low-resolution aggregated data, this system puts the high-resolution, out-of-band, per-component time-series data from OLCF systems into a job context to strategically target both online interactive consumption and accelerated offline batch analytics. Based on accumulated knowledge of handling power profiles in analytics contexts, the resulting output data was designed to fulfill many known use cases in the organization.

At the core, this system implements a data pipeline that merges system-level power and energy time-series data streams and the application-level job allocation log streams and performs aggregations in a streaming fashion. Aggregates are performed across the time axis (i.e., 1–10 second window) as well as the space axis (i.e., all components within the time step), thereby capturing statistics such as mins, maxes, means, medians, standard deviations, and inter-quartiles. Such aggregations are performed at a certain resolution, and a set of nodes and components are allocated for each job. To process this volume of data in real time, this pipeline relies on the structured streaming API provided by Apache Spark and supported by a 6-node (320-core) data analytics cluster dedicated to this purpose. By considering each data stream as an ever-appended table processed in mini-batches, data processing code was implemented once and applied to historical data and real-time data ingests, thereby simultaneously enabling flexibility in handling datasets.

The output of this pipeline results in a modest 1.5–2.0 TB per year of fully indexed data that covers job-level and component-level data and is ingested into two different data stores for consumption. For online, low-latency interactive use cases, the data is ingested into Apache Druid, which is a scalable time-series database that will back web services and REST (representational state transfer) APIs exposed to the organization. For offline long-term data analytics, the data is ingested into a data lake, which is backed by an S3 file system and Apache Parquet file format that will supply data to downstream data analytics workflows and training pipelines.

This system fills the large gap between data accumulation and consumption by providing low-latency interactive queries and high-throughput processing of contextualized power measurements. Despite the valuable insights from such interactions, delivering such insights to the organization was difficult due to the amount of data from large systems such as Summit and Frontier. Thanks to this effort, the OLCF can accelerate its innovations in interactive job power analytics software, power profile prediction models, facility cooling optimization tools, and its understanding of the systems. Figure 4-2 shows the total power utilization by two sampled jobs.

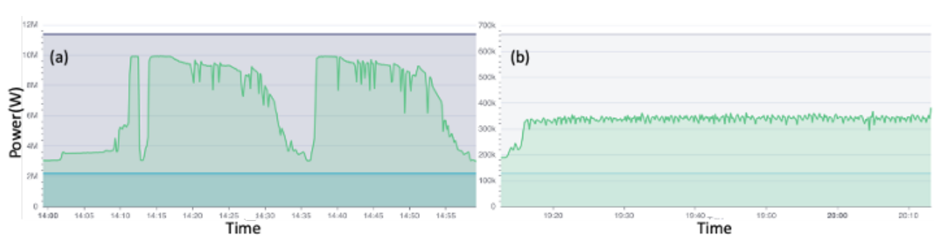

Figure 4-2. Figure showing power profiles of two jobs. The *y*-axis is the magnitude of power consumption in Watts, and the *x*-axis is time. The plot frequency is 10 seconds. (a) A matrix-solver job running on 4,374 nodes. (b) An ML job running on 64 nodes.

4.1.7 Federated Learning

Federated learning (FL) is a technique for sharing ML models without sharing raw data. FL-trained models generally converge faster and fit more closely to the data than independently trained models on smaller individual datasets. Depending on the FL algorithm, privacy-preserving properties can be proven, which may enable much better models to be trained in healthcare, genomics, and other privacy-sensitive fields.

In 2022, OLCF staff members chose to evaluate NVFlare as the supported FL environment on OLCF systems. Throughout that time, OLCF staff regularly engaged with senior scientists at ORNL to understand use cases for FL. This engagement helped OLCF staff drive necessary operational developments and the implementation of a secure deployment model for the distributed NVFlare components. NVFlare has a centralized control plane that must be managed securely by the OLCF, while end-user FL learning tasks are performed within individual project contexts. Figure 4-3 shows the overall NVFlare integration architecture at the OLCF.

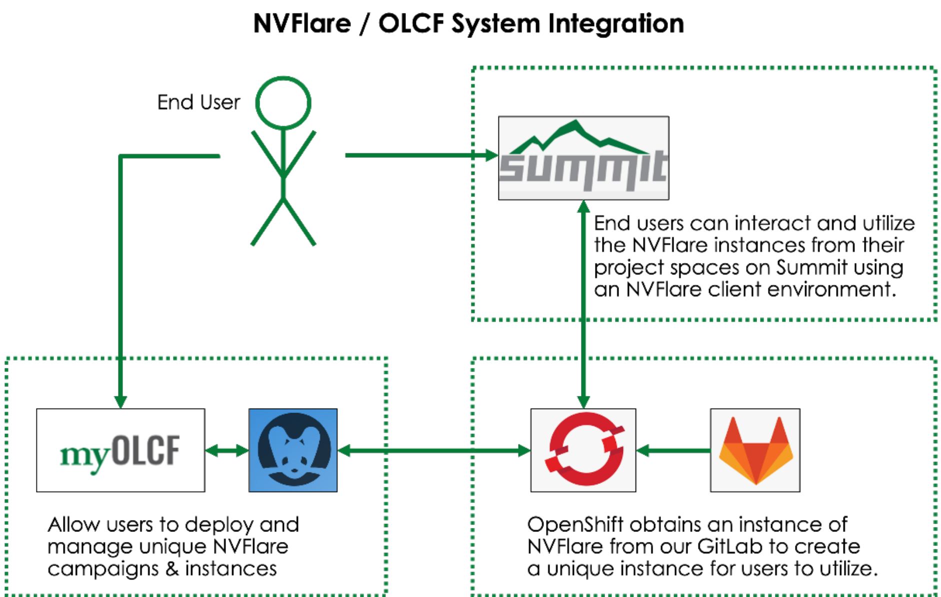

Figure 4-3. The overall NVFlare system integration at the OLCF.

In 2022, OLCF staff members also enabled the automatic deployment of the NVFlare server from OLCF's user portal (MyOLCF), provided NVFlare client environment on Summit, established a partnership with NVIDIA to further NVFlare capabilities on HPC systems, and demonstrated progress at SC22's FL workshop as an invited speaker.

4.2 OPERATIONAL INNOVATION – MANAGEMENT/WORKFORCE

OLCF understands the significance of maintaining a healthy management and workforce structure for ensuring successful and continuous operation. To this end, OLCF is pursuing a multipronged pipeline approach. The first stage of 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 HR and Communications to develop a robust recruitment campaign.

4.2.1 DOE's RENEW Initiative to Support Five Pathway Summer Schools for Students from Underrepresented and Underserved Groups in STEM

The DOE's Office of Science supports nearly 100 high schoolers, recent high school graduates, and early undergraduate students from underrepresented groups and underserved schools in STEM through awards for five Pathway Summer Schools at six national laboratories. OLCF benefits from this program because ORNL is one of the awardees (among Ames, Argonne, PNNL, Fermilab/BNL). The funding comes from the Office of Science's Reaching a New Energy Sciences Workforce (RENEW) initiative (https://science.osti.gov/initiatives/RENEW). Sponsored by the Office of Workforce Development for Teachers and Scientists (WDTS), the WDTS RENEW Pathway Summer Schools aim to diversify the STEM pipeline via hands-on learning opportunities at DOE national laboratories.

The WDTS RENEW Pathway Summer Schools will create a consortium of concerted and continued efforts across the DOE national laboratories to explore, initiate, and sustain inclusive pathways for young students to enter the science and technology workforce in support of the DOE mission. To recruit participants, the Pathways Summer Schools will leverage existing partnerships between the DOE national laboratories and local, regional, and national organizations that support underrepresented groups in STEM. Students will participate in a range of activities to increase STEM efficacy and promote STEM identity, including hands-on learning guided by a mentor or mentoring team, networking with mentors and other summer school participants, and professional development activities. The summer schools will be available during summer 2023.

4.2.2 Regional and On-Site Workforce Advancement

OLCF also develops parallel outreach plans to multiple underrepresented schools across the United States to bolster its student pipeline. OLCF has begun a program with Tennessee Technological University (TN Tech) as an academic alliance. Groundwork in 2022 is expected to lead to OLCF hosting about five interns in 2023. Combined with the Computing and Computational Sciences Directorate's goals to improve diversity, equity, and inclusion, OLCF identified Minority Serving Institutions and women's colleges in the southeast as ideal institutions from which to recruit interns and graduates. OLCF is developing a curriculum to start a targeted summer school program for interns from these schools to provide them with hands-on experience with emerging fields in HPC. The summer school will be a week-long, in-person program with topics and schedule as follows:

- Parallel systems and HPC for beginners (1 day)
- Parallel programming (1 day)
- Data analytics and deep learning at scale using multiple GPUs (1/2 day)
- Parallel I/O and file systems (1/2 day)
- Scientific data and metadata and HPC workflows (1 day)
- Career paths at ORNL and sister DOE national labs (1 day)

4.2.3 On-the-job Training Opportunities

The other stage of 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. We have instituted a staff development program for all new operators and currently have three early career system admins. As the Operations Control Room (OCR) ages out, our goal (just like we did in October 2021) 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 admin skills, and further education in IT (as needed). It is difficult to estimate their time in the OCR prior to promotion to early career admin because it is based on incoming skill set and degree. However, we have a 2–3-year goal for these candidates. As we hired the early career staff members, we looked at their skill sets and desires and placed them in section groups based on those criteria. Our goal is for them to develop rapidly and to use their abilities with their embedded groups to transfer tasks that can be delegated to the operators. Again, how long they are in that role depends on effort and professional progress (and potentially education). Our desire is 1 year at the earliest (unless advanced and other group leaders need their skills) but likely 2 years. Through the progression, our goal is to have a healthy mix of early career and seasoned OCR operators to increase the OCR abilities/skill set and to provide homegrown talent for the other groups. We are currently seeing success with our efforts and current staff. Because other groups across NCCS pick developed talent through this program, we are always looking for the next good entry-level person. Overall, this is a success for NCCS.

4.2.4 Frontier Recruitment Campaign

The OLCF, the Computing and Computational Sciences Directorate, and ORNL central communications teams collaborated with ORNL HR and Recruiting to develop a targeted digital recruitment campaign focused on career opportunities provided by the launch of the Frontier supercomputer. The team held focus group meetings, developed messaging, designed materials, and explored paid placements and advertising to grow visibility of Frontier-related job opportunities and increase the number of qualified applicants for these positions. With the campaign, we learned that Google ads based on keyword search and Facebook ads provide the best return on investment. We spent $30–$33 per application received compared to the US Industry average of $260. We found that paid ads for general job categories provide 50%–90% more clicks than unpaid posts on social media sites. We learned that trade publications provide less return on investment but are still valuable for increasing brand recognition. We observed that posts with videos received 55% more clicks than static images, which aligns with what we expected. We also observed that taglines focused on diversity and solving big problems were most effective. Finally, we had a few observations that need to be tested to confirm. For example, we found that LinkedIn is more expensive than other social media platforms but may be valuable for difficult-to-hire job categories or specific positions, and posts by staff about specific positions may be the most effective use of unpaid ads on social media sites. This experiment was a helpful exercise in coordinating our messages and building up our recruitment efforts.

4.3 POSTDOCTORAL FELLOWS

4.3.1 Program

DOE recognizes the need to train and retain computational scientists in a broad range of disciplines that support DOE and the nation’s critical mission needs 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 ensure a well-trained computational science workforce in academia and industry and at the national laboratories. In recognition of this need, DOE proposed that ASCR establish a postdoctoral training program at its user facilities, including the OLCF, ALCF, and NERSC, for future Computational Scientists for Energy, the Environment, and National Security (CSEEN). The objectives of this program are (1) to help 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) to 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) to 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 develop 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 occur in conjunction with a CAAR project (both OLCF-5 funded and ECP funded). This model offers the greatest potential for scientific breakthroughs through computing and provides ample opportunity to publish in the 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 2022, a total of 13 postdocs were members of the OLCF workforce. Of those 13, 8 were fully supported by OLCF funds (including 3 postdoctoral fellows at least partially supported by ECP Application Integration). Five of the postdocs were at least partially supported by sources outside the OLCF, including individual ECP AD and SciDAC projects. The background and current work of these postdocs in the Science Engagement section is described below.

4.3.1.1 David Eberius

David Eberius joined the Algorithms and Performance Analysis group in October 2020 after earning his PhD in computer science from the University of Tennessee, Knoxville, where he researched techniques for performance measurement and analysis of HPC software. During his graduate studies, he focused on message-passing communications software and collaborated as part of the team that develops the OpenMPI implementation of MPI. Since joining the OLCF, Eberius has been evaluating programming models (implementation and performance optimization) to enable performance portability for the ECP ExaBiome project—in particular, the tools and techniques for understanding the observed performance of code running on GPUs and identifying architectural feature(s) that bottleneck performance. David joined Intel in late 2022.

Publications:

D. Eberius, P. Roth, D. M. Rogers, “Understanding Strong Scaling on GPUs Using Empirical Performance Saturation Size,” 2022 IEEE/ACM International Workshop on Performance, Portability and Productivity in HPC (P3HPC), Dallas, TX, USA, Nov 2022, pp. 26–35, DOI: 10.1109/P3HPC56579.2022.00008.

D. Eberius, D. Boehme, and O. Pearce, “Did the GPU obfuscate the load imbalance in my MPI simulation?" 2021 IEEE/ACM International Workshop on Hierarchical Parallelism for Exascale Computing (HiPar), St. Louis, MO, USA, 2021, pp. 20–29, DOI: 10.1109/HiPar54615.2021.00008.

4.3.1.2 Paul Eller

Paul Eller joined the Algorithms and Performance Analysis Group in January 2020 after earning his PhD in computer science from the University of Illinois at Urbana-Champaign. During his doctoral studies, he focused on techniques for measuring, modeling, and improving the scalability and performance of blocking and non-blocking Krylov solvers. To support this research, he developed the Scalable Algorithm Test Bed, which is a collection of tools for measuring and analyzing the performance (and performance variation) of computational kernels and network communication. While at the OLCF, Eller collaborated with the CoMet comparative genomics project team to study the performance benefit of using reduced-precision numerical methods in computationally intensive GPU kernels. Paul left ORNL in January 2022 to work for General Dynamics at the DOD Army Research Lab in Aberdeen proving Ground, MD, where he will be working to optimize codes on DOD supercomputers.

Publications:

J. Lagergren, M. Cashman, V. Melesse Vergara, P. Eller, J.G.F.M. Gazella, H. Chhetri, J. Streich, S. Climer, P. Thornton, W. Joubert, and D. Jacobson, "Climatic clustering and longitudinal analysis with impacts on food, bioenergy, and pandemics," Phytobiomes Journal 6(3):1–34, Sep 2022, DOI: 10.1094/PBIOMES-02-22-0007-R

4.3.1.3 Nicholson Koukpaizan

Nicholson Koukpaizan joined the Algorithms and Performance Analysis Group in June 2020 after earning his PhD in aerospace, aeronautical, and astronautical engineering from the Georgia Institute of Technology. As a member of Georgia Tech's Nonlinear Computational Aeroelasticity Laboratory, he researched techniques for modeling fluidic actuation for aerodynamic flow control with a special focus on jet interaction fluidic oscillators sited on helicopter bodies. Since joining the OLCF, Koukpaizan has worked on evaluating programming models (implementation and performance optimization) to enable performance portability for ECP applications, particularly for the ExaSGD (Power Flow Optimization at Exascale) project. ExaGO and HiOp depend on RAJA and MAGMA to target GPUs. Koukpaizan is developing a GPU-accelerated computational fluid dynamics code for aerodynamic flow control applications; the code now supports OpenACC and OpenMP offloading to target the GPUs. In February 2022, Koukpaizan began as a staff member in the Advanced Computing for Life Sciences and Engineering Group as a Computational Scientist, Fundamental Turbulence.

Publications:

Nicholson K. Koukpaizan, Ari Glezer, and Marilyn J. Smith. (2021). Computational Characterization and Boundary Condition Models of a Jet Interaction Fluidic Oscillator. *AIAA Journal*, 59(9).

Koukpaizan, N. K., Glezer, A., and Smith, M. J. (2020). "Towards full-scale fuselage drag reduction computations using fluidic oscillators." Paper presented at the Vertical Flight Society's 76th Annual Forum and Technology Display.

4.3.1.4 Isaac Lyngaas

Isaac Lyngaas joined the Advanced Computing for Life Sciences and Engineering Group in September of 2019 after earning his PhD in computational science from Florida State University, where he worked on the development of new numerical approaches for solving nonlocal continuum models. During his time at ORNL, he has worked on refactoring a cloud-resolving model's Fortran code base to C++ using a performance portability library for single-source portability. Other work in this area includes implementing OpenMP, SYCL and CUDA backends in a C++ portability library for accelerated hardware and the development of Clang-based precompiler tools to eliminate GPU run time errors in a C++ Portability Library. He began as a staff member in the same group in March 2022. Lyngaas is currently collaborating with the MOSSAIC/CANDLE research group in CSED working to implement scalable transformer ML architectures on tasks related to document classification for cancer pathology reports. Lyngaas and several internal ORNL collaborators participated in the American Association of Physicists in Medicine grand challenge; this team won the grand challenge prize for the Truth-Based CT (TrueCT) reconstruction challenge. This effort to enter the imaging science space has also led to collaborations with a radiology research group at Duke University. He is currently working with mouse brain imaging data from this group, and his research efforts include work to train large-scale ML models for image segmentation on OLCF systems.

Publications:

Norman, M, Lyngaas, I., Bagusetty, A., and Berill, M. (2022) “Portable C++ Code that can Look and Feel Like Fortran Code With Yet Another Kernel Launcher (YAKL).” *International Journal of Parallel Programming*. DOI: s10766-022-00739-0

Other artifacts:

Member of team that was the Winner of the 2022 American Association of Physicists in Medicine Grand challenge: Truth-Based CT Reconstruction Challenge

Member of team that was a finalist for the 2021 COVID-19 Gordon Bell Prize at SC21

4.3.1.5 Murali Meena

Murali Meena is a computational scientist (Fundamental Turbulence) in the Advanced Computing for Life Sciences & Engineering group in the Science Engagement section. Prior to joining ORNL, Murali earned his PhD in mechanical engineering from the University of California, Los Angeles in 2020. His PhD research focused on using network (graph) theory to characterize, model, and control vortical interactions in turbulence and wake flows. In particular, he introduced a network community-based framework to identify closely interacting vortical elements that were used to formulate reduced-order models and perform flow-modification of complex laminar and turbulent vortical flows. Murali joined OLCF in July 2020 as a postdoctoral research associate at the Advanced Computing for Life Sciences and Engineering group. His research focused on using ML to formulate subgrid-scale turbulence closures for atmospheric flows. He began as a staff member in the same group in March 2022. He is currently continuing this work and additionally working with users of OLCF resources in projects that involve complex fluid flows and turbulence. Additionally, he is also using ML, graph theory, and other data-driven techniques to characterize and model complex systems in physical and biological sciences. They are targeted at applications such as identifying extreme climate events, geophysical flow reconstruction and modeling, development of clustering techniques in graph theory for applications in turbulent flows and geophysical applications, and identifying key interactions in various fungal communities for biofertilizer development and revealing microbe-host interactions.

Publications:

T. A. Rush, J. Tannous, M. J. Lane, M. Gopalakrishnan Meena, A. Carrell, J. Golan, M. Drott, S. Cottaz, S. Fort, J.-M. Ané, N. Keller, D. Pelletier, D. Jacobson, D. Kainer, P. Abraham, R. Giannone, and J. Labbé, (2022) “Lipo-chitooligosaccharides induce specialized fungal metabolite profiles that modulate bacterial growth,” mSystems, e01052-22.

M. Gopalakrishnan Meena, M. R. Norman, and D. M. Hall, “Subgrid-scale surrogate modeling of idealized atmospheric flows: A deep learned approach using high-resolution simulation data,” 12th International Symposium on Turbulence and Shear Flow Phenomena, Osaka, Japan, July 19–22 (Online), 2022 (TSFP 12 229) [peer-reviewed conference proceedings].

N. L. Hickmon, et al., “Artificial Intelligence for Earth System Predictability (AI4ESP) Workshop Report (Section: Knowledge Discovery and Statistical Learning),” No. ANL-22/54. Argonne National Laboratory, Argonne, IL (United States), 2022.

Other artifacts:

Member of team that won the 2022 American Association of Physicists in Medicine Grand challenge: Truth-Based CT Reconstruction Challenge

Lead of Invention disclosure sponsored by ORNL, UT-Battelle: M. Gopalakrishnan Meena, M. J. Lane, J. Tannous, and T. A. Rush, “Using bipartite networks to determine interactions between analytes and chemical treatments,” 2021 (provisional patent application filed)

4.3.1.6 Darren Hsu

Darren Hsu joined the Advanced Computing for Chemistry and Materials Group in May 2021. Previously, he had a postdoctoral appointment at AstraZeneca from October 2020 to April 2021. Hsu earned his PhD in chemistry from Northwestern University in 2020. His research focused on utilizing ultrafast, time-resolved x-ray scattering and molecular dynamics to track structural changes of small molecules and proteins. Since joining the OLFC, Hsu has been working on developing ML Boltzmann generators to sample particle configurations efficiently. In addition, he has been contributing to the molecular modeling effort by modifying mdgx.cuda code for high-throughput simulations and a transformer-based AI model for protein-ligand structure prediction. His postdoctoral work with his mentor Jens Glaser resulted in a Gordon-Bell Special Prize finalist award at SC22, “TwoFold: highly accurate structure and affinity prediction for protein-ligand complexes from sequences.” Hsu interviewed for a computational scientist, biophysics position in the Advanced Computing for Chemistry and Materials Group; he was a successful candidate but accepted a solution architect position at NVIDIA.

4.3.1.7 Swarnava Ghosh

Swarnava Ghosh joined the Advanced Computing for Chemistry and Materials Group in August 2020. Previously, he had a postdoctoral appointment from October 2016 to August 2020 in the Department of Mechanical and Civil Engineering at the California Institute of Technology. Ghosh earned his PhD from the Georgia Institute of Technology in August 2016 for his work on efficient, large-scale, real-space electronic structure calculations. His research at ORNL investigates the magnetism in disordered systems by using first principles, real-space multiple scattering calculations, such as the dependence of magnetism on disorder in Mn–Sb–Te quantum material systems and the influence of spin on the mechanical behavior of magnetic structural alloys. Ghosh contributes to the development of the ORNL Locally Self-Consistent Multiple Scattering (LSMS) code. In particular, he is extending the capability of this code to the efficient calculation of interatomic forces and non-spherical atomic potentials. Ghosh successfully interviewed for a computational scientist, materials position in the Advanced Computing in Chemistry and Materials Group and is presently liaison to two INCITE projects.

Publications:

Ghosh, S. “Towards Ab-Initio Simulations of Crystalline Defects at the Exascale Using Spectral Quadrature Density Functional Theory.” *Appl. Mech.* 2022, 3, 1080–1090. https://doi.org/10.3390/applmech3030061.

Ghosh S. and Bhattacharya K. “Spectral quadrature for the first principles study of crystal defects: Application to magnesium.” *Journal of Computational Physics* 456, 1 May 2022, 111035.

4.3.1.8 Mariia Karabin

Mariia Karabin joined the Advanced Computing for Chemistry and Materials Group in July 2020 after earning her PhD from Clemson University for her work on the application of statistical mechanics methods to the construction of interatomic potentials and sampling algorithms. Her research at ORNL focuses on the first principles investigation of the physics of high-entropy alloys based on first principles density functional calculations, developing multiple scattering-based methods, first-principles derived Hamiltonians to Monte Carlo simulations, and application of these methods to the phase transitions and atomic behavior of alloys. She is also working in the CAAR to prepare new capabilities (Lloyd's formula) and accelerate the LSMS first principles multiple scattering code for efficient use on Frontier and its combination with a model training and Monte Carlo workflow on this platform. Karabin's 3rd year extension expires in July; she is actively pursuing permanent staff positions in industry, academia, and national labs.

Publications:

M. Karabin et al. "Ab initio approaches to high-entropy alloys: a comparison of CPA, SQS, and supercell methods." *Journal of Materials Science* 57(23) 10677–10690.

4.3.1.9 Hector Corzo

Hector Corzo joined the Advanced Computing for Chemistry and Materials Group as an OLCF postdoctoral fellow in January 2022 after finishing his 2-year postdoctoral appointment at the Center for Chemical Computation and Theory Department of Chemistry & Biochemistry at the University of California, Merced. Corzo earned his PhD in theoretical and computational chemistry at Auburn University in 2018 for his work developing a novel multideterminant electron propagator for open-shell molecular systems. Corzo has extensive experience in highly accurate electronic structure methods and implementation on GPU-accelerated supercomputers. Under the supervision of his postdoc mentor, Dmytro Bykov, Corzo is implementing novel coupled cluster methods into the LS-Dalton HPC chemistry program. He is also using his time at ORNL to build skills in ML and QC.

Publications:

H. H. Corzo et al. "Using projection operators with maximum overlap methods to simplify challenging self-consistent field optimization." *Journal of Computational Chemistry* 43(6) 382–390.

A. A. Taka and H. H. Corzo. ""Good Vibrations: Calculating Excited-State Frequencies Using Ground-State Self-Consistent Field Models." 18(12) 7286–7297.

4.3.1.10 Baishan Hu

Baishan Hu joined the Advanced Computing for Nuclear, Particle, and Astrophysics group in Oct 2022 as a postdoctoral research associate. Baishan completed his PhD in theoretical nuclear physics from Peking University in July 2017. He then pursued a postdoc at Peking University in the following two years. Afterward, Baishan joined TRIUMF in Canada as a postdoctoral fellow from November 2019 to October 2022. His research focuses on ab initio calculations of atomic nuclei and related topics such as dark matter, fundamental symmetries, and nuclear astrophysics. Since joining the OLFC, Baishan has been working on state-of-the-art simulations for nuclear structure and reactions, both performing and analyzing simulations and developing the next generation of simulation code and analysis tools.

4.3.1.11 Charles Stapleford

Charles Stapleford joined the Advanced Computing for Nuclear, Particle, and Astrophysics group in October 2020. He previously earned a PhD in physics at North Carolina State University. His research has focused on neutrino oscillations and transport in astrophysical simulations. Since joining the OLCF, Charles has been working on the development of the neutrino transport portion of the first supernova code designed to run at exascale. In addition, he has been working to modernize a well-established 1D supernova code for HPC.

4.3.1.12 Tor Djarv

Tor Djarv earned his PhD in physics from Chalmers University of Technology in Gothenburg, Sweden, where his primary focus was on ab initio nuclear physics with the no-core shell model. More specifically, his research focused on how to include realistic chiral three-nucleon forces either directly or through approximations in the no-core shell model and to develop a new no-core-shell-model code named JupiterNCSM. His research at ORNL focuses on ab initio nuclear physics with coupled-cluster and weak processes such as neutrinoless double beta decay.

4.3.1.13 Henry Monge Camacho

Henry Monge Camacho joined the Advanced Computing for Nuclear, Particles, and Astrophysics Group in April 22. Previously, He held a postdoctoral appointment at the University of North Carolina at Chapel Hill from 2018 to 2020, and he was a Visiting Professor at the University of Costa Rica from 2020 to 2022. Henry earned his PhD in physics from the College of William and Mary. His research focuses on the calculation of nuclear matrix elements from Lattice QCD, which is a technique to numerically solve the effects of strong interactions between elementary particles. At ORNL, Henry is working on the improvement and further development of applications for Lattice QCD calculations of hadronic spectrum contractions to run on GPU devices. He is also involved in computing electromagnetic corrections to QCD using Lattice QCD.

2022 5 – Risk Management

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

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 2022 are listed and described in this section. Planned mitigations and implementations are included in the subsequent descriptions. As of this writing, the OLCF has zero high-priority operational risks, but that could change because the risk management approach continuously reviews and assesses OLCF operations for 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. The plan covers both OLCF operations and its various projects (OLCF-5 during CY 2022). Each project execution plan refers to the main Risk Management Plan but may also incorporate project-specific adjustments. 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, 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, 24 active entries are in the OLCF operations risk register that fall into two categories: risks for the entire facility and risks for a specific portion of the facility. Facility-wide risks are concerned with safety, funding, expenditures, and staffing. The more focused risks are concerned with reliability, availability, and use of the system or its components (e.g., the computing platforms, power and cooling infrastructure, storage, networks, software, and user support).

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 2022

Table 5-1 contains the major risks tracked for OLCF operations in 2022. The selected risks are all rated medium or high in impact.

Table 5-1. 2021 OLCF major risks.

Risk ID/description	Probability/impact	Action	Status
723: Safety – personal injury	Low/medium	Mitigating	We 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 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.
1063: Programming environment and tools may be inadequate for future architectures	Medium/medium	Mitigating	OLCF has continued to work with vendors, standards committees, and users to ensure adequate tools are available for users on future architectures. Additionally, we track several risks related to programming environment (both compilers and accelerator libraries) and tools as part of the OLCF-5 risk register and routinely review the status of those risks and work to mitigate or possibly retire them.
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.

New Risks in this Reporting Period

No risks were added 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

Frontier will enter user operations in 2023, and operations on both Frontier and Summit will be critical to the success of the OLCF in 2023. Risk ID 1245 (System unavailability due to mechanical/electrical system failure) directly impacts Summit. While this is mitigated through preventative maintenance, inspections, and monitoring, it remains in place because a trigger of that risk could have significant effects.

Major construction activities associated with OLCF-5 have wound down in and around the building; however, some activities are continuing. The construction requires occasional closure of some areas as well as ingress/egress paths for staff and visitors. This requires increased focus on safety in and around the work areas by all staff (Risk ID 723).

We anticipate adding several risks that have been tracked as part of the OLCF-5 project into the Operations Risk Register as Frontier enters operations. These fall into the broader category of system use, such as the use of newer programming models and tools and interoperability between different programming models. In addition to being tracked as part of the OLCF-5 project, this is somewhat addressed in Risk 1063 in which OLCF continues to work with hardware and software vendors, standards bodies, and the user community to deploy tools and programming environments to provide a smooth transition.

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

The risks listed in Table 5-2 were encountered and effectively mitigated in 2022. A short summary of the status and impact on the operations of the OLCF is included for each risk.

Table 5-2. Risks encountered and effectively mitigated in CY 2022.

Risk No. 408	System outages from external causes
Risk owner	James H. Rogers
Status	Accept
Probability	Low
Impact	Cost: Low Schedule: Low Scope/Tech: Low
Trigger Event	On May 27, 2022, a tree fell on a feeder unrelated to OLCF; however, as that line was re-energized, a lightning arrestor on a pole related to that feeder blew. This resulted in a quick loss to ground on a feeder for OLCF, which caused a 0.9 s power loss. This impacted 50% of Summit.
Mitigations	Fact-finding was conducted and found that this failure is not predictable and is rare. Discussions for preventable line maintenance as well as other monitoring to locate any suspect future failures was conducted as part of fact finding.
Triggers	Intelligence on potential or actual outage situations or conditions.

2022 6 – Environment Safety and Health

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

6. ENVIRONMENT, SAFETY, AND HEALTH

CHARGE QUESTION 6: Does the facility exhibit a culture of continuous improvement in environment, safety, and health (ES&H) practices to benefit staff, users, the public, and the environment? Has the Facility implemented appropriate ES&H measures?

OLCF RESPONSE: Yes. The OLCF is committed to providing a safe workplace for ORNL employees, subcontractors, and visitors. 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. In addition, 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 (SMEs), or a combination. In addition to utilizing the ORNL SMEs, 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 seek 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 2022, the normal day-to-day operations met the zero accident performance criteria and remained safe, efficient, and effective due to 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

The OLCF continued to operate 24-7-365 with operator support on each shift. However, the operator staff was short staffed due to illness, and the OLCF resorted to implementing our *fill-in* staff, which were roles added as a response to COVID-19. Previous years' efforts to create Emergency Fill-in Guides for emergency staffing proved valuable and ensured continuous operation.

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. Noise survey results were included in the RSS for the required reading of all staff.

6.1.2.3 Safety Talks

Safety 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 91 safety talks in CY 2022.

6.1.2.4 Walk Throughs

Walk throughs are conducted at set intervals for the OLCF operators. Walk throughs are performed at a set frequency and in set room locations. Operators focus on facility conditions and observation of work practices. The frequency of walk throughs was modified in CY 2022 to reflect the busy period. Over 5,500 walk throughs were conducted by the staff of five operators.

6.1.2.5 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 2022 large-scale workload included the completion of the Frontier installation, initial start up, break fix maintenance, troubleshooting and repair, subcontractor cooling loop water quality issues, pump replacements, large-scale receiving/shipping, and other center-related activities. The Frontier activities were challenging due to (1) the introduction of chemicals not typically used in a data center for cooling-loop cleaning, (2) the coordination and safe integration of conventional craft support, and (3) analyzing hazard/controls for troubleshooting and repairing new equipment. In additional to Frontier, the NOAA Gaea C5 machine was safely installed this year.

For CY 2022, 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, colocated ORNL employees, or the environment.

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

6.2.1.1 Chemical Hazards

The Frontier vendor's cooling-loop cleaning required chemicals not normally found in the hazard set for a data center. The center manager worked diligently with vendor ES&H support to analyze the hazards and to develop controls. Controls were developed to include requiring segregation of materials; proper PPE; proper waste disposal; unique marking/labeling of containers, pumps, and systems to indicate which chemical they were to handle; and separating this effort from other workers. Additionally, all systems for adding and removing coolant were carefully reviewed, and changes were made to reflect ORNL requirements. The result of ORNL's effort provided the vendor with a process that became the basis for their operations for cleaning cooling loops. As a result, these ES&H-related controls were implemented at other vendor installation sites.

In addition to the support in the facility, the center relied on additional ORNL SMEs to assist with proper waste characterization and laboratory analysis of samples. The SMEs also provided concurrence that the controls developed were adequate.

6.2.1.2 Material Handling

The chemical flushing and the parts received were analyzed for material handling hazards. ORNL insisted that the vendor purchase proper material handling equipment to move hundreds of 55-gallon drums and the spare parts. Material handling controls for these products were added to the hazard analysis for the associated tasks. Material handling was constantly used in pre-task briefings and daily stand up meetings. The vendor's original approach to the material handling did not include specialized equipment and relied on manual labor. ORNL provided some material handling equipment and insisted that the vendor reduce the manual risk and the vendor purchased specialized equipment including drum lifts, dollies, carts, etc. In addition, ORNL provided additional resources to temporarily store new products until ample space could be made in the facility.

6.2.1.3 Oversight

The large volume of work associated with the chemical flushing and Frontier start up resulted in expanded shifts. During this period, the Facility operated two shifts per day for 7 days a week. Additionally, at one period, the vendor worked three shifts for a 24-hour operation. To provide consistent oversight, center staff changed shifts, additional staff members were used, center staff worked extra days, and walk-through frequencies were increased. ORNL also required the vendor to ensure that each shift included a senior manager who was familiar to ORNL's operations. Traveling personnel and/or temporary hires were not allowed to work without senior level oversight.

6.2.1.4 Pre-Entry Briefing

All vendor personnel arriving on site were given a site briefing by the center manager prior to working in the center. The briefing covered the hazard analysis (tasks, hazards, and controls), the applicable safe conduct of research principles, human performance, applicable lessons learned, and stop/suspend work authorization. The briefing was approximately 1 hour, and over 125 different visitors have been briefed.

6.2.1.5 Safety Pauses

Safety pauses were issued several times by the center manager as he deemed necessary to refocus on safety. The short pauses were used to discuss recent leading indicators/non-compliances to procedures, hazard analysis, to re-stress safe work and the safe conduct of research principles, and to ensure that everyone was committed to safe execution of the work. Project managers, engineers, and staff were invited to attend, and all staff were asked directly if they had something to contribute. In some cases, staff members contributed suggestions for changes that were incorporated into the work process.

6.2.1.6 Post Task Feedback Meetings

The center manager held numerous post-task feedback meetings. These meetings were informal and conducted just after a new task was completed for the first time. The meetings asked several questions: What did we do well? What did we do poorly? What can we improve? Feedback was received from the staff, and the center manager provided feedback based on his observations. The process provided continuous improvement and an opportunity for the staff to critique their performance. The feedback was incorporated into the next iteration of the task, including any applicable changes to the hazard analysis or vendor-documented work control.

6.2.1.7 Subcontractor ES&H Engagement

The engagement of the subcontractor's ES&H support was critical to ensuring that the vendor was adequately taking ownership of their staff's safety and health and fulfilling subcontract ES&H requirements. This effort or engagement had not been past practice for the center or the vendor. The effort was successful in CY 2022 and will be expanded in CY 2023 to include site visits and routine meetings between vendor ES&H and the center manager.

2022 7 – Cybersecurity

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

7. SECURITY

CHARGE QUESTION 7:

- Does the Facility exhibit a culture of continual improvement in cybersecurity practices?
- Does the Facility have a valid cyber security plan and Authority to Operate?
- 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. The most recent OLCF Authority to Operate was granted on February 21, 2017 and is managed through an ongoing authorization process; no authorization termination date is set (Figure 7-1). Updates to the ORNL Cyber Security Program Plan as well as the Supercomputing security zone System Security Plan have been made in CY 2022 to update policies to align with the newest revision (5) of NIST SP 800-53. We expect to have a new Authority to Operate letter sometime in CY 2023 when the Authorizing Official has reviewed the updated package and assessment results.

The technical staff members track and monitor for existing threats and vulnerabilities to assess the risk profile of the OLCF operation. The Facility 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
ORNL Site Office
P.O. Box 2008
Oak Ridge, Tennessee 37831-6269

March 22, 2017

Mr. Kevin A. Kerr
Information Systems Security Manager
Oak Ridge National Laboratory
UT-Battelle, LLC
Post Office Box 2008
Oak Ridge, Tennessee 37831-6045

Dear Mr. Kerr:

AUTHORIZATION DECISION DOCUMENT FOR OAK RIDGE NATIONAL LABORATORY (ORNL) SUPERCOMPUTING ENCLAVE

Reference: Letter from Kevin A. Kerr to Johnny O. Moore, subject, *Contract DE-AC05-00OR22725, ORNL Supercomputing Enclave Approval to Operate*, dated February 21, 2017

As the Authorizing Official, I have reviewed the referenced request. The ORNL Supercomputing Enclave is authorized to operate. No additional conditions outside the substance of the request are required.

The information system is now being managed by an ongoing authorization process, thus an authorization termination date is not set. I accept the responsibility for performing all necessary activities associated with the ongoing authorization process.

If there are any questions or additional information is required, please contact John Young at (865) 576-7471 or youngjc1@ornl.gov.

Sincerely,

Johnny O. Moore, Manager
ORNL Site Office

Enclosure

cc w/enclosure:
Mike E. Bartell, ORNL
Amy D. Nuckols, ORNL
Neil Masincupp, SC-OR
Martha J. Kass, SC-OSO
John C. Young, SC-OSO

Document/Material Transmitted Contains Official Use Only Information.

When separated from enclosure, this document **does not** contain Official Use Only.

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 4, 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 of controls for protecting the confidentiality and integrity of user and system information, which authorizes the Facility 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 inter-group coordination and use of CI) 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 (HIPAA'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 where 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 led the development of a software development lifecycle framework plan in CY 2022 in collaboration with the Software Services Development and Platforms groups. The purpose is to identify stakeholders early in the lifecycle of a project and engage other non-security operational teams to gather requirements and influence the design to harden the eventual software product. Additionally, various software scanning tools were evaluated and tested, which will result in a new software scanning capability for OLCF-developed custom software to help identify software weaknesses and potential vulnerabilities. Significant amounts of technical debt were also paid off this year: major upgrades to vulnerability scanning software versions and authenticator synchronization application containerization reduced the amount of break/fix work required to maintain those systems.

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 2022, the OLCF security team tracked 41 above-baseline threats—none of which resulted in a compromise of OLCF systems by malicious attackers. We treat each of these events as a live incident 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.

At its core, security has recently become a data science problem. Network intrusion detection, host-based intrusion detection, system and firewall logging, vulnerability scanning, netflow collection, and databases (e.g., hardware inventory, software inventory, and user CRM) all have information to help an incident responder take appropriate action. The OLCF security team continues to develop Copacetic to help wrangle data from these different data sources and apply business rules to ensure that the actual state of the systems within the data center match the desired state. As zero-trust initiatives are completed in CY 2023, active monitoring of data sources to provide real-time access control decision-making will be required.

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 are defined in more detail 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]	UA[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 RPS screening.	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 screening.	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 All US citizens or lawful permanent residents go through RPS screening.	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] UA: 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 UA must be executed with each user's institution.
[4] Sensitive data rules: This form contains the user acknowledgment, which 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.

2022 8 – MISSION IMPACT, STRATEGIC PLANNING, AND STRATEGIC ENGAGEMENTS

HIGH PERFORMANCE COMPUTING FACILITY
2022 OPERATIONAL ASSESSMENT
OAK RIDGE LEADERSHIP COMPUTING FACILITY
April 2023

8. STRATEGIC RESULTS

CHARGE QUESTION 8:

- Are the methods and processes for monitoring scientific accomplishments effective?
- Is the facility collaborating with technology vendors and /or advancing research that will impact next generation HPC platforms?
- Has the Facility demonstrated effective engagements with critical stakeholders (e.g., the Office of Science's Science Programs, DOE Programs, DOE national laboratories, User Facilities, and/or other critical US 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 our understanding of nature and strengthen the connection between advances in fundamental science and technology innovation

8.1 SCIENCE HIGHLIGHTS AND ACCOMPLISHMENTS

The Facility collects and reports annually the number of refereed publications that result (at least in part) from the use of the Facility's 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 submitted or in preparation. This is a reported number, not a metric. In addition, 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 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 possibilities for features.

8.1.1 OLCF Publications Report

In 2022, 532 publications that resulted from the use of OLCF resources were published, based on a data collection completed on April 10, 2023. In this document, *year* refers to the calendar year (CY) unless it carries the prefix FY. In the 2021 OLCF OAR, 497 publications were reported (this number has been significantly revised upward; see below). A list of 2014–2022 publications is available on the OLCF website (https://www.olcf.ornl.gov/publications/). Guidance allows accepted and in-press publications to be reported, but the OLCF reports only publications that appear 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 2012–2022 period, which shows 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 2012–2022.

Year	Unique, confirmed OLCF publications	High-impact publications with JIF* >10
2022	532	17
2021	555	16
2020	534	19
2019	467	18
2018	497	20
2017	477	27
2016	468	33
2015	366	21
2014	315	16
2013	364	9
2012	342	20

*JIF = Journal impact factor

8.1.2 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 outreach. 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 2022, these accomplishments advance the state of the art in science and engineering R&D across diverse disciplines and advance DOE's science programs toward their targeted outcomes and mission goals. OLCF users published many breakthrough publications in high-impact journals in 2022, which is an additional indication of the breadth of these achievements. These publications are shown in Table 8-2.

Table 8-2. Publications in high-impact journals in 2022.

Journal	Number of publications
Science Translational Medicine	1
Science	1
Neuron	1
Nature Physics	1
Nature Materials	1
Nature Climate Change	2
Nature Chemical Biology	1
Nano Letters	1
Journal of the American Chemical Society	1
IEEE Transactions on Evolutionary Computation	1
Annual Review of Fluid Mechanics	1
Angewandte Chemie-International Edition	1
Advanced Materials	3
ACS Nano	1

Altogether in 2022, OLCF users published 49 papers in journals with a journal impact factor (JIF) of greater than 7 and 17 papers in journals with a JIF greater than 10.

8.1.3 Science Highlights

8.1.3.1 Pushing the Frontier in the Design of Laser-Based Electron Accelerators

The plasma simulation code's early runs on Frontier produced impressive results for a proposed laser-plasma particle accelerator.

PI: Jean-Luc Vay, LBNL
Allocation Program: DD, ECP

The Science

Experiments for high-energy physics require massive accelerators, such as the 16-mile-long Large Hadron Collider near Geneva, Switzerland, or the Spallation Neutron Source at ORNL, to generate the particle speeds necessary for scientific discovery. Plasma acceleration was first proposed more than 40 years ago to investigate more compact accelerators for high-energy physics experimentation. Scientists posited that shooting a charged particle beam or laser through a low-density plasma would displace the particles, thereby creating an electric field to accelerate electron or positron beams in a much shorter distance than in conventional radio-frequency accelerators. This process could result in smaller, cheaper accelerators for science. WarpX, the first mesh-refined, particle-in-cell code for kinetic plasma simulations that is optimized for parallel computing, has been developed mainly to simulate plasma-based particle accelerators. Researchers do not have analytical tools or experimental diagnostics that are accurate enough to capture the physical phenomena at play in the laser-matter interaction, so they need high-fidelity numerical tools such as WarpX. And because it involves the simulation of many plasma particles, these simulations are very costly and require the largest machines available.

The Impact

As part of the ECP, the team optimized WarpX to run on some of the world's fastest supercomputers—ORNL's Frontier and Summit, Berkeley Lab's Perlmutter, and Fugaku at the Riken Center for Computational Science in Kobe, Japan—despite their different architectures. Combined with the software's adaptive mesh refinement—the ability to refine the resolution only in a certain region of the simulation grid to increase the speed and accuracy of its calculations—this optimization makes WarpX capable of much faster, lower-cost, and higher-fidelity 3D models of laser-matter interactions than current codes.

Summary

The WarpX team produced a 3D simulation at scale of their own novel concept: a combined plasma particle injector and accelerator, which focuses a high-power femtosecond (1 quadrillionth of a second) laser onto a hybrid solid/gas target. The simulation's predictions were later validated by a proof-of-concept experiment performed on the Salle Jaune laser at Laboratoire d'Optique Appliquée in France by Adrien Leblanc and other CEA collaborators. The team reported that WarpX runs 500× faster than the previous version of the code, Warp, since they began working on their ECP project 6 years ago. They also measured their Gordon Bell paper's figure of merit (the quantity used to characterize a method's performance) in terms of how many updates per second each system achieved. Frontier tops the list at 1.1e13 per second, Fugaku came in second at 9.3e12 per second, and Summit's best result was third place at 3.4e12 per second. The WarpX team used Frontier's exascale power (1.1 billion-billion floating point operations per second of theoretical peak performance) for their submission to the Gordon Bell competition and won the Association for Computing Machinery's (ACM's) 2022 Gordon Bell Prize in November.

Funding

This research was supported by the ECP.

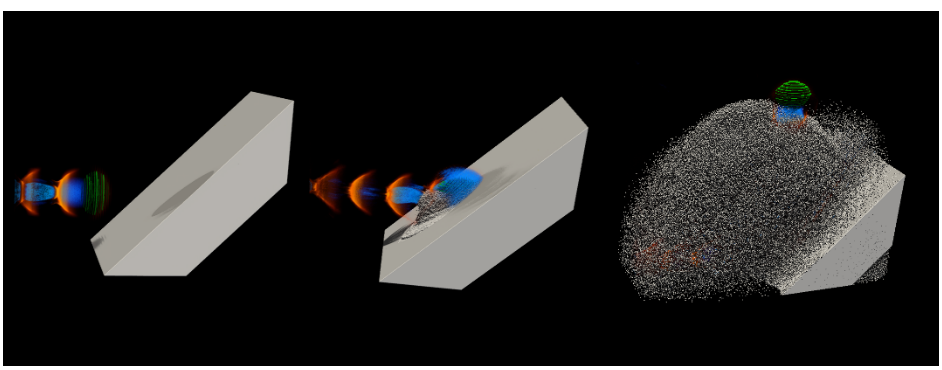

Figure 8-1. Using the Frontier supercomputer, the WarpX team produced a 3D simulation at scale of their own novel concept: a combined plasma particle injector and accelerator, which focuses a high-power femtosecond (1 quadrillionth of a second) laser onto a hybrid solid/gas target. The simulation's predictions were later validated by a proof-of-concept experiment performed on the Salle Jaune laser at Laboratoire d'Optique Appliquée in France by Adrien Leblanc and other CEA collaborators. Image credit: Dave Pugmire/ORNL

Publication: Luca Fedeli, et al., "Pushing the Frontier in the Design of Laser-Based Electron Accelerators with Groundbreaking Mesh-Refined Particle-In-Cell Simulations on Exascale-Class Supercomputers," Paper Presented at the International Journal of High Performance Computing Applications, 2022.

Related Links: "WarpX Named Gordon Bell Prize Finalist," OLCF News (October 28, 2022)

8.1.3.2 Distilling How Water Turns into Ice

Using AI, quantum mechanics, and the Summit supercomputer, a team of Princeton University researchers have produced the largest, most accurate water-to-ice simulation.

PI: Roberto Car, Princeton University
Allocation Program: ALCC

The Science

Using ORNL's Summit supercomputer, Princeton University researchers have taken trailblazing steps to study nucleation—how water turns into ice. Through a combination of quantum mechanics and AI, the team has simulated how water molecules transition into solid ice at a level of accuracy and at a system size unmatched by previous methods. Their study reveals the nucleation rates (how fast ice clusters form) in systems composed of hundreds of thousands of atoms.

The Impact

Such nucleation data could serve in a variety of initiatives, from improving the accuracy of weather and climate modeling to advancing the process of flash-freezing food. The Princeton team's nucleation study is also one of the most significant applications yet of DeePMD (named for deep potential molecular dynamics), a 2020 Gordon Bell Prize–winning simulation code optimized for HPC. DeePMD uses quantum mechanics to describe the motion of atoms and molecules and AI to make predictions of how they will interact. The team trained a neural network in small molecular systems to identify atomic configurations with their corresponding forces and then applied it to larger systems.

Summary

Princeton University researchers have taken trailblazing steps to study the nucleation phenomenon from the fundamental laws of nature by using the Summit supercomputer. Through a combination of quantum mechanics and AI, the team has simulated how water molecules transition into solid ice at a level of accuracy and at a system size unmatched by previous methods. Their study reveals the nucleation rates (how fast ice clusters form) in systems composed of hundreds of thousands of atoms. Such data could serve in a variety of initiatives, from improving the accuracy of weather and climate modeling to advancing the process of flash-freezing food. So far, the Princeton team's nucleation study is perhaps one of the most significant applications of DeePMD (named for deep potential molecular dynamics), a 2020 Gordon Bell Prize–winning simulation code optimized for high-performance computing. DeePMD uses quantum mechanics to describe the motion of atoms and molecules. Such first-principles simulations can produce more accurate results with greater predictive power than simpler, semi-empirical approaches but at the cost of requiring much greater computational power.

Funding

This work was conducted within Princeton's Chemistry in Solution and at Interfaces center, which is funded by the DOE.

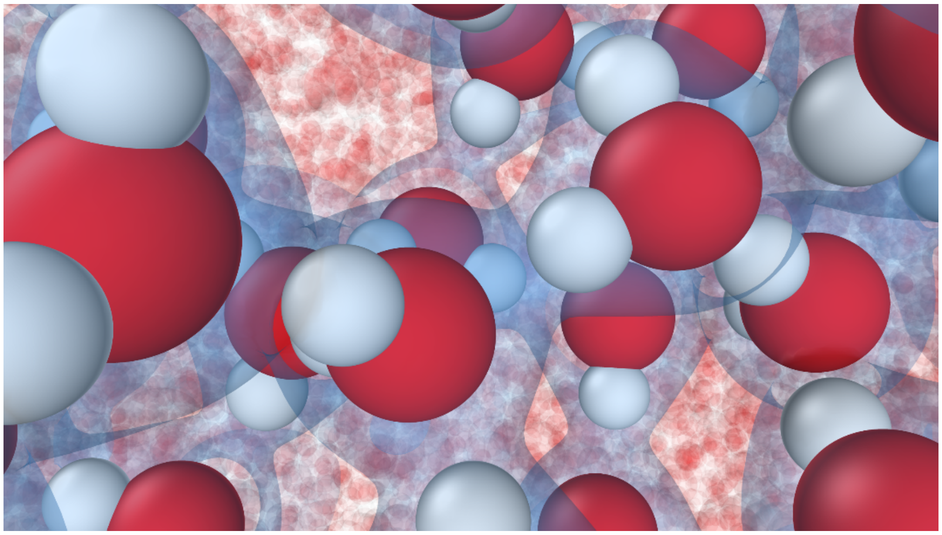

Figure 8-2. Using quantum mechanics to predict the physical forces felt by atoms and molecules in molecular dynamics simulations was originally posited in 1985 by physicists Roberto Car and Michele Parrinello, who named their approach ab initio (Latin for "from the beginning"). Now, with the increase in computing power and the advent of AI, the promise of ab initio molecular dynamics is being fulfilled. Image credit: Pablo Piaggi, Princeton University.

Publication: P. M. Piaggi, J. Weis, A. Z. Panagiotopoulos, P. G. Debenedetti, and R. Car, "Homogeneous Ice Nucleation in an Ab Initio Machine-Learning Model of Water," *Proceedings of the National Academy of Sciences* 119, no. 33 (August 2022)

Related Link: "Distilling How Water Turns into Ice," OLCF News (November 9, 2022)

8.1.3.3 Supercomputing for Swift Protein Modeling

Arizona State University team models molecular structures on ORNL's Summit.

PI: Abhishek Singharoy, Arizona State University
Allocation Program: INCITE

The Science

Uncovering the molecular structures of proteins can help scientists elucidate biological functions of proteins and search for drug candidates for diseases. To better understand proteins, scientists can use imaging techniques to probe their molecular structures. However, many imaging techniques fall short of capturing details at the atomic level. That's why a team led by Abhishek Singharoy at Arizona State University used the OLCF's IBM AC922 Summit supercomputer to model molecular structures in fine detail. The team recently simulated millions of structures and matched the results with experimental data to gain new insights into how proteins transition to different shapes, or conformations.

The Impact

The team's results can point scientists to rare conformations of molecular structures. The team discovered rare conformations of Flpp3, a protein involved in rabbit fever, which is a bacterial disease that can be fatal. They found that in the most probable structure for the protein, it remains closed. However, in rarer conformations, the protein has a binding pocket that opens. This could give scientists a better understanding of how to thwart the disease process. Simulations such as these can also help scientists realize the flexibility of certain structures and capture proteins' unfolding and symmetry breaking as they transition to different states.

Summary

A team led by Abhishek Singharoy at Arizona State University used the OLCF's IBM AC922 Summit supercomputer to model molecular structures of proteins in fine detail. Uncovering the molecular structures of proteins can help scientists elucidate biological functions of proteins and search for drug candidates for diseases. The team recently simulated millions of structures and matched the results with experimental data to gain new insights into how proteins transition to different conformations. The team's results can point scientists to rare conformations of molecular structures. In future work, the team will investigate DNA and RNA complexes. Additionally, they are creating complementary tools, such as incorporating the AlphaFold protein structure database with CryoFold.

Funding

This work was supported by funding from the SMS and Biodesign Center for Applied Structure Discovery at Arizona State University, CAREER award by NSF-MCB.

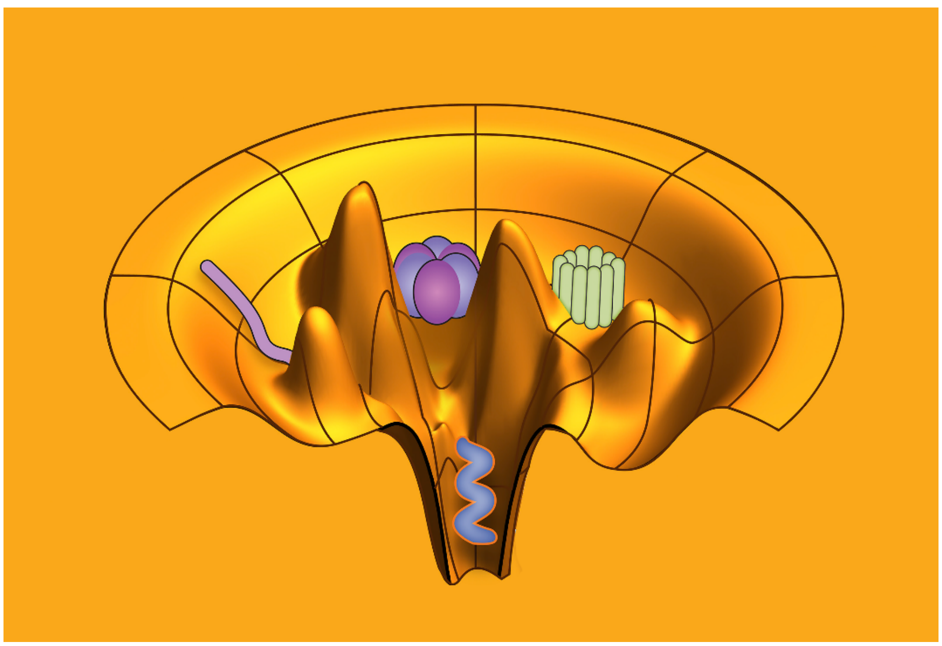

Figure 8-3. An ensemble refinement of molecular motors with CryoFold. Image Credit: Abhishek Singharoy, Arizona State University

Publication: Mrinal Shekhar et al., "CryoFold: Determining Protein Structures and Data-Guided Ensembles from Cryo-EM Density Maps," *Matter* 4, no. 10 (2021): 3195–216, https://doi.org/10.1016/j.matt.2021.09.004.

Related Link: "Supercomputing for Swift Protein Modeling" OLCF News (January 15, 2022)

8.1.3.4 Super Simulations for Superconducting

Summit study spins up new insights into correlated electron systems.

PI: Thomas Maier, ORNL
Allocation Program: INCITE

The Science

A study led by researchers at DOE's ORNL used the Summit supercomputer to close in on the answer to a central question of modern physics that could help conduct development of the next generation of energy technologies. The study used Summit to simulate interactions in a system of electrons within a solid. The simulations applied the Hubbard model, which is the most straightforward model of a system of interacting electrons in various dimensions, to explore how a class of copper alloys known as cuprates act as superconductors that transmit electricity with no loss of energy.

The Impact

Cuprates can be used in power transmission and generation, high-speed magnetic levitation (maglev) trains, and medical applications but generally display their full superconducting properties under extreme cold—typically hundreds of degrees below freezing. Explaining this superconductivity could crack the code to delivering superconductivity at room temperature and provide cheap, speedy, and sustainable energy. Maier's team received an allocation of 900,000 node hours on Summit to explore the model in depth. The results revealed new insights into the relationships between electron spin and charge stripes, including when stripes form as superconductivity develops. The findings point to targets for further study as researchers zero in on how superconducting occurs.

Summary

Using Summit, researchers at ORNL simulated interactions among a system of electrons within a solid to explore how cuprates act as superconductors that transmit electricity with no loss of energy. Cuprates can be used in power transmission and generation, high-speed maglev trains, and medical applications but generally display their full superconducting properties under extreme cold—typically hundreds of degrees below freezing. Explaining this superconductivity could crack the code to delivering superconductivity at room temperature and provide cheap, speedy, and sustainable energy. The results revealed new insights into the relationships between electron spin and charge stripes, including when stripes form as superconductivity develops. The researchers needed a machine as powerful as Summit to support computations for a system large enough to see the stripes.

Funding

This work was supported by the SciDAC program funded by the DOE Office of Science, ASCR and Basic Energy Sciences, Division of Materials Sciences and Engineering.

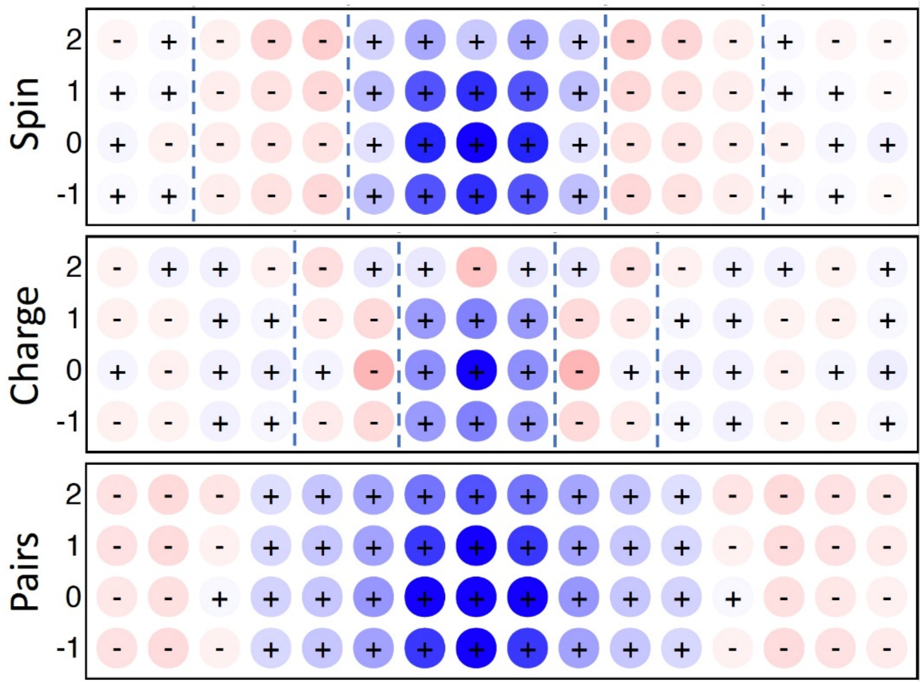

Figure 8-4. An international team of researchers used Summit to model spin, charge, and pair-density waves in cuprates to explore the material's superconducting properties. The results revealed new insights into the relationships between these dynamics as superconductivity develops. Image Credit: Thomas Maier, ORNL

Publication: Peizhi Mai, et al., "Intertwined Spin, Charge and Pair Correlations in the Two-Dimensional Hubbard Model in the Thermodynamic Limit," *Proceedings of the National Academy of Sciences* 119, no. 7 (2022): e2112806119, https://doi.org/10.1073/pnas.2112806119.

Related Link: "Super Simulations for Superconducting," OLCF News (February 18, 2022)

8.1.3.5 Where Worlds Collide

Team simulates collider physics on quantum computer.

PI: Benjamin Nachman, LBNL
Allocation Program: QCUP

The Science

Scientists aim to uncover insights about the smallest building blocks of nature by observing high-energy particle collisions in laboratory environments, such as at the Large Hadron Collider in Geneva, Switzerland. A Berkeley Lab team is exploring what happens in these collisions by using calculations to compare predictions with the actual collision debris. The team has leveraged an IBM Q quantum computer through the OLCF's QCUP to capture part of a calculation of two protons colliding. The calculation can show the probability that an outgoing particle will emit additional particles. The team used effective field theory to break down their full theory into components. Ultimately, they developed a quantum algorithm to allow the computation of some of these components on a quantum computer while leaving other computations for classical computers.

The Impact

Using a quantum computer alone to solve these kinds of calculations requires a number of qubits that is well beyond the quantum compute resources available today. The team aims to describe particles' properties theoretically and then implement a version of them on a quantum computer. Next, the team hopes to add more dimensions to their problem, break their space up into a smaller number of points and scale up the size of their problem. Eventually, they hope to make calculations on a quantum computer that are not possible with classical computers.

Summary

LBNL physicists Christian Bauer, Marat Freytsis, and Benjamin Nachman leveraged an IBM Q quantum computer through the OLCF's QCUP to capture part of a calculation of two protons colliding. The calculation can show the probability that an outgoing particle will emit additional particles. In a paper published in *Physical Review Letters*, the researchers describe how they used a method called effective field theory to break down their full theory into components. Ultimately, they developed a quantum algorithm to allow the computation of some of these components on a quantum computer while leaving other computations for classical computers. In the future, the team hopes to add more dimensions to their problem, and eventually make calculations on a quantum computer that are not possible with classical computers.

Funding

The team performed the IBM Q calculations with funding from the DOE Office of Science Office of High Energy Physics Program as part of the Quantum Information Science Enabled Discovery program (QuantISED).

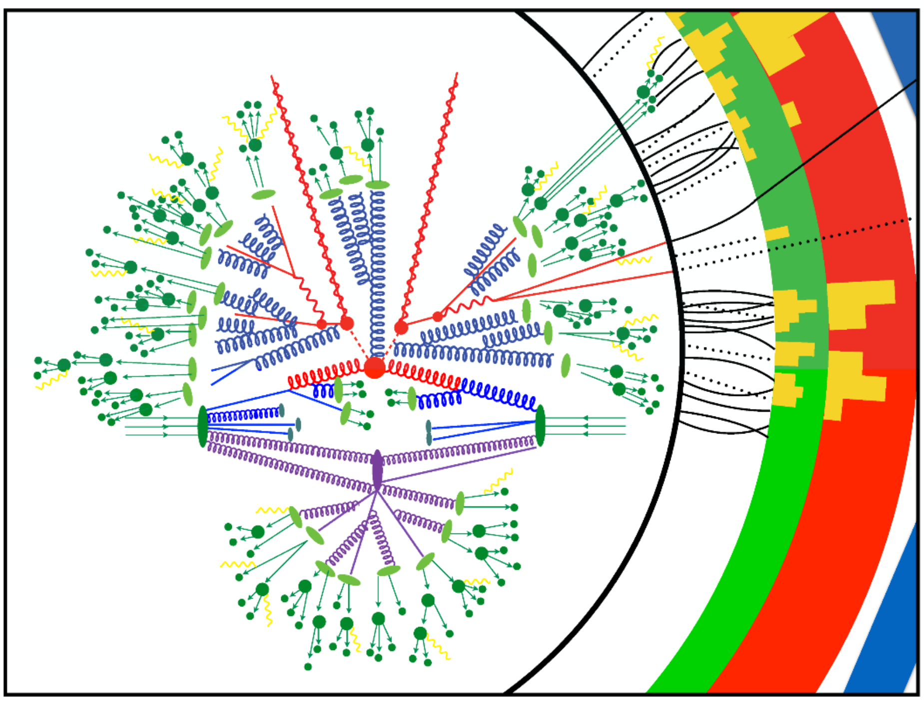

Figure 8-5. Exploring the smallest distance scales with particle colliders often requires detailed calculations of the spectra of outgoing particles (smallest filled green circles) based on the dynamics of the underlying quarks (straight lines) and gluons (curly lines). Image Credit: Benjamin Nachman, Berkeley Lab

Publication: Christian W. Bauer, Benjamin Nachman, and Marat Freytsis, "Simulating Collider Physics on Quantum Computers Using Effective Field Theories," *Physical Review Letters* 127, no. 21 (2021), https://doi.org/10.1103/PhysRevLett.127.212001.

Related Link: "Where Worlds Collide," OLCF News (April 12, 2022)

8.2 RESEARCH ACTIVITIES AND VENDOR ENGAGEMENT FOR FUTURE OPERATIONS

8.2.1 American Association for Physicists in Medicine

A team led by Xiao Wang in the Computational Science and Engineering Division competed in a Grand Challenge Competition hosted by the American Association for Physicists in Medicine. The team included Science Engagement Section researchers Tom Beck, Balint Joo, Isaac Lyngaas, Murali Gopalakrishnan Meena, and Matthew Norman as well as ORNL researchers Anuj Kapadia, Singanallur Venkatakrishnan, and Amir Ziabari, with Charles Bouman and Gregery Buzzard from Purdue University. Their team was one of 55 from around the world that competed to reconstruct the highest resolution medical images with the most accurate clinical diagnosis for a variety of pathological diseases by relying on simulated raw CT scanner data supplied by the Duke University Center for Virtual Imaging Trials. The ORNL/Purdue team won with a Bayesian estimation image reconstruction method called Model Based Iterative Reconstruction, which uses statistical and physics modeling to quantify uncertainties in the medical image acquisition and thus create a more advanced reconstruction. Their work helps bring scientists closer to the long-sought balance between acquiring high-quality detailed medical imagery and keeping radiation doses as low as possible. More detailed imagery enables an earlier diagnosis and more effective treatment, while lower radiation doses minimize harm to patients.
https://www.olcf.ornl.gov/2022/08/15/ornlpurdue-team-wins-ct-imaging-competition/

8.2.2 R&D 100 Award

OLCF scientists Austin Harris, Bronson Messer, and Tom Papatheodore received an R&D 100 Award for development of the Flash-X multiphysics simulation software as part of the ExaStar project within the DOE's ECP. Flash-X, a flexible and accessible open-source software instrument, uses a combination of partial and ordinary differential and algebraic equations to simulate various physical phenomena, including astrophysics, CFD, and cosmology. The codes within ExaStar are being used to model stellar explosions and the evolution of the heavy elements. A recent study used Flash-X to model a Type Ia supernova on the OLCF's Summit. Flash-X incorporates a performance portability layer that is language agnostic and thus compatible with a variety of computer systems. The open-source software features components that offer an easily customizable plug-and-play mode for most scientific applications. Flash-X divides the configuration of specific applications into smaller portions, so each individual configuration tool remains relatively simple. A previous version of the software, FLASH, enabled a variety of scientific discoveries over the past decade. That version is no longer fully compatible with state-of-the-art systems, particularly hybrid systems such as Frontier and Aurora that employ a combination of CPUs and GPUs.
https://www.olcf.ornl.gov/2022/09/08/olcf-research-team-wins-rd-100-award/

8.2.3 Fortran Committee of the International Committee for Information Technology Standards

Reuben Budiardja in the Science Engagement Section was named group chair of the Fortran Committee of the International Committee for Information Technology Standards (INCITS). The INCITS/Fortran Technical Committee is responsible for the development and interpretation of US and international standards for the Fortran programming language. Budiardja holds a PhD in computational astrophysics from the University of Tennessee, Knoxville. His research interests range from development of large-scale astrophysics simulation code, to investigating the mechanism of core-collapse supernovae, to

developing system tools to understand application motifs, usage, and performance on supercomputers. In his role with INCITS, Reuben helps set and maintain consensus-driven, market-relevant standards for technology on a worldwide scale to enable continued innovation across a variety of industries and institutions.
https://www.incits.org/committees/pl22.3

8.2.4 Electronic Imaging Symposium Best Paper

A Science Engagement-led team won a Best Paper Award at the 2023 Electronic Imaging Symposium in the High Performance Computing for Imaging. The team consisted of ORNL's Muralikrishnan Gopalakrishnan Meena, Amir K. Ziabari, Singanallur Venkatakrishnan, Isaac R. Lyngaas, Matthew R. Norman, Balint Joo, Tom Beck, Anuj Kapadia, and Xiao Wang along with Purdue University's Charles A. Bouman. The team introduced a deep learning image-based material decomposition method guided by physics and that requires no access to projection data and will advance precision medicine techniques by using OLCF resources. This approach uses a hybrid unsupervised and supervised learning technique to tackle material decomposition and leverages the unique x-ray absorption rate of calcium compared with body tissues. The method has been successfully demonstrated on simulated breast models to decompose contributions into calcification, adipose, fibroglandular, and air.
https://www.imaging.org/IST/Conferences/EI/EI2023/EI2023.aspx

8.2.5 Finalists for the ACM Gordon Bell Prize for High Performance Computing: Exaflops Biomedical Knowledge Graph Analytics

An ORNL-led team used Frontier to scan hundreds of thousands of biomedical concepts from millions of scientific publications in search of potential connections among symptoms, diseases, conditions, and treatments. The study earned the team a finalist nomination for the Association of Computing Machinery Gordon Bell Prize, which is awarded annually at SC for outstanding achievements in applying HPC to challenges in science, engineering, and large-scale data analytics. The team's project seeks to push the fast-forward button on drug discovery to streamline exploration for promising leads. The inspiration came in early 2020 during the COVID-19 pandemic, when scientists around the world turned their attention to searching for potential treatments. Ramki Kannan and Tom Potok of ORNL's Computer Science and Mathematics Division led a team of researchers from ORNL, including Hao Lu of the NCCS, and collaborators from AMD, the Georgia Institute of Technology and the University of California, San Francisco (UCSF). They developed the Distributed Accelerating Semiring All-Pairs Shortest Path algorithm (DSNAPSHOT) as a method to use AI to pinpoint potential links among millions of medical concepts across decades of scientific publications. The team targeted a dataset of COVID-19 and associated coronaviruses drawn from more than 800,000 papers. They used 9,200 of Frontier's more than 9,400 nodes to perform an initial search across a graph drawn from PubMed and the Scalable Precision Medicine Open Knowledge Engine (SPOKE), which is a comprehensive index of medical databases maintained by UCSF. The run reached a speed of 1 exaflops at single precision and took only 11.7 minutes to search more than 7 million data points drawn from 18 million publications. The team ultimately hopes to scale up the algorithm to scan the full depth of SPOKE and PubMed combined and to make the search as easily customizable as a Google query.
https://www.olcf.ornl.gov/2022/10/25/fast-tracking-medical-discovery/

8.2.6 Finalists for the ACM Gordon Bell Special Prize for High Performance Computing-Based COVID-19 Research

An ORNL team including NCCS researchers Feiyi Wang, Darren Hsu, Wayne Joubert, Aditya Kashi, Michael Matheson, Jens Glaser, and Hao Lu, and John Gounley of the Computer Sciences and Engineering Division, designed an ML software stack, TwoFold, that predicts how strongly a given drug

molecule will bind to a pathogen as well as the 3D structure of how the molecule will attach to the target. Their paper was chosen as a finalist for the ACM Gordon Bell Special Prize for High Performance Computing-Based COVID-19 Research and was published in the *International Journal of High-Performance Computing Applications*. TwoFold delivers both predictions at the same time, which is a drastic improvement compared with current methods. Predicting these properties could streamline and speed up the lengthy trial-and-error process of lab experimentation for viable treatments for diseases such as COVID-19. The team developed TwoFold by first using the OLCF's Summit to train a deep learning neural network on experimental data for drug-target interactions and receptor-bound structures and then using Frontier to solve a matrix of millions of predictive equations produced by that network. This forecasting power shortens time to solution and eliminates the need to painstakingly map the complicated protein structure of a virus or similar pathogen in the lab before launching the search for potential treatments.
https://www.olcf.ornl.gov/2022/11/11/predicting-protein-targets-and-the-molecules-that-bind-with-them

8.2.7 Summit on Summit

The Summit on Summit is an active collaboration of NVIDIA and computing facilities with NVIDA-accelerated computers, including ORNL, LLNL, and NERSC at LBNL. The engagement takes the form of technical work and discussions called *WorkStreams*, which are led by an NVIDIA representative and cover a wide range of topics important to the facilities. Through monthly review of progress made in each WorkStream, this organization has proven to be very effective.

Topic areas include compilers, math libraries, data center monitoring, development tools, portable memory, communication, data analytics, reduced precision, coding for performance, python, Monte Carlo algorithms, structured and unstructured mesh, particle methods, ADIOS, GPU-direct storage, HPC and AI, HPC at the Edge, ARM on NVIDIA GPUs, and multi-tenancy. During 2022, virtual meetings were held every 4 months in lieu of the in-person workshops hosted by NVIDIA in Santa Clara before the pandemic.

In the last meeting of institution leads with NVIDIA management, a decision was made to review the current WorkStream at an upcoming in-person workshop and select a new set of WorkStreams for this successful collaboration, which ultimately contributes to the success of the facilities, the facilities' vendors, and their users.

8.2.8 Accelerated Data Analytics and Computing (ADAC) Institute

In 2016, the Accelerated Data Analytics and Computing (ADAC) Institute was created as a consortium of HPC computer centers. As founding members, ORNL, Tokyo Tech, and ETH Zürich collaborate and leverage their respective investments in application software readiness to expand the breadth of applications capable of running on accelerated architectures. Since then, LLNL, Argonne, the National Computing Infrastructure at the Australian National University in Canberra, Australia, the University of Tokyo, the National Institute of Advanced Industrial Science and Technology in Tokyo, Japan, the RIKEN Center for Computational Science in Kobe, Japan, the Jülich Forschungs Zentrum in Jülich, Germany, and the IT Center for Science in Espoo, Finland, have joined the institute. All organizations manage HPC centers that run large, GPU-accelerated supercomputers and provide key HPC capabilities to academia, government, and industry to solve many of the world's most complex and pressing scientific problems.

In 2022, the activities were organized in three topic areas: applications, resource management, and performance portability. ADAC meets twice per year for an in-person workshop to discuss progress on joint activities and planning for the next 6-month period. The biannual meetings in 2022 were held virtually because of the COVID-19 pandemic. The ADAC Governance Board met during SC22 and decided to commence the in-person workshops starting in 2023.

8.2.9 Industry Engagement

ACCEL, OLCF's industrial partnership program, remained on track in 2022 as industry and the nation continued to emerge from the COVID-19 pandemic and develop their new normal work routines. Thirty-seven industry projects were underway on Summit during the year: 22 new projects launched and 15 that began prior to 2022 continued throughout the year.

- The 37 industrial projects underway in 2022 represented approximately 9% of the total projects provided to external user programs, INCITE, ALCC, and DD (including ECP).
- These projects used 5,489,538 Summit node hours, which represents approximately 16% of the total Summit hours.
- The total industrial project hours on Summit were allocated as follows: 65% INCITE, 30% ALCC, and 5% OLCF DD.
- The 22 new projects received their awards through INCITE (6 projects), ALCC (3 projects), and DD (13 projects).

8.2.9.1 Observations about the Industrial Projects

INCITE

This was another strong year for industry INCITE awards. Industry had six new awards along with three INCITE projects that began in 2021 and continued into 2022, for a total of nine INCITE projects. Industry and universities partnered in four of the new industry INCITE awards to ensure broader community engagement.

ALCC

Industry fared well in the ALCC program with three new awards. GE Aviation received one of these awards, which included the first industrial award on Frontier in addition to time on Summit. GE will begin work on Frontier when it is open for broad community access. In the interim, GE has time on Summit and Crusher.

OLCF DD

We had three new industrial users and one new industry that received DD awards. ZAP Energy, a commercial fusion energy small business supported by venture capital (VC) funding, including from Bill Gates's Breakthrough Energy Ventures, Chevron Technology Ventures, and Shell Ventures, received a DD award.

Whisper Aero, another VC-backed small business, and Joby Aviation in partnership with software firm Cascade Technologies also received a DD award. Whisper Aero and Joby Aviation are leads in the emerging eVTOL air taxi industry. The Joby/Cascade project team was awarded time on Crusher to prepare for access to Frontier when it opens for broad community access.

We continued to support the HPC4MFG project by providing a DD award to Raytheon in partnership with ORNL.

8.2.9.2 Industrial Project Highlight

Summit Fires up Predictive Breakthrough for Gas Turbines
Simulations Could Help Head Off Billion-Dollar Problem

PI: Yonduck Sung, Solar Turbines
Allocation Program: DD

The Science

Simulations performed on ORNL's Summit supercomputer promise to help solve a billion-dollar problem for the gas turbine industry by enabling engineers to understand and predict thermoacoustic oscillations early in the design process. The oscillations, caused by heat and noise, can damage parts, shut down operations, and shake the machinery apart under severe conditions.

This computational approach has been tried before, but previous efforts stalled due to the extraordinary computing power required to model the complicated flow of gases during combustion.

The Impact

Large turbines can take years to design and build, and the flaws that lead to potentially damaging oscillations might not show up until all the components are made and assembled at the final stage, thereby forcing costly do-overs. The simulations on Summit that use charLES, a high-fidelity CFD code developed by Cascade Technologies, demonstrated for the first time how engineers can head off the oscillations and promise to change the industry's design process.

Solar Turbines increased in-house computing resources by 230% with investments in hybrid computer architectures similar to Summit's and estimates the changes could eliminate as much as 3,800 metric tons of carbon dioxide emissions per turbine per year for a single product line.

Summary

Solar Turbines, a subsidiary of Caterpillar, manufactures a widely used family of mid-sized industrial gas turbines that serve a variety of industries, from natural gas and crude oil production, processing, and pipeline transmission, to generating heat and electricity. These industrial gas turbines rely on a delicately balanced mix of fuel and air to produce maximum energy from the minimum amount of fuel. Lean mixtures like those used in Solar Turbines products lower the fuel-to-air ratio even further to reduce toxic gases such as nitrogen oxides that contribute to air pollution. Combustion of these lean mixtures comes with a billion-dollar-a-year problem. Fluctuations in fuel injection and air flows along with other variables such as injection location, direction, and velocity can lead to noisy and potentially damaging thermoacoustic oscillations.

"At the extreme, these instabilities can shake a system apart, and they often don't appear until late in the design cycle when all the components are built and assembled, and a physical test of the full engine is performed," said Yonduck Sung, a senior principal engineer for Solar Turbines who led the study. "At that point, we may have to start all over."

Researchers from Solar Turbines and software partner Cascade Technologies teamed up to use large-scale simulation to better understand and reliably predict thermoacoustic oscillations early in the design phase without burning time and money on a cycle of repeated physical tests and redesigns. They applied for and received a DD allocation of computing time on Summit via the OLCF industrial partnership program ACCEL. Solar Turbines engineers could model only one burner unit using the company's in-house computing system. The power of Summit enabled simulation of a full combustion chamber, complete with a full annular combustion section and 14 operating burners. The team used charLES to perform large-eddy simulations (LESs), which capture turbulence, and study the interplay between the combustion and engine pressure.

"[LESs]would have taken months to run on a conventional computer," said Sanjeeb Bose, Cascade's chief technical officer. "Summit's world-class speeds shrank that time to a matter of days." Lee Shunn, a senior research scientist for Cascade who helped write the code added, "By using Summit and charLES, we were able to predict combustion-induced oscillation in our test cases for the first time."

The results are changing Solar Turbines' approach to designing turbine systems, Sung said. "Summit's fast turnaround showed us that we can produce results in an actionable time scale, thereby shortening the design-cycle time for our products and reducing costs associated with physical testing and redesign." The success with Summit also led the company to increase its in-house HPC resources by 230% with investments in hybrid computer architectures similar to Summit's.

"Thanks to Summit, we now know how to perform computations that once seemed impractical or impossible," said John Mason, director of gas turbine technology and new product development.

Funding

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

Figure 8-6. Gas turbines power industries around the world, but vibrations from heat and noise can shake the systems apart under the wrong conditions. Simulations performed on ORNL's Summit supercomputer could help head off that problem. Credit: Solar Turbines

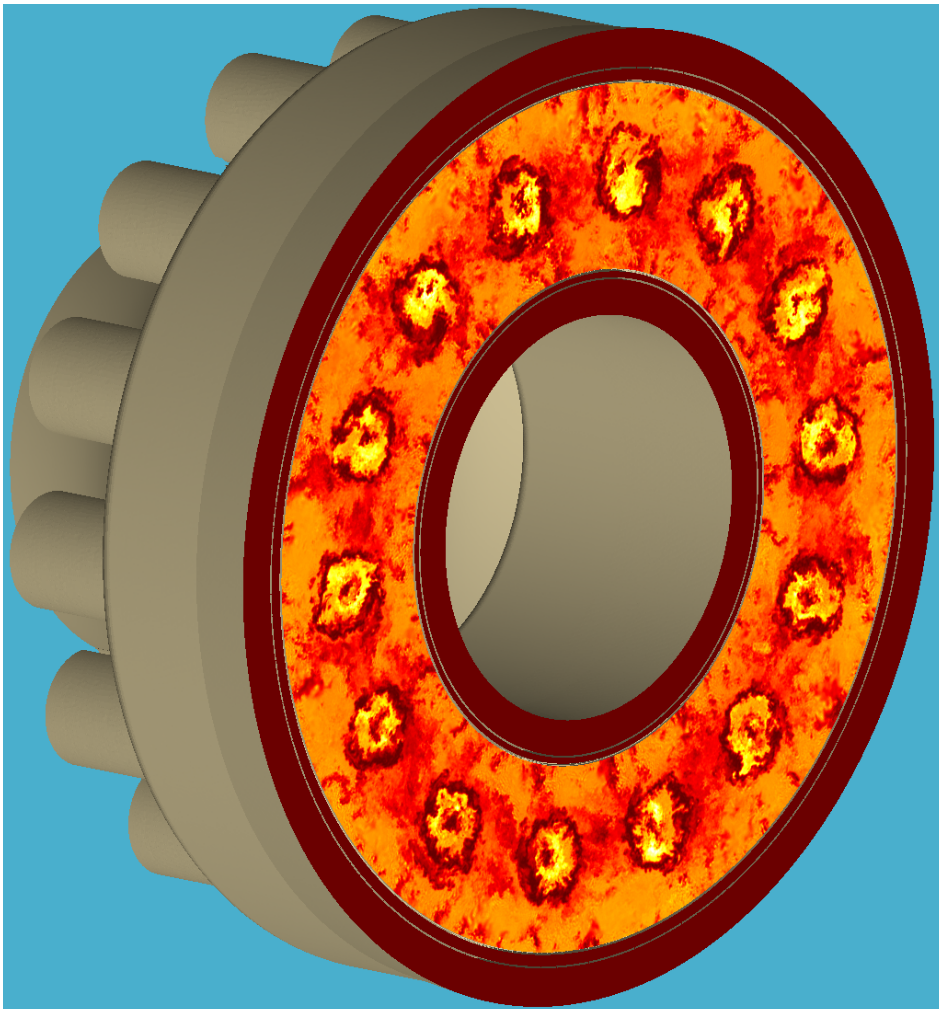

Figure 8-7. The computational power of ORNL's Summit supercomputer enabled Solar Turbines engineers to model a 14-burner gas combustion chamber. Credit: Solar Turbines.

Publication: In development

Related Link: "Summit Fires up Predictive Breakthrough for Gas Turbines," OLCF News (March 17, 2022).

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 increase 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 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 OLCF resources. Using institutional resources, the Office of Science's Biological and Environmental Research-sponsored investigators in the Atmospheric Radiation Monitoring 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. Basic Energy Sciences-sponsored scientists who use facilities such as ORNL's CNMS and the Spallation Neutron Source have successfully won discretionary time on Summit to perform AI-driven electron microscopy and design-of-experiment campaigns (e.g., porting their codes to GPUs). The movement of data from the facilities is supported by traditional data transfers and new tools to simplify the user experience (e.g., the DataFlow tool discussed in Section 4.1.4.3). The lessons learned from workflows for high-energy physics programs (e.g., ATLAS-supporting BigPanDA workflows), NP 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.

8.3.2 Ongoing Engagements with the Atmospheric Radiation Measurement Research Facility

In 2022, NCCS deployed the Cumulus-2 cluster, a 16,348 core Dell HPC system roughly 4× more powerful than the previous Cumulus-1 cluster. Cumulus-2 and its 4 PB parallel file system joined the Atmospheric Radiation Measurement Research Facility's Stratus cluster; together, these systems process PBs of atmospheric data collected from hundreds of instruments and observation facilities around the world. The OLCF offers data management support and expertise in archival storage, thereby ensuring access to data required for critical climate research and providing safe long-term storage.

Hosting these HPC resources and data at the OLCF offers several benefits to both the OLCF and the Atmospheric Radiation Measurement Research Facility. One of the primary projects actively using Stratus and Cumulus is the LES ARM Symbiotic Simulation and Observability (LASSO) project. The LASSO project has developed *data bundles* that combine observations and high-resolution model output, which are then made accessible to researchers through a searchable web interface. Supporting projects like this showcases the unique data and compute service capabilities that the OLCF provides to a global research community, using underlying infrastructure hosted in NCCS.

8.3.3 Engagement with the National Institutes of Health (NIH) and the National Cancer Institute (NCI)

The collaboration between ORNL and National Institutes of Health (NIH)/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 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 our models in real-world cancer registry settings; and (5) enabling large-scale transformer training on LCF 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 requires access to large HPC resources because these models must be trained on hundreds of Summit nodes using data parallelism. Additionally, Transformer models must be trained on the raw text of cancer pathology reports, which includes PHI.

To better support these efforts, the OLCF launched a security framework called CITADEL that allows researchers to use supercomputers for research using protected health data for the first time. 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, which are 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, thereby 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 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 2022, the MOSSAIC team began deploying the Case-Level Multi-Task Hierarchical Self Attention Network, which is an NLP algorithm that autocodes cancer pathology reports at participating SEER registries 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.

8.3.4 Engagement with the US Department of Veterans Affairs (VA)

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, extensive expertise in big data, and HPC to advance veterans' treatment and the VA's medical and genomic research. The Million Veteran Program is using Summit to identify genetic markers for early onset prostate cancer (before the age of 55). Supported by the CITADEL security framework, which protects sensitive personal health data so it can be studied directly on the supercomputer, the project is using computationally efficient transformer models on Summit to capture interactions within and between genes, a capability not viable without supercomputers.

These partnerships have resulted in unprecedented developments in security that enable the NIH and VA to leverage HPC for research that uses sensitive data and have strengthened the HPC ecosystem at ORNL.

8.3.5 Engagement with Air Force Weather (AFW) and NOAA

ORNL and Air Force Weather (AFW) launched Miller and Fawbush, two HPE Cray EX supercomputers to support the AFW's numerical weather modeling at much higher speed and fidelity. Thanks to these powerful systems, which reached Certification of Operational Readiness in February 2021, AFW is now 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. The early receipt of these systems, in advance of OLCF-5, provided our staff with critical decision points related to the full software stack, the tools for managing these systems, and invaluable experience for delivering production-capable systems based on the new EX platform. These lessons significantly benefitted the OLCF-5 Frontier project, supporting key decisions early in that project, in advance of the delivery of the initial test and development systems.

ORNL and NOAA established the National Climate-Computing Research Center (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. In 2021, based on success achieved over the program's lifespan, it was extended another five years. The NCCS hosts the program's primary computational resource, Gaea, which consists of a pair of HPE XC40 HPC systems providing 5.29 petaflops of computational power. In 2022, work began to procure and install a new climate modeling system with a peak performance of 10 petaflops. The NCRC supports and complements NOAA's climate mission to understand climate variability and change to 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, nurturing a highly trained computational workforce, and engaging the global user community.

The partnership allows the opportunity for collaborative R&D and 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 2022 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 during each effort benefit both NOAA and the OLCF, and the broadened perspective helps to harden and improve the end result for all programs in the NCCS.

8.3.6 ECP Engagement Summary

A primary and natural place for community engagement has been DOE's ECP, the goal of which is to develop software and applications and influence the development of hardware technology—all to facilitate the successful deployment and operation of capable exascale systems. The ECP continues to fund efforts in national labs, academia, and industry; these timely investments will significantly aid the OLCF in delivering capable exascale systems with robust system and application software that can address the science gaps immediately upon delivery and acceptance of the systems. The OLCF's engagement with ECP includes three primary thrust areas, described below, as well as many pairwise and other interactions (e.g., staff involvement in ECP AD and ST projects).

8.3.6.1 OLCF-ECP AD and ST Engagements

The OLCF has a long history of readying applications for its forthcoming architectures dating back to before the delivery of the OLCF-3 system (Titan), then with OLCF-4 (Summit), and now with OLCF-5 (Frontier). CAAR has served as a successful collaboration point for application teams, vendors, and tool developers to exploit hierarchical parallelism within applications in preparation for next-generation architectures. The OLCF has chosen to partner with the ECP to augment the OLCF CAAR portfolio with an additional 12 ECP AD teams. These teams were selected to diversify OLCF application readiness efforts funded through the OLCF-5 CAAR such that the OLCF will have a broad suite of applications ready to use on Frontier. These projects were also selected with an eye toward matching the architectural strengths of Frontier with the appropriate computational motifs and methods employed by these ECP applications. The teams that are currently partnering with the OLCF and their OLCF Scientific Computing liaison are listed in Table 8-3.

Table 8-3. Listing of OLCF ECP engagement applications, the ECP AD PI, and the OLCF Scientific Engagement liaison.

WBS / ECP AD Project	Project PI	OLCF Liaison
2.2.1.01 LatticeQCD	Andreas Kronfeld (Fermilab)	Balint Joo
2.2.1.02 NWChemEx	Theresa Windus (Ames Laboratory)	Dmitry Liakh
2.2.1.03 GAMESS	Mark Gordon (Iowa State University)	Dmytro Bykov
2.2.1.05 ExaAM	John Turner (ORNL)	Stephen Nichols
2.2.2.02 Combustion-PELE	Jackie Chen (Sandia)	Kalyana Gottiparthi
2.2.2.03 ExaSMR	Steven Hamilton (ORNL)	Mark Berrill
2.2.2.05 WDMApp	Amitava Bhattacharjee (PPPL)	Antigoni Georgiadou
2.2.3.01 ExaStar	Dan Kasen (LBNL)	Austin Harris
2.2.3.02 ExaSky	Salman Habib (Argonne)	Antigoni Georgiadou
2.2.3.05 E3SM-MMF	Mark Taylor (Sandia)	Matt Norman
2.2.4.02 ExaSGD	Slaven Peles (PNNL)	Philip Roth
2.2.4.04 ExaBiome	Kathy Yelick (LBNL)	Philip Roth

PPPL = Princeton Plasma Physics Laboratory

With ECP funding, these ECP AD teams have dedicated staff expertise from the OLCF Scientific Engagement section, access to the system vendor's Center of Excellence (COE) from both Cray and AMD, access to early test bed hardware provided by the Frontier vendor, and potential support from postdoctoral researchers through the CSEEN program as availability allows. In addition, because of the dependence of the AD project on the ECP ST, the ECP ST projects also have access to the COE resources and access to early test bed hardware.

This mutually beneficial partnership enables ECP to learn from the application readiness lessons learned and best practices developed during two prior instantiations of the CAAR program. Additionally, the expected OLCF application and software portfolio ready for Frontier has been further diversified by including these additional projects in the CAAR program.

The OLCF continued to contribute to the ECP project in the Hardware Integration area by leading or participating in efforts in application support, software integration, facility resource utilization, and training and productivity.

APPENDIX A. 2022

There were no recommendations received from the 2021 Operational Assessment Review committee.

APPENDIX B. 2022

Date	Title of science highlights submitted to DOE	Writer	URL
1/5/22	Thinking outside the Black Box	Matt Lakin	https://www.olcf.ornl.gov/2022/01/05/thinking-outside-the-black-box/
1/7/22	Firing Up Quantum Fidelity	Matt Lakin	https://www.olcf.ornl.gov/2022/01/07/firing-up-quantum-fidelity/
1/10/22	Scientists Use Summit Supercomputer, Deep Learning to Predict Protein Functions at Genome Scale	Kim Askey	https://www.olcf.ornl.gov/2022/01/10/scientists-use-summit-supercomputer-deep-learning-to-predict-protein-functions-at-genome-scale/
1/25/22	Supercomputing Exposes Potential Pathways for Inhibiting COVID-19	Elizabeth Rosenthal	https://www.olcf.ornl.gov/2022/01/25/supercomputing-exposes-potential-pathways-for-inhibiting-covid-19/
1/25/22	Artificially Altered Material Could Accelerate Neuromorphic Device Development	Elizabeth Rosenthal	https://www.olcf.ornl.gov/2022/01/25/artificially-altered-material-could-accelerate-neuromorphic-device-development/
1/25/22	Supercomputing for Swift Protein Modeling	Rachel McDowell	https://www.olcf.ornl.gov/2022/01/25/supercomputing-for-swift-protein-modeling/
1/25/22	Team Taps ORNL's DataFed to Gain Materials Insights	Rachel McDowell	https://www.olcf.ornl.gov/2022/01/25/team-taps-ornls-datafed-to-gain-materials-insights/
1/25/22	Pioneering Frontier: Standing Up Exascale	Rachel McDowell	https://www.olcf.ornl.gov/2022/01/25/pioneering-frontier-standing-up-exascale/
1/31/22	Philip Roth Named General Chair for SC24	Coury Turczyn	https://www.olcf.ornl.gov/2022/01/31/philip-roth-named-general-chair-for-sc24/
2/2/22	Decoding the Role of CSB Protein in DNA Repair	Coury Turczyn	https://www.olcf.ornl.gov/2022/02/02/decoding-the-role-of-csb-protein-in-dna-repair/
2/16/22	Demystifying the World of High-Performance Computing	Betsy Sonewald	https://www.olcf.ornl.gov/2022/02/16/demystifying-the-world-of-high-performance-computing/
2/24/22	COVID-19 in the Classroom: Simulating the Spread	Rachel McDowell	https://www.olcf.ornl.gov/2022/02/24/covid-19-in-the-classroom-simulating-the-spread/
3/17/22	From Summit to the Stars	Matt Lakin	https://www.olcf.ornl.gov/2022/03/17/from-summit-to-the-stars/
3/17/22	Summit Fires up Predictive Breakthrough for Gas Turbines	Matt Lakin	https://www.olcf.ornl.gov/2022/03/17/summit-fires-up-predictive-breakthrough-for-gas-turbines/
3/23/22	International Volunteer Team of Researchers Launch Groundbreaking Workflows Community Platform	Betsy Sonewald	https://www.olcf.ornl.gov/2022/03/23/international-volunteer-team-of-researchers-launch-groundbreaking-workflows-community-platform/
3/28/22	Forging Ahead with Frontier: Ready to Crush Science	Rachel McDowell	https://www.olcf.ornl.gov/2022/03/28/forging-ahead-with-frontier-ready-to-crush-science/
3/31/22	Pioneering Frontier: Stepping Up CANDLE	Rachel McDowell	https://www.olcf.ornl.gov/2022/03/31/pioneering-frontier-stepping-up-candle/
4/6/22	Pioneering Frontier: Optimizing Materials Models	Rachel McDowell	https://www.olcf.ornl.gov/2022/04/06/pioneering-frontier-optimizing-materials-models/

Date	Title of science highlights submitted to DOE	Writer	URL
4/12/22	Where Worlds Collide	Rachel McDowell	https://www.olcf.ornl.gov/2022/04/12/where-worlds-collide/
4/22/22	New Deep Learning Techniques Lead to Materials Imaging Breakthrough	Elizabeth Rosenthal	https://www.olcf.ornl.gov/2022/04/22/new-deep-learning-techniques-lead-to-materials-imaging-breakthrough/
4/26/22	OLCF's Ashley Barker Takes on New Role, Steps Down from ECP	Coury Turczyn	https://www.exascaleproject.org/olcfs-ashley-barker-takes-on-new-role/
4/28/22	US Department of Energy's INCITE program seeks proposals for 2023	Katie Bethea	https://www.olcf.ornl.gov/2022/04/28/us-department-of-energys-incite-program-seeks-proposals-for-2023/
5/2/22	OLCF's Ashley Barker Takes on New Role	Coury Turczyn	https://www.olcf.ornl.gov/2022/05/02/olcfs-ashley-barker-takes-on-new-role/
5/4/22	Register for Lustre User Group Virtual Conference 2022	Betsy Sonewald	https://www.olcf.ornl.gov/2022/05/04/register-for-lustre-user-group-virtual-conference-2022/
5/5/22	Summit, neutrons crack code to uranium compound's signature vibes	Matt Lakin	https://www.olcf.ornl.gov/2022/05/05/summit-neutrons-crack-code-to-uranium-compounds-signature-vibes/
5/12/22	Layered Perovskite Power	Coury Turczyn	https://www.olcf.ornl.gov/2022/05/12/layered-perovskite-power/
5/29/22	Frontier Supercomputer Debuts as World's Fastest, Breaking Exascale Barrier	Katie Bethea	https://www.olcf.ornl.gov/2022/05/29/frontier-supercomputer-debuts-as-worlds-fastest-breaking-exascale-barrier/
5/31/22	Pioneering Frontier: Navigating a New Era	Rachel McDowell	https://www.olcf.ornl.gov/2022/05/31/pioneering-frontier-navigating-a-new-era/
6/8/22	Pushing the new Frontier	Matt Lakin	https://www.olcf.ornl.gov/2022/06/08/pushing-the-new-frontier/
6/13/22	Speeding up simulations	Matt Lakin	https://www.olcf.ornl.gov/2022/06/13/speeding-up-simulations/
6/13/22	Weaving wonders	Matt Lakin	https://www.olcf.ornl.gov/2022/06/13/weaving-wonders/
7/5/22	ECP Advances the Science of Atmospheric Convection Modeling	Coury Turczyn	https://www.olcf.ornl.gov/2022/07/05/ecp-advances-the-science-of-atmospheric-convection-modeling/
7/5/22	ORNL and US Air Force Hold Ribbon-Cutting Ceremony for Weather Forecasting System	Betsy Sonewald	https://www.olcf.ornl.gov/2022/07/05/ornl-and-us-air-force-hold-ribbon-cutting-ceremony-for-weather-forecasting-system/
7/6/22	ECP Advances the Science of Atmospheric Convection Modeling	Coury Turczyn	https://www.exascaleproject.org/ecp-advances-the-science-of-atmospheric-convection-modeling/
7/13/22	Sky's the Limit for Cumulus-2	Coury Turczyn	https://www.olcf.ornl.gov/2022/07/13/skys-the-limit-for-cumulus-2/
7/25/22	Rafael Ferreira da Silva Elevated to IEEE Senior Member	Betsy Sonewald	https://www.olcf.ornl.gov/2022/07/25/rafael-ferreira-da-silva-elevated-to-ieee-senior-member/

Date	Title of science highlights submitted to DOE	Writer	URL
7/25/22	TRITON: A Powerful Toolkit for Modern Flood Modeling	Betsy Sonewald	https://www.olcf.ornl.gov/2022/07/25/new-model-harnesses-supercomputing-power-for-more-accurate-flood-simulations/
7/25/22	Pioneering Frontier: Automating at Exascale	Betsy Sonewald	https://www.olcf.ornl.gov/2022/07/25/pioneering-frontier-automating-at-exascale/
7/29/22	OLCF, ECMWF Accepting Proposals for New Virtual Hackathon	Betsy Sonewald	https://www.olcf.ornl.gov/2022/07/29/olcf-ecmwf-accepting-proposals-for-new-virtual-hackathon/
8/1/22	OLCF Launches Program for Hybrid Computing Allocations	Coury Turczyn	https://www.olcf.ornl.gov/2022/08/01/olcf-launches-program-for-hybrid-computing-allocations/
8/2/22	ECP Launches Programs to Extend DOE's HPC Workforce Culture	Coury Turczyn	https://www.exascaleproject.org/ecp-launches-programs-to-extend-does-hpc-workforce-culture/
8/15/22	ORNL/Purdue Team Wins CT Imaging Competition	Coury Turczyn	https://www.olcf.ornl.gov/2022/08/15/ornlpurdue-team-wins-ct-imaging-competition/
8/17/22	ORNL Celebrates Launch of Frontier – the World's Fastest Supercomputer	Matt Lakin	https://www.olcf.ornl.gov/2022/08/17/ornl-celebrates-launch-of-frontier-the-worlds-fastest-supercomputer/
8/24/22	Summer of High-Performance Computing	Coury Turczyn	https://www.olcf.ornl.gov/2022/08/24/summer-of-high-performance-computing/
9/8/22	OLCF researchers win R&D 100 award	Matt Lakin	https://www.olcf.ornl.gov/2022/09/08/olcf-research-team-wins-rd-100-award/
9/13/22	Lab Experiments Support the COVID-19 Bradykinin Storm Theory	Coury Turczyn	https://www.olcf.ornl.gov/2022/09/13/lab-experiments-support-the-covid-19-bradykinin-storm-theory/
9/19/22	New Storage System Complements Existing Capabilities at the OLCF	Betsy Sonewald	https://www.olcf.ornl.gov/2022/09/19/new-storage-system-complements-existing-capabilities-at-the-olcf/
9/28/22	ORNL Researchers Earn R&D 100 Award for AI Technology	Coury Turczyn	https://www.olcf.ornl.gov/2022/09/28/ornl-researchers-earn-rd-100-award-for-ai-technology/
10/3/22	Spinning up Quantum Fidelity	Matt Lakin	https://www.olcf.ornl.gov/2022/10/03/spinning-up-quantum-fidelity/
10/4/22	ORNL staff join industry, government mentors in student cluster competition	Betsy Sonewald	https://www.olcf.ornl.gov/2022/10/04/ornl-staff-join-industry-government-mentors-in-student-cluster-competition/
10/13/22	2022 OLCF User Group Executive Board Election	Bill Renaud	https://www.olcf.ornl.gov/2022/10/13/2022-olcf-user-group-executive-board-election/
10/10/22	Pioneering Frontier: Bringing the Science	Coury Turczyn	https://www.olcf.ornl.gov/2022/10/10/pioneering-frontier-bringing-the-science/
10/17/22	Exascale Day 2022	Betsy Sonewald	https://www.olcf.ornl.gov/2022/10/17/exascale-day-2022/
10/25/22	Fast-tracking medical discovery	Matt Lakin	https://www.olcf.ornl.gov/2022/10/25/fast-tracking-medical-discovery/
10/27/22	WarpX Named Gordon Bell Prize Finalist	Coury Turczyn	https://www.olcf.ornl.gov/2022/10/27/warpx-named-gordon-bell-prize-finalist/

Date	Title of science highlights submitted to DOE	Writer	URL
10/28/22	Solving the protein puzzle	Matt Lakin	https://www.olcf.ornl.gov/2022/10/28/solving-the-protein-puzzle/
11/1/22	LLNL, ORNL unite for UnifyFS	Betsy Sonewald	https://www.olcf.ornl.gov/2022/11/01/llnl-ornl-unite-for-unifyfs/
11/8/22	Discovering New Materials at Exascale Speed	Coury Turczyn	https://www.exascaleproject.org/qmcpack/
11/9/22	Distilling How Water Turns into Ice	Coury Turczyn	https://www.olcf.ornl.gov/2022/11/09/distilling-how-water-turns-into-ice/
11/11/22	Predicting Protein Targets and the Molecules that Bind with Them	Coury Turczyn	https://www.olcf.ornl.gov/2022/11/11/predicting-protein-targets-and-the-molecules-that-bind-with-them/
11/14/22	INCITE program awards supercomputing time to 56 projects to accelerate science and engineering research	Katie Bethea	https://www.olcf.ornl.gov/2022/11/14/incite-program-awards-supercomputing-time-to-56-projects-to-accelerate-science-and-engineering-research/
11/16/22	Charting the new Frontier	Matt Lakin	https://www.olcf.ornl.gov/2022/11/16/charting-the-new-frontier/
11/17/22	Plasma Simulation Code Wins 2022 ACM Gordon Bell Prize	Coury Turczyn	https://www.olcf.ornl.gov/2022/11/17/plasma-simulation-code-wins-2022-acm-gordon-bell-prize/
11/18/22	Exascale acceleration	Matt Lakin	https://www.olcf.ornl.gov/2022/11/18/exascale-acceleration/
12/1/22	Data Transfer Tool Makes New Connections at ORNL	Betsy Sonewald	https://www.olcf.ornl.gov/2022/12/01/data-transfer-tool-makes-new-connections-at-ornl/
12/1/22	Secure Science with CITADEL	Betsy Sonewald	https://www.olcf.ornl.gov/2022/12/01/secure-science-with-citadel/
12/5/22	The Climate in a Container	Betsy Sonewald	https://www.olcf.ornl.gov/2022/12/05/the-climate-in-a-container/
12/4/22	Supercomputers Lead to Super Learning Experiences	Betsy Sonewald	https://www.olcf.ornl.gov/2022/12/04/supercomputers-lead-to-super-learning-experiences/
12/13/22	OLCF's 2022 User Meeting Held Virtually on Heels of Frontier Launch	Quinn Burkhardt	https://www.olcf.ornl.gov/2022/12/13/olcfs-2022-user-meeting-held-virtually-on-heels-of-frontier-launch/
12/21/22	Advancing the Additive Manufacturing Revolution	Coury Turczyn	https://www.exascaleproject.org/advancing-the-additive-manufacturing-revolution/

APPENDIX C. 2022

The Oak Ridge Leadership Computing Facility (OLCF) provided the Summit computational resource (Table C-1) and the Alpine and HPSS data resources for production use in 2022. 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.

IBM AC922 (SUMMIT) RESOURCE SUMMARY

The OLCF installed and deployed an IBM AC922 system named Summit, which became available for full production on January 1, 2019. Summit is composed of 4,662 high-density compute nodes, each equipped with two IBM POWER9 CPUs and six NVIDIA Volta GPUs. 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 (HPL) and Conjugate Gradient (HPCG) 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 will provide spare parts for Summit if an optional 6th year is exercised in the contract.

Table C-1. Summit 2022

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)	4,662	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 100G InfiniBand (Non-blocking Fat Tree)

GPFS 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 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 gigabytes per second of performance.

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 256GB 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. The data analytics cluster is currently connected to the OLCF's 250 PB high-performance GPFS file system, Alpine.

HIGH-PERFORMANCE STORAGE SYSTEM RESOURCE SUMMARY

The OLCF provides a long-term storage archive system based on the High-Performance Storage System (HPSS) software product co-developed by IBM, Los Alamos National Laboratory (LANL), Sandia National Laboratories (Sandia), Lawrence Livermore National Laboratory (LLNL), Lawrence Berkeley National Laboratory (LBNL), and Oak Ridge National Laboratory (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.

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 room 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 will expand by 11 PB of Spectrum Scale and 7 PB of tape in late Q1 CY 2023 with additional expansions to both disk and tape components slated for 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 file system, 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-2. 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-2. OLCF HPC system production dates, 2008–present.

System	Type	Production date[a]	Performance end date[b]	Notes
Summit	IBM AC922	March 19, 2021	Present	Production with 4,680 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.
Summit	IBM AC922	July 1, 2020	Present	Production with 4,626 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	Present	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.
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.

Table C-2. OLCF HPC system production dates, 2008–present (continued).

System	Type	Production date[a]	Performance end date[b]	Notes
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. 2022

2022 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 (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]).

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

APPENDIX E. 2022

DIRECTOR'S DISCRETIONARY PROJECTS (ENHANCED ENCLAVE) ENABLED (AT ANY POINT) IN CY 2022

Project ID	PI	Institution	Most recent summit allocation	Summit usage	Project name
ARD144_MDE	Sriram Shankaran	GE Aviation	20,000	0	High Pressure-Low Pressure Turbine Interaction Dynamics
BIF134_MDE	Joe Lake	ORNL	0	0	SPI Beta Testing with Cancer Data
BIF138_MDE	Raquel Dias	University of Florida	5,000	0	Deep learning for accurate and cost-effective imputation of genotypes at whole genome level
CSC466_MDE	Jeremy Cohen	ORNL	0	0	SPI Beta Testing with Sparse PPMI Matrix
CSC487_MDE	Joseph Lake	ORNL	30,000	16,481	Population Health Research with Sensitive Data
CSC489_MDE	Jeremy Cohen	ORNL	20,000	1	VA Prostate Cancer
GEO141_MDE	Dalton Lunga	ORNL	10,000	0	Global Scale Gravitational Modeling with AI
MED118_MDE	Silvia Crivelli	LBNL	20,000	0	Develop language models for EHR data
SYB109_MDE	Michael Garvin	ORNL	20,000	0	Genomic Structural Variation
SYB110_MDE	Daniel Jacobson	ORNL, UTK	20,000	0	Combinatorial Associations: From EHR or Epistasis

DIRECTOR'S DISCRETIONARY PROJECTS (MODERATE ENCLAVE) ENABLED (AT ANY POINT) IN CY 2022

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
ARD143	Wesley Brewer	ORNL	10,000	11	iBench Extensions for ML Surrogate Model Inferencing At Petascale
ARD147	Wesley Harris	MIT	20,000	12,324	Direct numerical simulations of hypersonic boundary layer transition on a flat plate
ARD153	Zhi Jian Wang	University of Kansas	20,000	10,017	Wall-modeled large eddy simulation of high lift configuration
ARD154	Frank Ham	Cascade Technologies	0	0	Predicting broadband noise signatures of urban air mobility vehicles
ARD155	Vineet Ahuja	Whisper Aero Inc.	20,000	483	CFD Simulations of Novel Airframe Configurations with Distributed Electric Propulsors

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
AST031	Pierre Ocvirk	Universite de Strasbourg, Strasbourg Astronomical Observatory	20,000	81,319	Reionization and Its Impact on the Local Universe: Witnessing Our Own Cosmic Dawn
AST146	Brian O'Shea	Michigan State University	20,000	22	Measuring Performance and Scaling of Kokkos-accelerated Athena++ on Summit
AST154	Philipp Moesta	University of Amsterdam	20,000	7	Dynamical Space-Time GRMHD Simulations of Neutron-star Mergers and Remnants
AST163	Christian Cardall	ORNL	20,000	3,346	INCITE Preparation for 3D+1D Core-Collapse Supernova Simulations with GenASiS
AST166	Gaurav Khanna	UMass Dartmouth	5,000	0	Mixed-Precision WENO Method for Hyperbolic PDE Solutions
AST169	Brant Robertson	UC Santa Cruz	70,000	0	Preparing Cholla for Extreme-Scale Cosmological Simulations using Summit
AST170	Yuran Chen	WashU, WUSTL	20,000	20,506	Simulations of the Magnetospheres of Neutron Stars and Black Holes
AST176	Eliu Huerta Escudero	Argonne, UIUC	15,000	7,579	Geometric deep learning for gravitational wave astrophysics
ATM112	Wei Zhang	ORNL	60,000	4,517	Aim High: Air Force R&D Collaboration
ATM121	Xin-Zhong Liang	U. of Maryland	15,000	0	Dashboard for Agricultural Water use and Nutrient management (DAWN)
ATM122	Branko Kosovic	UCAR, NCAR	20,000	13,073	GPU-Resident Real-Time Large-eddy Simulations of Stably Stratified Atmospheric Boundary Layers
ATM127	Ngoc-Cuong Nguyen	MIT	20,000	3,167	High-Fidelity Space Weather Modeling
ATM129	Veerabhadra Kotamarthi	Argonne	20,000	5	Distributed Wind Obstacle Model
BIE108	Peter St. John	NREL	10,000	623	Enzyme Engineering Directly from Protein Primary Sequence Via Natural Language Processing
BIE119	Abhishek Singharoy	Arizona State University	40,000	45,582	Membrane models of biological energy transfer
BIF113	Kjiersten Fagnan	LBNL	0	0	JGI Data Archive
BIF117	Yanling Liu	FNLCR, Frederick National Laboratory	20,000	0	Systematic Annotation Creation on H&E WSIs for Automated Digital Pathology Informatics
BIF122	Ramu Anandakrishnan	VCOM	15,000	0	Identifying Multi-Hit Combinations of Genetic Mutations in Cancer

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
BIF139	Chongle Pan	OU	20,000	2,868	Biolearning
BIF140	Dali Wang	ORNL	20,000	0	Efficient prediction of lineage tree and interactions with graph neural networks and deep reinforcement learning
BIF141	Thomas Beck	ORNL	15,000	19	A Multimodal Atlas of Cells and Circuits in the Mouse Brain
BIP109	Harel Weinstein	Weill Cornell Medicine	80,000	32,808	Lipid shuttling molecular machines enabling functions of human cell membranes
BIP167	Philip Kurian	Howard University	20,000	0	Computing Many-Body Van Der Waals Dispersion Effects in Biomacromolecules
BIP169	Martin Karplus	Harvard University	20,000	0	Computational Design of HIV Vaccination Schedule
BIP198	Ada Sedova	ORNL	10,000	5,312	Genes to Interactomes
BIP208	James Gumbart	Georgia Tech	80,000	777,059	Determining the Contribution of Glycosylation to SARS-CoV-2 S-Protein Conformational Dynamics
BIP211	Rommie Amaro	UC San Diego	50,000	26,559	SARS-CoV-2 in Respiratory Aerosols
BIP212	Abhishek Singharoy	Arizona State University	21,000	0	Molecular Simulations of Energy Transfer in Supercomplexes
BIP213	Juan Perilla	U. of Delaware, UIUC	20,000	15,407	Molecular Basis of the Native SARS-CoV-2 Liposome Physical Properties From All-Atom Simulations
BIP215	Julie Mitchell	ORNL	65,000	3,187	Training an Advanced Model for Structure-Based Proteomics
BIP216	Arvind Ramanathan	Argonne	40,000	419	Understanding SARS-CoV-2 Replication-Transcription Complex by Integrating Experiments, Simulations, and ML Techniques
BIP217	Alan Hicks	ORNL	20,000	13,059	Sampling Closed-Open Transition of SARS-CoV-2 Nsp15 in Real Time
BIP218	Michael Kiebish	Berg	20,000	8,668	Supercomputing-Guided Drug Discovery and Artificial Intelligence in the Pharmaceutical Sciences
BIP221	Ahmet Yildiz	UC Berkeley	30,000	5,942	Identification and Reengineering of Effective Nanobodies Against SARS-CoV-2 Omicron Variant
BIP222	Ruth Nussinov	Leidos, Inc.	15,000	0	Modeling microbial proteins and their interactions with the human proteins

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
BIP225	Harel Weinstein	Weill Cornell Medicine	120,000	127,877	Small Molecule Inhibitors of the Membrane Fusion Machinery of SARS-CoV-2.
BIP227	Giulia Palermo	UC Riverside	20,000	0	Dynamics and Mechanism of Transposon-Encoded CRISPR-Cas Systems
BIP229	Mahmoud Moradi	University of Arkansas, Fayetteville, UIUC	10,000	12,921	Conformation Free Energy Landscapes of Viral Glycoproteins
BIP230	Yinglong Miao	University of Kansas	19,800	2,518	Accelerated Molecular Simulations of Protein Interactions with SARS-CoV-2 Spike
BIP232	Pin-Kuang Lai	Stevens Institute of Technology	20,000	2,780	Integrating molecular dynamics simulations and machine learning to accelerate antibody development
BIP233	Dilnoza Amirkulova	Proctor & Gamble Company	20,000	0	Molecular modeling of surfactant interactions with phospholipids bilayers mimicking corneal epithelium
CFD124	Bamin Khomami	UTK	200,000	113,039	Elucidating the Molecular Rheology of Entangled Polymeric Fluids via Direct Comparison of NEMD Simulations and Model Predictions
CFD136	John Gounley	ORNL	15,000	24	Performant High-Order Lattice Boltzmann for Exascale Applications
CFD142	Balaji Jayaraman	GE	20,000	0	Characterizing Coastal Low-Level Jets and their Impact on Offshore Wind Farms
CFD144	Saumil Patel	Argonne	18,000	2	Scaling Internal Combustion Engine Simulations Using nekRS
CFD150	Fazle Hussain	Texas Tech University	20,000	240	Direct Numerical Simulation of Turbulent Pipe Flow Using Nek-OpenMP/OpenACC
CFD154	Spencer Bryngelson	Georgia Tech	8,000	630	Accelerated Sub-Grid Multi-Component Flow Physics
CFD159	Vittorio Badalassi	ORNL	5,000	259	Fusion Energy Reactor Models Integrator (FERMI)
CFD161	Mark Kostuk	General Atomics	18,000	5,204	ALMA for SNS
CFD163	Ioannis Nompelis	University of Minnesota	20,000	0	Porting of an implicit finite volume solver for compressible flow to GPUs
CFD165	Federico Municchi	Colorado School of Mines	20,000	0	CFD simulations of heat and mass transport in solar powered membrane distillation systems
CHM155	Peter Coveney	University College London	80,000	65,857	COMPBIO

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
CHM156	Remco Havenith	U. of Groningen	7,500	187	TURTLE
CHM160	Andre Severo Pereira Gomes	CNRS	20,000	1	PRECISE: Predictive Electronic Structure Modeling of Heavy Elements
CHM174	Monojoy Goswami	ORNL, UTK	30,000	34,838	Multi-scale/multi-physics molecular simulations at the Chemical Sciences Division
CHM175	Thomas Miller	Entos	12,000	9,770	Accelerating COVID-19 Drug Discovery through Leadership-scale Graph Neural Net Machine Learning
CHM180	Aditya Savara	ORNL	25,000	21,797	Simulating Mesoscale Structural Dynamics of Vitrimers
CHM181	Kurt Mikkelsen	University of Copenhagen	15,000	10,990	Massively parallel, GPU-enabled cluster perturbation methods
CHM183	Boris Kozinsky	Harvard University	20,000	38,972	Large-scale reactive catalytic dynamics from machine learning
CHM184	Darrin York	Rutgers	20,000	2	AMBER Drug Discovery
CHM186	Anna Gudmundsdottir	University of Cincinnati	10,000	11	Photopopping Crystals
CHM187	Vyacheslav Bryantsev	ORNL	20,000	0	Molten salts modeling with ab initio accuracy
CHM188	Thanh Do	UTK	20,000	164	Seeking novel N-methylated macrocyclic peptide conformers while optimizing MD methodology
CHM189	Alex Ivanov	ORNL	20,000	0	Pm complex chemistry relativistic simulations
CHP115	Roberto Car	Princeton	20,000	17,703	Million Atom Chemical Dynamics at Heterogeneous Aqueous Interfaces
CLI137	Forrest Hoffman	ORNL	10,000	33	Earth System Grid Federation 2 (ESGF2)
CLI138	Moetasim Ashfaq	ORNL	0	0	Analytical Frameworks for Sub-Seasonal to Multi-Decadal Climate Predictions and Impact Assessments
CLI143	Nathaniel Collier	ORNL	5,000	0	Identifying Ecosystems Vulnerable to Climate Change
CLI144	Peter Thornton	ORNL	20,000	263	Ultra-high Resolution Land Process Simulation Using GPUs and OpenACC on Summit
CLI146	Jiafu Mao	ORNL	15,000	117	Ecosystem Resilience to Thermal Extremes: Urbanization Impacts
CLI147	Valentine Anantharaj	ORNL	0	0	A petascale data hackathon for exploring the baseline for a digital twin of the earth

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
CLI900	Valentine Anantharaj	ORNL	5,000	83	Provisioning of Climate Data
CMB103	Jacqueline Chen	Sandia	20,000	6,116	Mitigating Climate Change Through Ammonia/Hydrogen Combustion
CMB124	Ronald Grover	GM	20,000	4,284	Prediction of Engine Knock in a Gasoline Direct Injection Engine
CMB147	Joseph Oefelein	Georgia Tech	0	0	Analysis of Combustion and Wave Dynamics in Rotating Detonation Engines
CMB148	Aditya Konduri	IISc, IISc Bangalore	15,000	0	Scalable Mathematically Asynchronous Algorithms for Flow Solvers
CPH005	Dario Alfe	University College London	20,000	1,082	New Frontiers for Material Modeling via Machine Learning Techniques with Quantum
CPH111	Andreas Glatz	Argonne	18,000	10,288	Genetic Algorithms for Tailored Superconducting Materials
CPH130	Jihong Ma	UVM	20,000	184	Conductive Polymer
CPH137	Emanuel Gull	U. of Michigan	10,000	7,807	Exploration of Time-Dependent Second Order Perturbation Theory
CPH138	Sumit Sharma	OU	20,000	4	Computational Studies of Interactions of Small Molecules with Biological and Polar Interfaces
CPH141	Ivan Stich	Slovak Academy of Sciences	20,000	995	QMCPACK performance comparison and tests
CSC143	Norbert Podhorszki	ORNL	20,000	31,668	ADIOS – The Adaptable IO System
CSC343	Sergey Panitkin	University of Montreal	5,000	10	Porting the ATLAS Experiment Software and Workload Management System
CSC345	Ramakrishnan Kannan	ORNL	20,000	143	Parallel Low-rank Approximation with Nonnegative Constraints (PLANC)
CSC357	Laxmikant Kale	UIUC	15,000	2,166	CharmRTS
CSC362	Wen-mei Hwu	UIUC	20,000	804	Petascale 3D Image Reconstruction with Ptycho-Tomography
CSC369	Dmitry Pekurovsky	UC San Diego	15,000	102	Scalable Software Framework for Multidimensional Fourier Transforms
CSC377	Allan Grosvenor	Microsurgeonbot	20,000	16,340	Expertise-as-a-Service via Scalable Hybrid Learning: R&D Supporting Improvements to Hybrid Intelligence Agent Tooling
CSC380	Bronson Messer	ORNL	150,000	90,496	CAAR for Frontier

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
CSC382	Catherine Schuman	UTK	5,000	1,019	Scalable Neuromorphic Simulation and Training
CSC397	Michael McCarty	NVIDIA	15,000	24,852	Testing Legate Deployment
CSC401	Barbara Chapman	SUNY Stony Brook	10,000	1,461	Support for the Deployment of OpenMP in Large-Scale Scientific Applications
CSC418	Tjerk Straatsma	ORNL	60,000	0	ADAC Applications Readiness
CSC427	Michela Taufer	UTK	15,000	3,134	A4NN: Accelerating Scientific AI Leveraging Open Data and Open Models
CSC434	Bronson Messer	ORNL	50,000	30	Portable Performance on Exascale Hybrid Architectures
CSC436	John Paul Walters	USC	15,000	0	CASPER: Compiler Abstractions Supporting high Performance on Extreme-scale Resources
CSC443	Seung-Hwan Lim	ORNL	20,000	0	Generalized Scalable Parallel Training of Deep Graph Neural Networks
CSC452	Abhinav Bhatele	U. of Maryland, LLNL	20,000	15,990	Performance Analysis and Tuning of HPC and AI Applications
CSC457	Massimiliano Lupo Pasini	ORNL	20,000	48	Scalable stable numerical optimization for artificial intelligence applied to computed tomography
CSC462	Antonios Kougkas	Illinois Tech	15,000	0	Hermes: Extending the HDF Library to Support Intelligent I/O Buffering for Deep Memory and Storage Hierarchy System
CSC467	Andrew Gallo	GE	5,000	0	Genesis CI/CD
CSC470	Christopher Stanley	ORNL	20,000	11	Flexible Privacy-Enabled Platform for Sensitive Applications
CSC471	Rishi Khan	Extreme Scale Solutions	10,000	1,445	Scaling Distributed GPU FFT to Exascale
CSC489	Jeremy Cohen	ORNL	0	0	VA Prostate Cancer Companion
CSC491	Terry Jones	ORNL	15,000	0	WEAVABLE
CSC492	John Cohn	IBM	10,000	0	Structure Transformer for Code Generation
CSC495	Hong Liu	ORNL	5,000	0	SMCDC2022
CSC499	Irina Rish	University of Montreal	20,000	9,296	Scalable Foundational Models for Transferable Generalist AI
CSC505	Mark Coletti	ORNL	20,000	5	Increasing search effectiveness for MENNDL via variable-length representation
CSC506	Johannes Blaschke	LBNL	10,000	0	NESAP

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
CSC511	Geoffrey Lentner	Purdue University	10,000	1	HyperShell Development and Scaling
CSC519	Maciej Cytowski	Pawsey Supercomputing Center	15,000	0	Pawsey SC access for early testing purposes
ENG123	Ramesh Balakrishnan	Argonne	10,000	18,836	Direct Numerical Simulation of Separated Flow over a Speed Bump at Higher Reynolds Numbers
ENG130	Zhiting Tian	Cornell	18,000	10,356	Understanding Heat Transport in Complex, Ultra-Low Conductivity Materials
ENG131	Pablo Seleson	ORNL	20,000	606	A Scalable Cabana based Peridynamics Fracture Simulator
FUS137	William Fox	Princeton Plasma Physics Laboratory	20,000	11,374	Energetics of Collisionless Plasmas in the Laboratory and Space
FUS138	Axel Huebl	LBNL	0	0	Elucidating Acceleration Mechanisms in Laser-Plasma Ion Accelerators
FUS144	Predrag Krstic	SUNY Stony Brook	20,000	0	Classical and quantum-classical molecular dynamics for the fusion plasma-material interface
FUS147	Noah Reddell	HPE	20,000	0	Study of High Energy Density Z-Pinch Plasma Stability by Kinetic Model on GPU
GEN026	Doug Lattman	ORNL	0	0	DC-DataCenterPowerResearch
GEN042	Kevin Thach	ORNL	1,000	0	Platforms – Managed Work Activities
GEN156	Olga Kuchar	ORNL	0	0	DataFed
GEO112	Philip Maechling	USC	75,000	99,734	Extreme-Scale Simulations for Advanced Seismic Ground Motion and Hazard Modeling
GEO139	Sangkeun Lee	ORNL	5,000	135	URBAN-NET
GEO142	Dalton Lunga	ORNL	15,000	1	Scalable GeoAI Workflows for Satellite Image Analytics
HEP134	Dirk Hufnagel	Fermilab	2,000	0	HEPCloud-FNAL
LGT100	Andre Walker-Loud	LBNL	60,000	38,871	The Structure and Interactions of Nucleons from the Standard Model
LGT107	Rajan Gupta	LANL	20,000	63,907	Nucleon Matrix Elements: Probes of New Physics
LGT114	Amy Nicholson	University of North Carolina at Chapel Hill	10,000	7,724	Electromagnetic corrections to the nucleon axial charge
LGT120	Amy Nicholson	University of North Carolina at Chapel Hill	10,000	0	Electromagnetic corrections to the nucleon axial charge

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
LRN009	George Siopsis	UTK	5,000	2	Hybrid Quantum-Classical Reinforcement Learning in Controlled Quantum Networks
LRN018	Vikram Jadhao	Indiana University	20,000	0	Training Machine Learning Surrogates for Simulations using Adaptive Methods
LRN025	Steven Young	ORNL	20,000	0	Model Parallel Neural Architecture Search
LSC115	Vipin Sachdeva	Silicon Therapeutics	20,000	1,209	Weighted Ensemble Simulations for In Silico Modeling of Targeted Protein Degradation
LSC117	Qun Liu	BNL	5,000	4,167	Investigation of SARS-CoV-2 Host Interactions Using AlphaFold
MAT198	Ivan Oleynik	University of South Florida	50,000	15,289	Predictive Simulations of Phase Transitions in Dynamically Compressed Materials
MAT201	Panchapakesan Ganesh	ORNL	100,000	4,157	Center for Nanophase Materials Sciences
MAT223	Neil Gershenfeld	MIT	10,000	0	Performance Scaling of Particle Systems for Discovery of Multiphysics Models
MAT226	Jan Michael Carrillo	ORNL, UTK	20,000	16,813	Molecular Dynamic Simulations of Amphiphilic Oligomer Membranes: Design Rules Towards Stable Membranes Capable of Learning and Memory
MAT230	Vikram Gavini	U. of Michigan	20,000	0	DFT-FE
MAT231	Zachary Ulissi	Carnegie Mellon University	35,000	0	Deep Learning for Electrocatalyst Applications with the Open Catalyst 2020 Dataset
MAT233	Ying Wai Li	LANL	15,000	22	Data Driven Modeling of Non-Equilibrium Dynamics in Chemical and Materials Systems
MAT235	Guannan Zhang	ORNL	20,000	2	AI-based Scalable Feature Extraction for Real-time Processing and Analytics of Neutron Scattering Data
MAT240	Yi Yao	Duke	20,000	24,295	GPU Acceleration of an All-Electron Full-Potential G0W0 Method
MAT242	Yangyang Wang	ORNL	10,000	1,526	Resolving Flow-Induced Mesoscopic Structures in Polymeric Materials
MAT243	Aliya Tychengulova	Satbayev University	5,000	0	Study on the catalytic mechanism and properties of materials for water-splitting in natural and artificial photosynthesis

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
MAT250	Massimiliano Lupo Pasini	ORNL	20,000	2,976	Scalable accelerated training of physics informed deep learning models for material science
MAT253	Jonathan Schwartz	U. of Michigan	15,000	1,059	Fused Multi-Modal Electron Tomography
MED112	Sumitra Muralidhar	VA, Veterans Administration, BVARI,	140,000	185,453	Genome-Wide Phenome-Wide Association Study in the Million Veteran Program
MED114	James Christian	ORNL	2,000	0	Construction of Environmental Determinants of Health Data for VA and NIH
MED115	Hong-Jun Yoon	ORNL	15,000	7,915	Studies of Scalable Machine Learning Algorithms for Histopathological Image Analysis
MED121	Hong-Jun Yoon	ORNL	20,000	0	Skin Cancer Classification Systems with Mobile Devices
NME100	Ram Mohan	North Carolina Agriculture and Technology State University	1,000	0	Nano to Engineering Scale Modeling – Materials, Mechanics, and Manufacturing
NPH136	Rick Archibald	ORNL	20,000	0	Development of Design Optimization Code for the Transformational Challenge Reactor
NPH143	Kenneth McElvain	UC Berkeley	20,000	855	Neutron Star Interiors: Nuclear-matter Calculations from Chiral Effective Field Theory
NPH145	Kehfei Liu	University of Kentucky	20,000	0	Beyond Standard Model Physics from Lattice QCD
NPH152	Keh-Fei Liu	University of Kentucky	20,000	18,798	Gravitational Form Factors of Hadrons
NRO107	Shinjae Yoo	BNL	10,000	0	Spatiotemporal learning
NTI114	Michael Shirts	U. of Colorado	19,000	0	Nanoscale phase behavior of and transport in lyotropic liquid crystal membranes
PHY129	Alexander Tchekhovskoy	Northwestern University	20,000	42,379	Simulating Neutron Star Binary Merger Remnant Disks and Tilted Thin Disk
PHY130	Zhihong Lin	UC Irvine	20,000	32,357	Integrated Simulation of Energetic Particles in Burning Plasmas
PHY149	Markus Eisenbach	ORNL	20,000	5,083	Pushing the Limits of Classical Simulation of Hard Quantum Circuits via Novel Tensor Network Algorithms and Accelerated High Performance Computing
PHY158	Phillip Lotshaw	ORNL	25,000	0	QAOA simulations

Project ID	PI	Institution	Most recent Summit allocation	Summit usage	Project name
PHY166	Deborah Levin	UIUC	20,000	5,148	High Fidelity Kinetic Simulations of Plasma-Based Flows Using Heterogeneous Computational Strategies
PHY171	Michael Wilking	SUNY Stony Brook	20,000	155	Markov Chain Monte Carlo Development for Neutrino Physics
PSS101	Jonathan Jara-Almonte	Princeton Plasma Physics Laboratory	20,000	0	Reconnection Turbulence in the Magnetotail
SYB105	Daniel Jacobson	ORNL	100,000	68,116	CBI Contribution in Kind Allocation
TUR120	Pui Kuen Yeung	Georgia Tech	60,000	0	Extreme-scale phenomena in turbulent dispersion and mixing
TUR137	Justin Sirignano	Oxford	20,000	11,175	Deep Learning Closure Models for Large-Eddy Simulation of Unsteady External Aerodynamics
TUR138	Ramesh Balakrishnan	Argonne	73,173	21,844	Simulating the Transmission and Dispersion of Aerosol in Turbulent Flows
TUR140	Shanti Bhushan	Mississippi State University	10,000	0	Neural Network Framework to Enhance Unsteady Turbulent Flow Predictions