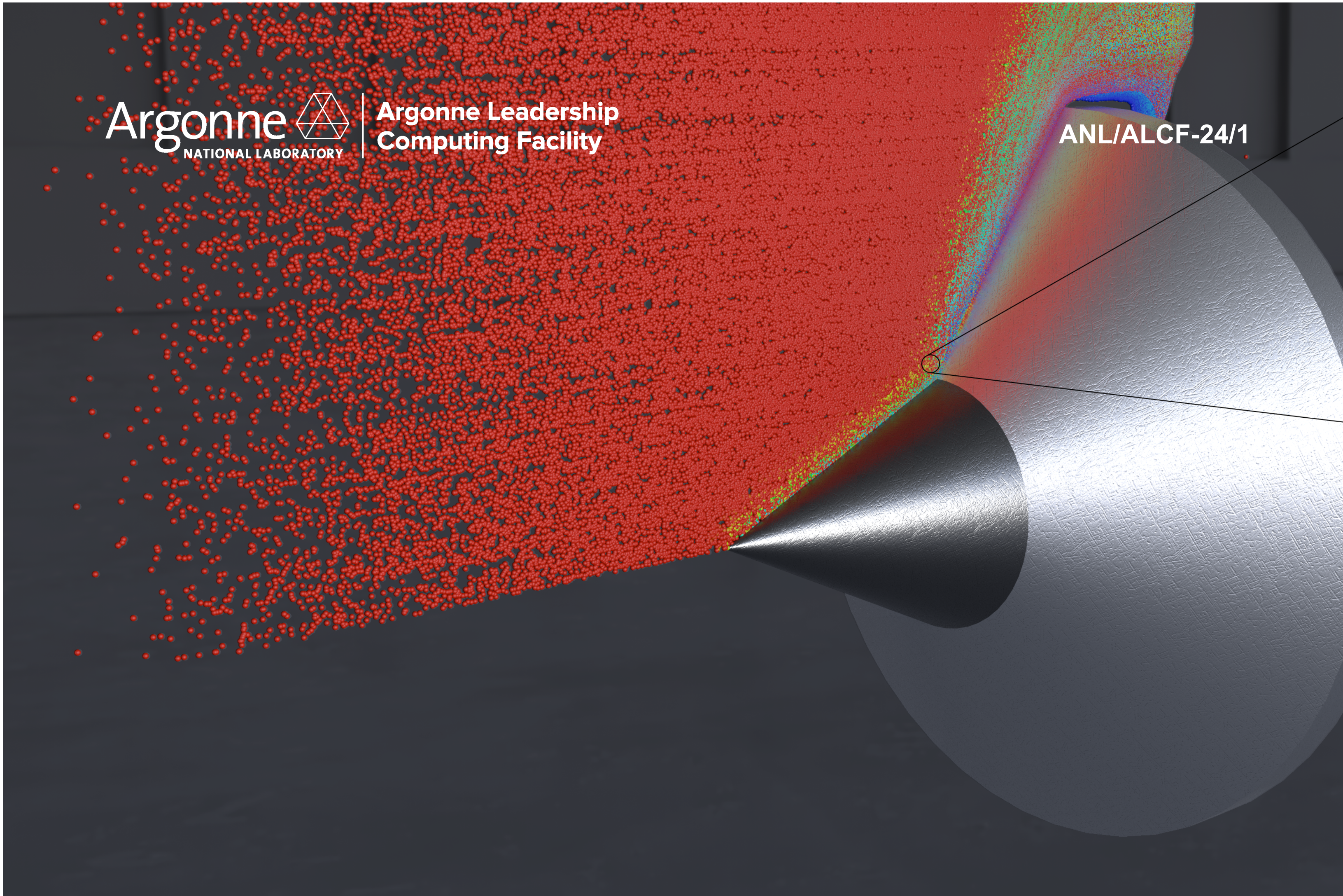
Argonne NATIONAL LABORATORY | Argonne Leadership Computing Facility

ANL/ALCF-24/1



2023 Operational Assessment Report

On the cover: *Researchers from the University of Dayton Research Institute and the Air Force Research Laboratory are using ALCF supercomputers to carry out quantum mechanically guided simulations of hypersonic flows in thermal and chemical non-equilibrium. This image shows the velocity field from a calculation to simulate a Mach 8.2 experiment.*

Image credit: *ALCF Visualization and Data Analytics Team and University of Dayton Research Institute.*

Contents

Contents (Cont.)

Contents (Cont.)

Contents (Cont.)

Contents (Cont.)

Figures

Figures (Cont.)

Figures (Cont.)

Tables

Tables (Cont.)

Background

In 2004, the U.S. Department of Energy's (DOE's) Advanced Scientific Computing Research (ASCR) program established the Leadership Computing Facility (LCF) with a mission to provide the world's most advanced computational resources to the open science community. The LCF is a huge investment in the nation's scientific and technological future, inspired by a growing demand for large-scale computing and its impact on science and engineering.

The LCF operates two world-class centers in support of open science at Argonne National Laboratory (Argonne) and at Oak Ridge National Laboratory (Oak Ridge) and deploys diverse machines that are among the most powerful systems in the world today. The LCF ranks among the top U.S. scientific facilities delivering impactful science. The work performed at these centers informs policy decisions and advances innovations in far-reaching areas such as energy assurance, ecological sustainability, and global security.

The leadership-class systems at Argonne and Oak Ridge operate around the clock every day of the year. The high level of services these centers provide and the exceptional science they produce justify their existence to the DOE Office of Science and the U.S. Congress.

This page intentionally left blank.

Executive Summary

This Operational Assessment Report describes how the Argonne Leadership Computing Facility (ALCF) met or exceeded every goal set by DOE for the calendar year (CY) 2023.

In CY 2023, ALCF operated Theta, an Intel-based Cray XC40 system (11.7 petaflops) augmented with 24 NVIDIA DGX A100-based nodes (3.9 petaflops), that supports diverse workloads, integrating data analytics with artificial intelligence (AI) training and learning in a single platform; and Polaris (44 petaflops) AMD and NVIDIA-based HPE Apollo 6500 Gen10+ system that provides a powerful new platform for preparing applications and workloads for Aurora, Argonne National Laboratory's (Argonne's) upcoming Intel-Hewlett Packard Enterprise (HPE) exascale supercomputer.

In CY 2023, Theta delivered a total of 20.4 million node-hours to 13 Innovative and Novel Computational Impact on Theory and Experiment (INCITE) projects and 5.4 million node-hours to ASCR Leadership Computing Challenge (ALCC) projects (10 awarded during the 2022–2023 ALCC year and 9 awarded during the 2023–2024 ALCC year), as well as providing substantial time for Director's Discretionary (DD) projects (5.8 million node-hours). As Table ES.1 shows, Theta performed exceptionally well in terms of overall availability (93.9 percent), scheduled availability (99.4 percent), and utilization (89.8 percent; Table 2.1).

Since going into production in August 2022, Polaris has supported research teams via the U.S. Department of Energy's (DOE's) Exascale Computing Project (ECP) and the ALCF's Aurora Early Science Program (ESP). In CY 2023, it also supported 13 INCITE projects (1888.6K node-hours) and 6 ALCC projects awarded during the 2022–2023 ALCC year (507.1K node-hours) and 9 awarded during the 2023–2024 ALCC year (217.3K node-hours), and numerous DD projects (1192K node-hours). As Table ES.1 shows, Polaris performed extremely well in terms of overall availability (93.8 percent), scheduled availability (99.2 percent), and utilization (81.1 percent; Table 2.1). In CY 2023, ALCF supported more than 1,600 users, including those who came for training. As of the submission date of this document, ALCF's users have published 240 papers in peer-reviewed journals and technical proceedings. Other CY 2023 highlights include:

- In June, ALCF completed the installation of Aurora's 10,624th and final blade. The system was opened to Aurora ESP users in November.
- The results of high-performance Linpack runs for a part of Aurora secured the second spot on the TOP500 list announced in November.
- The Distributed Asynchronous Object Store (DAOS) is a major component of the Aurora system. During SC23, a partial DAOS run topped the IO500 list by a significant margin.
- The resources available in the AI Testbed were significantly expanded, and it continues to be extremely useful and popular with users.
- ALCF stood up a set of integrated tools and resources aimed at meeting the needs of experimental facilities and providing a comprehensive approach to integrated research infrastructure (IRI) by employing several new facility capabilities.

- ALCF staff members piloted an immersive HPC training program for 60 science, technology, engineering, and math (STEM) students. The program focused on using cutting-edge computational and data science tools to work on energy justice projects.
- ALCF conducted 28 training activities, reaching more than 1,500 participants in events where participation was tracked.
- In a separate workforce development effort, ALCF expanded its *Intro to AI-driven Science of Supercomputers* series curriculum to include the basics of large-language models (LLMs) and their applications to science.

Table ES.1 Summary of the Target and Actual Data for the Previous Year (2023) Metrics

Area	Metric	2023 Target	2023 Actual
User Results	User Survey – **Overall Satisfaction**	3.5/5.0	4.5/5.0
	User Survey – **User Support**	3.5/5.0	4.5/5.0
	User Survey – Problem Resolution	3.5/5.0	4.5/5.0
	User Survey – **Response Rate**	25.0%	37.6%
	% User Problems Addressed within Three Working Days	80.0%	93.6%
Business Results	Polaris Overall Availability	90.0%	93.8%
	Polaris Scheduled Availability	90.0%	99.2%
	Theta with expansion Overall Availability	90.0%	93.9%
	Theta with expansion Scheduled Availability	90.0%	99.4%
	% of INCITE node hours from jobs run on 20.0% or more of Polaris (99 - 532 nodes)	20.0%	56.7%
	% of INCITE node hours from jobs run on 60.0% or more of Polaris (297 - 532 nodes)	N/A[a]	13.7%
	% of INCITE node hours from jobs run on 20.0% or more of Theta (800 - 4008 nodes)	20.0%	88.3%
	% of INCITE node hours from jobs run on 60.0% or more of Theta (2400 - 4008 nodes)	N/A	17.8%
	Theta-fs0 Overall Availability	90.0%	94.2%
	Theta-fs0 Scheduled Availability	90.0%	99.9%
	Grand Overall Availability	90.0%	93.2%
	Grand Scheduled Availability	90.0%	99.5%
	Eagle Overall Availability	90.0%	94.2%
	Eagle Scheduled Availability	90.0%	99.9%
	HPSS[b] Overall Availability	90.0%	93.8%
	HPSS Scheduled Availability	90.0%	99.6%

[a] N/A = not applicable.

[b] HPSS = high-performance storage system.

Soon, a fully capable Aurora will be used for many AI and computational science applications and to train a 1-trillion-parameter large language model for scientific research, changing the computing landscape once again. ALCF will continue to engage in strategic activities that push the boundaries of what's possible in computational science and engineering and allow ALCF to deliver science on day one.

This page intentionally left blank.

Response to Recommendation from the PY OAR

The Facility should list recommendations from the previous year's OAR and the Facility's responses to them. If prior year recommendations were directed to DOE/ASCR, that point should be clearly stated. The Facility should include the responses from DOE/ASCR in the report.

The Facility may also comment on any plans or actions that may have changed since these original responses.

NOTE: This data is for informational purposes, and is not formally a part of the review. It will allow the reviewers to place certain actions in context.

In the previous year's OAR, the ALCF did not receive any recommendations.

This page intentionally left blank.

Section 1. User Support Results

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

ALCF Response

ALCF has processes in place to effectively support its customers, to resolve user problems, and to conduct outreach. The 2023 annual user survey measured overall satisfaction, user support, and problem resolution, and serves both to mark progress and to identify areas for improvement (Table 1.1). User satisfaction with ALCF services remains consistently high, as evidenced by survey response data. The following sections describe ALCF events and processes, consider the effectiveness of those processes, and note the improvements made to those processes during calendar year (CY) 2023.

Table 1.1 All 2023 User Support Metrics and Results [a]

		2022 Target	2022 Actual	2023 Target	2023 Actual
Number Surveyed		N/A [c]	1,446	N/A [c]	1,495
Number of Respondents (Response Rate)		25.0%	555 (38.4%)	25.0%	562 (37.6%)
Overall Satisfaction	Mean	3.5	4.5	3.5	4.5
	Variance	N/A	0.6	N/A	0.6
	Standard Deviation	N/A	0.8	N/A	0.6
Problem Resolution	Mean	3.5	4.5	3.5	4.5
	Variance	N/A	0.5	N/A	0.5
	Standard Deviation	N/A	0.7	N/A	0.7
User Support	Mean	3.5	4.5	3.5	4.5
	Variance	N/A	0.5	N/A	0.5
	Standard Deviation	N/A	0.7	N/A	0.7
		2022 Target	**2022 Actual**	**2023 Target**	**2023 Actual**
% of Problems Addressed Within Three Working Days [b]		80.0%	93.3%	80.0%	93.6%

[a] In September 2015, all Advanced Scientific Computing Research (ASCR) facilities adopted a new definition of a facility user based on guidance from the U.S. Department of Energy's (DOE's) Office of Science. Under this definition, a user must have logged in to an ALCF resource during the period in question. This definition of a user provides the basis for all survey results.

[b] The statistical population represented in this metric includes problem tickets submitted from all users. Note that this is a larger population than those who qualify under the September 2015 definition of a facility user mentioned in note *a* above.

[c] N/A = not applicable.

Survey Approach

In 2017, the ALCF worked with a consultancy to revise and shorten its annual user survey: omitting the workshop-related questions, requiring responses only for those questions that comprise the DOE metrics for the Operational Assessment Report (OAR), and making every other question optional. The facility polls workshop attendees separately.

The 2023 user survey closely resembled the 2022 survey, with minor modifications to infrastructure, performance, and debugging tools/services listed in various questions. Two new questions were added: one that captured the use or planned use of workflow tools on ALCF systems and another that checked which accelerator programming model was employed in users' codes. Two questions were removed: one that captured users' enthusiasm for an ALCF Slack channel and another that gauged their interest in participating in an annual meeting.

The 2023 survey was administered through a contract with Statistical Consulting Services at the Department of Statistics and Actuarial Science at Northern Illinois University and consisted of 6 required questions and 24 optional questions. The survey and associated e-mail campaign ran from November 16, 2023, through December 31, 2023. Each reminder e-mail included a user-specific link to the online survey. Most respondents were able to complete the survey in 10 minutes or less.

Likert Scale and Numeric Mapping

Almost all questions in the ALCF survey use a six-point Likert Scale. This is a standard way to rate user responses for surveys because (1) it provides a symmetric agree-disagree scale; (2) it can be mapped to a numeric scale; and (3) given a certain sample size, it can be used with a normal distribution to obtain useful statistical results. The method also allows for use of off-the-shelf statistics functions to determine variance and standard deviation.

ALCF follows a standard practice and maps the Likert Scale in this way or similar:

Statement	Numeric
Strongly Agree	5
Agree	4
Neutral	3
Disagree	2
Strongly Disagree	1
N/A [a]	(No Value)

[a] N/A = not applicable.

The Overall Satisfaction question applied a different point scale, as follows:

Statement	Numeric
Excellent	5
Above Average	4
Average	3
Below Average	2
Poor	1

Beginning in 2017, some of the non-metric survey questions were revised to capture sentiments about various aspects of the ALCF's user services that used the options below:

Select all that apply.
Praise
Suggestions for Improvement
Average
Below Average
Poor

1.1 User Support Metrics

Everyone who met the definition of a facility user during the fiscal year (FY) 2023—that is, users who would be counted under the Facility's annual user statistics submission to the Office of Science—was asked to complete a user survey, 1,624 individuals in total. Of those individuals, 43 did not receive the email due to undeliverable messages and 86 chose to opt-out of the survey. Of the 1,495 remaining users, 562 responded, for a 37.6 percent response rate.

Table 1.2 shows user survey results grouped by allocation program. While Innovative and Novel Computational Impact on Theory and Experiment (INCITE) and ASCR Leadership Computing Challenge (ALCC) users reported higher average Overall Satisfaction than Director's Discretionary (DD) users, the variations are not statistically significant. Other metrics are comparable, in that the variations are not statistically significant.

Table 1.2 User Survey Results in 2023 by Allocation Program

Metrics by Program		INCITE	ALCC	INCITE + ALCC	DD	All
Number Surveyed		238	83	321	1,174	1,495
Number of Respondents		93	37	130	432	562
Response Rate		39.1%	44.6	40.5%	36.8%	37.6%
Overall Satisfaction	Mean	4.6	4.3	4.5	4.5	4.5
	Variance	0.5	1.0	0.7	0.6	0.6
	Standard Deviation	0.7	1.0	0.8	0.7	0.8
Problem Resolution	Mean	4.6	4.6	4.6	4.5	4.5
	Variance	0.5	0.5	0.5	0.4	0.5
	Standard Deviation	0.7	0.7	0.7	0.7	0.7
User Support	Mean	4.6	4.6	4.6	4.5	4.5
	Variance	0.5	0.5	0.5	0.5	0.5
	Standard Deviation	0.7	0.7	0.7	0.7	0.7
All Questions	Mean	4.6	4.5	4.6	4.5	4.5
	Variance	0.5	0.6	0.5	0.5	0.5
	Standard Deviation	0.7	0.8	0.7	0.7	0.7

As Table 1.3 shows, in CY 2023, the ALCF again exceeded the targets for overall satisfaction and user support.

Table 1.3 User Support Metrics

Survey Area	2022 Target	2022 Actual	2023 Target	2023 Actual
Overall Satisfaction Rating	3.5/5.0	4.5/5.0	3.5/5.0	4.5/5.0
Avg. of User Support Ratings	3.5/5.0	4.5/5.0	3.5/5.0	4.5/5.0

1.2 Problem Resolution Metrics

Table 1.4 shows the target set for the percentage of user problem tickets or "trouble tickets" addressed in three days or less, which ALCF exceeded. A trouble ticket, which encompasses incidents, problems, and service requests, is considered "addressed" once (1) a staff member accepts the ticket, (2) the problem is identified, (3) the user is notified, and (4) the problem is solved, or it is in the process of being solved.

Table 1.4 Problem Resolution Metrics

	2022 Target	2022 Actual	2023 Target	2023 Actual
% of Problems Addressed Within Three Working Days [a]	80.0%	93.3%	80.0%	93.6%
Avg. of Problem Resolution Ratings	3.5/5.0	4.5/5.0	3.5/5.0	4.5/5.0

[a] The statistical population represented in this metric includes problem tickets submitted from all users. Note that this is a larger population than that of DOE users.

1.3 User Support and User Engagement

1.3.1 Tier 1 Support

1.3.1.1 Phone, Slack, and E-mail Support

In CY 2023, ALCF addressed and categorized 6,838 user support tickets, an 18 percent increase from the previous year. The biggest increases were in the Access, Accounts, and Allocations categories (Table 1.5). This is possibly due to the increased number of project allocations. Polaris became more utilized, and the AI testbed was opened to users.

In CY 2023, ALCF enhanced our categorization of Accounts tickets to understand better how support issues are distributed across its resources. This detailed breakdown is an initial step in detecting trends and patterns within the Accounts category. For OAR 2024, ALCF plans to implement a more refined categorization scheme.

The significant number of tickets for Resource Theta is likely due to it being the designated default for multi-resource projects, in addition to handling Theta-specific queries. Other contributing factors include the discontinuation of Theta's operation on December 31, 2023, and several scheduled and unscheduled outages, as detailed in Section 2.4.1.1. These factors may have led to an increase in ticket volume.

Below are the detailed categorization results for Accounts tickets:

Resource Theta dominates with 79.9 percent of the tickets. The Miscellaneous category is a collection of various tickets that are not connected to any resource, for example, tickets that are not intended for ALCF or misdirected e-mails, comprising 9.5 percent of tickets. Resource Polaris is associated with 6.0 percent of tickets, Resource Aurora accounts for 1.8 percent, and Resource Sunspot for 1.7 percent of tickets. For all the remaining resources – Eagle, Grand, Theta-fs0, ThetaGPU, Cooley, Website, and those in the Testbed category – each contributes a smaller portion, with ticket volumes ranging from 0.1–0.5 percent.

Table 1.5 Ticket Categorization for 2022 and 2023

Category	2022	2023
Access	915 (15%)	1,285 (19%)
Accounts [a]	2,490 (42%)	2,841 (42%)
Allocations	1,080 (18%)	1,196 (17%)
Applications Software	201 (3%)	271 (4%)
Bounced Emails	88 (1%)	29 (0%)
Compilers	35 (1%)	32 (0%)
Data Transfer	49 (1%)	58 (1%)
Debuggers	7 (0%)	8 (0%)
File System	66 (1%)	87 (1%)
HPSS [b]	156 (2%)	17 (0%)
I/O and Storage	0 (0%)	176 (3%)
Libraries	20 (1%)	25 (0%)
Miscellaneous	119 (2%)	157 (2%)
Network	8 (0%)	11 (0%)
Performance and Performance Tools	9 (0%)	27 (0%)
Reports	146 (2%)	169 (2%)
Scheduling	337 (5%)	319 (5%)
System	73 (1%)	129 (2%)
Visualization	2 (0%)	1 (0%)
TOTAL TICKETS	5,801 (100%)	6,838 (100%)

[a] In CY 2023, 79.9% of tickets related to Theta, 9.5% to Miscellaneous, 6.0% to Polaris, 1.8% to Aurora, 1.7% to Sunspot, and 1.1% to all other resources combined. This was the simplest way to break down the Accounts category. We are investigating other methods.

[b] HPSS = high-performance storage system.

1.3.1.2 Suballocations Management for APS Projects

Suballocations let principal investigators (PIs) control who on their team can run jobs, how much of the allocation they are allowed to consume, and the date range in which the allocation is valid. ALCF piloted suballocations as part of a crosscutting project with the Advanced Photon Source (APS) user facility at Argonne demonstrating fully automated, "no human in the loop" data analysis of beamline data. ALCF provided a "master allocation" to the APS, which was then subdivided among the beamlines participating in the pilot.

1.3.1.3 Userbase Improvements

Automated Testing Increased Productivity for Userbase3 Software

The Web Development team increased testing productivity by incorporating a continuous improvement test server, which runs all automated tests every night. The output is tracked and posted to Slack and GitLab, so test results can be analyzed to verify that any changes to the codebase have not caused failures in other areas.

User Interface Improvements

Improved instructions on the Userbase3 home page for Account Request and Account Reactivation helped guide users to the appropriate area on our website. This work was driven by analysis of user tickets, which identified this as a significant point of confusion for users: they would often request a new account rather than request reactivation of an old account.

Release of the Polaris Status Page ("the Gronk")

ALCF is committed to providing our users with useful tools. To this end, we released the Polaris Gronk – the web-based graphical interface that displays Polaris's running and queued jobs. With Polaris Gronk, our users can easily keep track of all running jobs, queued jobs, and upcoming reservations on the Polaris system. This display is located at https://status.alcf.anl.gov/#/polaris.

File Upload/Download Capability Added to Userbase3

Work was completed to allow users to upload and download files (images, reports, etc.) to Box secure cloud management through ALCF's Userbase3 user interface. This work will allow ALCF to upload and download project reports for use in the future. The Laboratory Computing Resource Center is currently using this Userbase3 feature.

1.3.2 Application Support for Individual Projects

1.3.2.1 Identifying Reaction Coordinates for Complex Protein Dynamics

The primary goal of protein science is to understand how proteins function, which requires understanding the functional dynamics responsible for transitions between different functional structures of a protein. Using the recently developed generalized work functional (GWF), a biomedical engineering group at the University of Illinois Chicago reported the discovery of the exact reaction coordinates (RCs) for a large-scale functional process of an important protein system. The RCs were leveraged to explore the protein reaction mechanism on Theta with individual calculations typically running on 128–512 nodes for 3–8 hours. ALCF staff worked with the project PI in helping to build Gromacs code, assisting in optimal usage of the resources by coordinating the running of production jobs, and optimizing Gromacs performance by tuning simulation parameters. The team applied the GWF method to study the functional process of HIV-1 protease, the major drug target of AIDS, and identified six RCs involved in this process.

1.3.2.2 Electronic Excitation Response of DNA to High-Energy Proton Radiation in Water

Studying the mechanism of the electronic excitation response of DNA with high-energy proton radiation can help to elucidate damage during proton radiotherapy, or under extreme conditions such as those experienced by astronauts in space. The team used the Qb@ll branch of the Qbox code for real-time, time-dependent density functional theory calculations to simulate DNA surrounded by water. The simulations included 3,991 atoms and 11,172 electrons, with six different proton kinetic energies sampled for each proton path. Two proton paths were considered: one directly through the center of the DNA and another along a sugar-phosphate side chain. Batch job submissions used up to 4,096 nodes on Theta. ALCF staff assisted with switching to optimized libraries and compiling issues. ALCF also provided consulting for using GPUs and improving INQ code on Polaris. The results of the simulations showed that the damage to DNA occurs in the sugar-phosphate backbone rather than in the nucleobases.

1.3.2.3 Argonne Nexus

Experimental DOE user facilities are increasing their data collection rates and have expanded needs for experiment-time data analysis and processing. ALCF stood up a set of integrated tools and resources aimed at meeting the needs of experimental facilities, providing a comprehensive approach to integrated research infrastructure (IRI). Called Nexus, it employs several new facility capabilities: a demand queue on Polaris that will preempt running jobs to give experiments access to compute nodes for immediate computation; the Eagle file system that allows experimental facilities to share data with collaboration members that do not have ALCF accounts through Globus Guest Collections; implementations of Globus flows that integrate file transfers with application execution on Polaris and machines at other ASCR facilities; the ALCF Community Data Co-op (ACDC), a web portal for running analysis and displaying resulting data products; and in select cases, service accounts specific to an experiment apparatus that can be shared among different user-operators. In CY 2023, many Nexus components were tested in live runs with APS X-ray beamlines, where experiment-time analysis was performed using the Polaris demand queue with Globus tools. Results were displayed in the ACDC web portal for an X-ray photon correlation spectroscopy application. APS is currently down for upgrades, but in the past year, over eight APS beamlines have implemented Nexus workflows or are in the process of doing so. ALCF is also discussing with other X-ray light sources, including ALS at Berkeley Lab, NSLS at Brookhaven Lab, and LCLS-II at Stanford Linear Accelerator Center (SLAC), to extend this approach to their facility workloads. ALCF has also worked with the nuclear fusion tokamak DIII-D at General Atomics to create a Globus Flows workflow that will deploy their applications on Polaris and automatically redirect jobs to Perlmutter or Summit if Polaris is unavailable. ALCF is currently exploring other applications and data transfer strategies for the DIII-D case and plans to run live with their experiment when they come back from maintenance in late spring 2024.

1.3.3 User Engagement

1.3.3.1 General Outreach

Diversity, Equity, Inclusion, and Accessibility

Through participation in annual Argonne-sponsored outreach events, such as Introduce a Girl to Engineering Day and Science Careers in Search of Women, ALCF staff members connect with young women to introduce them to potential career paths in STEM. ALCF also promotes STEM careers to women through participating in Argonne's Women in Science and Technology program, AnitaB.org's Top Companies for Women Technologists program, and the Grace

Hopper Celebration of Women in Computing. ALCF staff also attend the Richard Tapia Celebration of Diversity in Computing Conference to recruit from a diverse set of backgrounds and ethnicities and volunteer as mentors for participants of the National Association for the Advancement of Colored People's (NAACP's) Afro-Academic, Cultural, Technological, and Scientific Olympics (ACT-SO), a year-long achievement program designed to recruit, stimulate, and encourage high academic and cultural achievement among African American high school students. In CY 2023, ALCF staff members stepped in as founding members of the Chicago chapter of Women in High-Performance Computing (WHPC), bringing together women and other underrepresented students and professionals in the Chicago area to promote equity and offer resources in the field of HPC.

Facility Tours

Visitors to Argonne can request a tour of the ALCF. ALCF visitors include student groups, congressional representatives and other government officials, industry representatives, summer research students, visiting researchers, and journalists. Various staff members tours that can include the machine room, the visualization lab, and the Aurora installation space, where guests can see Aurora being built.

Various staff members hosted more than 250 groups of visitors in 2023. Members of ALCF's leadership team welcomed District Supervisor Andrea Bailey on May 24, 2023; U.S. Representative Marcy Kaptur on September 1, 2023; and ACCESS DOE-VTO Program Manager Haiyan Croft on October 26, 2023.

User Advisory Council

ALCF's User Advisory Council (UAC) provides guidance on proposed policy changes and services and gives feedback on the experiences and needs of the user community in general. The ALCF Director appoints members who have expert knowledge of the tasks and requirements of specific applications or domain areas. In CY 2023, council member Aiichiro Nakano (University of Southern California) was reappointed, and six new members were appointed. The new members are Yongjun Choi (TAE Technologies Inc.), Olexandr Isayev (Carnegie Mellon University), Eric Johnsen (University of Michigan), Johan Larsson (University of Maryland), Ravi Madduri (Argonne National Laboratory), and Vikram Mulligan (Flatiron Institute). An announcement was made in May 2023 and an initial meeting was held on June 7, 2023, with Ravi Madduri chairing. Quarterly meetings will resume in 2024 to discuss agenda topics such as ALCF's Data Policy.

1.3.3.2 Training Activities

ALCF offers workshops and webinars on various tools, systems, and frameworks. These hands-on training programs are designed to help PIs, their project members, and future users take advantage of leadership-class computers available at ALCF and enhance the performance and productivity of their research. ALCF also collaborates with peer institutions and vendor partners to offer training that strengthens community competencies and promotes best practices. In CY 2023, ALCF conducted 28 training activities, reaching more than 1,500 participants in events where participation was tracked. Below is a list of ALCF 2023 training activities:

ALCF INCITE GPU Hackathon (hybrid)

ALCF, NVIDIA, and OpenACC organized a hybrid hackathon to help researchers prepare for the 2024 INCITE Call for Proposals. The hackathon provided an opportunity for 11 research teams to accelerate their AI or HPC research under the guidance of expert mentors to achieve performance gains and speedups using various programming models, libraries, and tools. The goal was for computational scientists to port, accelerate, and optimize their scientific applications to modern computer architectures, including GPUs. Five teams subsequently submitted 2024 INCITE proposals. (Dates: April 18, April 25, and May 2–4, 2023)

Intro to High-Performance Computing Bootcamp

The Intro to HPC Bootcamp is an immersive program designed to provide undergraduate and graduate students in STEM with hands-on experience working on energy justice projects using state-of-the-art computational and data science tools and techniques. The boot camp hosted 60 students, supported by 10 peer mentors and 20 facilitators, including staff from ALCF, OLCF, and NERSC, as well as not-for-profit and academic partners. This pilot boot camp was a 5-day experience led by ALCF staff and hosted at Lawrence Berkeley National Laboratory. (Dates: August 7–11, 2023)

ALCF Hands-on HPC Workshop

The ALCF hosted an in-person, Hands-on HPC Workshop at Argonne attended by 121 participants, with 10 from the University of Chicago and 36 other external users. The morning session covered a variety of technical talks, and afternoons were dedicated to hands-on sessions on Polaris. In attendance was a high school student from IMSA, as well as the 2023 ATPESC awardees. (Dates: October 10–12, 2023)

2023 INCITE Proposal Writing Webinars (virtual)

The INCITE program, ALCF, and the Oak Ridge Leadership Computing Facility (OLCF) jointly hosted two webinars on effective strategies for writing an INCITE proposal. (Dates: April 25, 2023, and May 2, 2023)

Argonne Training Program on Extreme-Scale Computing (ATPESC)

ATPESC, founded by Argonne and now part of the Exascale Computing Project, is an intensive, two-week program focusing on the key skills and tools needed to use supercomputers for science. The program features talks from leading computer scientists and HPC experts; and hands-on training using DOE leadership-class systems at ALCF, OLCF, and NERSC. In 2023, ATPESC attracted 79 attendees from 58 different institutions worldwide. Video recordings of ATPESC sessions are available on Argonne's YouTube training channel. (Dates: July 31–August 12, 2023)

Aurora Early Science Program (ESP) Dungeons, Hackathons, and Workshops

In 2023, the Intel Center of Excellence (COE), in collaboration with ALCF's Early Science Program (ESP), held multiday events where select ESP and ECP project teams worked on developing, porting, and profiling their codes on Sunspot with help from Intel and Argonne experts. The 2023 program included the following activities:

- May 17–18, 2023: Hackathon 20 (CFDML/PHASTA ML) (virtual)
- June 13–15, 2023: Workshop #5 (Sunspot)
- July 11–13, 18–20, 25–27 and August 8–10, 2023: Dungeon #5 (Sunspot) (virtual)

ALCF Webinars (virtual)

The 2023 ALCF webinar program consisted of two tracks: *ALCF Developer Sessions* and the *Aurora Early Adopter Series*. The ALCF Developer Sessions were focused on those writing code for Aurora. The Aurora Early Adopter Series was focused on public discussions related to Aurora. All talks are posted to ALCF's YouTube channel, and the associated training materials can be found on the ALCF Events website. ALCF also participates in useful community events, the IDEAS productivity project webinar series, and Intel webinars. The 2023 webinar program was as follows:

- January 25, 2023: Introduction to the Migrating to SYCL Portal
- February 6–10, 2023: ECP Tutorial Days
- March 29, 2023: Accelerate Python Loops with the Intel AI Analytics Toolkit
- April 24, 2023: Workflow Software for Managing Large-Scale Job Campaigns at ALCF
- May 24, 2023: A Tale of Two Apps: Preparing XGC and HACC to run on Aurora
- June 21, 2023: An Overview of Aurora's Hardware and Software
- July 19, 2023: Getting Started Boot Camp on Polaris
- August 30, 2023: MPI on Aurora and Sunspot-Examples and Best Practices
- September 27, 2023: Optimizing Workloads on Aurora and Sunspot: Examples with SYCL
- October 25, 2023: Interactive High-Performance Computing with ALCF JupyterHub

AI Testbed Workshops

ALCF offered hands-on training for all four AI testbed systems now in production and available to users. Each workshop was a two-day, hands-on opportunity for ALCF users to familiarize themselves with the systems and learn how they can advance next-generation AI for science.

- SambaNova AI Training Workshop (July 11–12, 2023)
- Cerebras AI Training Workshop (July 17–18, 2023)
- Graphcore AI Training Workshop (July 25–26, 2023)
- Groq AI Training Workshop (December 6–7, 2023)

1.3.3.3 Community Outreach

In CY 2023, ALCF hosted 8 community outreach events, reaching 2,500+ students, both in their classrooms and on-site, and engaging with over 9,000 members of the public. These activities ranged from giving tours to industry groups and DOE leadership to participating in STEM efforts and classroom visits with K-12 students. ALCF held several summer coding camps, and ALCF staff participated in computer science education events such as the Hour of Code. ALCF staff contribute to a wide range of activities aimed at sparking students' interest in scientific computing and promoting career possibilities in STEM fields, including at Argonne's Open House. Additionally, the ALCF's annual summer student program allows college students to work side-by-side with staff members on real-world research projects and to utilize some of the world's most powerful supercomputers, working in areas like computational science, system administration, and data science.

STEM Activities

2023 Argonne Open House

Argonne hosted a lab-wide Open House for the first time in 5 years, welcoming over 9,000 people from the community to visit Argonne and learn about the amazing science and impressive facilities hosted on-site. ALCF staff members led interactive activities introducing both children and adults to basic concepts in computing, as well as giving tours of Aurora and the Visualization Laboratory. Over 1,000 Argonne staff supported the event, including a large percentage of the ALCF staff. (Date: May 20, 2023)

Summer 2023 Research Internships

ALCF hosted 37 student interns through various programs, including DOE's Science Undergraduate Laboratory Internships (SULI) program, Argonne's Research Aide program, and the Community College Internship (CCI) program. During the summer, these young researchers worked closely with their mentors from ALCF and engaged in various activities in the field of scientific computing, leading up to a final presentation to the ALCF community. The student projects included using advanced edge computing, benchmarking graph neural networks for science on AI accelerators, bi-directional in-situ analysis and visualization using virtual reality, and advancing research in brain algorithms.

2023 Introduce a Girl to Engineering Day

Argonne hosted 118 eighth-grade girls for the annual Argonne Introduce a Girl to Engineering Day (IGED). Seventy-five Argonne mentors, including ALCF staff members, volunteered for an engaging day of presentations and hands-on activities focused on STEM careers. (Date: February 17, 2023)

2023 Science Careers in Search of Women

Argonne's 2023 Science Careers in Search of Women conference was held in person with the support of 80 Argonne volunteers. This annual Lab-sponsored event offered female high school students an opportunity to explore the world of STEM research through various interactions with Argonne's women scientists and engineers, including ALCF staff, with activities such as career panel discussions, career exhibits, and tours of Argonne, including Aurora and the Visualization Lab. A total of 240 girls and 40 of their teachers participated in this event. (Date: April 11, 2023)

2023 CodeGirls @ Argonne Camp (virtual)

Argonne held its seventh annual CodeGirls@Argonne summer camp, a weeklong STEM course for 6th- and 7th-grade girls taught by Argonne Learning Center and ALCF staff. Twenty-two girls learned Python coding fundamentals (prior programming experience was not a pre-requisite), experimented with robotics, and met women scientists who use code to solve problems. The group also toured the ALCF machine room and learned about the future Aurora supercomputer. (Dates: August 7–11, 2023)

2023 Coding for Science Camp

Coding for Science Camp was a five-day enrichment experience for high school freshmen and sophomores new to coding. The camp curriculum promotes problem-solving and teamwork skills through hands-on coding activities, such as coding with Python and programming a robot, and interactions with Argonne staff members working in HPC and visualization. This camp, a joint initiative of Argonne's Educational Programs Office and ALCF hosted 27 students. (Dates: July 24–28, 2023)

Intel OneAPI Academic Hands-on Workshop

The second OneAPI Academic Hands-on Workshop was hosted at the University of Illinois Chicago in a collaboration of Intel, ALCF, and academic partners from around Chicago. This event hosted 100 students from Chicago universities for a day-long workshop during which they learned the fundamentals of oneAPI, with hands-on work using Intel's Developer Cloud. (Date: October 21, 2023)

Hour of Code

During Computer Science Education Week (CSEdWeek) in December, Argonne computer scientists visited Chicagoland schools and assisted teachers in celebrating the Hour of Code, a global movement to introduce students to computer programming in a fun way. In 2023, 37 Argonne staff volunteers, including 9 from ALCF, visited 32 elementary, middle, and high schools in and around Chicago to give short tutorials and interact with students on coding activities, reaching an estimated 2,000+ students. (Dates: December 4–8, 2023)

1.3.4 Communications

1.3.4.1 Website Support Center Continuous Improvement

ALCF's online Support Center contains a wide range of resources, from onboarding guides to community announcements to video training tutorials. The Support Center is maintained by ALCF's media team and technical staff and undergoes internal content reviews, resulting in continuous improvements in the form of new web features and redesigned web pages.

In CY 2023, the team migrated the ALCF AI Testbed and Aurora documentation into ALCF user guides in GitHub. The team also developed a new video channel planned for the summer of 2024 and created videos to help users get started on Polaris and the AI Testbed systems.

1.3.4.2 Consistent Cadence of ALCF Impact on Exascale and AI Efforts

ALCF's communications team continued the two Aurora article series: *Best Practices for GPU Code Development*, which highlighted Aurora's ESP and ECP code optimization efforts, and the Aurora Software Development series, which highlighted the activities and collaborations guiding the facility and its users into the next era of scientific computing.

In CY 2023, ALCF continued to collaborate with Intel on several Aurora Early Science Program articles, "Code Together" podcasts, and Intel product launches. ALCF also worked with Groq and Graphcore to introduce GroqRack and Bow IPU to the research community.

1.3.4.3 Communicating Scientific Impact

ALCF produced science stories and articles and promoted HPC training opportunities throughout CY 2023. Furthermore, ALCF planned marketing campaigns around major annual HPC conferences and events such as the ECP Annual Meeting, International Supercomputing Conference (ISC), Exascale Day, and Supercomputing Conference (SC).

In CY 2023, ALCF placed 69 original science stories in various news outlets in coordination with Argonne's Communications & Public Affairs (CPA) Division and other ALCF direct relationships. ALCF tracked media hits through its media monitoring service, Meltwater. In CY 2023, Meltwater reported 1,208 unique ALCF media hits (a decrease of 190 from 2022), 175 of which were chronicled on the ALCF website, and an audience reach of 1.06 billion (an increase of 192.57 million). Note: Meltwater defines "reach" as estimating the potential viewership of any article based on the number of visitors to the specific source on both desktops and mobile devices.

ALCF also produced various publications that describe aspects of the facility's mission and summarize its research achievements (Table 1.6). Most of these documents are available for download on the ALCF website, and this year, the team published its digital ALCF Annual Report.

Table 1.6 Publications Designed for Print

Publication	Frequency	When
Press and Visitor Packets	As Needed	As Needed
Industry Brochure	As Needed	As Needed
Computing Resources	As Needed	As Needed
Annual Report	Yearly	March
Science Report	Yearly	October
Fact Sheets	Yearly	November
INCITE Posters	Yearly	December

1.3.4.4 Messaging for Users and Community

ALCF maintained several communication channels, including direct e-mail campaigns, scriptable e-mail messages, social media postings (Facebook, Twitter, and LinkedIn), and website postings (Table 1.7). Target audiences are identified in Table 1.8. Users could opt out of the system notification and newsletter mailing lists.

ALCF's monthly e-newsletter, *Newsbytes*, highlighted ALCF-supported research advancements, promoted training events and allocation program announcements, and linked to relevant news stories. Special announcements about certain training opportunities and fellowships were sent throughout the year as needed.

Table 1.7 Primary Communication Channels

Channel Name	Description	When Used/Updated
Newsbytes	HTML-formatted newsletter featuring science, facility news, recent news hits, and upcoming training events.	Monthly
Special Announcements	E-mail newsletter and text-format with information on conferences, training events, etc.—both ALCF and non-ALCF opportunities.	Ad hoc
Weekly Digest	Plain-text weekly rollup of events affecting ALCF systems and software, upcoming deadlines, and training opportunities.	Weekly
Social Media	Social media used to promote ALCF news and events.	Frequently
ALCF Website	An integrated information hub for user documentation, program and resources descriptions, user-centric events, feature stories about users, and related news.	Frequently
Custom E-mail Messages	Notification of machine status or facility availability, typically in a text-based format per user and channel preference.	As needed

Table 1.8 Target Audiences

Channel	Target Audience(s)
Newsbytes	Users, scientific communities, students, the public
Special Announcements	Users, scientific communities, students, the public
Weekly Digest	Current users on the systems with accounts
Social Media	Users, followers of the ALCF, collaborators, students, scientific communities, the public
ALCF Website	Users, collaborators, students, scientific communities, the public
Custom E-mail Messages	Specific projects, user groups, PIs/proxies, individual users

Conclusion

Maximizing the use of ALCF resources is at the forefront of all improvements to ALCF's customer service. In CY 2023, ALCF continued to support project teams in adapting their codes to new architectures and to prepare others to apply for major project allocations. ECP and ESP users were provided access to Aurora. ALCF continued to partner with other national laboratories and the ECP and to present work in premier scientific journals and at top professional meetings and conferences. ALCF users gained access to the AI testbed systems and GitLab-CI on Polaris. ALCF made multiple improvements to account and project management software, which helped users and administrative staff.

This page intentionally left blank.

Section 2. Operational Performance

Did the facility's operational performance meet established targets?

ALCF Response

ALCF has exceeded the metrics target for system availability. For the reportable areas, such as Mean Time to Interrupt (MTTI), Mean Time to Failure (MTTF), and system utilization, ALCF is on par with the other DOE facilities and has demonstrated exceptional performance. To assist in meeting these objectives and improving overall operations, ALCF tracks hardware and software failures and analyzes their impact on user jobs and metrics as a significant part of its improvement efforts.

Tables 2.1 and 2.2 summarize all operational performance metrics of HPC computational and storage systems reported in this section.

Table 2.1 Summary of All Metrics Reported in the Operational Performance Section

	Theta with expansion Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 Theta expansion (NVIDIA DGX): 24-node 24 TB of DDR4 RAM 7.68 TB of GPU memory				Polaris (HPE Apollo 6500 Gen 10+) 560-node 17920-core 2240 Nvidia A100 GPU 287 TB DDR4 90 TB HBM2			
	CY 2022		CY 2023		CY 2022		CY 2023	
	Target	Actual	Target	Actual	Target	Actual	Target	Actual
Scheduled Availability	90.0%	98.6%	90.0%	99.4%	90.0%	98.2%	90.0%	99.2%
Overall Availability	90.0%	95.7%	90.0%	93.9%	90.0%	94.7%	90.0%	93.8%
System MTTI	N/A [e]	10.62 days	N/A	13.20 days	N/A	9.86 days	N/A	13.72 days
System MTTF	90.0%	32.87 days	N/A	72.67 days	N/A	47.78 days	N/A	91.03 days
Expansion System MTTI	N/A	12.53 days	N/A	14.31 days	N/A	N/A	N/A	N/A
Expansion System MTTF	N/A	60.39 days	N/A	121.27 days	N/A	N/A	N/A	N/A
INCITE Usage	17.8M	20.4M	17.8M	20.4M	1322.0K	589.6K	2000.0K	1764.3K
Total Usage	N/A	33.2M	N/A	31.6M	N/A	1163.6K	N/A	3680.7K
System Utilization	N/A	97.0%	N/A	89.8%	N/A	67.3%	N/A	81.1%
INCITE Overall Capability [a,c]	40.0%	85.8%	40.0%	88.3%	20.0%	47.7%	20.0%	56.7%
INCITE High Capability [b,d]	N/A	27.2%	N/A	17.8%	N/A	22.3%	N/A	13.7%

[a] Theta with expansion Overall Capability = Jobs using ≥ 20.0 percent (800 nodes).

[b] Theta with expansion High Capability = Jobs using ≥ 60.0 percent (2400 nodes).

[c] Polaris Overall Capability = Jobs using ≥ 20.0 percent (99 nodes).

[d] Polaris High Capability = Jobs using ≥ 60.0 percent (297 nodes).

[e] N/A = not applicable.

Table 2.2 Summary of Operational Performance of HPC Storage Systems

Theta-fs0 File System Cray Sonexion 2000 with 9.2 PB of usable storage				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
Scheduled Availability	90.0%	99.3%	90.0%	99.9%
Overall Availability	90.0%	96.3%	90.0%	94.2%
System MTTI	N/A	13.52 days	N/A	14.94 days
System MTTF	N/A	72.51 days	N/A	182.30 days
Grand File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
Scheduled Availability	N/A	99.5%	90.0%	99.5%
Overall Availability	N/A	96.6%	90.0%	93.2%
System MTTI	N/A	14.68 days	N/A	13.08 days
System MTTF	N/A	121.13 days	N/A	72.68 days
Eagle File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
Scheduled Availability	N/A	99.5%	90.0%	99.9%
Overall Availability	N/A	96.6%	90.0%	94.2%
System MTTI	N/A	14.68 days	N/A	14.94 days
System MTTF	N/A	121.13 days	N/A	182.30 days
HPSS Archive LTO8 tape drives and tape with 305 PB of storage capacity				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
Scheduled Availability	95.0%	99.7%	95.0%	99.6%
Overall Availability	90.0%	96.7%	90.0%	93.8%
System MTTI	N/A	14.70 days	N/A	12.69 days
System MTTF	N/A	121.28 days	N/A	60.58 days

[a] N/A = not applicable.

2.1 ALCF Production Resources Overview

During CY 2023, the ALCF operated several production resources.

- Polaris is a 560-node, ~18K core, HPE Apollo 6500 Gen 10+ with 287 TB of RAM and 2240 NVIDIA A100 GPUs. It went into production on August 9, 2022, and supported ALCC and INCITE campaigns all year.
- Theta is a 4,392-node, ~281K core, 11.69-PF Cray XC40 with 892 TB of RAM. Theta supported both ALCC and INCITE campaigns all year.
- Theta's expansion consisting of 24 NVIDIA DGX A100-based nodes. The DGX A100 comprises eight NVIDIA A100 GPUs with AMD EPYC 7742 CPUs and a total of 24 TB of DDR4 RAM and 7.68 TB of GPU memory.
- Theta-fs0 is a Sonexion 3000 Lustre filesystem mounted by Theta with 9.2 PB of usable space.
- Grand and Eagle are each 100PB Lustre filesystems and are mounted facility-wide. The only difference between them is that Eagle has Globus sharing enabled and is used as a community file system.
- The facility-wide HPSS (high-performance storage system) tape archive is comprised of three 10,000-slot libraries with LTO8 drives and tapes and some legacy LTO6 drives and tapes. It has a maximum storage capacity of 305 PB.

2.2 Definitions

- *Overall availability* is the percentage of time a system is available to users. Outage time reflects both scheduled and unscheduled outages. For HPC facilities, scheduled availability is the percentage of time a designated level of resource is available to users, excluding scheduled downtime for maintenance and upgrades. To be considered a scheduled outage, the user community must be notified of the need for a maintenance event no less than 24 hours in advance of the outage (emergency fixes). 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 seven calendar days prior. If regularly scheduled maintenance is not needed, 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 timeframe is categorized as an unscheduled outage. A significant event that delays a return to scheduled production will be counted as an adjacent unscheduled outage if the return to service is four or more hours later than the scheduled end time. For storage resources the system is considered available if any user can read and write any portion of the disk space. The availability metric provides measures that are indicative of the stability of the systems and the quality of the maintenance procedures.
- *MTTI, mean time to interrupt,* is the time, on average, to any outage on the system, whether unscheduled or scheduled. Also known as MTBI (Mean Time Between Interrupt).

- *MTTF, mean time to failure,* is the time, on average, to an unscheduled outage of the system.
- *Usage* is defined as resources consumed in units of node-hours.
- *Utilization* is the percentage of the available node-hours used (i.e., a measure of how busy the system was when it was available).
- *Total System Utilization* is the percent of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user group, including staff and vendors.

2.3 Polaris

2.3.1 Scheduled and Overall Availability

Polaris entered full production on August 9, 2022. In consultation with the DOE Program Manager, ALCF has agreed to a target of 90 percent overall availability and a target of 90 percent scheduled availability. (ASCR requested that all user facilities use a target of 90 percent for scheduled availability for the lifetime of a production resource.) Table 2.3 summarizes the availability results for Polaris.

Table 2.3 Availability Results

Polaris (HPE Apollo 6500 Gen 10+) 560-node 17920-core 2240 Nvidia A100 GPU 287 TB DDR4 90 TB HBM2				
	CY 2022		CY 2023	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	90.0	98.2%	90.0	99.2%
Overall Availability	90.0	94.7%	90.0	93.8%

The remainder of this section covers significant availability losses, and responses to them, for both scheduled and overall availability data. Details on the calculations can be found in Appendix A.

2.3.1.1 Explanation of Significant Availability Losses

This section briefly describes the causes of major availability losses from January 1, 2023, through December 31, 2023, as shown in Figure 2.1.

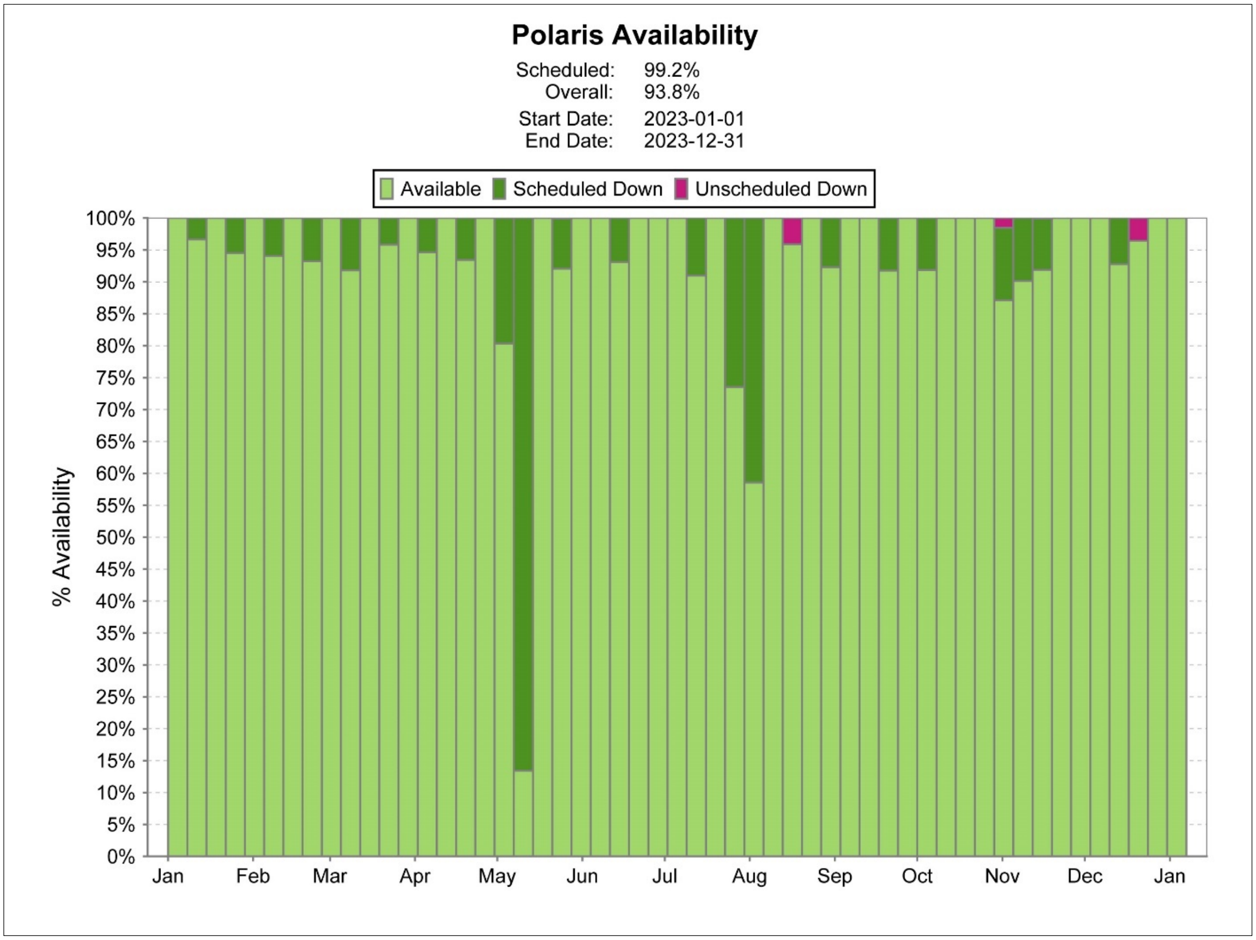


Figure 2.1 Polaris Weekly Availability for CY 2023

Graph Description: Each bar in Figure 2.1 represents the percentage of the machine available for seven days. Each bar accounts for all the time in one of three categories. The pale green portion represents available node-hours, the darker green represents scheduled downtime for that week, and the red represents unscheduled downtime. Significant loss events are described in detail below.

May 5, 2023: Scheduled outage – Data center cooling system upgrade
The data center had a 10-day complete cooling outage and a partial power outage due to a cooling upgrade project for Aurora.

July 28, 2023: Scheduled outage – Data center power outage
As part of the effort to stand up Aurora, the data center experienced a 4-day power outage.

August 14, 2023: Unscheduled outage – User notice not sent
ALCF missed sending out a scheduled outage notice within 24 hours of the outage, which converted it to an unscheduled outage.

October 30, November 6, 2023: Scheduled outages for Slingshot 11 Upgrade
Two scheduled outages were taken for network maintenance, upgrading to Slingshot 11 from Slingshot 10.

November 2, 2023: Unscheduled outage – Slingshot 11 Bandwidth issue
Problems with Slingshot 11 caused substantial bandwidth issues and an outage was required to apply a fix.

December 12, 2023: Unscheduled outage – Network issue
Network ports went into an unconfigured state due to cable swaps. This created a cascading effect as links reset. An unscheduled outage was needed to reset and reboot all slingshot switches and compute nodes.

2.3.2 System Mean Time to Interrupt (MTTI) and System Mean Time to Failure (MTTF)

2.3.2.1 MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.4 summarizes Polaris's current MTTI and MTTF values, respectively.

Table 2.4 MTTI and MTTF Results

Polaris (HPE Apollo 6500 Gen 10+) 560-node 17920-core 2240 Nvidia A100 GPU 287 TB DDR4 90 TB HBM2				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System MTTI	N/A	9.86 days	N/A	13.72 days
System MTTF	N/A	47.78 days	N/A	91.03 days

[a] N/A = not applicable.

Polaris currently has a biweekly maintenance outage scheduled to perform upgrades, hardware replacements, OS upgrades, etc. ALCF uses these preventative maintenance (PM) opportunities to schedule other potentially disruptive maintenance such as facility power and cooling work and storage system upgrades and patching. Although Polaris's biweekly maintenance schedule does not directly affect MTTF, it tends to cap MTTI at 14 days.

2.3.3 Resource Utilization

The following sections discuss system allocation and usage, system utilization percentage, and capability usage.

2.3.3.1 System Utilization

System utilization is a reportable value with no specific target. A rate of 80 percent or higher is generally considered acceptable for a leadership-class system. Table 2.5 summarizes ALCF system utilization results, and Figure 2.2 shows system utilization over time by program.

Table 2.5 System Utilization Results

Polaris (HPE Apollo 6500 Gen 10+) 560-node 17920-core 2240 Nvidia A100 GPU 287 TB DDR4 90 TB HBM2				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System Utilization	N/A	67.3%	N/A	81.1%

[a] N/A = not applicable.

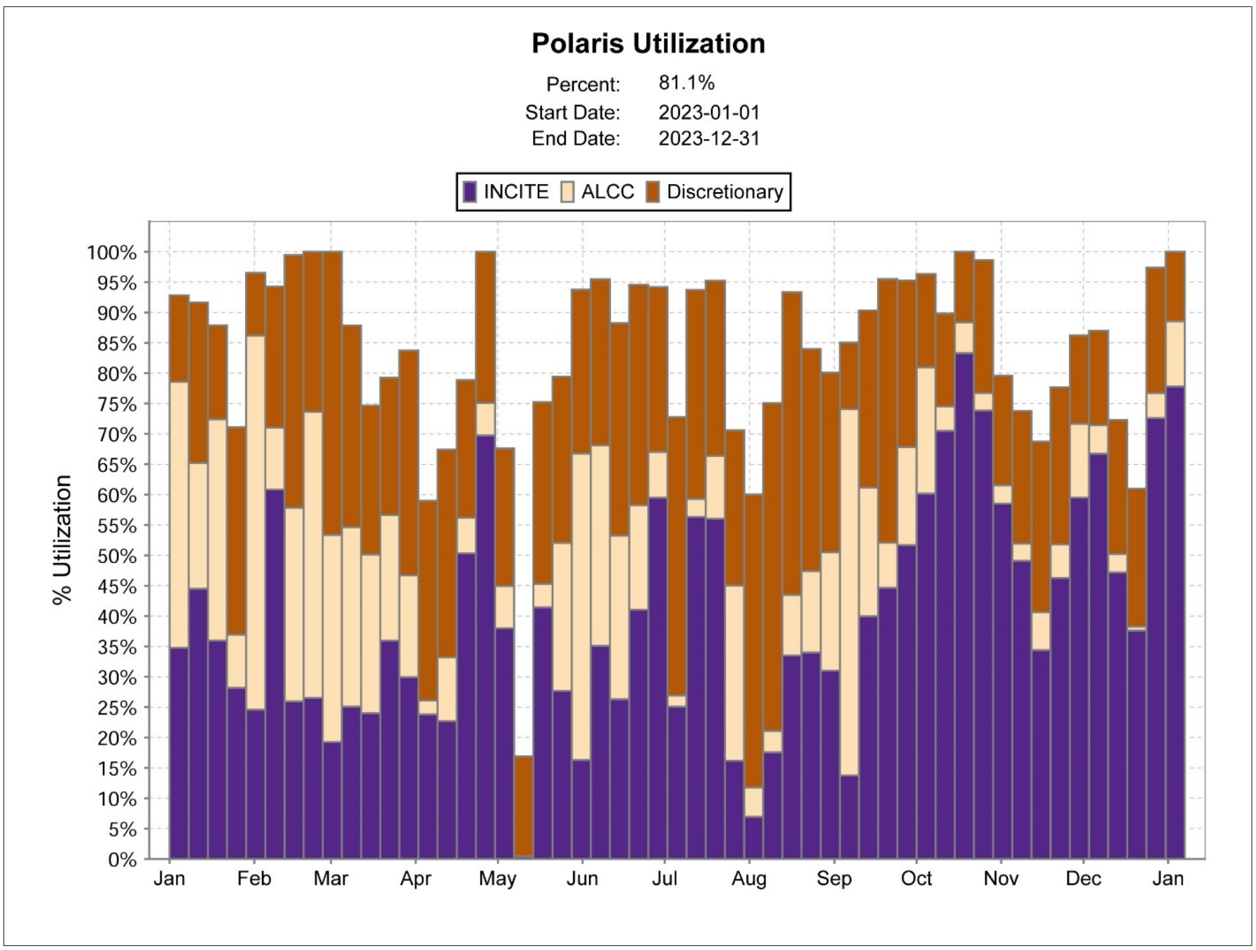


Figure 2.2 Polaris System Utilization over Time by Program

The system utilization for Polaris was 81.1 percent for its 2023 production period of January 1, 2023, through December 31, 2023. Polaris's utilization was higher in CY 2023 than in CY 2022.

Table 2.6 shows how Polaris's system hours were allocated and used by the program. Multiplying the theoretical hours by availability and system utilization values agreed upon with ALCF's DOE Program Manager determines the available hours. Of the hours available, 60 percent were allocated to the INCITE program, up to 20 percent to the ALCC program, 10 percent to the ECP program, and 10 percent to the DD program. The allocated values for the DD allocations appear higher than expected because they represent a rolling allocation. Most DD projects are exploratory investigations, so the time allocations are often not used in full. DD allocations are discussed in detail in Section 3.2. In CY 2023, Polaris delivered a total of 3,680,653 node-hours.

Table 2.6 Node-Hours Allocated and Used by Program

Polaris (HPE Apollo 6500 Gen 10+) 560-node 17920-core 2240 Nvidia A100 GPU 287 TB DDR4 90 TB HBM2						
	CY 2022			CY 2023		
	Allocated	Used		Allocated	Used	
	Node-hours	Node-hours	%	Node-hours	Node-hours	%
INCITE	1322.0K	589.6K	50.7%	2000.0K	1764.3K	47.9%
ALCC	355.8K	92.6K	8.0%	943.4K	724.3K	19.7%
DD	543.6K	481.5K	41.3%	2895.8K	1192.0K	32.4%
Total	2221.4K	1163.6K	100.0%	5839.2K	3680.7K	100.0%

Summary: For CY 2023, the system usage for INCITE was below expectations, while ALCC and DD were in line with expectations. System utilization values were in line with expectations. More details on the usage of INCITE, ALCC, and DD can be found in sections 3.1 and 3.2. The calculations for system utilization are described in Appendix A.

2.3.3.2 Capability Utilization

Polaris has a total of 560 nodes of which 56 were purchased as extra nodes and 8 are debug nodes. Therefore, ALCF used 496 nodes to calculate capability. On Polaris, capability is defined as using greater than 20 percent of the 496 nodes, or 99 nodes, and high capability is defined as using greater than 60 percent of the 496 nodes, or 297 nodes. See Table A.6 in Appendix A for more detail on the capability calculation. Table 2.7 and Figure 2.3 show that ALCF has substantially exceeded 20 percent capability on Polaris for all allocation programs. Figure 2.4 shows the three programs' utilization of total node hours (from Table 2.7) over time, and Figure 2.5 shows the overall distribution of job sizes over time.

Table 2.7 Capability Results

Polaris (HPE Apollo 6500 Gen 10+) 560-node 17.9K-core 2240 Nvidia A100 GPU 287 TB DDR4 90 TB HBM2						
	CY 2022			CY 2023		
Capability Usage	**Total Hours**	**Capability Hours**	**Percent Capability**	**Total Hours**	**Capability Hours**	**Percent Capability**
INCITE Overall [a]	589.6K	281.1K	47.7%	1764.3K	999.6K	56.7%
INCITE High [b]	589.6K	131.5K	22.3%	1764.3K	241.4K	13.7%
ALCC Overall	92.6K	34.7K	37.5%	724.3K	443.6K	61.2%
ALCC High	92.6K	3.6K	3.9%	724.3K	275.3K	38.0%
Director's Discretionary Overall	481.5K	200.3K	41.6%	1192.0K	493.1K	41.4%
Director's Discretionary High	481.5K	17.3K	3.6%	1192.0K	118.2K	9.9%
TOTAL Overall	1163.73K	516.1K	44.3%	3680.7K	1936.3K	52.6%
TOTAL High	1163.7K	152.4K	13.1%	3680.7K	634.9K	17.2%

[a] Polaris Overall Capability = Jobs using ≥ 20.0 percent (99 nodes).

[b] Polaris High Capability = Jobs using ≥ 60.0 percent (297 nodes).

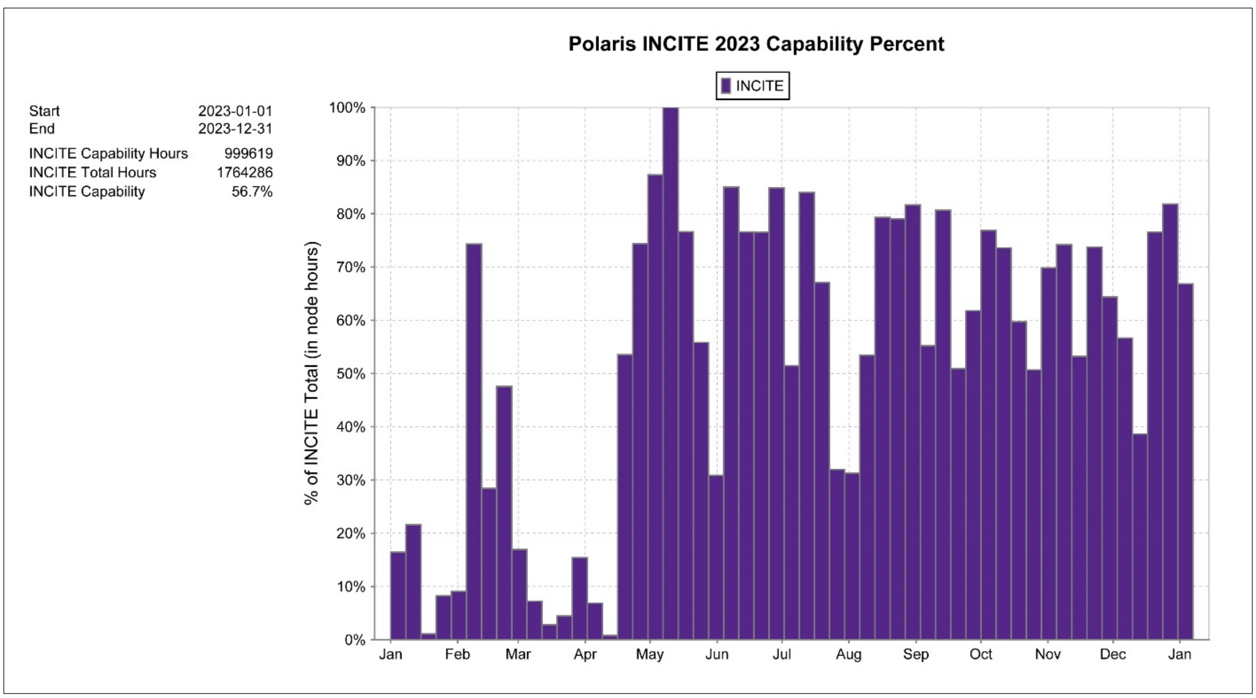


Figure 2.3 Polaris INCITE Overall Capability

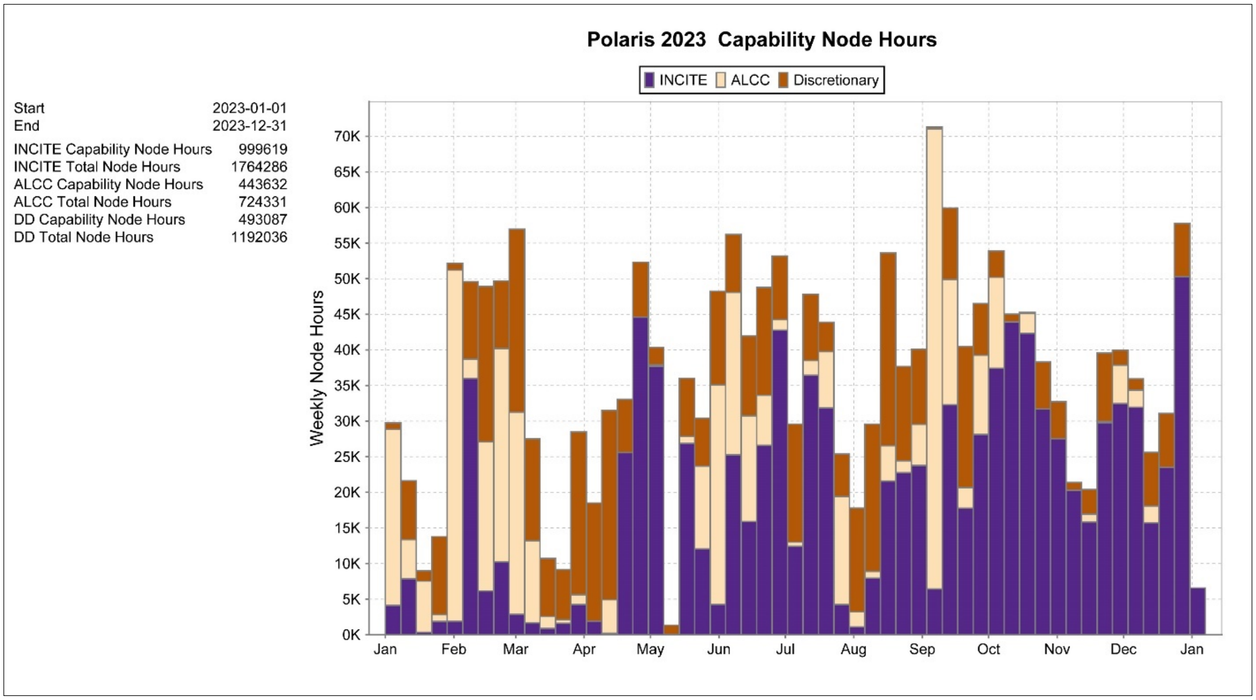


Figure 2.4 Polaris Capability Node-Hours by Program

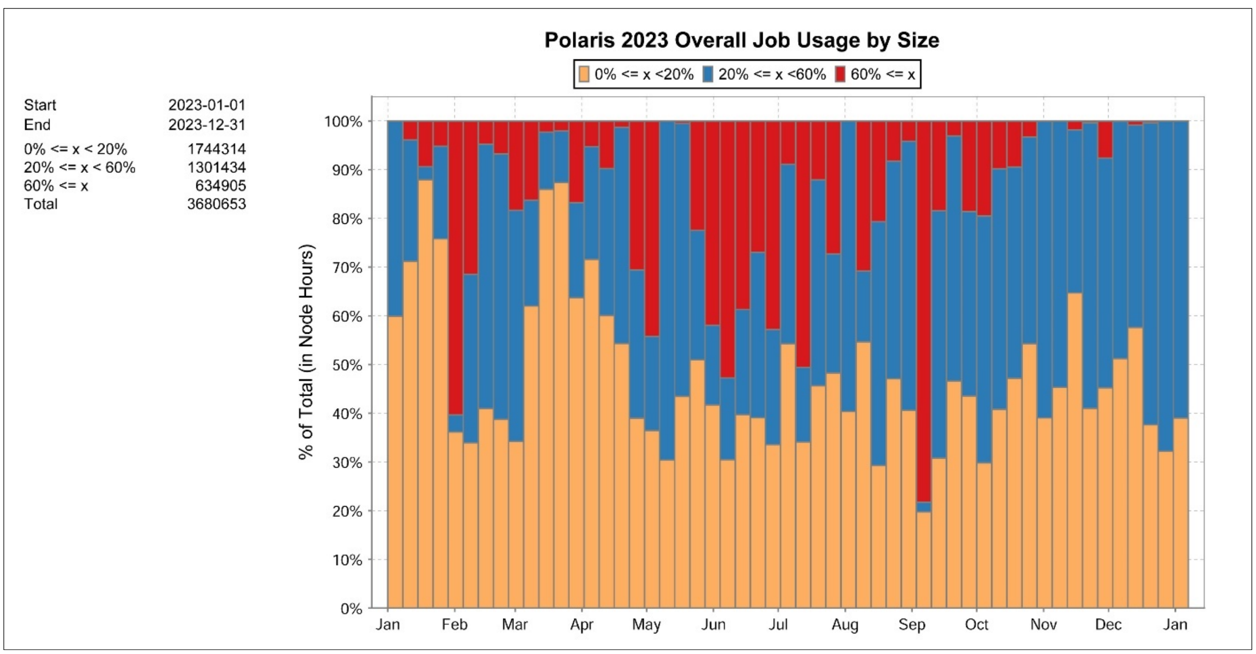


Figure 2.5 Polaris Job Usage by Size

2.4 Theta

2.4.1 Scheduled and Overall Availability

Theta entered production on July 1, 2017, and ended production on December 31, 2023. The GPU expansion to Theta entered production on January 1, 2022, and ended production on December 31, 2023. In consultation with the DOE Program Manager, ALCF agreed to a target of 90 percent overall availability and a target of 90 percent scheduled availability (ASCR requested that all user facilities use a target of 90 percent for scheduled availability for the lifetime of the production resources). Table 2.8 summarizes the availability results for Theta with expansion.

Table 2.8 Availability Results

Theta with expansion Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 Theta expansion (NVIDIA DGX): 24-node 24 TB of DDR4 RAM 7.68 TB of GPU memory				
	CY 2022		CY 2023	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	90.0	98.6	90.0	99.4
Overall Availability	90.0	95.7	90.0	93.9

The remainder of this section covers significant availability losses and responses to them for both scheduled and overall availability data. Details on the calculations can be found in Appendix A.

2.4.1.1 Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability for the period January 1, 2023, through December 31, 2023, as shown in Figure 2.6.

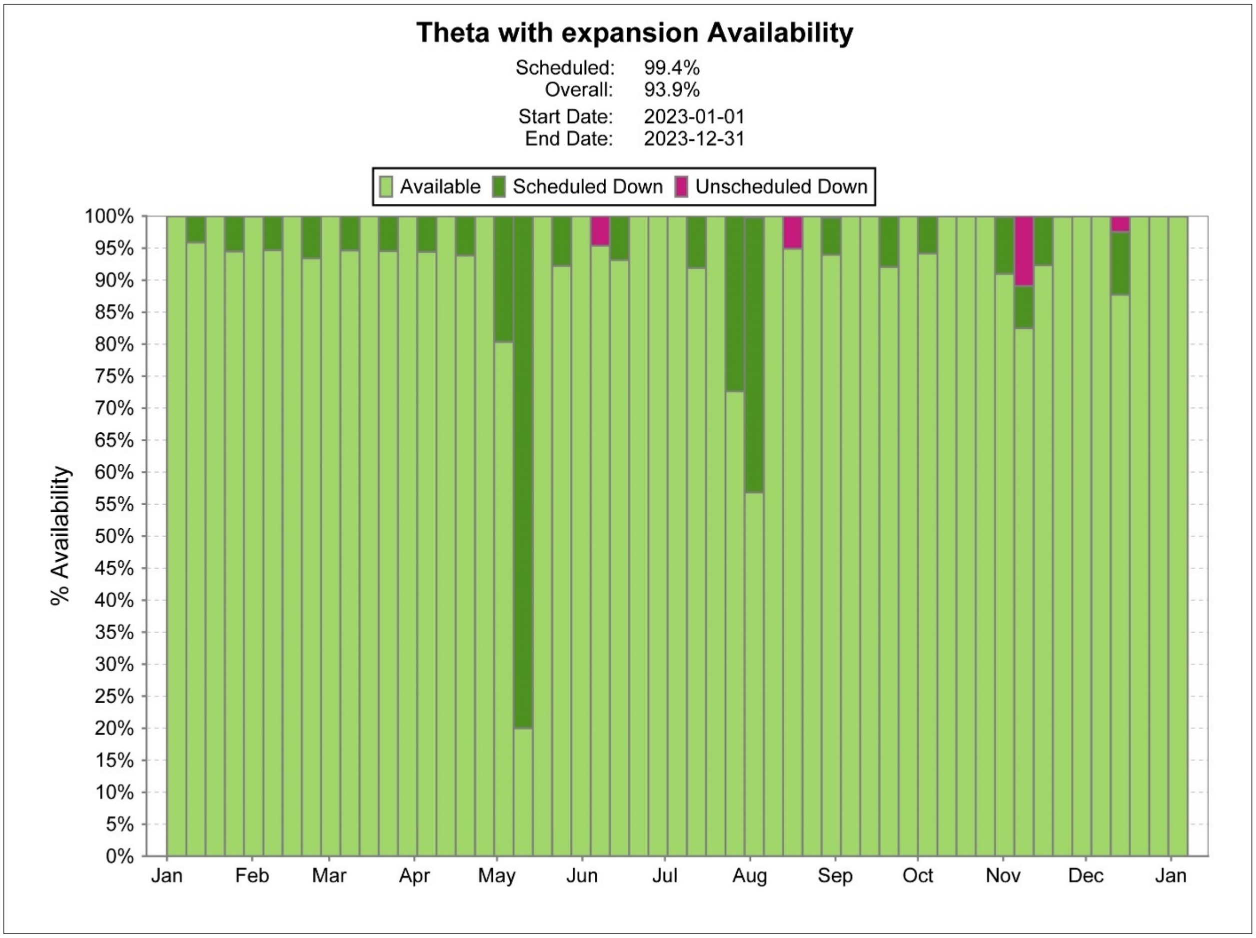


Figure 2.6 Theta with expansion Weekly Availability for CY 2023

Graph Description: Each bar in Figure 2.6 represents the percentage of the machine available for seven days. Each bar accounts for all the time in one of three categories. The pale green portion represents available node-hours, the darker green represents scheduled downtime for that week, and the red represents unscheduled downtime. Significant loss events are described in detail below.

May 5, 2023: Scheduled outage – Data center cooling system upgrade
The data center had a 10-day complete cooling outage and a partial power outage due to a cooling upgrade project for Aurora.

June 5, 2023: Unscheduled outage – Theta: maintenance reservation
A preventative maintenance reservation was not deleted for a canceled maintenance period. The reservation caused the scheduler to stop all jobs during the reservation time period. The scheduler resumed scheduling jobs when the PM reservation was canceled.

July 28, 2023: Scheduled outage – Data center power outage
As part of the effort to stand up Aurora, the data center experienced a 4-day power outage.

August 14, 2023: Unscheduled outage – Theta, Theta Expansion: User notice not sent
ALCF missed sending out a scheduled outage notice within 24 hours of the outage, which converted it to an unscheduled outage.

October 30, November 6, 2023: Scheduled outages for Slingshot 11 Upgrade
Two scheduled outages were taken for network maintenance, upgrading to Slingshot 11 from Slingshot 10.

November 7, 2023: Unscheduled outage – Theta: node errors after maintenance
During the November 6, 2023, scheduled preventative maintenance window, Theta encountered unusually slow node start-up times, timeouts, and mount errors, which affected the scheduler. This was resolved with a system restart.

December 13, 2023: Unscheduled outage – Theta: add Eagle file system
Due to the GNI system being used, a reboot was necessary to add the Eagle file system to this platform. The network architecture mandates that any significant changes made to the network configuration, such as Lustre additions or a substantial increase in the number of nodes connected, require retraining and rebooting.

2.4.2 System Mean Time to Interrupt (MTTI) and System Mean Time to Failure (MTTF)

2.4.2.1 MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.9 summarizes the current MTTI and MTTF values, respectively, for Theta and Theta expansion.

Table 2.9 MTTI and MTTF Results

Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System MTTI	N/A [a]	10.62 days	N/A	13.20 days
System MTTF	N/A	32.87 days	N/A	72.67 days

Theta expansion (NVIDIA DGX): 24-node 24 TB of DDR4 RAM 7.68 TB of GPU memory				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System MTTI	N/A	12.53 days	N/A	14.31 days
System MTTF	N/A	60.39 days	N/A	121.27 days

[a] N/A = not applicable.

Theta currently functions on a biweekly maintenance schedule to perform Cray driver upgrades, hardware replacements, OS upgrades, etc. ALCF uses these preventative maintenance (PM) opportunities to schedule other potentially disruptive maintenance such as facility power and cooling work and storage system upgrades and patching. Although Theta's biweekly maintenance schedule does not directly affect MTTF, it tends to cap MTTI at 14 days.

2.4.3 Resource Utilization

The following sections discuss system allocation and usage, system utilization percentage, and capability usage.

2.4.3.1 System Utilization

System utilization is a reportable value with no specific target. A rate of 80 percent or higher is generally considered acceptable for a leadership-class system. Table 2.10 summarizes ALCF system utilization results, and Figure 2.7 shows system utilization over time by program.

Table 2.10 System Utilization Results

Theta with expansion Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 Theta expansion (NVIDIA DGX): 24-node 24 TB of DDR4 RAM 7.68 TB of GPU memory				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System Utilization	N/A [a]	98.1%	N/A	89.8%

[a] N/A = not applicable.

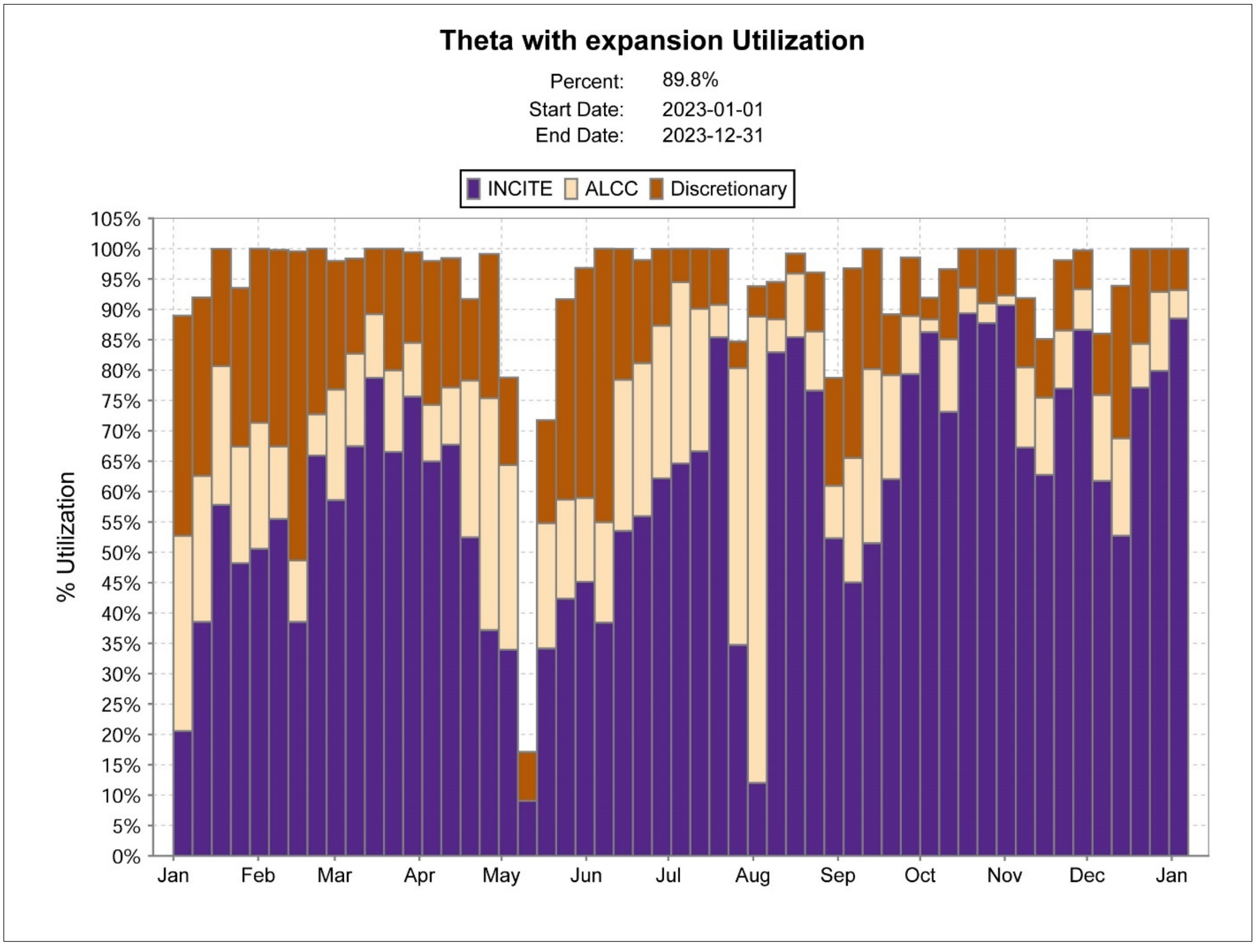


Figure 2.7 System Utilization over Time by Program

The system utilization for Theta with expansion was 89.8 percent for its 2023 production period of January 1, 2023, through December 31, 2023.

Table 2.11 shows how Theta with expansion's system hours were allocated and used by allocation source. Multiplying the theoretical hours by availability and system utilization values that were agreed upon with ALCF's DOE Program Manager determines the hours available. Of the hours available, 60 percent were allocated to the INCITE program, up to 20 percent to ALCC program allocations, 10 percent to the ECP program, and 10 percent to DD allocations. The allocated values for the DD allocations appear higher than expected because they represent a rolling allocation. Most DD projects are exploratory investigations, so the time allocations are often not used in full. DD allocations are discussed in detail in Section 3.2. In CY 2023, Theta delivered a total of 31,620,254 node-hours.

Table 2.11 Node-Hours Allocated and Used by Program

Theta with expansion Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 Theta expansion (NVIDIA DGX): 24-node 24 TB of DDR4 RAM 7.68 TB of GPU memory						
	CY 2022			CY 2023		
	Allocated	Used		Allocated	Used	
	Node-hours	Node-hours	%	Node-hours	Node-hours	%
INCITE	17.8M	20.4M	61.6%	17.8M [a]	20.4M	64.5%
ALCC	6.4M	7.4M	22.4%	5.7M [a]	5.4M	17.2%
DD	7.1M	5.3M	16.0%	15.7M	5.8M	18.3%
Total	31.3M	33.2M	100.0%	39.3M	31.6M	100.0%

[a] There were no INCITE or ALCC award allocations on Theta with expansion in 2023.

Summary: For CY 2023, the system usage and system utilization values were in line with expectations. The calculations for system utilization are described in Appendix A.

2.4.3.2 Capability Utilization

On Theta, capability is defined as using greater than 20 percent of the machine, or 800 nodes, and high capability is defined as using greater than 60 percent of the machine, or 2,400 nodes. See Table A.6 in Appendix A for more detail on the capability calculation. Table 2.12 and Figure 2.8 show that ALCF has substantially exceeded these metrics set for INCITE. Although no targets are set, data are also provided in the table for ALCC and DD projects as reference. Figure 2.9 shows the three programs' utilization of total core hours (from Table 2.12) over time, and Figure 2.10 shows the overall distribution of job sizes over time.

Table 2.12 Capability Results

Theta with expansion Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 Theta expansion (NVIDIA DGX): 24-node 24 TB of DDR4 RAM 7.68 TB of GPU memory						
	CY 2022			CY 2023		
Capability Usage	**Total Hours**	**Capability Hours**	**Percent Capability**	**Total Hours**	**Capability Hours**	**Percent Capability**
INCITE Overall [a]	20.4M	17.5M	85.8%	20.4M	18.0M	88.3%
INCITE High [b]	20.4M	5.6M	27.2%	20.4M	3.6M	17.8%
ALCC Overall	7.4M	4.6M	61.7%	5.4M	3.5M	64.4%
ALCC High	7.4M	0.5M	6.6%	5.4M	0.0M	0.1%
Director's Discretionary Overall	5.3M	0.8M	15.5%	5.8M	2.0M	33.9%
Director's Discretionary High	5.3M	0.1M	1.8%	5.8M	0.4M	7.5%
TOTAL Overall	33.2M	22.9M	69.1%	31.6M	23.5M	74.2%
TOTAL High	33.2M	6.1M	18.5%	31.6M	4.1M	12.8%

[a] Theta with expansion Overall Capability = Jobs using ≥ 20.0 percent (800 nodes) of Theta.

[b] Theta with expansion High Capability = Jobs using ≥ 60.0 percent (2400 nodes) of Theta.

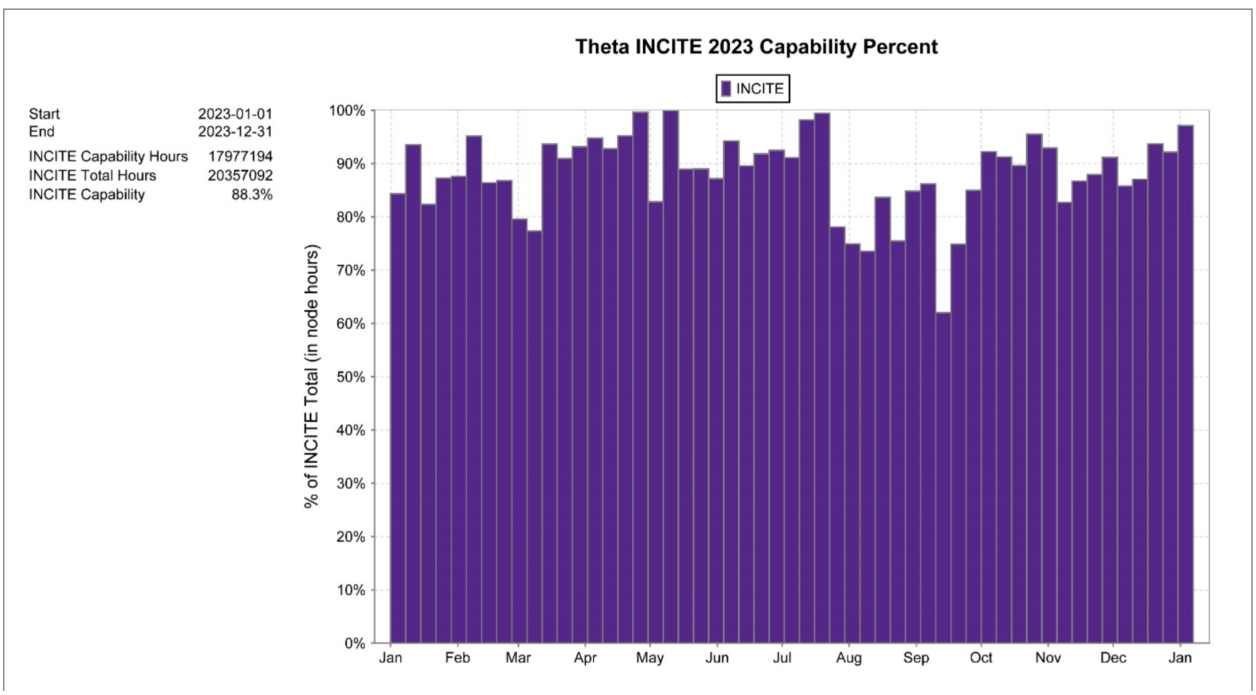


Figure 2.8 Theta INCITE Overall Capability

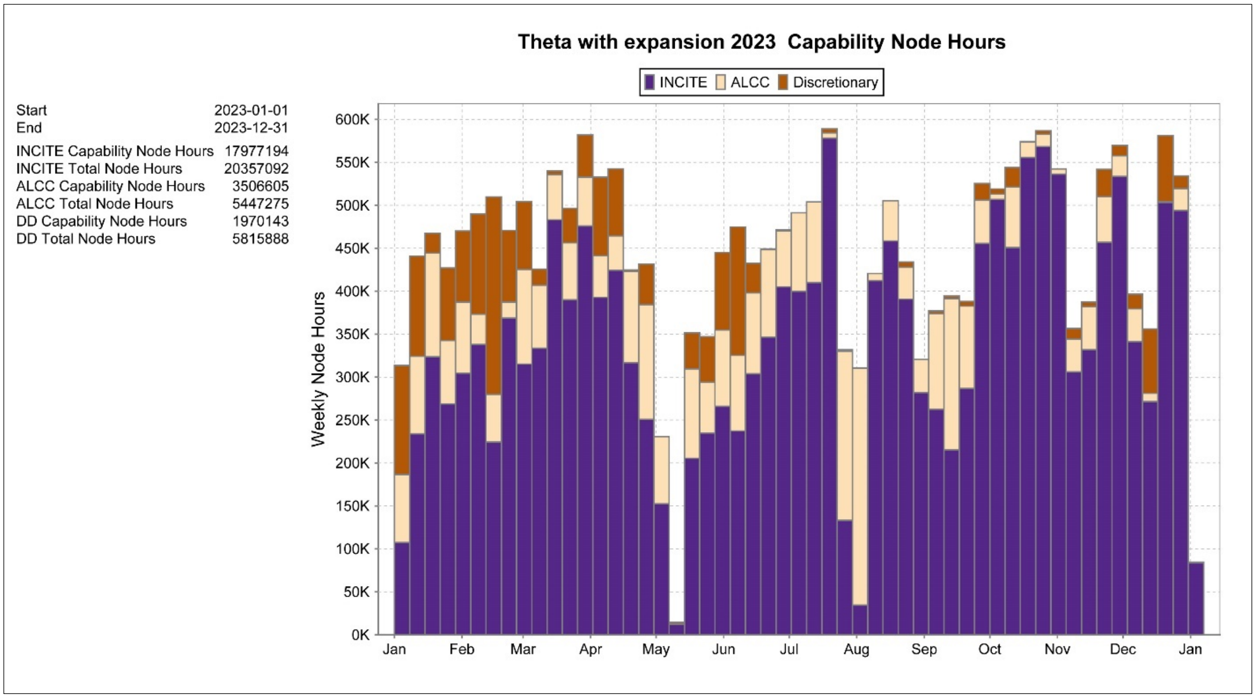


Figure 2.9 Theta with Expansion Capability Node-Hours by Program

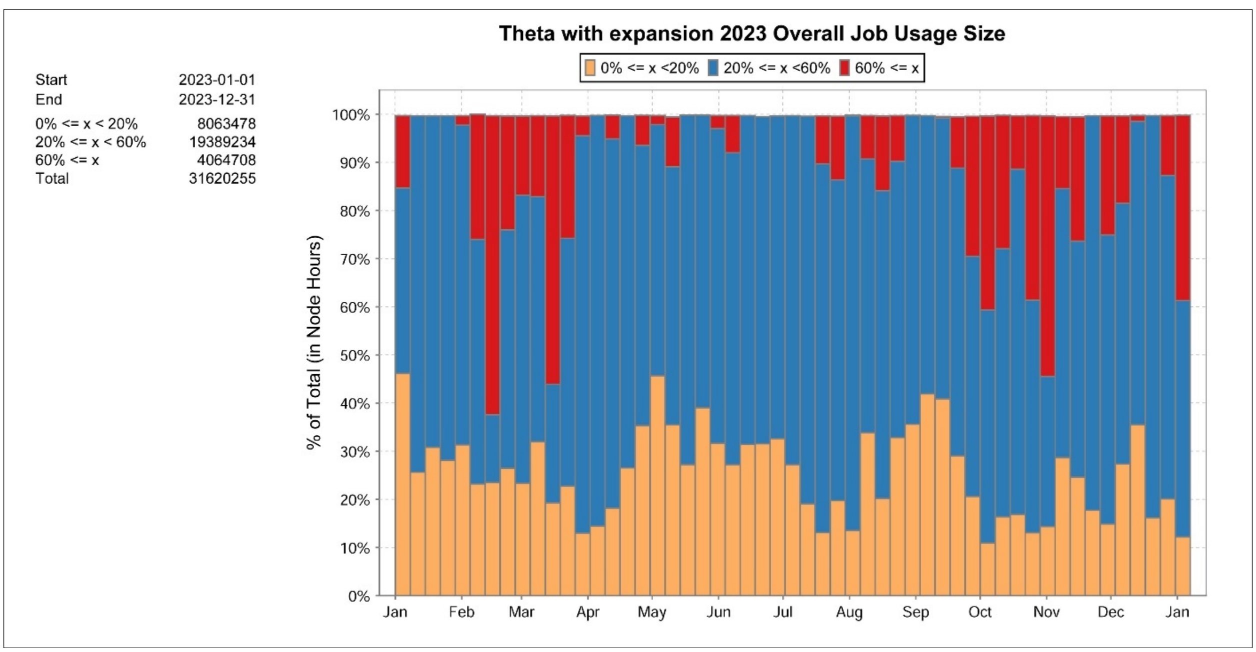


Figure 2.10 Theta with Expansion Job Usage by Size

2.5 Storage

This section covers availability and MTTI/F metrics for the production storage resources.

2.5.1 Theta-fs0 Lustre File System

Theta-fs0, is a Cray Sonexion 2000 Lustre file system that is mounted by Theta and its expansion with 9.2 PB of usable space. Theta and Theta-fs0 were installed together and entered production on July 1, 2017. Theta-fs0 and Theta were not previously treated as separate entities.

2.5.1.1 Scheduled and Overall Availability

ALCF uses the target metrics of 90 percent overall availability and 95 percent scheduled availability, as proposed in the CY 2022 OAR. This follows ASCR's request that all user facilities use a target of 90 percent for scheduled availability for the lifetime of the production resources. Table 2.13 summarizes the availability results.

Table 2.13 Availability Results

Theta-fs0 File System Cray Sonexion 2000 with 9.2 PB of usable storage				
	CY 2022		CY 2023	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A	99.3	N/A	99.9
Overall Availability	N/A	96.3	N/A	94.2

[a] N/A = not applicable.

The remainder of this section covers significant availability losses and responses to them for both scheduled and overall availability data. Details on the calculations can be found in Appendix A.

Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability of the Sonexion for the period of January 1, 2023, through December 31, 2023, as annotated in Figure 2.11.

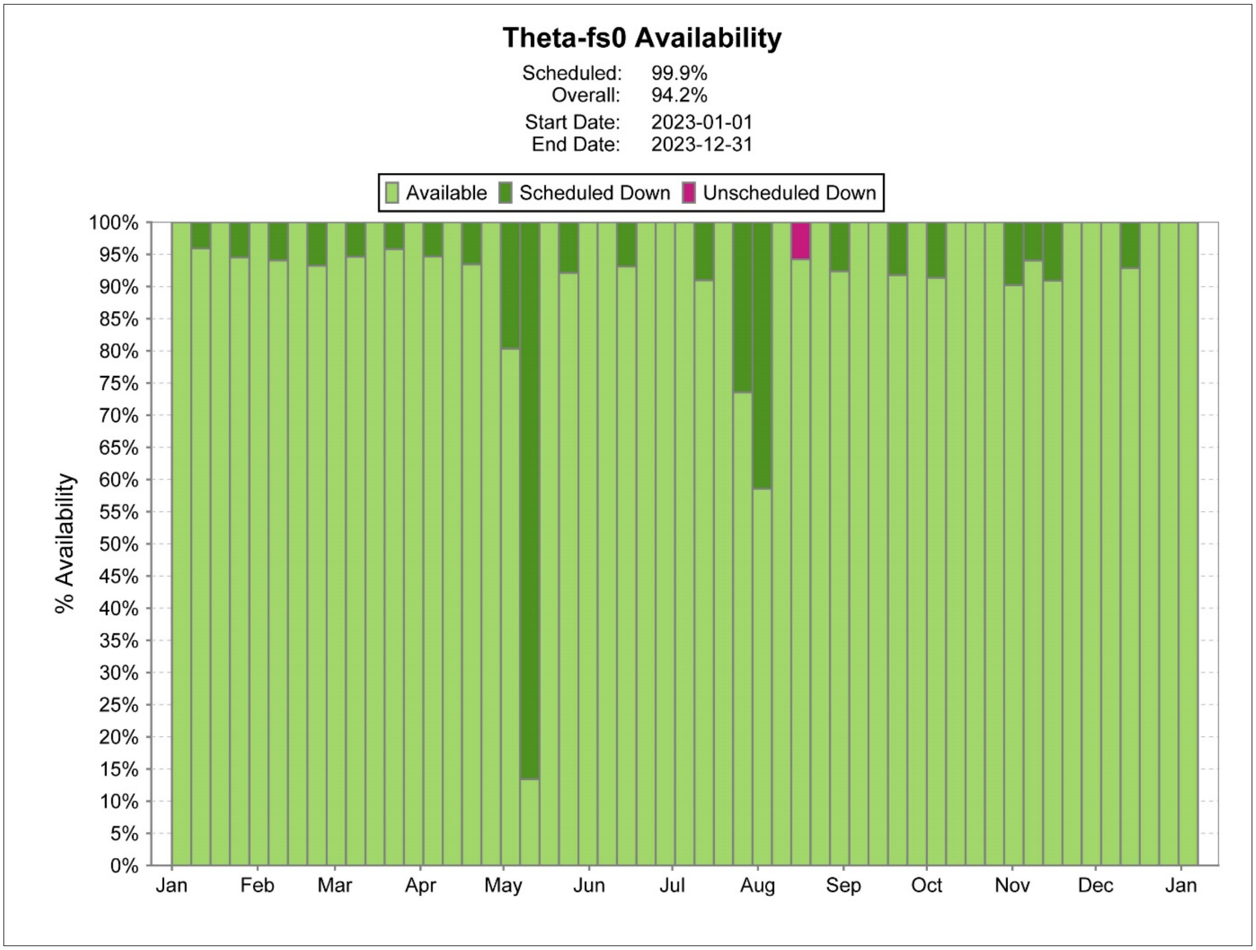


Figure 2.11 Theta-fs0 Weekly Availability for CY 2023

Graph Description: Each bar in Figure 2.11 represents the percentage of the machine available for seven days. Each bar accounts for all the time in one of three categories. The pale green portion represents available node hours, the darker green represents scheduled downtime for that week, and the red represents unscheduled downtime. Each of the significant loss events is described in detail below (these also appeared in the Theta section).

May 5, 2023: Scheduled outage – Data center cooling system upgrade
The data center had a 10-day complete cooling outage and a partial power outage due to a cooling upgrade project for Aurora.

July 28, 2023: Scheduled outage – Data center power outage
As part of the effort to stand up Aurora, the data center experienced a 4-day power outage.

August 14, 2023: Unscheduled outage –User notice not sent
ALCF missed sending out a scheduled outage notice within 24 hours of the outage, which converted it to an unscheduled outage.

October 30, November 6, 2023: Scheduled outages for Slingshot 11 Upgrade
Two scheduled outages were taken for network maintenance, upgrading to Slingshot 11 from Slingshot 10.

2.5.1.2 MTTI and MTTF

MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.14 summarizes the current MTTI and MTTF values.

Table 2.14 MTTI and MTTF Results

Theta-fs0 File System Cray Sonexion 2000 with 9.2 PB of usable storage				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System MTTI	N/A [a]	13.52 days	N/A	14.94 days
System MTTF	N/A	72.51 days	N/A	182.30 days

[a] N/A = not applicable.

Theta Lustre currently follows the Theta biweekly maintenance schedule. Theta-fs0 is not necessarily unavailable when Theta is in PM, but the PMs are often used to apply upgrades and patches.

2.5.2 Grand and Eagle Lustre File Systems

The ALCF installed a new set of Lustre file systems in 2022 running a Cray E1000 storage solution. Grand and Eagle each offer 100 PB of storage at 650 GB/s and provide availability protection if one fails. Additionally, the file systems have the capability of sharing via Globus, a move towards providing a community file system. The file systems went into production starting January 1, 2022.

ALCF uses target metrics of 90 percent overall availability and 90 percent scheduled availability since these file systems are tightly integrated with Theta and, at ASCR's request, that all user facilities use a target of 90 percent for scheduled availability for the lifetime of the production resources.

2.5.2.1 Grand Scheduled and Overall Availability

Table 2.15 summarizes the availability results for the Grand file system.

Table 2.15 Availability Results - Grand

Grand File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2022		CY 2023	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A	99.5	N/A	99.5
Overall Availability	N/A	96.6	N/A	93.2

[a] N/A = not applicable.

The remainder of this section covers significant availability losses and responses to them for both scheduled and overall availability data. Details on the calculations can be found in Appendix A.

Grand – Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability of Grand for the period of January 1, 2023, through December 31, 2023, as annotated in Figure 2.12.

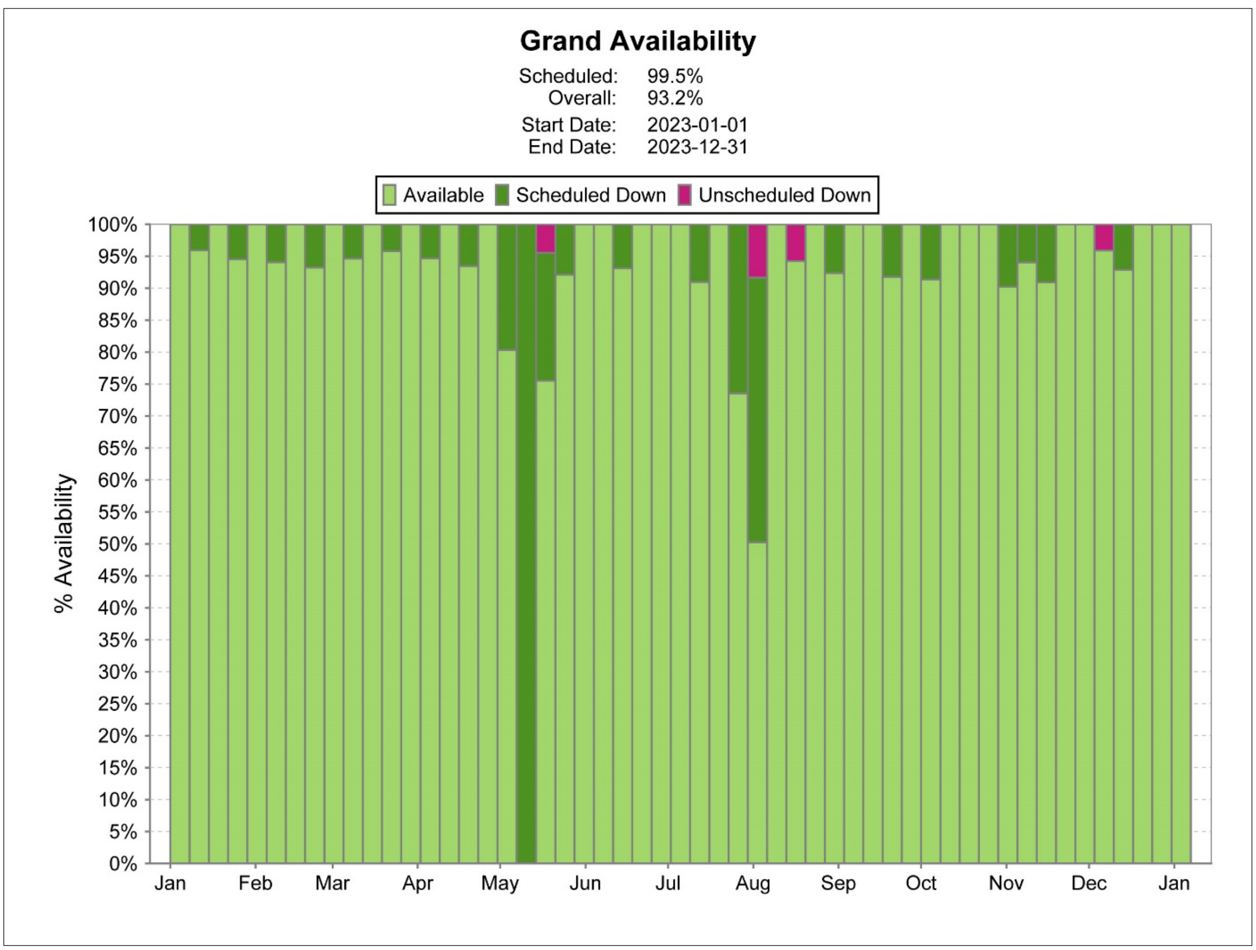


Figure 2.12 Grand Weekly Availability for CY 2023

Graph Description: Each bar in Figure 2.12 represents the percentage of the machine available for seven days. Each bar accounts for time in one of three categories. The pale green portion represents available core hours, the darker green represents scheduled downtime for that week, and the red represents unscheduled downtime. Each of the significant loss events is described in detail below (these also appeared in the Theta section.)

May 5, 2023: Scheduled outage – Data center cooling system upgrade
The data center had a 10-day complete cooling outage and a partial power outage due to a cooling upgrade project for Aurora.

May 15, 2023: Unscheduled outage – Multiple disk failures
Multiple disks from the array failed. The OST was kept offline while the disks were replaced.

July 28, 2023: Scheduled outage – Data center power outage
As part of the effort to stand up Aurora, the data center experienced a 4-day power outage.

August 3, 2023: Unscheduled outage – File system mount issue
The file system had an issue with distributed locks. Resolution: An HPE patch was installed.

August 14, 2023: Unscheduled outage –User notice not sent
ALCF missed sending out a scheduled outage notice within 24 hours of the outage, which converted it to an unscheduled outage.

October 30, November 6, 2023: Scheduled outages for Slingshot 11 Upgrade
Two scheduled outages were taken for network maintenance, upgrading to Slingshot 11 from Slingshot 10.

December 4, 2023: Unscheduled outage – HA node pair failure
An HA node pair failed. Access to OSTs was blocked until the issues were resolved. Additionally, Globus for Grand was paused.

Grand – MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.16 summarizes the current MTTI and MTTF values.

Table 2.16 MTTI and MTTF Results - Grand

Grand File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System MTTI	N/A [a]	14.68 days	N/A	13.08 days
System MTTF	N/A	121.13 days	N/A	72.68 days

[a] N/A = not applicable.

Grand generally follows Polaris's biweekly maintenance schedule. Grand is not necessarily unavailable when Polaris is in maintenance, but the maintenance windows are often used to apply upgrades and patches.

2.5.2.2 Eagle – Scheduled and Overall Availability

Table 2.17 summarizes the availability results for the Eagle file system.

Table 2.17 Availability Results - Eagle

Eagle File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2022		CY 2023	CY 2023
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A [a]	99.5	N/A	99.9
Overall Availability	N/A	96.6	N/A	94.2

[a] N/A = not applicable.

The remainder of this section covers significant availability losses and responses to them for both scheduled and overall availability data. Details on the calculations can be found in Appendix A.

Eagle – Explanation of Significant Availability Losses

This section briefly describes the causes of major losses in Eagle's availability from January 1, 2023, through December 31, 2023, as shown in Figure 2.13.

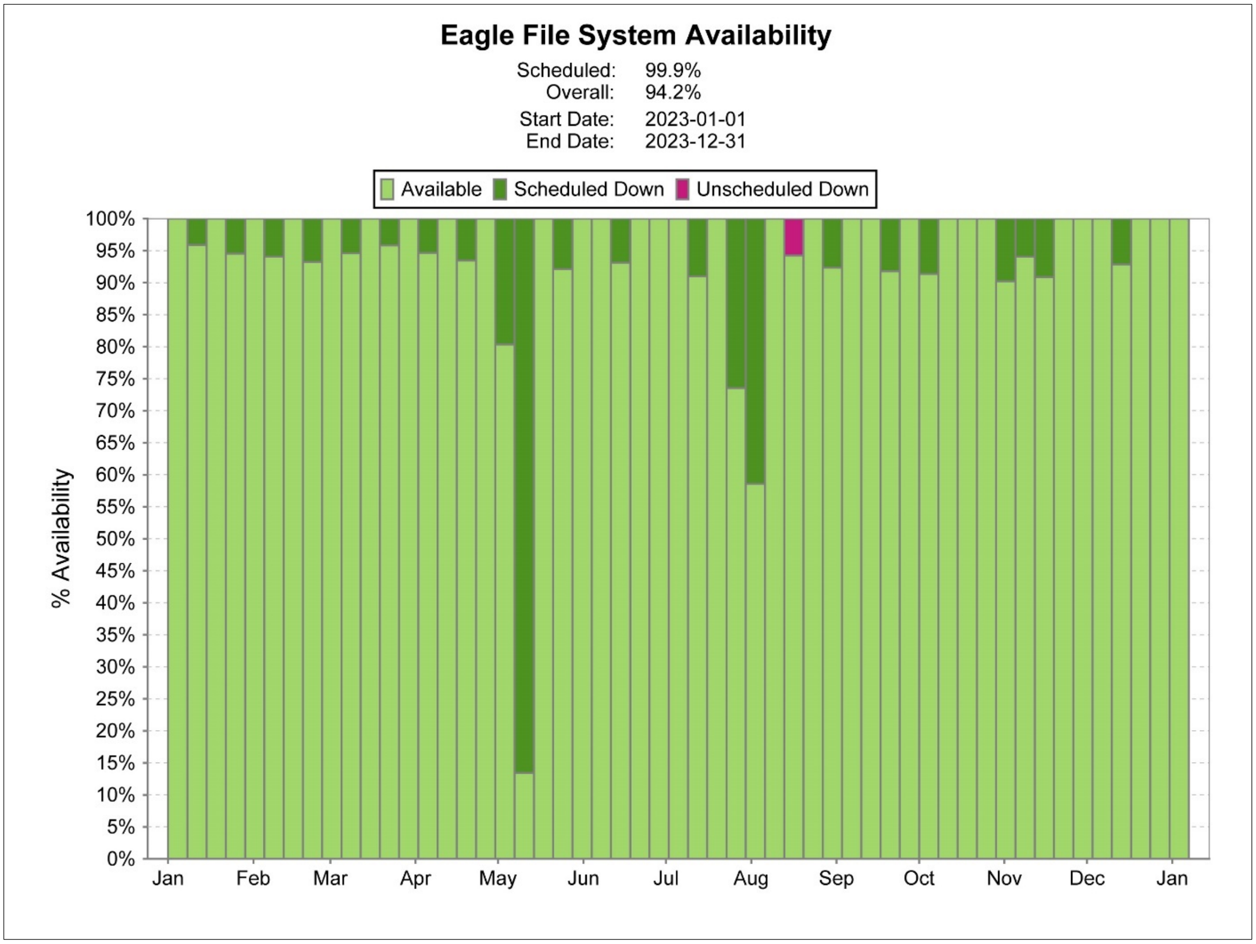


Figure 2.13 Eagle Weekly Availability for CY 2023

Graph Description: Each bar in Figure 2.13 represents the percentage of the machine available for seven days. Each bar accounts for time in one of three categories. The pale green portion represents available core hours, the darker green represents scheduled downtime for that week, and the red represents unscheduled downtime. Each of the significant loss events is described in detail below (these also appeared in the Theta section.)

May 5, 2023: Scheduled outage – Data center cooling system upgrade
The data center had a 10-day complete cooling outage and a partial power outage due to a cooling upgrade project for Aurora.

July 28, 2023: Scheduled outage – Data center power outage
As part of the effort to stand up Aurora, the data center experienced a 4-day power outage.

August 14, 2023: Unscheduled outage –User notice not sent
ALCF missed sending out a scheduled outage notice within 24 hours of the outage, which converted it to an unscheduled outage.

October 30, November 6, 2023: Scheduled outages for Slingshot 11 Upgrade
Two scheduled outages were taken for network maintenance, upgrading to Slingshot 11 from Slingshot 10.

Eagle – MTTI and MTTF Summary
MTTI and MTTF are reportable values with no specific targets. Table 2.18 summarizes the current MTTI and MTTF values.

Table 2.18 MTTI and MTTF Results - Eagle

Eagle File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System MTTI	N/A	14.68 days	N/A	14.94 days
System MTTF	N/A	121.13 days	N/A	182.30 days

Eagle generally follows Polaris's biweekly maintenance schedule. Eagle is not necessarily unavailable when Polaris is in maintenance, but the maintenance windows are often used to apply upgrades and patches.

2.5.3 Tape Storage

The facility-wide HPSS (high-performance storage system) tape archive was available to all ALCF users from all compute resources in CY 2023 as it has been in previous years. The tape storage is comprised of three 10,000-slot libraries with LTO8 tape drives and LTO8 tapes, with some legacy LTO6 drives and tapes remaining. The first tape library went into production in 2009 in the old ISSF data center, and the second followed in 2010 in the TCS data center, while the third library went into production in 2016. In 2019, all the tape libraries were moved to another building to provide separation of the archive data from the data center while also permanently vacating the ISSF datacenter. The HPSS disk cache and data movers are in the TCS data center. With the LTO8 drives and tape technology, the tape libraries have a maximum storage capacity of 305 PB.

2.5.3.1 Scheduled and Overall Availability

ALCF uses the target metrics of 90 percent overall availability and 95 percent scheduled availability, as proposed in the OAR for 2022. Table 2.19 summarizes the availability results.

Table 2.19 Availability Results - HPSS

HPSS Archive LTO8 tape drives and tape with 305 PB storage capacity				
	CY 2022		CY 2023	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A [a]	99.7	N/A	99.6
Overall Availability	N/A	96.7	N/A	93.8

[a] N/A = not applicable.

Note that HPSS is considered unavailable when users can't retrieve or access files via logins or data transfer nodes even though the HPSS libraries were unaffected during the scheduled maintenance periods, and still could do system functions like data migration. Therefore, HPSS's overall availability will reflect that users could not access it during scheduled maintenance.

Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability of HPSS for the period of January 1, 2023, through December 31, 2023, as shown in Figure 2.14.

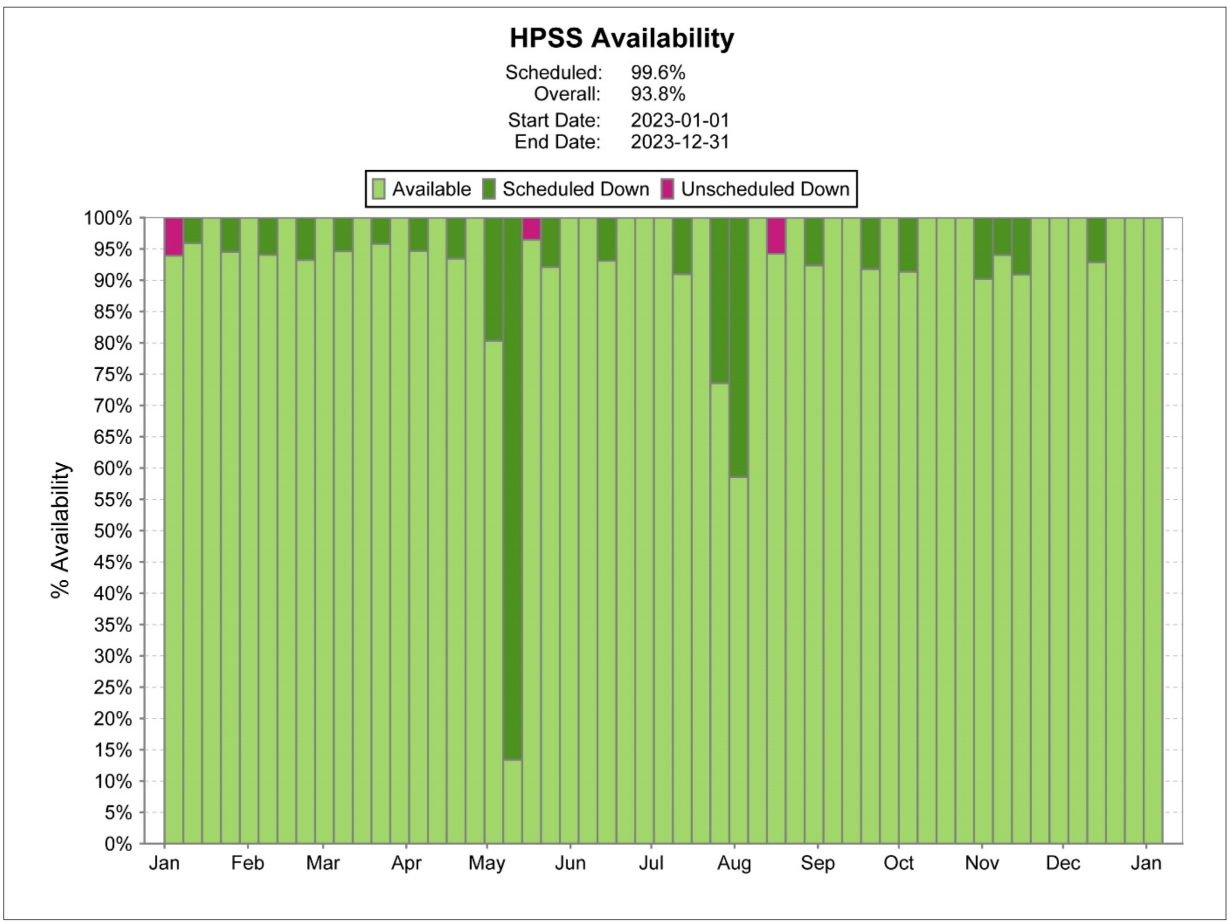


Figure 2.14 HPSS Weekly Availability for CY 2023

January 3, 2023: Unscheduled outage – Disk movers
HPSS was unavailable due to an issue with the disk movers.

May 5, 2023: Scheduled outage – Data center cooling system upgrade
The data center had a 10-day complete cooling outage and a partial power outage due to a cooling upgrade project for Aurora.

May 16, 2023: Unscheduled outage – Power Outage
Due to a power outage in Building 221 (the robotic tape libraries are located in a separate building), ALCF shut down HPSS while the room was on UPS power. It remained down until commercial power was restored.

July 28, 2023: Scheduled outage – Data center power outage
As part of the effort to stand up Aurora, the data center experienced a 4-day power outage.

August 14, 2023: Unscheduled outage –User notice not sent
ALCF missed sending out a scheduled outage notice within 24 hours of the outage, which converted it to an unscheduled outage.

October 30, November 6, 2023: Scheduled outages for Slingshot 11 Upgrade
Two scheduled outages were taken for network maintenance, upgrading to Slingshot 11 from Slingshot 10.

2.5.3.2 MTTI and MTTF

MTTI and MTTF Summary
MTTI and MTTF are reportable values with no specific targets. Table 2.20 summarizes the current MTTI and MTTF values.

Table 2.20 MTTI and MTTF Results

HPSS Archive LTO8 tape drives and tape with 305 PB storage capacity				
	CY 2022		CY 2023	
	Target	Actual	Target	Actual
System MTTI	N/A	14.70 days	N/A	12.69 days
System MTTF	N/A	121.28 days	N/A	60.58 days

[a] N/A = not applicable.

HPSS maintenance is not regular but is typically aligned with Theta's maintenance schedule. HPSS is often available even though other resources may be in preventative maintenance.

2.6 Center-Wide Operational Highlights

2.6.1 Polaris / Sirius

Polaris is the current ALCF workhorse for delivering compute cycles for science. Sirius is the Test and Development System (TDS) for Polaris. A considerable amount of work has been invested in enhancing the upgrade workflows. The goal is to ensure that any changes are carried out in a consistent and reproducible manner. System changes are first tested on Sirius, and once they are thoroughly inspected and approved, they are then replicated onto Polaris. It's similar to the "staging" system that's used in a software development pipeline. A major HPCM and slingshot upgrade is planned for Polaris in 2024, and testing started in CY 23 on Sirius.

The focus from DOE on Integrated Research Infrastructure, which is intended to facilitate DOE facilities working together in a more "holistic" fashion, will be important for science moving forward. The increased focus is new, but the idea is not. It is something that all the facilities have recognized and worked on over the years. In CY 2023, ALCF, working in partnership with APS and Globus, achieved an important milestone by running computations for beamlines on Polaris in a completely automated fashion with no human in the loop. When the APS upgrade is complete in 2024, the number of beamlines working in this mode is expected to grow, with the goal of making this the norm for beamlines that need additional computational power.

2.6.2 AI Testbed

The AI testbed continues to be extremely useful and popular with users and is available via the DD allocation program. In 2023, there was a significant expansion to the resources available in the testbed:

- The Cerebras CS-2 system was upgraded to include the MemX and SwarmX technologies. This enables users to create larger models and improves the processing of models and data, improving AI model training.
- The SambaNova SN-30 DataScale system was accepted and made available to users. The system has been upgraded to the second-generation processor SN-30. It consists of eight nodes, each featuring eight interconnected Reconfigurable Dataflow Units (RDUs) to enable model and data parallelism.
- A Graphcore Bow Pod64 system was installed and made available to users. The Bow Pod64 system, with 22 petaflops of AI compute, is the latest-generation accelerator from Graphcore. This is a single-rack system consisting of 64 Bow-class IPUs with a custom interconnect. The Graphcore software stack includes support for TensorFlow and PyTorch and the Poplar SDK used by ML frameworks.
- A GroqRack system consisting of the Groq Tensor Streaming Processor (TSP) was made available to users. Groq TSP provides a scalable, programmable processing core and memory building block able to achieve 250 TFlops in FP16 and 1 PetaOp/s in INT8 performance. The GroqRack consists of nine Groq nodes, each consisting of eight GroqChip TSP accelerators. GroqChip accelerators are connected via a proprietary chip-to-chip interconnect to enable larger models and data parallelism. The Groq system is highly optimized for inference.

2.6.3 Aurora / Sunspot

While Aurora is not yet in production, it has been a significant effort for the Operations team, which has been working in concert with Intel and HPE on system stabilization efforts and providing PBS support as needed. While many of the details are considered business sensitive, there was one instance where the Operations team was able to work with HPE on an upgrade to the HPCM software stack that saved significant downtime. The directions provided by HPE would have resulted in the system being completely unavailable for the entire duration of the upgrade. Working with the HPE HPCM team, a modified procedure was developed that allowed for downtime only during the upgrade of the "HPCM core" but then returned the compute nodes to service while work on all the other upgrades occurred in parallel.

2.6.4 Storage

Data migration was a significant topic this year. Due to the theta-fs0 decommissioning mentioned below, users needed to migrate their data to the Grand or Eagle file systems. To enable a sufficiently powerful file system for Aurora before acceptance, all user data is being migrated off the Grand hardware onto the Eagle hardware.

The ALCF HPSS system received a major upgrade that significantly improved the capacity and performance of the disk cache while also making needed OS updates to ensure security compliance.

The Distributed Asynchronous Object Store (DAOS) is a major component of the Aurora system. During SC23, a partial DAOS run topped the IO500 list by a significant margin.

2.6.5 Other Facility Notes

Networking: The ALCF is currently connected to ESnet at 4 × 100 Gbps, as well as having a dedicated 4 × 100 Gbps path to the APS. Working with the Computing, Environment, and Life Sciences Directorate (ALCF's parent organization), a more capable, flexible, and resilient network architecture was designed for the ALCF external networking. When implemented, the 4 × 100 GigE links to ESnet will be converted to a redundant pair of 400 GigE connections. The dedicated 4 × 100 GigE links to the APS will be converted to 4 × 400 GigE with the option to double that in the future. The equipment has been received on-site, and the upgrade will occur in 2024.

Software Development Best Practices: The ALCF web development team made software development changes to improve efficiency in operations. A standardized Linux development environment was created inside Argonne's virtual machines (VM) environment, with each developer having a dedicated VM. Much of the software development moved off individual laptops and into those virtual machines, reducing the requirements of the laptop and the risk if the laptop is stolen. This also made automated testing much more consistent, avoiding timeouts due to underpowered laptops. Services were moved into Podman containers, and ALCF is working on fully automated deployment in production via the containers.

2.6.6 Decommissioned Systems

On December 31, 2023, at 23:59:59 UTC, Theta, theta-fs0, and the K80-based visualization cluster called Cooley delivered their last cycles/bytes to science. Over six years, Theta delivered 200 million node-hours and supported more than 600 projects. Beyond all the impactful science that was accomplished on these systems, it also marked the end of the “many-core” era at ALCF.

Conclusion

ALCF maximized the use of its HPC systems and other resources consistent with its mission. ALCF has exceeded the metrics of system availability and capability hours delivered. For the reportable areas—MTTI, MTTF, and utilization—ALCF is on par with OLCF and NERSC, and the values reported are reasonable. These measures are summarized in Table 2.1. ALCF closely tracks hardware and software failures and their impact on user jobs and metrics. This data is used to select troubleshooting efforts and improvement projects. In CY 2023, this regular failure analysis has continued to drive code changes to ALCF’s software infrastructure and has provided details to support the debugging of storage system problems.

This page intentionally left blank.

Section 3. Allocation of Resources

(a) Did the allocation of computing resources conform with ASCR's published allocation policies (i.e., ratio of resources allocated between INCITE, ALCC, Director's Discretionary, ECP)?

(b) Was the allocation of Director's Discretionary 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?

ALCF Response

The allocation of resources is consistent with ASCR's requested allocation policies. On Theta, the breakdown of allocations available to INCITE, ALCC, DD, and ECP is 60 percent, 20 percent, 10 percent, and 10 percent, respectively. The 30 percent of the facility's resources available to ASCR is provided through the 20 percent to ALCC and 10 percent to ECP. INCITE was allocated 50 percent on Polaris in CY 2023 so that 10 percent could be used for projects in the Early Science Program (ESP) for Aurora, supporting the portion of Polaris's role as a testbed for Aurora readiness.

The INCITE program fully allocates all of the time available to it. The ALCC program is managed by ASCR and typically allocates all of the time available to it. Section 3.3 describes ALCF's mechanisms to promote user allocation utilization, including ALCF's overburn policy, which allows INCITE and ALCC projects that have used 100 percent of their allocations to continue running capability-sized jobs until they have used 125 percent of their allocations. The DD time is overallocated, but substantially underused due to the exploratory nature of the projects in the DD program.

As the results show in Section 8, these are reasonable allocations of resources. Below are a few areas ALCF considers when analyzing usage statistics for the allocation programs.

3.1 Usage of the INCITE and ALCC Hours

The 2023 INCITE program allocated 17.8M node-hours to 13 projects on Theta, and a total of 20.4M node-hours were delivered to INCITE projects on Theta (Table 3.1). Figure 3.1 shows that of these 13 projects, 10 were able to use at least 90 percent of their allocation, including 5 that used more than 100 percent of their initial allocation. Of the remaining 3 projects, 2 used at least 77 percent of their allocation, and the remaining project used 44 percent of its allocation.

The 2023 INCITE program allocated 2,000.0K node-hours to 13 projects on Polaris, and a total of 1888.6K node-hours were delivered to INCITE projects on Polaris (Table 3.2). Figure 3.2 shows that of these 13 projects, 8 were able to use at least 85 percent of their allocation, including 4 that used more than 100 percent of their initial allocation. Of the remaining 5 projects, 4 used between 48 and 58 percent of their allocation, and the remaining project used 12 percent of its allocation. The Polaris INCITE projects did not use 100 percent of the time allocated to INCITE in 2023. This was an unfortunate combination of factors that all happened at once. Several projects were slowed down due to software bugs, and many teams were focusing on fulfilling their end-of-project ECP requirements on Frontier. Finally, a necessary last-quarter update on Polaris slowed down progress in the final push. In addition to ALCF's usual mechanisms for promoting user allocation utilization described in Section 3.3, projects that did not have an INCITE project on Polaris that continued into 2024 were allowed to continue running using their 2023 allocation until January 31, 2024, which helped several projects with their milestone progress. ALCF staff work closely with projects throughout the year. For 2024, ALCF has accelerated the timeline for the ALCF Director to engage with PIs and has overallocated resources within INCITE to reduce risk.

For the 2022–2023 ALCC year, 10 projects had allocations on Theta with a total of 3.4M node-hours used, and 6 projects had allocations on Polaris with a total of 507.1K node-hours used. Figure 3.3 shows that all 10 projects on Theta used at least 97 percent of their allocation, and 9 projects used more than 100 percent of their initial allocation. Figure 3.5 shows that of the 6 projects on Polaris, 3 used more than 90 percent of their allocation (including 1 that used more than 100 percent of its initial allocation), 2 used between 56 and 60 percent of their allocation, and the remaining allocation used 15 percent of its allocation. For various reasons, several ALCC 2022–2023 projects on Theta and Polaris requested extensions, and these projects were permitted to continue using their allocations through July 31, 2023.

The 2023–2024 ALCC year was approximately halfway through its allocation cycle at the end of CY 2023. At that point, 9 projects received allocations totaling 2.8M node-hours on Theta, and 9 projects received allocations totaling 970K node-hours on Polaris. The 2023–2024 ALCC projects on Theta used a total of 2.1M node-hours from July 1, 2023, through December 31, 2023, and their allocation usage is shown in Figure 3.4. Due to the decommissioning of Theta in December 2023, the 2023–2024 ALCC projects on Theta were only half-year projects and ended on December 31, 2023. The 2023–2024 ALCC projects on Polaris used a total of 217.3K node-hours from July 1, 2023, through December 31, 2023, and their allocation usage is shown in Figure 3.6.

Table 3.3 summarizes the ALCC node-hours allocated and used on Theta in CY 2023. The 5.7M ALCC node-hours allocated are calculated by adjusting the 2022–2023 and 2023–2024 ALCC year allocations on Theta by the percentage of their award cycle occurring in CY 2023, then summing these two values. The total 5.4M ALCC node-hours used is the sum of all node-hours used by any ALCC project on Theta in CY 2023.

Table 3.4 summarizes the ALCC node-hours allocated and used on Polaris in CY 2023. The 943.4K ALCC node hours allocated are calculated by adjusting the 2022–2023 and 2023–2024 Polaris ALCC year allocations by the percentage of their award cycle occurring in CY 2023. The total 724.3K ALCC node-hours used is the sum of all node-hours used by any Polaris ALCC project in CY 2023.

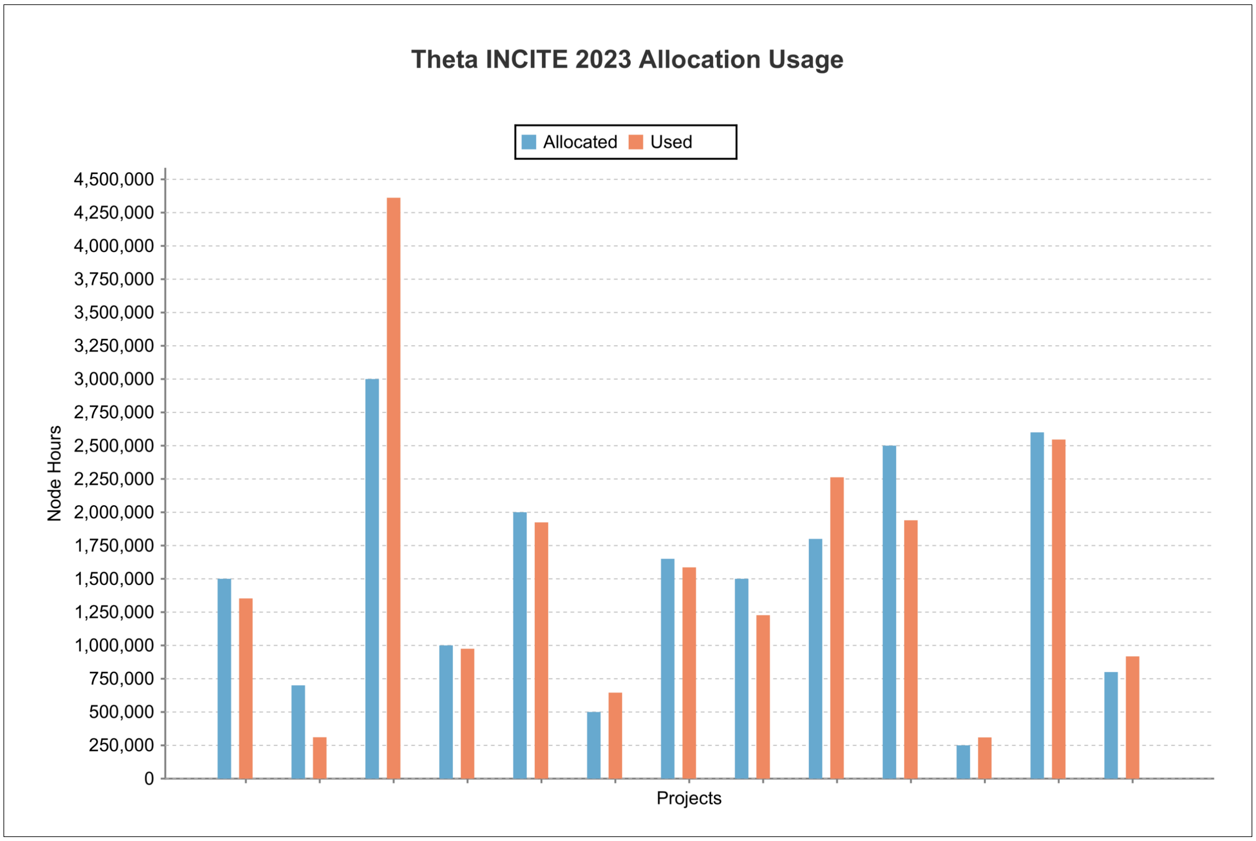


Figure 3.1 Theta INCITE 2023 Allocation Usage (Note: Projects are randomly ordered.)

Table 3.1 INCITE 2023 Time Allocated and Used on Theta

Projects	Theta
Allocated Node-Hours	17.8M
Used Node-Hours	20.4M

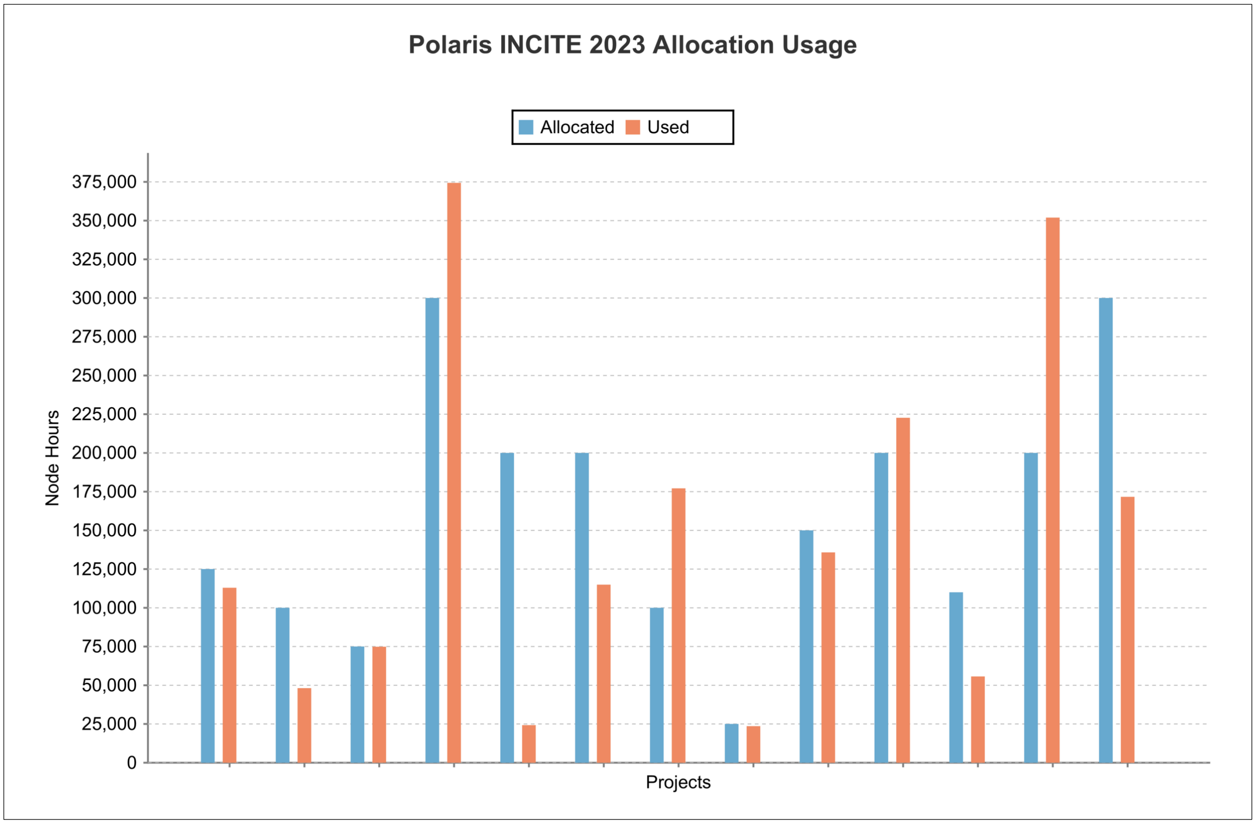


Figure 3.2 Polaris INCITE 2023 Allocation Usage (Note: Projects are randomly ordered.)

Table 3.2 INCITE 2023 Time Allocated and Used on Polaris

Projects	Polaris
Allocated Node-Hours	2000.0K
Used Node-Hours [a]	1888.6K

[a] Polaris INCITE usage includes usage from jobs of projects with an extension into January 2024.

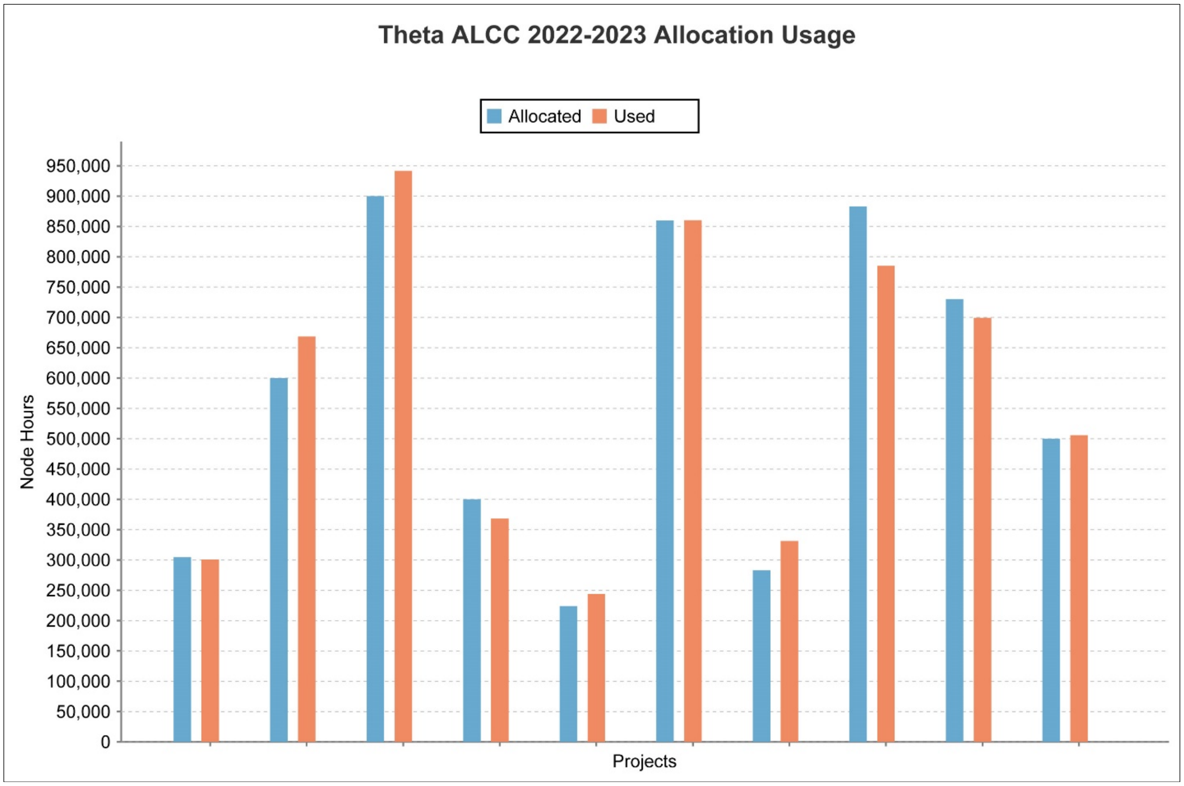


Figure 3.3 Theta ALCC 2022–2023 Allocation Usage (Note: Projects are randomly ordered.)

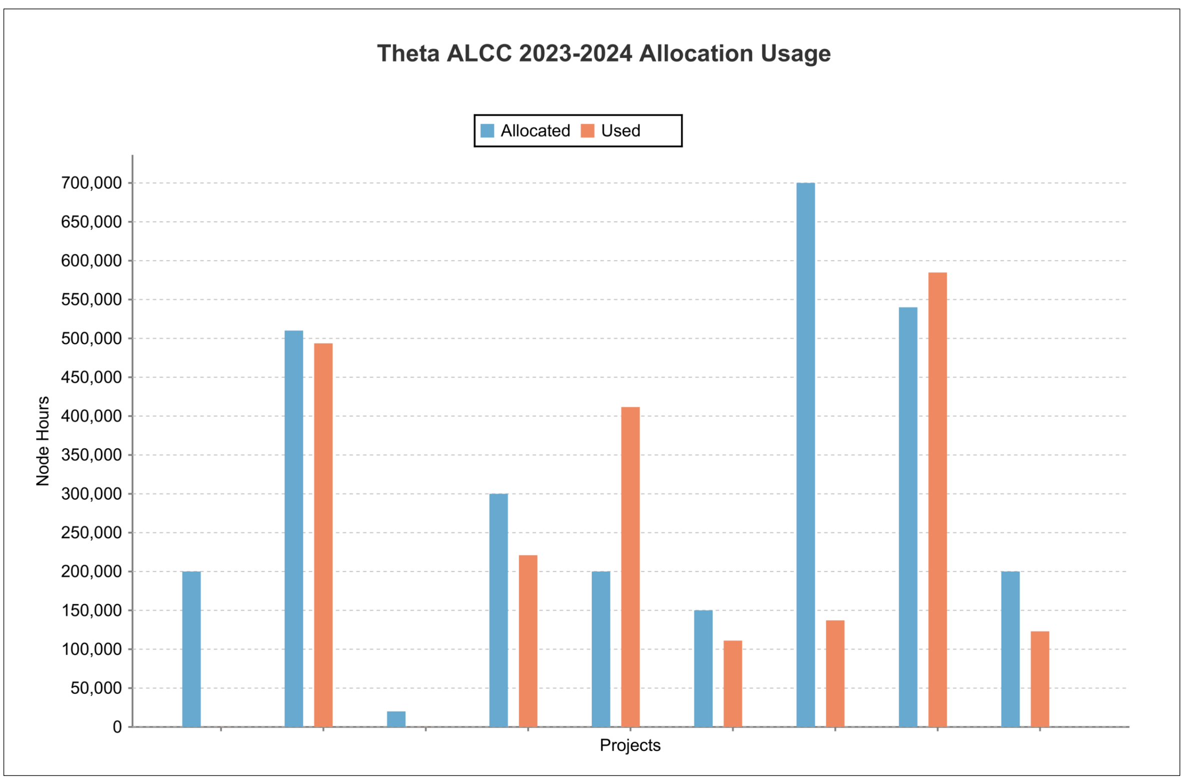


Figure 3.4 Theta ALCC 2023–2024 Allocation Usage as of December 31, 2023 (Note: Projects are randomly ordered.)

Table 3.3 ALCC Time Allocated and Used on Theta CY 2023

Projects	Theta
Allocated Node-Hours	5.7M [a]
Used Node-Hours	5.4M [b]

[a] Allocation total is obtained by adjusting each of the ALCC cycle allocations (2022–2023, 2023–2024) to prorate for the amount of time allocated in CY 2023, then summing.

[b] Usage total is the number of node-hours charged for jobs run against any ALCC allocation in CY 2023.

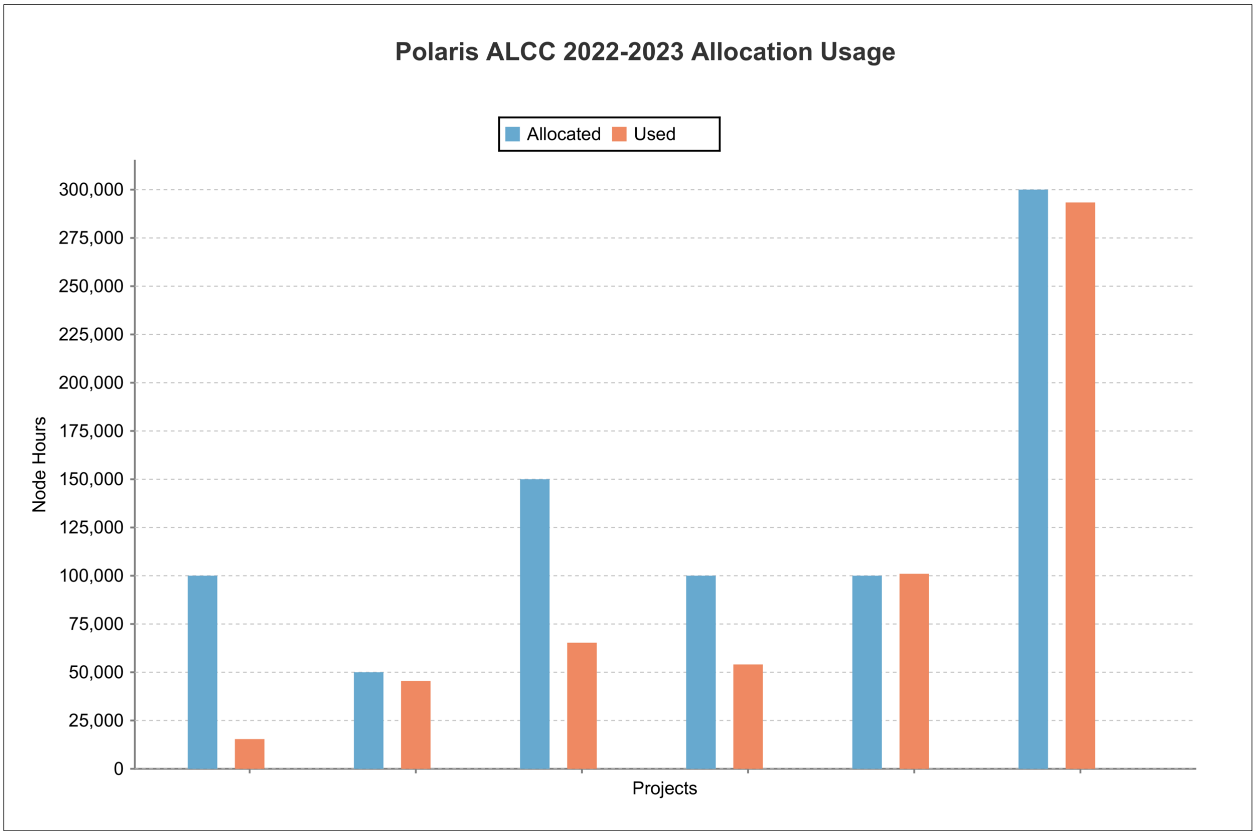


Figure 3.5 Polaris ALCC 2022–2023 Allocation Usage (Note: Projects are randomly ordered.)

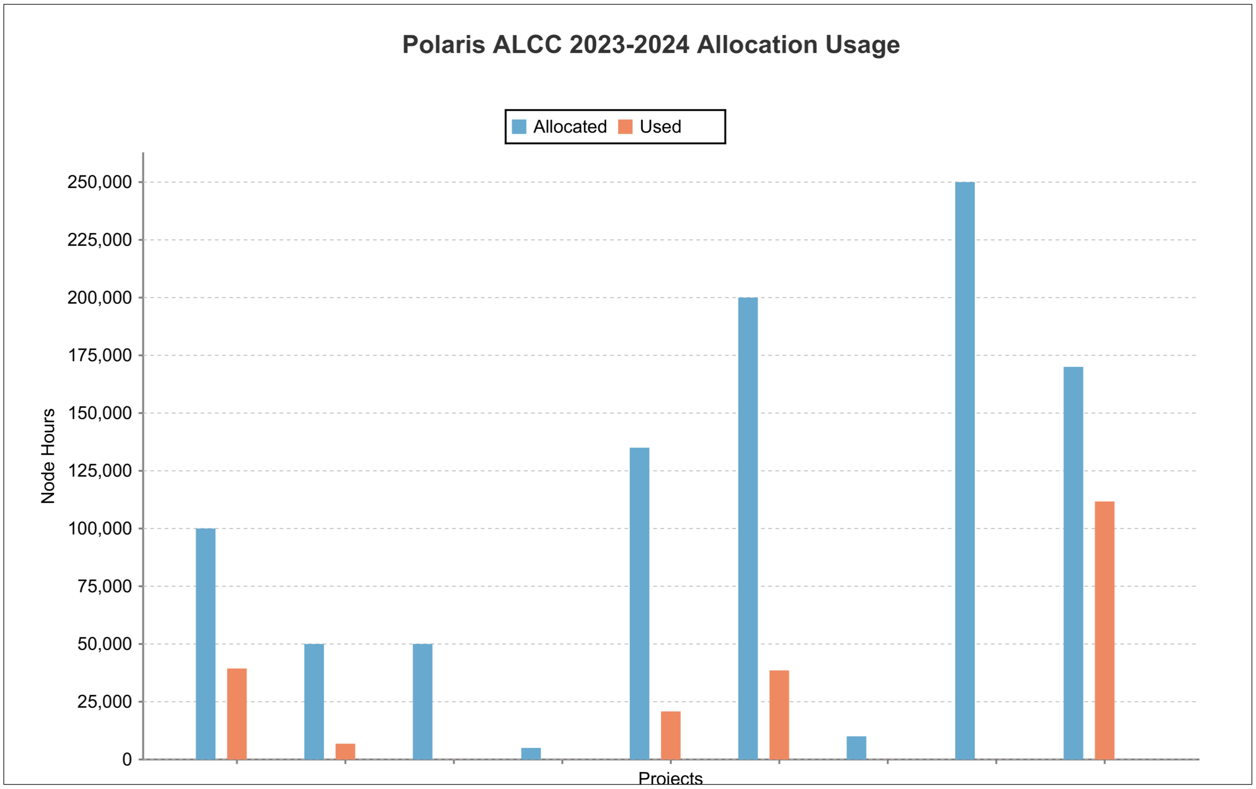


Figure 3.6 Polaris ALCC 2023–2024 Allocation Usage as of December 31, 2023 (Note: Projects are randomly ordered.)

Table 3.4 ALCC Time Allocated and Used on Polaris in CY 2023

Projects	Polaris
Allocated Node-Hours	943.4K [a]
Used Node-Hours	724.3K [b]

[a] Allocation total is obtained by adjusting the 2023–2024 ALCC cycle allocations to prorate for the amount of time allocated in CY 2023.

[b] Usage total is the number of node hours charged for jobs run against any ALCC allocation in CY 2023.

3.2 Facility Director's Discretionary Reserve Time

The DD program serves members of the HPC community who are interested in testing science and applications on leadership-class resources. Projects are allocated in five categories:

1) INCITE or ALCC proposal preparation
2) Code support and/or development
3) Strategic science
4) Internal/support
5) ECP support

INCITE and ALCC proposal preparation allocations are offered for projects that aim to submit an ALCC or INCITE proposal. These projects can involve short-term preparation (e.g., a run of scaling tests for their computational readiness) or long-term development and testing.

Code support and/or development allocations are used by teams porting and optimizing codes or projects developing new capabilities. This category includes the development, testing, and runs required for competitions such as the Gordon Bell Prize. Projects in this category have been responsible for bringing new capabilities to ALCF.

ALCF also allocates time to projects that might still be some time away from submitting an INCITE proposal, or that offer a "strategic science" problem worth pursuing. Examples include supporting projects from DOE's Scientific Discovery through Advanced Computing (SciDAC) program, industry research efforts, and emerging use cases, such as coupling experimental and computing facilities.

Internal/support projects are devoted to supporting the ALCF mission. ALCF does not reserve node-hours for internal activities. All internal use comes out of the DD allocation pool. This category regularly includes projects that help the staff support the users and maintain the system, such as diagnostics and testing of tools and applications.

To support the dynamic needs of ECP, the ECP time was moved from ALCC to DD starting in 2019, but the 10 percent allocation is still part of the overall ASCR fraction of the system. As a result, the discretionary pool grew to 20 percent of the system to support ECP. ECP and the computing facilities run a Resource Allocations Council (RAC) that meets monthly to discuss the computing needs of ECP projects, allocating up to 10 percent of the system.

DD allocations are requested through the ALCF website and are reviewed by the Allocations Committee (which includes representatives from Operations, User Experience, and the Catalyst teams). The committee collects additional input from ALCF staff, where appropriate. Allocations are reviewed on their readiness to use the resources and their goals for the allocations and are awarded time on a quarterly basis. The DD allocation program has high demand, and often the requested amount cannot be accommodated.

Tables 3.5 and 3.6 show the total time allocated and used in the DD program on Theta with expansion and Polaris, respectively, during 2023. By its very nature, the DD program is amenable to over-allocation because time is often left unused; however, it should be noted that these totals do not represent available allocations for the entire calendar year. A project might have a 1,000-node-hour allocation that only persists for three months, but that 1,000-node-hour allocation is counted entirely in the annual total node-hour number. Projects are not guaranteed the time allocated. DD projects run at a lower priority than INCITE or ALCC projects, which could reduce the amount of time available for their use. Exceptions are made for some internal projects that support acceptance of new hardware or support of users, which is in line with the ALCF core mission.

Table 3.5 DD Time Allocated and Used on Theta with expansion in CY 2023

Projects	Theta with expansion
Allocated Node-Hours	15.7M
Used Node-Hours	5.8M

Table 3.6 DD Time Allocated and Used on Polaris in CY 2023

Projects	Polaris
Allocated Node-Hours	2895.8K
Used Node-Hours	1192.0K

Lists of the CY 2023 DD projects on Theta with expansion and Polaris are provided in Appendix B.

Figures 3.7 and 3.8 provide breakdowns of the CY 2023 DD allocations by domain for Theta with expansion and Polaris, respectively.

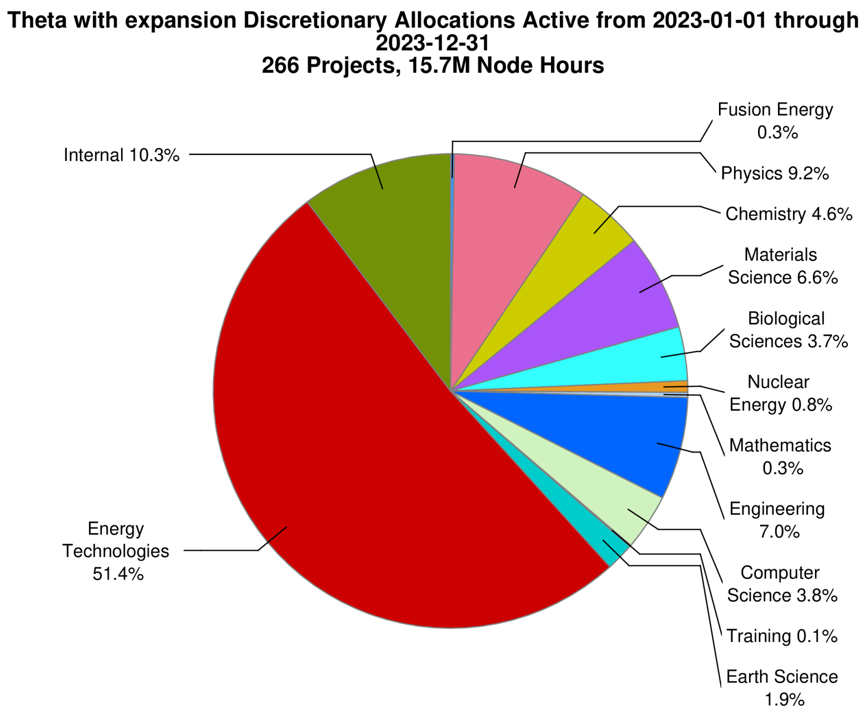


Figure 3.7 Theta with expansion CY 2023 DD Allocations by Domain

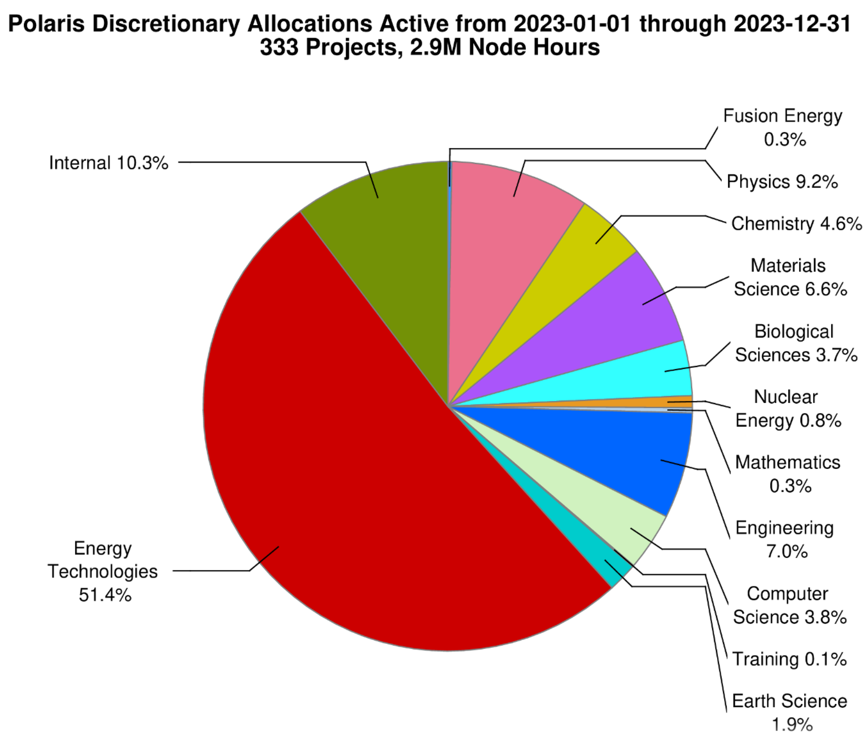


Figure 3.8 Polaris CY 2023 DD Allocations by Domain

3.3 User Allocation Utilization

Inevitably, some projects will use less time than allocated and others will want to use more. The ALCF has multiple methods of managing under- and over-utilization of user allocations. The overall goal of all the policies is to ensure that user projects have the greatest chance to accomplish their science goals.

To rebalance some of the allocated time across projects to ensure optimal utilization of resources, an overburn policy is in effect for INCITE and ALCC projects, which permits high-utilization projects to continue using the machine effectively for capability jobs. If an INCITE or ALCC project has exhausted its allocation in the first 11 months of its allocation year, it is eligible for overburn time. Currently, capability jobs submitted by INCITE and ALCC projects will run in the default queue (instead of backfill) for the first 11 months of the allocation year until 125 percent of the project allocation has been consumed.

Should additional overburn hours be needed, INCITE and ALCC projects may provide the ALCF with a short description of what the project plans to do with the additional hours, highlighting specific goals or milestones and the time needed to accomplish them. The scheduling committee, allocations committee, and ALCF management review these requests. Non-capability jobs from projects that have exhausted their allocation will run in the backfill queue.

This overburn policy does not constitute a guarantee of extra time, and ALCF reserves the right to prioritize jobs submitted by projects that have not yet used 100 percent of their allocation. The earlier that an INCITE or ALCC project exhausts its allocation, the more likely they are to be able to take full advantage of the overburn policy.

As noted in the discussions for Figures 3.1 and 3.2, in CY 2023, nine INCITE projects went beyond 100 percent of their awarded time, using a total of 2.5M node-hours beyond their allocations. Similarly, a total of 10 ALCC projects went beyond 100 percent of their awarded time (Figures 3.3 and 3.5) for the 2022–2023 ALCC year, using a total of 0.3M node-hours beyond their allocations. (Note that these totals include all mechanisms available to projects to use time beyond their awards, including what is enabled through this overburn policy as outlined above.)

Under-utilization earlier in the allocation year is primarily managed through personal contact with the projects to understand issues and assist with resolving any problems. Examples of these challenges can include data movement, scheduling challenges, porting problems, and bugs. Most of the time, this approach is effective, but if significant under-utilization persists further into the allocation year, then ALCF may apply its stated policies for pulling back time from INCITE or ALCC projects to be redistributed among other projects in that program. For INCITE, utilization is assessed, and a pullback of time can be applied at the beginning of the fifth and ninth months of the allocation year. For ALCC, utilization assessment, and possible pullback occur after the seventh month of the allocation year.

3.4 Other Large-Scale Managed Resources

In addition to its HPC systems, ALCF supports its user base with the Grand and Eagle file systems, which are accessible from multiple ALCF systems and constitute a vital part of the ALCF Community Data Co-Op. Every new project awarded compute time is also given a storage allocation on Grand or Eagle. Supplementary information on the technical specifications and availability for Grand and Eagle can be found in section 2.5.2, and a description of significant changes in the storage space undertaken in 2023 can be found in section 2.6.4. As part of the mission to support collaborative and data-driven scientific discoveries, ALCF provides researchers with services to securely share and reliably transfer data using Globus. In 2023, ALCF supported 37 data-only projects using these services. The Eagle filesystem with Globus sharing is a key component in ongoing efforts to automate the process of taking data at APS, processing that data with on-demand computing resources on Polaris, and finally sharing that processed data with collaborators and the scientific community.

Conclusion

The ALCF delivered the following node-hours to the allocation programs in CY 2023: 22.3 million to INCITE, 6.1 million to ALCC, and 7.0 million to DD. The DD program has been used not only to develop INCITE and ALCC proposals but also to conduct real science of strategic importance and to drive development and scaling of potential INCITE and ALCC projects. Excellent ALCF support and reliable, high-performing ALCF resources have enabled INCITE and ALCC projects to run jobs efficiently and to achieve science goals that could not otherwise have been reached.

This page intentionally left blank.

Section 4. Operational Innovation

(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?

ALCF Response

Listed below are the innovations and best practices carried out at ALCF during CY 2023. ALCF innovations and best practices have helped to prepare for future systems, have enabled more efficient operations, and have strengthened collaboration and engagement, both across ASCR facilities and beyond. As outlined in section 4.3, ALCF is effectively utilizing its postdoctoral fellows.

4.1 Operational Innovation – Technical

The ALCF has undertaken several projects to improve the operations of ALCF and to better respond to user needs.

4.1.1 GitLab CI

Challenge: ALCF has made a GitLab instance available to users, but it is challenging to get users to use it. Many potential users are unfamiliar with GitLab and continuous integration (CI) tools, so comprehensive training is essential.

Approach: To address this challenge, funded by ECP, ALCF staff developed the "Intro to GitLab CI Training Series," consisting of five sections hosted on the GitHub that provide:

- Comprehensive training: The training series offers in-depth education on GitLab and CI concepts, ensuring that users have the knowledge and skills necessary to leverage the GitLab instances.
- Accessibility: By providing a tiered learning path, the series caters to users with varying levels of expertise, making GitLab and CI accessible to beginners while offering advanced content for more experienced users.
- Practical examples: Along with use cases, these demonstrate the real-world applications of GitLab CI, motivating users to apply their knowledge in utilizing the GitLab instances.

Impact/Status: The outcomes of this project include the following:

- Increased utilization: The aim is for more users to embrace the GitLab instances at ALCF for their development and collaboration needs.

- Reduced barriers: The training series lowers the barrier to entry for users who were previously hesitant to use GitLab and CI tools, fostering a more inclusive environment.
- Enhanced collaboration: The increased adoption of GitLab instances may improve collaboration among ALCF users and collaborators, leading to more streamlined development processes.
- Continuous improvement: The tutorial remains an ongoing project, with regular updates and additions to meet the evolving needs of users.

By facilitating the adoption of GitLab instances, this effort not only improves the individual workflows of users but also contributes to a more collaborative and efficient environment at ALCF. The ALCF plans to publish and promote the videos associated with the training series. Once published, the associated links will be added to the GitHub page. Promotion will include other ASCR facilities that may wish to adopt this training approach.

4.1.2 Coupling HPC Simulations and AI with CFDML and NekRS-ML

Challenge: The field of AI/ML for science is rapidly expanding, with more and more researchers trying to augment or substitute traditional HPC simulations with ML models. This requires a new programming environment that offers seamless integration of simulation codes with deep learning frameworks, thus creating complex workflows, in-memory coupling of components, in situ or in transit data transfer, and inference serving. Several factors pose challenges to this new type of workflow, namely, the difference in programming languages used by traditional HPC simulations (Fortran, C, C++) and AI/ML (Python, TensorFlow, PyTorch), file system avoidance to prevent I/O bottlenecks and efficient scaling to leadership-class systems.

Approach: To overcome these challenges, ALCF, with contributions from other Argonne staff in other divisions, created two open-source software projects that are available on GitHub. The first is computational fluid dynamics for machine learning (CFDML), which was developed in response to the needs of a data science and ML project in the Aurora Early Science Program. CFDML leverages the open-source libraries SmartSim and SmartRedis to concurrently deploy large-scale CFD simulations, a database, and distributed ML applications to perform in situ (online) training and inference of ML models for turbulent flows. Making use of the database, the simulation and ML components can run concurrently on the HPC system, have access to asynchronous two-way communication between all components, and can share training data, metadata, and model checkpoints throughout the run. Moreover, by adopting a collocated deployment strategy, the components can utilize the computational resources (CPU and GPU) on the same set of nodes and limit all data transfer to on-node operations. With this strategy, CFDML demonstrated near-perfect scaling efficiency on up to 512 nodes of Polaris and imposed a negligible overhead on the simulation. The code repository for CFDML is hosted on GitHub.

The other project is NekRS-ML, which is a sandbox for different methods of coupling simulation and ML designed to provide tutorials and examples for ALCF users and the CFD community. It is based on the highly efficient and scalable ECP code NekRS, which was featured as one of the Gordon Bell finalists at SC23. NekRS-ML can utilize the workflow strategy developed in CFDML to perform online training of graph neural networks (GNNs) constructed directly on the partitioned CFD mesh and can also perform surrogate modeling of fluid flows. The GNN work incorporates CFD technologies for exchanging information at the part boundaries into the GNN layers, such that training and inference with these networks are independent of the number and shape of the graph partitions. This new development for GNN allows distributed training on very large graphs at scale and consistent solutions regardless of the partition strategy, therefore being very well suited for integration with numerical codes that rely on domain decomposition. The code repository for NekRS-ML is hosted on GitHub.

Impact/Status: Having a solution to the problem of coupling large-scale HPC simulations with distributed training will facilitate the development of accurate surrogate models for complex and realistic scientific problems. The usual approach of running the simulation *a priori*, saving all the training data to file, and then performing model training can cause significant I/O and storage bottlenecks in the exascale era. CFDML and NekRS-ML offer useful solutions for training surrogate models concurrently with the simulation. Both of these open-source projects have made a significant impact on ALCF users. CFDML is currently being used by an Aurora Early Science Project performing closure modeling for large eddy simulation of complex aerodynamic flows and was also the focus of a 2024 INCITE award for Polaris and Aurora. NekRS-ML was presented at the 2023 ALCF Hands-On User Workshop, resulting in a new collaboration between ALCF and Argonne's Nuclear Engineering division to perform active learning and optimization of Molten Salt Reactor geometries. In the coming years, NekRS-ML will continue to be developed and provide users with new tutorials and demonstrations of coupled workflows.

Outcomes: Development of NekRS-ML and CFDML has so far been confined to Argonne and not yet been adopted by other ASCR facilities, however there is interest in this innovation from Argonne's Nuclear Engineering (NE) and Transportation and Power Systems (TAPS) divisions, both of which use NekRS code to carry out fluid dynamics simulations, and could include collaborators from outside Argonne in the future.

4.1.3 DL Benchmarking across Architectures

Challenge: As Python has become the go-to programming language for the broader AI/ML community, it is important to know how portable and performant these Python models are across different hardware with minimal change to the source code and proper leveraging of vendor specifications. ALCF focused on the GPU accelerators employed in DOE-funded supercomputers such as Aurora (Intel), Frontier (AMD), and Perlmutter (NVIDIA).

Approach: ALCF, in collaboration with Intel, worked on applications chosen from diverse scientific backgrounds. ALCF executed these applications in an environment that utilizes vendor-specified techniques to maximize performance on specific hardware. A couple of representative applications were used for benchmarking:

- CosmicTagger (originates from high-energy neutrino physics): This application identifies neutrino interaction events using AI/ML methods. The scheme was implemented using Python frameworks PyTorch and TensorFlow with various precision data types.
- Flood Filling Network (FFN, comes from neuroscience): This application segments animal brain images to identify different components like brain tissue, blood vessels, etc.,

using AI/ML methods. The scheme was implemented using TensorFlow with various precision data types.

The performance analysis has included:

- Benchmarks on NVIDIA A100 (on Polaris), AMD MI250 (on JLSE), and Intel Data Center Max 1550 (on Sunspot) and identifying best practices.
- Testing of various precision data types: FP32, TF32, FP16, BF16.
- Exploration of methods to increase throughput of relatively smaller AI/ML workloads (both memory usage and compute capability).

Impact/Status: This work lays the foundation for future efforts in several promising directions:

- Management and deployment strategies for scientific workloads involving AI/ML at scale at DOE facilities.
- Demonstration of best practices through presentations, workshops, and shared code repositories. For example, this work was presented at the High-Performance Python for Science at Scale (HPPSS) workshop at SC23.
- Performance benchmarking at large scales. This effort focused on the performance at the smallest compute units (single device and single node). (The other facilities likely use similar benchmarking methodologies.) Code repositories are here and here.

4.1.4 ALCF's Container Registry: Streamlining HPC/AI Application Deployment

Challenge: HPC and AI applications at the ALCF demand an environment that can handle complex, resource-intensive tasks. Due to varying software environments, traditional deployment methods can lead to compatibility and performance issues. The primary challenge is to enhance the portability, isolation, and consistency of these applications across different computing systems such as Polaris, Theta, and Sunspot.

Approach: A suite of containers was developed by ALCF and made available on ALCF's GitHub space. These containers are specifically tailored for running traditional HPC/AI applications at scale and utilize Singularity, a container technology favored for its compatibility with HPC systems. The container suite includes a variety of applications, ranging from databases and MPI codes to data science workflows with popular frameworks like TensorFlow and PyTorch. This range ensures that a wide array of computational tasks can be supported efficiently. To facilitate collaboration and customization, ALCF provides detailed guidance for teams to create and maintain custom container registries on GitHub. This enables teams to collaborate effectively, manage version control, and ensure that all members work in consistent environments. The containers are designed to be deployed on various ALCF systems, ensuring researchers can transition between different computational resources without altering their application environment. While containers and registries exist at NERSC (for example, NERSC uses open-source Harbor as its registry), ALCF's unique approach to use GitHub as a registry sets it apart from other facilities.

Impact/Status: The container registry has significantly improved the portability and consistency of application deployment across different ALCF systems. Singularity containers provide an isolated environment for each application, which is crucial for testing and ensures that applications do not interfere with each other, leading to more reliable computational results. The guidance on creating custom registries on GitHub can empower teams to collaborate more effectively. Shared container environments mean teams can work together without discrepancies in software versions or environments.

4.1.5 HIP on Aurora: chipStar

Challenge: With current and upcoming DOE machines—Frontier and El Capitan—featuring AMD GPUs, HIP is a likely programming model for applications and runtimes to target. This raises the question of how to port those codes to Aurora, which has SYCL as a main programming model.

Approach: Bringing in HIP code using the natively supported SYCL programming model is a possibility, but offering native HIP support may also be a viable alternative, which would avoid increasing application complexity and maintenance costs. ALCF's HIP on Aurora project, called chipStar, was an ECP-funded, cross-laboratory effort in collaboration with Georgia Institute of Technology, University of Tampere in Finland, Intel, AMD, ORNL, and NERSC. chipStar, which ALCF now maintains on its own, is a HIP implementation built on top of OpenCL and Intel Level-Zero programming models. Critical HIP libraries are implemented on top of Intel oneMKL to offer maximum performance to applications on Aurora and leverage the work that Intel has done to support ECP applications.

Impact/Status: chipStar 1.0 was released in July 2023, with support for CP2K applications, libCEED, and GAMESS-libcchem. HIP libraries such as hipRTC, hipBLAS, or hipSolver are also supported. By leveraging the HecBench benchmark suite, the HIP on Aurora team embarked on improvements to software quality and performance after July and throughout the rest of 2023, which led to four interim releases between November 10, 2023, and January 10, 2024, with the release of chipStar 1.1 occurring on January 22, 2024. The chipStar ecosystem continues to expand, with support for porting the ExaBiome and HARVEY applications, as well as new libraries—hipFFT and hipRand—which are being implemented.

Outcomes: The codes, chipStar and accompany libraries, are hosted on GitHub. In CY 2023, ALCF staff published two publications about chipStar, one in *Concurrency and Computation Practice and Experience* (https://doi.org/10.1002/cpe.7866), and the other in *Euro-Par 2022: Parallel Processing Workshops: Euro-Par 2022 International Workshops* (https://doi.org/10.1007/978-3-031-31209-0_15).

4.2 Operational Innovation – Management/Workforce

ALCF works to prepare for next-generation systems through collaboration with vendors and other DOE facilities. ALCF reports on participation in research projects in Section 1.3.2 and professional community activities in Section 8.3.1.

4.2.1 HPC Training

Challenge: The landscape of computing sciences, particularly in HPC and AI, presents a challenge in recruiting and retaining a skilled workforce. Traditional academic settings often lack comprehensive integration of these advanced topics, limiting accessibility, particularly for historically underrepresented groups. Bridging this gap is essential to fostering accessibility and innovation in the field, ensuring a well-equipped workforce to tackle global challenges.

Approach: Various initiatives at ALCF work to address this challenge through innovative programs. One initiative, developed in collaboration with OLCF and NERSC, is the Intro to HPC Bootcamp, which offers a week-long immersive experience, providing hands-on training in energy justice projects using cutting-edge computational tools. The program prioritizes accessibility, offering travel support, stipends, and inclusive structures to encourage participation from underrepresented groups. Similarly, created in partnership with Intel and Chicago academic institutions, the Intel oneAPI Student Workshop directly exposes students to advanced tools and concepts needed for current HPC systems. Finally, building off of current broad interest in large-language models (LLMs), ALCF is developing curriculum to expand the Intro to AI-driven Science of Supercomputers series, first run in Fall 2022, to include the function of LLMs and their applications to science.

Impact/Status: These initiatives have yielded significant outcomes. The Intro to HPC Bootcamp saw overwhelming interest, with 300+ applications. A majority of participants identified as underrepresented, and many are now pursuing internship opportunities at DOE national laboratories, indicating increased engagement and interest in the field. Additionally, the Intel oneAPI Student Workshop garnered positive feedback, bolstering student confidence and fostering a clearer path toward careers in HPC and AI. Plans for future iterations and integration into existing educational programs underscore the sustainability and scalability of these efforts in shaping the future workforce in computational sciences.

4.2.2 Optimizing Log Management for Different Systems with Kafka Integration

Challenge: Managing and aggregating logs across various systems and components is a complex task at ALCF. With a multitude of nodes and subsystems like Aurora and Polaris, the challenge lies in efficiently collecting, processing, and monitoring logs in a scalable way. Traditional log-gathering methods often lead to fragmented data, slow retrieval, and difficulty in analysis, particularly in a high-performance computing environment where real-time data processing is critical.

Approach:

- Deployment of Kafka clusters: Kafka, a distributed event streaming platform, was deployed to enhance log management. Multiple Kafka clusters were set up within the ALCF infrastructure to facilitate log collection from various nodes and subsystems, including systems like Aurora and Polaris.
- Integration with existing systems: The Kafka clusters were integrated with existing ALCF systems, allowing for efficient log transmission from the compute nodes to the

Kafka brokers. This includes setting up direct links from systems to the respective Kafka nodes designated for log aggregation.

- Use of Filebeat for log forwarding: Filebeat, a lightweight shipper for forwarding and centralizing log data, was utilized to send logs from the compute nodes to the Kafka clusters. This allowed for the real-time processing of logs and simplified the management of log data flow.
- Scalable Kafka infrastructure: The Kafka infrastructure was designed to be scalable, accommodating the addition of more nodes to the Kafka clusters as ALCF's systems and log data volumes grow.

Impact/Status:

- Improved log aggregation: Because Kafka is distributed, logs from all over the ALCF infrastructure can be aggregated efficiently, ensuring that the data is centralized and accessible.
- Real-time processing: Kafka can manage a huge amount of high-throughput data, which enables real-time log processing. This capability is crucial for prompt monitoring and analysis and allows for quicker responses to potential issues.
- Enhanced analysis capabilities: The structured aggregation of logs via Kafka clusters has paved the way for more sophisticated analysis tools and algorithms to be applied, improving insights into system performance and user behavior.
- Lessons learned: This approach to log management could be valuable to other ASCR facilities.

4.2.3 DOE Cross-facility Workflow Training

Challenge: ALCF operates some of the largest supercomputers in the world. Users often need assistance or additional tools to deploy their codes at scale in workflows of very large size or complexity. ALCF constantly works to better support users in effectively using the appropriate tools and techniques. ALCF is not alone in this pursuit, as the other ASCR facilities have similar challenges. Some users even wish to deploy cross-facility workflows.

Approach: A team of staff and postdocs at NERSC, ALCF, and OLCF came together to organize a joint lab workshop targeted at users of the three facilities to provide demos of how to deploy four workflow tools at the facilities: GNU Parallel, Fireworks, Parsl, and Balsam. Organizers created an hour-long tutorial for each tool and a repository with code demonstrating how to run each tutorial on machines at NERSC, ALCF, and OLCF. The workshop was delivered on April 12, 2023, with tutorials delivered live on Zoom using machines at each of the three facilities, including Polaris at ALCF.

Impact/Status: The workshop had 197 signups, of whom 62 were ALCF users. Over 100 attendees participated on Zoom, and the participant level remained above 50 throughout the day. Feedback in the post-workshop survey was positive. Many users expressed a wish for more workflow training content in the future.

4.3 Postdoctoral Fellows

ALCF supports a steady-state postdoctoral fellowship program. Within this program, ALCF supports one named postdoctoral fellow, the Margaret Butler for Computational Science Fellow. Postdocs are awarded a one-year term with an option to renew for an additional year (this is typically the case), with a similar option for a third year. The program's major goal is to either convert the postdocs to a regular staff position, place them at another DOE laboratory, or support their efforts to find an academic or industry position. In all cases, the objective is for these postdocs to continue as lifelong users of DOE compute resources.

Argonne's Postdoctoral Program Office handles applications for postdoctoral positions. In 2023, ALCF supported 22 postdoctoral researchers (fellows), including 7 new hires, representing a range of scientific domains. During the year, three postdocs transitioned from ALCF, one of whom is now an Argonne Scholar in another directorate, one returned to his home division, and one joined academia. Five postdocs were part of ESP projects, two were part of ECP projects, and 15 were part of ALCF's steady-state operations.

After hiring, each postdoc was assigned a direct research supervisor and an Argonne staff mentor. The mentor, initially selected by the division or the supervisor, could be changed by the postdoc. The supervisors met with the postdoc weekly and engaged in the postdoc's research efforts. The supervisor evaluated the progress and completed a yearly standardized review submitted to ALCF management for consideration of appointment renewals. The mentor was responsible for meeting with the postdoc to discuss career development milestones and personal goals. This interaction happened as needed, but no less than once a quarter. The guidance for these discussions included key skills the postdoc should focus on over the next year, opportunities for development, and, if entering the third year, what skills or experience will be most beneficial to enabling a smooth career transition. The ALCF Director also met monthly with the postdocs to hear progress updates, address any issues specific to the postdoc community, and solicit general feedback.

ALCF supported the following 22 postdoctoral researchers in CY 2023:

Riccardo Balin (Ph.D., Aerospace Engineering, University of Colorado Boulder). **Hired:** January 2021. Converted to an assistant computational scientist role in ALCF in January 2024. **Research area:** Coupling of simulations with scientific machine learning with applications to computational fluid dynamics (CFD) and turbulence modeling. **Projects:** (1) *CFDM*: a software package that leverages the open-source libraries SmartSim and SmartRedis to deploy large-scale CFD simulations, a database concurrently, and distributed ML applications to perform in situ (online) training and inference of ML models for turbulent flows; (2) *NekRS-ML*: an open source sandbox for different methods of coupling simulation and ML designed to provide tutorials and examples for ALCF users and the CFD community; (3) assisting PI Ken Jansen's project to perform online training inferencing of ML models for turbulence closures from ongoing CFD simulations. **Accomplishments:** (1) Using a collocated deployment strategy, CFDML was demonstrated to achieve near-perfect scaling efficiency up to 512 nodes of Polaris and to impose a negligible overhead on the simulation; (2) worked with Intel to profile and verify correctness of popular ML inferencing libraries on Aurora; (3) demonstrated performance portability of scientific ML model training utilizing vendor framework libraries on leading GPU hardware;

(4) published two arXiv papers, two workshop papers, one conference paper, and produced a tutorial for ALCF's 2023 hands-on HPC workshop.

Shivam Barwey (Ph.D., Aerospace Engineering, University of Michigan). **Hired:** August 2022. **Termed:** October 1, 2023. **Current role:** Argonne Energy Technology and Security (AETS) Fellow at Argonne National Laboratory. **Research area:** Reduced-order modeling of high dimensional fluid dynamical systems, graph-based learning, data clustering, HPC. **Project:** Interpretable data-based surrogate modeling for unsteady fluid flows in complex geometries using graph neural networks (GNNs). **Scientific goal:** To develop neural network-based surrogate models that (a) are compatible with unstructured/skewed meshes and geometric perturbations, and (b) produce interpretable latent spaces. **Accomplishments:** Current work is focused on the model development side (i.e., developing the interpretable GNN architecture, ensuring stability in predictions, compatibility with skewed meshes, etc.). Demonstrations are performed on data sourced from high-Re turbulent flow over backward-facing step configurations. Training is performed on Polaris nodes.

Soloman Bekele (Ph.D., Computer Engineering, Indian Institute of Technology Delhi). **Hired:** February 2023. **Research area:** performance-energy tradeoff in heterogeneous HPC. **Projects:** (1) evaluation and review of the Aurora NRE deliverable milestones of GEOPM, an open-source framework for the exploration of power and energy optimization on heterogeneous platforms; (2) researching power and energy optimization strategies in heterogeneous HPC; and (3) ML-guided hardware reconfiguration in chip multiprocessors. **Accomplishments:** Expanded the tracing framework THAPI to encompass the comprehensive collection of device telemetry, including power consumption, frequency, and metrics such as GPU utilization; developed a visualization framework for THAPI powered by Google's Perfetto; and conducted frequency-scaled experiments using GEOPM and gathered the performance data for various applications.

Krishna Teja Chitty-Venkata (Ph.D., Computer Engineering, Iowa State University). **Hired:** August 2023. **Research area:** deep learning inference, optimization of neural networks on hardware platforms, AI for Science applications. **Project:** structured pruning for efficient finetuning and inference of large language models (LLMs) on AI accelerators. **Accomplishments:** Developed a novel pruning methodology to compress the pre-trained LLMs for efficient inference and finetuning; accelerated the pruned LLMs using the proposed pruning methodology on A100, Cerebras, and Graphcore systems; and integrated the structurally pruned models into the DeepSpeed inference framework, Low-rank finetuning framework, and Graphcore's PopXL framework.

Farah Ferdaus (Ph.D., Electrical and Computer Engineering, Florida International University). **Hired:** February 2023. **Research area:** Energy efficiency of AI accelerators. **Project:** Conducting experiments to measure the power consumption of the AI accelerators available in the AI Testbeds. The aim is to investigate the impact of hardware and software libraries on the energy consumption of the DL frameworks for both training and inference using a set of diverse and standard DL workloads. MLPerf benchmarks will be used to evaluate the energy consumption of the DL frameworks. **Accomplishments:** (1) Has developed a power monitoring tool using *NVML, PYHLML*, and *gcipuinfo* APIs to record power from NVIDIA GPU, Habana Gaudi-1, and Graphcore IPU, respectively, as part of this ongoing work. (2) Presented major findings internally at the postdoc research blitz session in April, the Power/Energy Efficiency Argonne Working Group meeting in September, and during CELS priority week in December.

Raymundo Hernandez-Esparza (Ph.D., Chemical Sciences, Metropolitan Autonomous University, Mexico City, Mexico). **Hired:** September 2022. **Research area:** computational chemistry. **Project:** chemical catalysis at exascale. **Accomplishments:** (1) Adapted Pynta, an automated workflow code to enable the calculation of thermochemistry and rate coefficients for reactions involving metallic surfaces, to use Polaris, Sunspot, and Aurora. Other improvements to Pynta enabled simultaneous simulations, at capability, in Polaris, Theta, and Sunspot, and reduced execution time from days to hours; (2) integrated Pynta code with PWDFT, an electronic structure code funded by ECP; and (3) presented these outcomes at two international conferences and the Exascale Catalytic Chemistry Annual Meeting, a multi-year BES project funded by the DOE Office of Science.

Khalid Hossain (Ph.D., Physics, Washington State University). **Hired:** January 2023. **Research area:** AI/ML for science, performance benchmarking with Python. **Projects:** (1) performance benchmarking of several AI for science applications as part of Aurora NRE project; (2) scaling out effort of CosmicTagger (an application originating in neutrino physics, part of Aurora NRE project); (3) testing and implementation of AI/ML framework JAX in the context of nuclear ground-state calculation; and (4) worked with the Parsl team to set initial application runs for effective execution of workflow methods. **Accomplishments:** (1) published a workshop paper and gave an HPPSS talk at SC23; (2) presented a poster at the 2023 Argonne Postdoc Symposium; and (3) participated in the first-ever Argonne Postdoc Slam.

Chunyong Jung (Ph.D., Atmospheric Sciences, North Carolina State University). **Hired:** December 2021. **Termed:** December 5, 2023. **Current role:** Postdoctoral researcher in the Environmental Science (EVS) Division at Argonne National Laboratory. **Research area:** high-resolution regional climate modeling and high-impact weather systems, examining their response to climate change and associated impacts. **Projects:** (1) High-resolution climate modeling for North America: In this project, we generate convection-permitting past and projected future climate analyses at a 4 km spatial resolution covering all of North America. (2) Tropical and extratropical storm impacts on future offshore wind energy (TREXO-Wind): In this project, we study the impact of hurricanes on offshore wind turbines under a warming climate to provide a more informed standard for wind turbine design and sitting. **Accomplishments:** (1) The resulting datasets have been analyzed relative to existing climate studies over various regions of North America, including the contiguous United States, Alaska, and Puerto Rico. This analysis evaluates the overall improvement achieved by the high-resolution modeling in comparison with their forcing data, as well as their effects on extremes. Findings from this research have resulted in one paper publication in *Sci Rep* and have been presented at numerous conferences. (2) We have developed a new fully coupled (atmosphere-ocean-wave) model, called C-WFS, and analyzed the impact of ocean waves on hurricane development through the new model. Findings from this analysis have been presented at multiple conferences.

Bharat Kale (Ph.D., Computer Science, Northern Illinois University). **Hired:** September 2023. **Research area:** visual analytics and information visualization. **Projects:** (1) using LLMs for inferring viral-human protein-protein interactions at scale; (2) visualizing embedding space of the viral proteins and the ESMFold structures for NIH's Bacterial and Viral Bioinformatics Resource Center (BV-BRC); (3) visualizing the embedding space of LLMs; and (4) interactive navigation of chemical space using AI models in the background. **Accomplishments:** (1) augmented an NVIDIA outreach activity for adding the genome-scale language model (GenSLM) to NVIDIA's playground for AI foundational models. This work involved embedding interactive visualizations that show the role of visual analytics in explaining the science behind these models; (2) published a paper in *Computer Graphics Forum*.

Geng Liu (Ph.D., Mechanical Engineering, City College of New York). **Hired:** October 2021. **Research area:** parallel computing. **Project:** Aurora ESP project *Extreme-Scale In-Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations* (PI: Amanda Randles). **Accomplishments:** (1) worked with multiple programming models on HARVEY, which is the code used in the ESP project, developed the Kokkos version of the cell part of HARVEY, and applied optimizations to it to improve the performances of several kernels, memory accessing and data transfers; (2) co-authored an SC23 workshop paper; (3) presented a poster at the 2023 Argonne Postdoctoral Research and Career Symposium; and (4) presented his performance evaluation and optimization of HARVEY work at Argonne's Aurora Applications Meeting.

Nathan Nichols (Ph.D., Material Sciences, University of Vermont). **Hired:** September 2021. **Research area:** SYCL portability framework development and performance optimization of scientific applications for exascale computing, with significant contributions to the MadGraph software and development of path integral quantum Monte Carlo and analytic continuation algorithms for Aurora. **Projects:** (1) development and optimization of MadGraph utilizing the SYCL portability framework; (2) collaborated on an effort to develop a workflow for MadGraph aimed at meeting science readiness goals; (3) executed scaling runs of MadGraph on Aurora, Sunspot, and Polaris, targeting GPU optimization; (4) adapted analytic continuation algorithm to SYCL, expanding its application to fermionic matter; and (5) initiated the development of conformance tests for the SYCL specification regarding hierarchical data-parallel kernels. **Accomplishments:** (1) publications in *Nature Communications* and *Phys. Rev. E.;* (2) presented an ATLAS ESP project progress update at the 2023 Aurora ESP Meeting, and developments in performance and portability for MadGraph at various meetings throughout the year; and (3) collaborated with internal teams at Argonne, CERN, Intel, and UTK focusing on high energy physics, improvements to the SYCL specification, and quantum Monte Carlo simulations.

Nwamaka Okafor (Ph.D., Electrical and Electronic Engineering, University College Dublin). **Hired:** December 2022. **Research area:** AI and ML applied to HPC log analysis. **Current projects:** Developing methods to effectively analyze a diverse set of ALCF system logs and building robust and interpretable AI models that would help to improve systems operations and efficiencies. This work involves analyzing how job characteristics impact wait times on Theta and Polaris, using machine learning to predict job wait times, and studying patterns in hardware errors that lead to job failures. **Accomplishments:** presented a lightning talk at the Women in HPC (WHPC) Workshop at SC23 and presented a poster at the 2023 Argonne Postdoctoral Symposium.

Siddhisanket (Sid) Raskar (Ph.D., Electrical and Computer Engineering, University of Delaware). **Hired:** June 2021. **Research areas:** dataflow models and architectures, HPC, ML, AI accelerators, performance modeling. **Projects:** (1) MLPerf HPC submission; (2) evaluation of AI accelerators (GNN, LLM, mixed precision, sparsity); (3) transfer learning for efficient program generation; (porting and optimization of HPC workloads on AI accelerators; (4) Codelet Model: fault tolerance and resiliency, dataflow software pipelining; and (5) PI of a 2023 LDRD expedition evaluating OpenMP on Graphcore. **Accomplishments:** (1) published in *Applied Sciences Journal* and in the proceedings of SIGMETRICS 2024, CARLA 2023, Euro-Par 2023, ICPE 2023, ExHET 2023, and PMAM 2023; (2) conducted a comprehensive performance study of LLMs on novel AI accelerators; and (3) gave invited talks at ALCF training events (ATPESC and ALCF Hands-on HPC Workshop), a tutorial at SC23, and presented LDRD Expedition results at LDRD Expedition Day 2023.

Shilpika (Ph.D., Computer Science, University of California, Davis). **Hired:** September 2023. **Research area:** Data visualization and analysis of HPC systems. Visualization and interpretation of AI for science to enable informed decision-making in AI workflows. **Projects:** (1) visualization and interpretation of AI for Simple Smiles Transformer (SST); (2) online 3D network graph displaying up to 70K protein-protein interactions (PPI) identified by a backend distributed LLM workflow; (3) visualization and analysis of error logs in HPC systems. **Accomplishments:** (1) built an online 3D network graph displaying up to 70K PPI identified by a backend-distributed LLM workflow; (2) identified bottlenecks and proposed improvements to the SST AI model that predicts molecular docking scores on proteins.

Mathi Thavappiragasam (dual Ph.D. in Electrical and Computer Engineering and in Computational Mathematics, Science and Engineering, Michigan State University). **Hired:** June 2022. **Research area:** Enhancing the performance and portability of application software using novel architecture features. Conducts studies to form specifications for the features of future architectures. **Current projects:** (1) ExaStar (porting the multi-physics astrophysics simulation modules of ExaStar to Intel systems, focusing on the higher-order neutrino radiation hydrodynamics module Thornado); (2) MiniMDock (porting the particle grid-based molecular docking proxy app to Intel systems using OpenMP-offloading as well as SYCL. Also, evaluating the miniMDock's HIP code-base portability on Intel systems using the chipStar HIP implementation); (3) high-level heterogeneous programming model research (studying possible data transfer strategies between host and devices for different data structures using programming models). **Accomplishments:** (1) successfully built ExaStar-Thornado and Flash-X on Intel GPU systems with the support of the Intel-Argonne COE, Intel exascale performance codesign team, and Intel compiler teams, and submitted a manuscript to ISC-HPC 24; (2) published an article in the proceedings of the 30th IEEE International Conference on High Performance Computing, Data, and Analytics, and Data Science (HiPC); and (3) gave an invited talk on performance portability enhancement on heterogeneous systems at University of Jaffna.

Umesh Unnikrishnan (Ph.D., Aerospace Engineering, Georgia Institute of Technology). **Hired:** November 2021. **Research area:** HPC, performance engineering, software portability, CFD. **Current projects:** (1) hardware integration and optimization on LCF resources; (2) application development to support Aurora ESP project PHASTA; (3) application development to attract usage on LCF resources. **Accomplishments:** (1) Performance benchmarking of ECP-CEED libParanumal software library on Sunspot; investigating and implementing optimization for compute-intensive kernels in the CNS-libParanumal application;

performance assessment and multi-node scaling of the application on Polaris and Sunspot with and without GPU-aware MPI. (2) Implementation of a SYCL backend for the libCEED library used in PHASTA; performance analysis and optimization of core functions in PHASTA; and continuous performance testing of PHASTA with different compiler versions, PETSc updates, and environment/compiler options. (3) Development of the CNS-libParanumal applications for high-speed applications, including implementation of shock-capturing schemes and reacting flow capabilities. (4) Gave two paper talks at the American Physical Society's 76th Annual Meeting of the Division of Fluid Dynamics and one at the Intel DevSummit for AI and HPC 2023.

Madhurima Vardhan (Ph.D., Biomedical Engineering, Duke University). **Hired:** September 2022. **Award:** Margaret Butler Fellow. **Research area:** Biomedicine. **Projects:** (1) establishing LLMs as high throughput synthetic Electronic Health Record data generators; (2) evaluating LLMs against traditional DL and state-of-art diffusion models on multi-dimensional axes of accuracy, utility, privacy-preserving, and augmentation axes; (3) enabling LLMs to function as personalized health coach via In-context learning; (4) developing a Personalized and Automated Healthcare Assistant by fine-tuning LLMs with deidentified structured and unstructured clinical datasets; and (5) evaluating the performance of generative models on HPC resources, Polaris and specialized AI accelerators, Cerebras and SambaNova. **Accomplishments:** (1) had two conference papers accepted (AI Health 2024, IEEE CAI 2024) and has one under review (IEEE EMBC 2024), had one journal paper accepted (*Internal Journal of Molecular Sciences*) and has another under review (*American Journal of Cardiology)*; (2) gave three invited lectures, attended the Rising Stars in Computational and Data Sciences in Austin, Texas, and was invited to the 10th Heidelberg Laureate Forum as a Young Researcher; and (3) made code improvements including generalizing a transformer-based code with respect to LLM models and clinical datatypes, and evaluated other generative models such as VAE, GAN, and DDPM on Polaris to compare against LLMs.

Archit Vasan (Ph.D., Biophysics, University of Illinois Urbana-Champaign). **Hired:** February 2023. **Research area:** machine learning for drug discovery. **Project:** Large-scale virtual screening for the CANDLE Early Science Project. **Accomplishments:** (1) ported ML framework to Aurora and ran large-scale virtual screening on vast chemical databases of billions of compounds; (2) Published a paper in the proceedings of SC23; and (3) gave talks at the Ninth Computational Approaches for Cancer Workshop (CCAFCW23) at SC23, at the Large Language Models (LLMs): Tutorial Workshop at SC23, and at the ALCF Hands-on HPC Workshop.

Kaushik Velusamy (Ph.D., Computer Science, University of Maryland, Baltimore). **Hired:** January 2022. **Research area:** data management, storage, and I/O. **Projects:** HDF5, ExaIO, Kove RAN, DAOS, DLIO. **Accomplishments:** (1) Evaluated the parallel I/O performance on DAOS, which will be a major filesystem on Aurora, with two standard I/O benchmark suites (IOR and H5Bench) to see DAOS's performance benefit to HPC and AI applications. Demonstrated 449x performance improvement for write operations and 307x for read operations on DAOS, compared to the Lustre file system on Polaris of the same scale. Also provided the documentation and a tutorial on how to use DAOS; (2) In the ExaIO ECP project, developed and implemented a topology-aware algorithm in the placement of aggregators in the HDF5 I/O layer called CCIO (custom collective I/O) to significantly improve the collective performance in large scale jobs. Performed systematic performance evaluations on the I/O subsystems in Theta, Polaris, and Summit, and demonstrated that CCIO improved the collective I/O by 5x-20x, especially for small messages; (3) worked with a summer intern to investigate the potential benefits of a disaggregated memory system (DMS) for deep learning workloads, using Kove DMS as an example implementation. For small and random I/O, demonstrated up to 20x read performance improvement compared to the Lustre file system. Showed that Kove DMS could potentially serve as a caching storage layer for Lustre to improve DL training time; (4) contributed to the design and development of ALCF Deep Learning I/O benchmark and LLNL Deep Learning Profiler, and submitted a publication on this work to CCGrid 24.

Brendan Wallace (Ph.D., Atmospheric Science, SUNY-Albany). **Hired:** August 2022. **Research area:** atmospheric science/climate dynamics. **Projects:** (1) Large-scale MCS environments in a dynamically downscaled regional climate model (supported by the NSF), which sought to qualify the uncertainty associated with future projections of deep convective storms by determining the impact of changes in the large-scale dynamic and thermodynamic environments on the overall response; (2) differing characteristics of the end-of-century rainfall response between a global climate model and dynamically downscaled regional climate model (supported by the NSF), which aimed to determine the sensitivity of projected changes in rainfall across the United States between a global climate model with coarse horizontal grid spacing and a dynamically downscaled regional climate model with convection-permitting grid spacing; (3) storm tracking in dynamically downscaled regional climate simulations, whose aim was to understand how characteristics of future storms may differ in size, speed, and lifetime, with a secondary goal of testing the sensitivity of this response to choice of tracking algorithm. **Accomplishments:** Was the lead author of two publications, one appearing in *Earth and Space* and the other in *Springer Climate Dynamics*.

Azton Wells (Ph.D., Physics, University of California, San Diego). **Hired:** September 2022. **Research area:** deep learning for cosmological simulations, genomics, and language. **Projects:** (1) Aurora GPT: co-lead of post-pretraining team and contributor to evaluation and dataset preparation teams. Focused on creating scientifically motivated human-feedback datasets for reinforcement learning and preference optimization; (2) foundation models of genomics using the BV-BRC genomes database; (3) extending context windows of transformers for genomics on AI testbeds. **Accomplishments:** (1) gave an invited talk at Georgia Tech's AstroML series and various divisional seminar presentations and a workshop; (2) PI of a 2023 LDRD Expedition to study large-context transformer models on the Cerebras platform and co-PI of a 2023 LDRD Expedition to study surrogate models of large cosmological volumes on SambaNova platform (PI Nesar Ramachandra); (3) co-PI of an RDI grant to study evolutional modeling from genomic models (PI Tom Brettin); (4) co-PI of an INCITE award "Foundation models for epidemiology" (PI Arvind Ramanathan); (5) created a new scalable benchmarking infrastructure for genomic models using state-of-the-art fine-tuning methods; and (6) performance analysis and modification of Megatron-DeepSpeed on Polaris for training genomic models with extreme context window.

Zhen Xie (Ph.D., Computer Science and Technology, University of Chinese Academy of Sciences). **Hired:** August 2021. **Termed:** August 22, 2023. **Current role:** Assistant Professor of Computer Science at Binghampton University. **Research area:** HPC with a focus on the interaction between ML algorithms and system-level performance optimization.

Projects: (1) building modern ML/DL algorithms and systems on heterogeneous and parallel HPC architectures (e.g., GPUs and AI accelerators); (2) automatic performance optimization on HPC applications with the aid of machine learning; (3) accelerating HPC applications using ML-based approximation. **Accomplishments:** (1) published papers at the 29th International European Conference on Parallel and Distributed Computing (Euro-Par'23), the 27th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming (PPoPP'23), the 2nd International Workshop on Extreme Heterogeneity Solutions (ExHET'23), and two workshop papers at SC23; and (2) awarded two Impact Argonne awards in recognition of the contribution to AI for science for High Performance Computing and Enhancement of Argonne's Reputation.

This page intentionally left blank.

Section 5. Risk Management

(a) Does the Facility demonstrate effective risk management practices?

ALCF Response

The overview of the risk management process that ALCF follows (and laid out in Section 5.1) clearly demonstrates that ALCF successfully managed its project and operational risks in 2023. As part of the ALCF's Risk Management Plan (RMP), all risks (proposed, open, and retired) are tracked, along with their triggers and mitigations (proposed, in progress, and completed), in a risk register managed by risk managers. All risk ratings in this report are post-mitigation. ALCF currently has 37 open risks, with one high operational risk: Funding/Budget Shortfalls, which is managed by careful planning with the DOE program manager and the implementation of austerity measures as necessary; and two medium operational risks: (1) Staff Recruitment Challenges, which is managed by ongoing recruiting and re-tasking of current staff as needed; and (2) Staff Retention Challenges, which is managed by making salaries as competitive as possible, developing flexible schedules, identifying promotion opportunities, and training new staff. The major risks tracked for 2023 are listed in Section 5.2, along with the details of the risks in Table 5.1. Risks encountered in 2023 and the mitigations for them are described in greater detail in Section 5.3. Section 5.6 and Table 5.1 provide details on the major risks that will be tracked in 2024.

5.1 Risk Management Process Overview

ALCF uses documented risk management processes, first implemented in June 2006, and outlined in its RMP, for operational and project risk management. The RMP is annually reviewed and updated as needed throughout the year to reflect changes, to incorporate new risk management techniques as they are adopted, and to incorporate best practices from other facilities. Risk management is part of ALCF's culture, and RMP processes are part of normal operations and all projects, such as the ALCF-3 project launched in CY 2013.

Risk management is an iterative process that includes identifying and analyzing risks, performing response planning, and monitoring and controlling risks as shown in Figure 5.1.

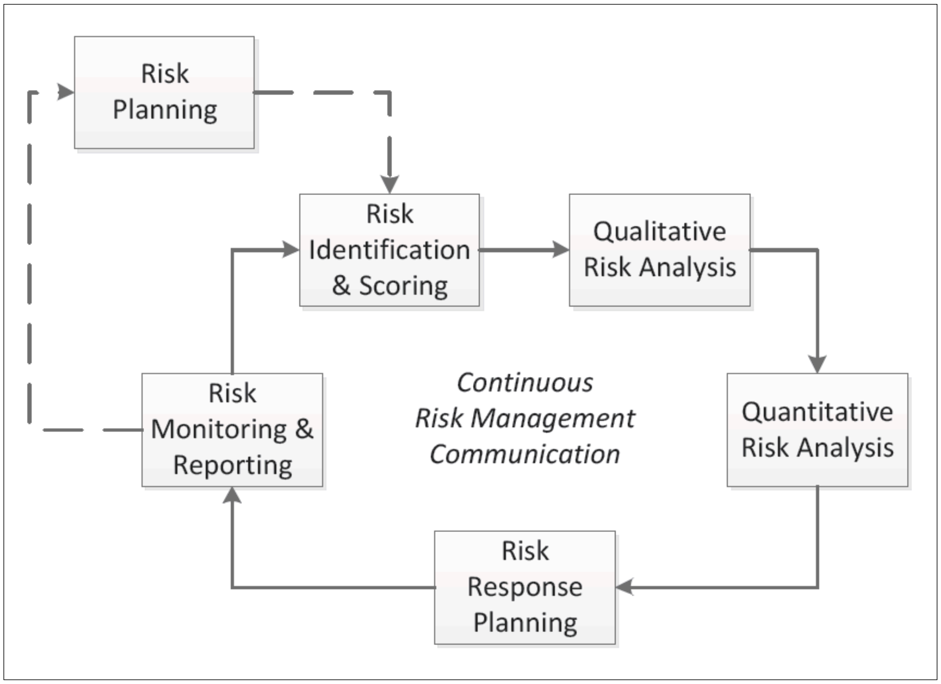


Figure 5.1 Illustration of the Iterative Risk Management Process from ALCF's Risk Management Plan

The ALCF risk management process consists of the following steps, which are performed on a continual basis in all normal operations and in all ALCF projects:

1. Plan, implement, and revise the RMP.
2. Identify threats and opportunities to cost, schedule, and technical objectives.
3. Analyze the impact of identified threats and opportunities to the cost, schedule, and technical baselines; and develop risk management strategies to manage and mitigate the risks.
4. Monitor risks, mitigation plans, and management reserve and contingency until the risks are retired or the project is closed.

A key part of this process is to identify potential threats and opportunities as early as possible so that the most critical risks can be assessed, the triggers effectively monitored, and the amount of management reserve/contingency needed to moderate the risks determined.

Risks are tracked using a secure, shared cloud-based storage system; risk forms and the risk register are formatted using Excel. Risk owners continually monitor the risks they own and submit monthly reports on all risks through the ALCF online risk reporting system.

5.1.1 Risk Review Board

The ALCF employs a five-person Risk Review Board with representatives from senior management, the operations team, the science team, industry outreach, and the financial services team. The board serves in an advisory capacity to ALCF management. The board meets as needed and offers recommendations regarding steady-state risk management issues. The RMP is consulted at all risk meetings. At each meeting, the board:

- Reviews proposed new risks and makes recommendations on whether to add proposed risks to the steady-state risk register.
- Monitors open risks and, for each open risk, reviews any new information on the risk provided by the risk owner or the steady-state risk managers and:
  - Determines whether the risk needs to be recharacterized.
  - Considers whether the risk has been managed and should be closed.
  - Reviews the mitigation strategies for the risk and considers whether any of the strategies need updating for any reason, such as a change in the technology landscape.
  - Works with the risk owner to modify the risk statement if needed to risk mitigation strategies, risk triggers, or risk scope.
  - Decide if a risk ownership should change.
- Reviews and identifies any risks to retire.
- Reviews the risks encountered in the past 18 months to discuss potential actions.
- Discusses risks encountered at other facilities and how they might apply to ALCF.

5.1.2 Risk Management in Day-to-Day Operations

ALCF currently has 37 open risks in the facility operations risk register and uses post-mitigated risk scoring to score the risks. The risks include general facility risks (such as funding uncertainties, staffing issues, and safety concerns) and specific risks (such as system component failures, availability of resources, and cost of electricity), so risks are distributed throughout the division. Risks are owned by group leads or managers. In ALCF operations, subject matter experts estimate risk mitigation costs for management consideration. In addition to formal Risk Review Board meetings, ALCF has many formal and informal individual risk meetings as needed. The Risk Review Board met once in CY 2023.

Risks are identified and evaluated, and mitigation actions are developed for all changes that occur at ALCF—from installing a new piece of hardware to changing the scheduling policy to upgrading software. For example, new risks are created anytime a resource goes from project to steady-state, or when a resource is decommissioned. If the risks are short-term or minor, they are not added to the registry. New significant risks identified during the individual meetings are added to the registry and reviewed at the next Risk Review Board meeting. Any change made to an existing risk—whether recharacterizing it, changing the factors affecting it, or retiring it entirely—requires a formal meeting, which any risk owners can propose.

Other tools besides the risk register are used for managing risks in day-to-day operational risk. An example is the use of Work Planning and Controls (WPCs) and Job Hazard Questionnaires (JHQs) to manage risks for activities where safety is a concern. WPCs are developed in consultation with safety and subject matter experts. JHQs are used for all staff and all contractors and cover all work. During planning meetings for any activities, staff members review the planned actions and identify possible safety concerns. If a potential risk is identified, detailed discussions with the safety experts are scheduled, and procedures for mitigating the risks are developed and documented in the WPC. The WPC is then used during the activity to guide the work.

In addition to machine operations, risk management is used in such diverse ways as evaluating and managing INCITE and ALCC proposal risks (e.g., the risk of too few proposals, the risk of a lack of diversity across science domains, the risk of too few capability proposals, etc.), safety risks in staff offices, leasing risks, support risks (e.g., including the opportunity risk that electricity costs could be lower than budgeted, etc.).

5.1.3 Continuation of the ALCF-3 Project

The project to procure and deploy ALCF's supercomputer Aurora, known as ALCF-3, continued in 2023. Risk Register managers continue to maintain a project risk register and track a set of detailed risks. Risk mitigation costs are developed using bottom-up cost analysis and then input to the commercial project risk analysis tool Oracle Primavera Risk Analysis (OPRA) to manage the contingency pool. These risks are not included in the risk numbers covered in this document and are not discussed further.

5.2 Major Risks Tracked

The ALCF operated both Theta and Polaris during CY 2023 and planned the growth of the staff and the budget to bring the facility to full strength. Because of this, ALCF continues to monitor a number of major risks for the facility. No risks were retired in 2023.

Five major operations risks were tracked for 2023, two with a risk rating of High, one Moderate, and two Low. None of these were encountered, and all of them were managed. They are described in Table 5.1. All risk ratings shown are post-mitigation. The risks are color-coded as follows:

- Red risks are Moderate or High.
- Orange risks are Low.

Table 5.1 Major Risks Tracked for CY 2023

ID	Title	Encountered	Rating	Notes
1059	Funding/Budget Shortfalls	No	High	ALCF regularly worked with the program office to plan a budget for handling the impact of a Continuing Resolution in FY 2023 and FY2024, new hires, and changes in the laboratory indirect expense rate. This risk remains a major concern as the facility moves forward with Theta, Polaris, and ALCF-3 in CY 2023.
25	Staff Recruitment Challenges	No	High	The ALCF added six new staff members in CY 2023: Four of those were external new hires, one was converted from a postdoc to the staff, and another was converted from a contractor to the staff. ALCF continues to have staff available who can be re-tasked as needed. With difficulty competing with industry for new hires, staff hiring remains a concern.
1049	Staff Retention	No	Moderate	ALCF did not lose any staff members in CY 2023. Budget concerns at Argonne and the growth in high-paying industry jobs for system administrators and programmers with HPC expertise make staff retention in future years a continuing concern.
1091	Injury to Workers/Overall Safety of the Division	No	Low	ALCF had increased data center activity in CY 2023, but zero incidents given the safety culture of ALCF.
0990	Electricity cost could increase beyond the budget	No	Low	ALCF monitored electricity usage data and worked with laboratory management.

5.3 Risks Encountered in the Review Year and Their Mitigations

ALCF encountered two risks in CY 2023. The risk owners are identified below, along with an assessment of the risk's probability and impacts, a brief description of what transpired, and how the risk was ultimately managed. Both risks have a residual rating of Very Low.

5.3.1 System Stability Issues due to Upgrades

1056: System stability issues due to upgrades	
Risk Owner	HPC Infrastructure Team Lead
Probability	Very Low
Impact	Cost: Very Low; Technical Scope: Medium
Risk Rating	Very Low
Primary Management Strategies	1) Perform upgrades on non-critical systems first when feasible. 2) Have a rollback plan in place. 3) Monitor performance closely following upgrade. 4) Work with the vendor to understand the upgrade(s) and the quality control processes. 5) Deep test on Test and Development Systems.
Triggers	1) Planned system upgrades. 2) System instability observed following system upgrades.

Description

On November 2nd, around 14:00 Central Daylight Time (CDT), ALCF took an unscheduled maintenance to conduct regression testing in support of the Slingshot 11 upgrade in progress on Polaris.

Evaluation

This regression testing was needed to determine whether to proceed with the planned further conversion of Polaris to Slingshot 11 on the following Monday. The regression testing required resetting ASICs (Application-Specific Integrated Circuits) in the Slingshot 11 cards, which is a disruptive action to the network.

Management

The Slingshot 11 fabric was validated, the regression tests completed, and the system was released to users.

5.3.2 Unable to Meet INCITE OMB Metric

1044: Unable to meet INCITE OMB metric	
Risk Owner	Computational Science Division Deputy Director
Probability	Very Low
Impact	Cost: Very Low; Technical Scope: Very High
Risk Rating	Very Low
Primary Management Strategies	1) Continuously monitor usage and inform PIs of their status and impact of project status via bimonthly reports and Catalyst weekly reports. 2) Encourage projects to use time earlier. 3) Enable scheduler changes to improve throughput. 4) Include readiness with award. 5) Activate swat team to work with highest-priority projects. 6) Discuss the current production machine with the community as a stable, known resource that codes are already ready for and more reflective of previous generation with less complexity. 7) ALCF staff work with the community to understand the sentiments about this machine. 8) New motivations for the use of an older machine considered. 9) If necessary, use a secondary ALCF machine for overwhelming end-of-year demand.
Triggers	1) Determination that the time left before the end of an INCITE or ALCC cycle is approaching the point of being insufficient for the remaining projects allocations. 2) Communication from the PI indicates project will not use all allocated core-hours. 3) Lack of response from PIs late in the cycle. 4) Conversations with project teams who express lowered interest in the ALCF production machine. 5) Drop in INCITE requests for the ALCF machine. 6) Drop in usage of the ALCF machine.

Description

INCITE projects on Polaris did not use 100% of the hours allocated in CY 2023.

Management

ALCF catalyst team members worked with their projects to help increase the usage of project allocations in CY 2023. The ALCF's INCITE overburn policy was implemented to allow projects that used 100% of their allocation to continue running jobs on Polaris, thus progressing towards the overall INCITE usage goal. A select group of projects were given an additional month beyond the end of the calendar year (13th month of INCITE) to continue running jobs to accomplish their science goals.

5.4 Retired Risks

No risks were retired in CY 2023.

5.5 New and Recharacterized Risks

There are no new risks and no recharacterized risks to report in 2023.

5.6 Top Operating Risks Monitored Closely for the Next Year

Table 5.2 lists the top operating risks that will be closely monitored in CY 2024, and the current risk rating and management strategies for each risk. The first three are the risks that experience has shown are most likely to be encountered in any fiscal year. We removed Staff Recruitment Challenges from our watchlist because ALCF was successful in hiring for open positions in CY 2023.

For CY 2024, we closely monitor two additional risks: ***Users perceive scheduling performance as poor*** and ***System Stability Issues***. ALCF is concerned that workloads related to the Integrated Research Infrastructure (IRI) community of users require a different mode of operation than those from other user programs like INCITE and ALCC. The new need could result in scheduling issues significantly as experiment-time workloads increase the demand for ALCF compute resources. As a direct result, ALCF will be closely monitoring risk 1019. With Aurora being operational in CY 2024 and nearing its transition from the ALCF-3 project into ALCF operations, ALCF is closely monitoring the system stability of this new machine, as reflected in risk 1078.

Table 5.2 Top Operating Risks Monitored for CY 2024

ID	Title	Rating	Management Strategies
1059	Funding/Budget Shortfalls	High	Develop austerity measures. Work closely with DOE sponsors to manage expectations and scope. Plan carefully, in conjunction with program office, for handling Continuing Resolution, leasing costs, and hires. Forward-pay lease to reduce overall leasing costs.
1049	Staff Retention	Moderate	Make salaries as competitive as feasible. Identify promotion opportunities. Develop flexible work schedules. Implement flexibility in work assignments.
1091	Injury to Workers/Overall Safety of the Division	Low	Promote safety culture at all levels of the division. Follow Argonne Integrated Safety Management Plan. Monitor work areas for potential safety concerns. Enforce use of personal protective equipment.
1019	Users perceive scheduling performance as poor	Low	Document scheduling policy. Revise policy as user needs evolve. Communicate the policy to the users as it changes. Regular reports about what projects are seeing long wait times. Schedule dedicated time for specialized jobs (e.g., "Big Run Mondays"). Discuss changes with scheduling committee.
1078	System Stability Issues	Low	Conduct regression testing to monitor system health. Monitor and analyze system failure data.

Conclusion

ALCF uses a proven risk management strategy that is documented in the RMP. Risks are reviewed and updated to reflect the dynamic nature of risk management, as well as to document new lessons learned and best practices captured from other facilities. Risk management is part of ALCF's culture and used by all staff, from senior management to summer students. A formal risk assessment is performed for every major activity within ALCF, with informal assessments used for smaller activities. Risks are monitored and tracked using a secure and shared cloud-based storage system. Many other tools are used to manage risks at ALCF, particularly in the area of safety. ALCF's effective risk management plan has contributed to the successful management of all significant risks encountered in CY 2023.

This page intentionally left blank.

Section 6. Environment, Safety, and Health

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

ALCF Response

ALCF maintains its commitment to health and safety at Argonne, solidifying its position as a leader in this domain. ALCF is guided by the laboratory Worker Safety and Health Program, which forms the basis for implementing the Integrated Safety Management System (ISM) to comply with 10 CFR 851. ALCF has established and implemented policies and procedures at the directorate and division levels that align with the Laboratory's overarching safety framework, ensuring the execution of work in a safe and effective manner.

6.1 Work Planning

6.1.1 Work Control Documents

ALCF utilizes the principle outlined in its work planning and control (WPC) manual to plan, review, and document all work activities. This manual integrates guidance from DOE-HDBK-1211-2014: *Activity-Level Work Planning and Control Implementation*, outlining how the division incorporates ISM principles into work planning and execution.

ALCF's day-to-day activities are assessed and documented through Work Control Documents (WCDs). The primary WCD, *Operations Skill of the Worker Standard Tasks*, provides a comprehensive overview of common low-risk operations tasks within the data center, specifying what, how, and where they occur. The division also employs task-specific WCDs for work with limited scope, unique activities, or specific skill requirements. For example, ALCF has a dedicated WCD for the replacement of rectifiers, a task reserved for skilled Qualified Electrical Workers (QEWs).

The approval and authorization of WCDs follow a graded approach, with more complex or hazardous work involving higher levels of oversight from laboratory management or subject matter experts (SMEs). WCDs undergo annual reviews to ensure they encompass all potential hazards and controls encountered by workers in the data center. New workers are added to the WCD once they have completed their training and demonstrated competence. Feedback from workers and annual reviews are integrated into revised WCDs as part of the ISM cycle.

6.1.2 Contractor Work Planning

ALCF utilizes subcontractors to perform various maintenance and repair work in the data center. Work planning for ALCF subcontractors and vendors is captured using Job Safety Analysis (JSA) plans. In CY 2023, ALCF developed 25 JSAs to identify and plan work activities. Each JSA outlines the collaborative planning and hazard identification process for the contractor's work activities and are closely managed by ALCF to ensure compliance with scope limits and hazard mitigation measures. Like WCDs, JSAs employ a graded approach, involving higher levels of management and SMEs for riskier or more complex work.

6.1.3 Training

ALCF personnel—including employees, students, and subcontractors at all levels—must be qualified and trained to perform their work assignments and demonstrate proficiency to perform those duties. Qualification and training criteria are subject to ongoing review by management for adequacy and effectiveness. ALCF staff members utilize the Job Hazard Questionnaire (JHQ) to capture the staff's roles and associated training. The JHQ is reviewed annually or whenever an employee's job assignment changes. Additionally, training requirements are laid out in WCDs and JSAs, and work can begin only if workers are qualified.

ALCF outlines its training requirements, including On-the-Job Training (OJT), in its WPC manual. Several task-specific jobs, such as using lift equipment, require ALCF to document as OJT and perform before an employee is qualified. Some very specific tasks, such as those requiring a QEW, require formal technical training that the laboratory SMEs would review prior to the start of work for both ALCF staff and subcontractors.

6.2 Safety Performance

6.2.1 Walkthroughs and Assessments

ALCF adheres to the plan-do-check-act (PDCA) cycle to integrate ISM functions, with management assessments and inspections serving as the check phase for feedback in line with ISM principles.

ALCF conducts biannual health and safety inspections led by the ALCF Director and the Environmental, Safety and Health (ESH) Coordinator. Inspection findings, issues, and improvement opportunities are tracked in software programs called Cority and Prism. Trending safety data from Cority is used to help identify risks and near misses to determine focus areas for further assessment.

Prism is the laboratory's performance management software that helps track issues and items from inspections, management assessments, and events (incidents or off-normal). The graded approach is utilized in Prism to identify, evaluate, correct, review, and analyze trends. ALCF uses both systems to ensure the integrity and quality of the work its personnel perform.

6.2.2 Ergonomics

Argonne has a robust ergonomics program with on-site physical therapy and an ergonomics SME. The CELS directorate has its own ergo room and Ergonomics Coordinator who works closely with the laboratory's physical therapist to perform ergonomic evaluations ALCF staff. With a local ergo room, the coordinator loans out ergonomic task chairs, keyboards, mice, and fatigue mats after an evaluation to find the best fit for each employee. Recommendations and a follow-up report are then sent out, so both the employee and supervisor have informed information for purchasing products. In 2023, 7 of the 26 CELS office ergonomic evaluations were for ALCF staff, and an additional 5 ALCF staff members had at-home office evaluations.

6.2.3 Safety Shares

Safety shares are provided at every biweekly all-hands meeting (AHM) of the Operations team. In addition to the safety shares, work planning, and upcoming projects are discussed, with attendance and input from the ESH Coordinator and Data Center Manager.

Safety protocols are reviewed with all ALCF visitors. Tours of the data center are planned and coordinated to ensure that visitors and tour groups receive important safety information, such as concerning proper attire and continued COVID controls, in advance of their arrival.

6.3 Incidents and Safety Pauses

In CY 2023, ALCF experienced zero Days Away Restricted or Transferred (DART) cases, zero first aid cases, and zero Total Recordable Cases (TRCs). The recorded cases over the past five years are summarized in Table 6.1.

Table 6.1 ALCF Injury and Illness Cases FY 2019 thru FY 2023

	DART	First Aid	TRC
2019	0	1	0
2020	0	1	0
2021	0	0	0
2022	0	0	1
2023	0	0	0

6.3.1 Aurora Install

The 2023 install of the Aurora compute blades took place without incident. Management of all system component deliveries and placement was planned, executed, and coordinated by ALCF staff, the Project Management Organization (PMO), TCS building management, and the subcontractors.

- Two safety pauses occurred during the delivery period to address the following concerns:
  - The CPU swap scope changed on January 23, 2023. Before resuming work, the subcontractor worked with ALCF staff to confirm that the procedure fell within their current scope of work.
  - On March 24, 2023, work paused to update and review a JSA with subcontractors to account for work in phases 2 and 3 being performed simultaneously.

This page intentionally left blank.

Section 7. Security

(a) Has the Facility demonstrated effective cyber security practices?

(b) Does the Facility have a valid cyber security plan and Authority to Operate?

(c) Does the Facility have effective processes for compliance with applicable national security policies related to Export Controls and foreign visitor access?

ALCF Response

ALCF demonstrates effective cybersecurity practices through our commitment to improving cybersecurity practices and expanding staff involvement in cybersecurity efforts. This can be seen in the efforts ALCF is putting into formalizing a cybersecurity program within ALCF and expanding cybersecurity working groups to incorporate staff from other groups. This resulted in ALCF having zero security incidents this year and allowed ALCF to reduce security risks. ALCF has a valid Authority to Operate (ATO) and an effective process for compliance with export controls and foreign visitor access.

7.1 Demonstrating Effective Cybersecurity Practices

ALCF maintains a strong cybersecurity posture and is vigilant in watching for evolving threats, vulnerabilities, and potential security incidents. ALCF's security team partners with OLCF, NERSC, and other facilities to jointly assess the risks of security incidents unique to HPC facilities. Working together, the facilities can quickly identify, coordinate, and put mitigations in place for new vulnerabilities in HPC systems.

In 2022, security team members from ALCF, OLCF, NERSC, and ESnet formalized the Secure ASCR Facilities (SECAF) working group. In 2023, ALCF continued the process of formalizing groups and practices by participating in two newly formed cybersecurity working groups. The first is at the CELS directorate level, bringing all individuals with a cybersecurity responsibility together. The second group is an ALCF internal security working group where we bring together at least one member from each team in Operations to contribute to improving cybersecurity at ALCF. This year, the ALCF security working group has started efforts to update and modernize our incident response plan and create a cybersecurity-focused risk registry. This is in addition to our cybersecurity staff being included in the Argonne Cyber team's meetings. These efforts to establish and foster cybersecurity efforts among staff show ALCF's commitment to building and demonstrating effective cyber practices. SECAF also worked with Program Managers to put on a cybersecurity-focused tabletop exercise. The focus of this was to identify gaps in communication that may happen during an incident and work to remediate gaps so incident responses can have clear lines of communication in the future.

In CY 2023, ALCF continued to improve its cybersecurity practices to ensure an effective cybersecurity program. ALCF also continued to improve tools, both locally and at the laboratory level, to increase cybersecurity capabilities. ALCF's security staff and Argonne's CSPO staff have a good working relationship and regularly discuss future initiatives. The following sections include examples of tools, capabilities, and procedures that demonstrate ALCF's effective cyber security practices.

In CY 2023, there were zero cybersecurity incidents on ALCF systems. ALCF's cybersecurity personnel take a proactive approach to problem management. Examples include efforts to:

1. Conduct privileged access reviews across the environment to help ensure that everyone has the appropriate level of access.
2. Conduct reviews of how and where data are stored to help ensure that data are accessible only to those with proper authorization.
3. Educate users and staff about how to prevent password exposure.
4. Educate developers on secure coding best practices via internal discussions/reviews and external training courses.
5. Integrate security auditing into developer workflows to identify security issues early in the development life cycle.
6. Keep the National Institute of Standards and Technology (NIST) certification package up to date, including NIST 800-53, 800-34, 800-30, and 800-18 compliance documents.
7. Archive and delete obsolete data.
8. Set password rotation policies with complexity requirements for ALCF systems and verify compliance.
9. Monitor new vulnerabilities to ALCF systems.
10. Conduct penetration testing of both internal- and external-facing web applications and recommend security improvements.
11. Expand Continuous Integration testing pipeline work, moving toward the time when workflows will be required to pass testing prior to merging code in.
12. Work with ALCF's scheduler vendor, Altair, to contribute security fixes upstream to their code base to ensure that ALCF's systems running PBS are more secure.
13. Work on developing frameworks for web-based or interrupt-driven HPC workloads to provide the best service to scientific users while ensuring systems are kept secure.
14. Work with system admin teams to collaborate on configuration management tooling and configurations of various systems.

Additionally, ALCF takes a layered approach to security in areas that fall within Argonne's CSPO-managed cybersecurity domain. The laboratory makes the following security services available to ALCF:

1. Cross-laboratory data sharing that CSPO integrates into automatic network blocking rules to keep systems secure.
2. Log analysis capabilities that allow CSPO to block IP addresses or bad actors. This provides ALCF additional protection from attacks on Argonne networks as they are detected.

3. CSPO provides Cloudflare, a powerful web application firewall (WAF), which protects ALCF's public sites from known attacks and ensures encryption requirements are met.
4. CSPO provides a Tenable security center (Tenable.sc) instance that ALCF can tie Nessus scanners into and allows leveraging CSPO's public and internal network scanners. This provides a view of network vulnerabilities in ALCF's environment.
5. A laboratory-managed network border firewall protects ALCF applications from both the public Internet and Argonne's internal networks.
6. CSPO helps to manage publicly visible vulnerabilities by automatically alerting ALCF when new vulnerabilities are detected and provides guidance for patching the system or removing public conduits.
7. CSPO maintains lab-wide security policy documentation under the Cyber Security Program Plan. ALCF, as a NIST Major Application, also maintains security documentation tailored to the facility, which it inherits from Argonne's standard security documentation.
8. Access to NetFlow data to help troubleshoot and investigate network-related issues across the laboratory's networks.
9. Access to cybersecurity tools such as Axonius for asset management and maintaining an inventory of installed software.
10. CSPO is running a cybersecurity hygiene improvement program across the lab that kicked off in CY 2023.

Some ALCF proactive measures revealed security vulnerabilities that were promptly addressed and fixed, usually within days of their discovery. Immediately after detection, ALCF staff investigated all relevant logs to determine whether the security vulnerability had been exploited. In CY 2023, none of the issues investigated were found to have been exploited. Examples of the security issues that were detected and their ensuing mitigations are as follows:

1. **Issue:** Passwords in some applications were found to be stored insecurely.
   Mitigation: Evaluation of the application data integrity showed no unauthorized access. ALCF changed the way passwords were stored and changed all related passwords as a precaution against potential exploitation.
2. **Issue:** An application run by ALCF had the possibility of logging secrets to a world-readable log file in a specific debug configuration.
   Mitigation: ALCF staff worked with the vendor to communicate the issue and ensure it was fixed in future releases. As an immediate mitigation, ALCF locked down the log files and audited the systems to see if any secrets were logged. The audit revealed that no secrets were exposed.
3. **Issue:** An individual was attempting to social engineer their way onto ALCF resources.
   Mitigation: ALCF staff noticed some attempts to create accounts that were not legitimate, so they rejected those requests. Upon investigation, ALCF was able to identify the person behind these attempts, and they were informed that their actions were not acceptable. The individual claimed to be working under the DOE bug bounty program, but it was clear that their actions were beyond the program's scope.

4. **Issue:** After moving an ALCF service behind Cloudflare, IIS IP-based ACLs were not properly applied, exposing what was intended to be an internal-only service to the internet.
   Mitigation: ALCF staff identified the issue and worked with Argonne's Business and Information Services (BIS) Division to resolve it and reinstate proper IO ACLs for the service.

ALCF will continue to proactively investigate security issues and monitor and respond to all vulnerabilities. Plans for improving the security of ALCF resources include:

1. Retiring obsolete services and data.
2. Verifying that strong encryption is used everywhere in the ALCF environment, and that plain text protocols are not used for production systems.
3. Improving real-time log analysis techniques.
4. Increasing the visibility of required security updates on systems.

ALCF is also planning to formalize cybersecurity practices to further fortify the cybersecurity program. This is being done through a variety of efforts, including:

1. Starting an ALCF-specific cybersecurity working group to bring all operational teams together in order to ensure cybersecurity coverage in all aspects of ALCF operations.
2. Creating a modernized incident response plan specific to ALCF that will align with the Argonne-wide incident response plan.
3. Developing a framework for a cybersecurity-specific risk registry for ALCF that will use risk modeling frameworks to quantify the level of risk.
4. Actively participating in the new directorate-wide cyber security working groups.
5. Expanding in-house security expertise by hiring two new roles within the CELS directorate. The first is a Cybersecurity Lead at the CELS level who will be able to assist ALCF. The second is an Application Security Engineer who will be dedicated to helping improve security in ALCF-developed software.

CSPO conducts an annual internal audit called a Division Site Assist Visit (DSAV) that typically assesses divisional compliance with NIST-800-53 controls, with each year's DSAV covering roughly one-third of the controls. In CY 2023, CSPO took an alternative approach to respond to presidential Executive Order EO 14028, OBM Memorandum M-22-09, and CISA Zero Trust Maturity Model by hiring a third-party assessment team to conduct a Zero Trust Architecture gap analysis of ALCF. The intent is to use this assessment to get a baseline understanding of where ALCF aligns in its zero-trust maturity journey and use this to understand what future efforts toward zero-trust would look like. This assessment was conducted using a methodology that follows CISA's guidance and Zero Trust Maturity Model. Overall, ALCF was assessed to be near an initial maturity level for zero trust. This shows that while ALCF has not put a focus on zero trust historically, the security methodologies put in place are moving ALCF in the right direction. Some recommendations to come out of this report included:

1. Investigate expanding into software-defined computing and containerization for rapid scalability and stability that would come with solutions like Kubernetes.
2. Investigate how Data Loss Prevention (DLP) tools could play a role in ALCF to help protect user data while ensuring we don't restrict the ability of users to move data in and out of the facility as needed.
3. Analyze ALCF's disaster recovery plan to ensure that it has detailed plans for how and when to restore systems. Additionally, ALCF should ensure that backups are regularly tested to ensure that the processes and procedures are understood by all staff that would need to be involved.
4. Investigate the use of micro-segmentation to enhance network security and ensure that in the event of an incident, a breach is contained to a small network segment.
5. Investigate improvements to existing log management solution to assist ALCF staff in understanding logs and gaining operational insight into the environment.

CSPO has committed to continuing the DSAV process in 2024. ALCF will continue to work with CSPO to verify that all Argonne security standards and practices are met.

7.2 Cybersecurity Plan

Argonne's Authority To Operate (ATO) includes the ALCF as a major application and was renewed on April 26, 2023. It is valid as long as Argonne maintains robust, continuous monitoring of its Cyber Security Program as detailed in the ATO letter, which is included at the end of this section.

7.3 Foreign Visitor Access and Export Controls

ALCF follows all DOE security policies and guidelines related to export controls and foreign visitor access.

Argonne is a controlled-access facility, and anyone entering the site or accessing Argonne resources remotely must be authorized. ALCF follows Argonne procedures for collecting information about foreign nationals who require site access or remote (only) computer access. All foreign nationals are required to have an active and approved ANL-593 in order to have an active ALCF account. Users can access ALCF resources only with an active ALCF account.

To apply for an ALCF account, the user fills out a secure webform in the ALCF Account and Project Management system, Userbase 3 (UB3), providing details such as legal name, a valid e-mail address, work address, phone number, and country of citizenship. They also identify the ALCF project with which they are associated. In addition, all foreign nationals (non-U.S. citizens) are required to fill out their personal, employer, demographic, and immigration/U.S. Citizenship and Immigration Services (USCIS) information in Argonne's Visitor Registration system, which is integrated with UB3. After the user submits their account application request, an e-mail is sent to the user's project PI for approval. Once the ALCF Accounts team receives the approval from the project PI, if the user is a foreign national, the user's details are electronically attached to an ANL-593 form and submitted to the Foreign

National Access Program (FNAP) Office for review. The FNAP Office is responsible for overseeing compliance with laboratory rules and DOE directives.

The ANL-593 form records the type of work the user will perform, including the sensitivity of the data used and generated. The ANL-593 must be approved by Argonne Cyber Security, FNAP, the Argonne Office of Counterintelligence, and the Argonne Export Control Office. Argonne's foreign visitor and assignments process integrates with the DOE Foreign Access Central Tracking System (FACTS), which documents and tracks access control records of international visits, assignments, and employment at DOE facilities and contractor sites. Once the ANL-593 form for the user is approved, the UB3 database is automatically updated with the user's ANL-593 start and end dates. The ALCF Accounts team then creates the user account and notifies the user. Any changes to the ANL-593 dates are automatically updated in UB3. Accounts are suspended if the user's ANL-593 expires.

ALCF allows only a limited subset of export control data on ALCF systems. ALCF works closely with Argonne's Export Control Office to complete a detailed security plan for what export control classifications are allowed and what security measurements are required for each instance of export-controlled data. If, at any time, the ALCF wants to allow new classifications of export control data on its systems, a new security plan must be created and approved by Argonne's Export Control Office and Argonne Cyber Security.

Department of Energy
Office of Science
Argonne Site Office
9800 South Cass Avenue
Lemont, Illinois 60439

April 26, 2023

VIA ELECTRONIC MAIL

Dr. Paul K. Kearns
Director, Argonne National Laboratory
UChicago Argonne, LLC
9700 South Cass Avenue
Lemont, Illinois 60439

SUBJECT: AUTHORITY TO OPERATE FOR THE ARGONNE NATIONAL LABORATORY INFORMATION TECHNOLOGY INFRASTRUCTURE

Dear Dr. Kearns:

Since the last renewal of this Authority to Operate (ATO) in April 2022, the Laboratory has conducted regular, continuous monitoring briefings and has kept me informed of changes in cybersecurity risk in accordance with the Risk Management Framework. The Laboratory has been testing at least 60 security controls annually on a rotating basis as part of the self-assessment program. Argonne National Laboratory has initiated an effort to bring its Cybersecurity Program Plan to the NIST SP 800-53 Revision 5 standard. Progress will be reported to me through regular, continuous monitoring briefings.

This has demonstrated that the Laboratory's IT Infrastructure is operating at an acceptable level of risk, and I am, therefore, the Authorizing Official, renewing the Authority to Operate for the General Computing - Low enclave and the General Computing - Moderate enclave. The IT Infrastructure continues to contain the following major applications, which have components in both enclaves:

- Accelerator Control Systems (APS and ATLAS)
- Argonne Leadership Computing Facility
- Business Systems
- Sensitive Information
- Cyber Federated Model (CFM)

This ATO will remain in effect as long as the Laboratory carries out continuous monitoring under the Risk Management Framework and there are no significant changes to Argonne's IT Infrastructure. The Laboratory should retain a copy of this letter with the security authorization package.

If you have any further questions, please contact me or have your staff contact Michael Wisniewski at michael.wisniewski@science.doe.gov or at (630) 252-8778.

A Component of the Office of Science

Sincerely,

Roger E. Snyder Digitally signed by Roger E. Snyder Date: 2023.05.08 10:07:43 -05'00'

Roger E. Snyder
Acting Manager

cc: R. Aker, ASO
M. Wisniewski, OIM
F. Healy, OIM
S. Hannay, ANL
M. Skwarek, ANL
R. Denney, ANL
D. Loda, ANL
J. Volmer, ANL

File: SEC.08
ASODOCLOG2023-138

Section 8. Mission Impact, Strategic Planning, and Strategic Engagements

(a) Are the methods and processes for monitoring scientific accomplishments effective?

(b) Has the Facility demonstrated effective engagements with technology vendors and /or engaged in effective research that will impact next generation technology relevant to the facility's mission?

(c) Has the Facility demonstrated effective engagements with critical stakeholders (such as the SC Science Programs, DOE Programs, DOE National Laboratories, SC User Facilities, and/or other critical U.S. Government stakeholders (if applicable)) to both enable mission priorities and gain insight into future user requirements?

ALCF Response

The science accomplishments of INCITE, ALCC, and DD projects demonstrate ALCF's impact in enabling scientific breakthroughs. ALCF staff members have worked effectively with individual project teams to adapt their simulation, data analytics, and machine learning codes to run efficiently in the ALCF environment, enabling scientific achievements that would not have been possible otherwise.

In this section, ALCF reports:

- Scientific highlights and accomplishments;
- Research activities/vendor engagements for future operations; and
- Stakeholder engagement.

8.1 Science Highlights and Accomplishments

ALCF employs various methods and processes for monitoring scientific accomplishments. Monthly scientific highlights (mostly from the catalyst team) are collected and documented in a quarterly report. The determination and coordination of scientific highlights are managed by ALCF's applications team, comprised of members of the catalyst team, the data science team, and the performance engineering team, and in consultation with ALCF's Director of Science. Other sources of scientific highlights include technical communications between ALCF staff members and a project PI or co-PI, significant findings reported in a high-impact publication or conference presentation, and a catalyst's own involvement in a publication.

ALCF tracks and annually reports the number of peer-reviewed publications resulting (in whole or in part) from using the facility's resources. For ALCF, tracking takes place during a period of five years following the project's use of the facility. This may include publications in press or accepted but does not include papers submitted or in preparation. The count is a reported number, not a metric. The facility may report other publications where appropriate. Methods used for gathering publication data include asking users to verify or update ALCF's online publications database and conducting Google Scholar and Crossref searches.

Table 8.1 shows the breakdown of refereed publications based, in whole or in part, on the use of ALCF resources, and highlights those appearing in major journals and proceedings. These include one publication in *Nature*, six in *Nature Communications*, two in *npj Climate and Atmospheric Science*, three in *npj Computational Materials*, three in *Scientific Data*, and two in *Scientific Reports* (combined in the Nature journals category in the table below); six in *Physical Review Letters*; one in 38th International Conference, ISC High Performance 2023 (ISC); and three in the proceedings of the 2023 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC). Table 8.2 shows updated publication counts from prior years based on new information received after the prior year's OAR deadline.

Table 8.1 Summary of Users' Peer-Reviewed Publications in CY 2023

Nature Journals	Physical Review Letters	ISC	SC	Total 2023 Publications
17	6	1	3	240

Table 8.2 Summary of Users' Peer-Reviewed Publications for 5-year Moving Window

OAR Year	CY 2019	CY 2020	CY 2021	CY 2022	CY 2023
Total Publications	288	257	249	213	240

Science Highlights

Scientific highlights are short narratives that illustrate the user facility's contribution to advancing science. Highlights may describe a research accomplishment or significant finding from either a current project or from a project originating in a previous year, as data analysis may occur several months after the computational campaign has been completed.

Each project highlight includes a figure and a bar graph showing time allocated and time used: the first number in the graph title is the allocation total and the second (in parentheses) shows how much time the project used. The individual bars represent the percentage of time used on the fraction of the machine shown below the bar, which are "no capability," "low capability," and "high capability" from left to right.

8.1.1 Allegro-Legato: Scalable, Fast, and Robust Neural-Network Quantum Molecular Dynamics via Sharpness-Aware Minimization

The Science

Neural-network quantum molecular dynamics (NNQMD) simulations based on machine learning are revolutionizing atomistic simulations of materials. Their state-of-the-art quantum-mechanical accuracy in the latest NNQMD model, Allegro, has been combined with orders-of-magnitude greater speed based on Allegro's spatially localized descriptors. The resulting Allegro-Legato (fast-"smooth") model was more stable while maintaining the same inference speed and accuracy.

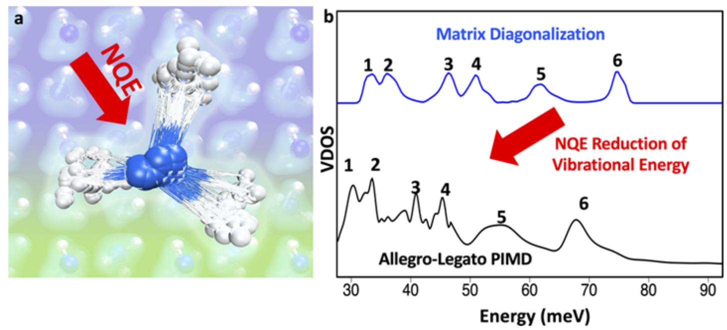
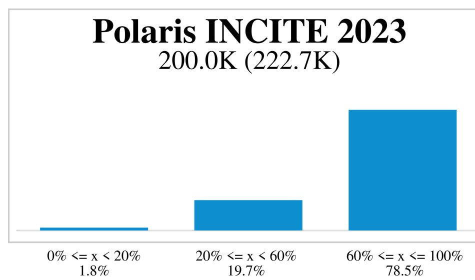


Figure 8.1 Computed vibrational spectra of ammonia: (a) representative image of ammonia molecule in a path-integral molecular dynamics (PIMD) simulation; and (b) vibrational spectrum computed at (top) zero temperature without nuclear quantum effects (NQEs) and (bottom) at finite temperature with the Allegro-Legato PIMD simulation.

Image courtesy of Thomas Linker (USC).

Ammonia (NH_3) has a higher energy density than liquid hydrogen and can be stored at a much lower and less energy-intensive temperature. Thus, NH_3-based fuel technologies are well positioned with infrastructure already in place for annual NH_3 production. However, accurate prediction of thermodynamic behavior for crystals and liquids is critical for their applications in biological, pharmaceutical, and energy systems. In Allegro-Legato simulations of NH_3, the expected softening of high-energy vibrational modes was observed at finite temperature with inclusion of nuclear quantum effects (NQEs) (Figure 8.1); the USC team found this result to be consistent with high-end neutron experiments performed at ORNL.



The Impact

Allegro-Legato allows much larger spatiotemporal-scale NNQMD simulations than are otherwise possible. Unlike typical MD methods for effective long-time sampling (e.g., for protein folding), which can skew dynamical information, Allegro-Legato enables "true" long-time dynamics that can be directly compared with high-resolution spectroscopic experiments and thus overcomes the prohibitive computational demand of accounting for subtle NQEs in simulations, which is essential for accurately calculating thermodynamic properties. Achieving synergy between the most advanced neutron experiment and leadership-scale NNQMD simulations lays a foundation for a green and sustainable ammonia-based fuel technology.

Summary

The Allegro model provides much higher accuracy and speed but suffers a fidelity-scaling problem where the number of unphysical predictions grows with the number of particles, making it difficult to run stable simulations for long. The USC team implemented a sharpness-aware minimization (SAM) solution to this problem, resulting in Allegro-Legato's state-of-the-art accuracy and speed with drastically improved fidelity scaling and systematically delayed time-to-failure — essential to extracting meaningful scientific knowledge from large-scale simulations on leadership-scale parallel computers. Allegro-Legato's scalable parallel implementation has excellent computational scaling; and GPU acceleration combining accuracy, speed, robustness, and scalability allows large spatiotemporal-scale NNQMD simulations on exascale computing platforms. Implemented in the NNQMD code RXMD-NN, Allegro-Legato achieved a weak-scaling parallel efficiency of 0.91 on 480 nodes of Polaris for a system with ~13 million particles. On a single Polaris node, the code achieves a 7.6× speedup using four GPUs compared to the single 32-core CPU. Efforts are underway to ready the RXMD-NN for Aurora. Plus, with the inclusion of NQEs, Allegro-Legato path-integral molecular dynamics (PIMD) correctly showed a softening of NH_3's high-energy, intermolecular modes in the 30- to 90-meV region at finite temperature, explaining the high-end neutron-scattering observations performed at ORNL.

ALCF Contribution: The ALCF catalyst contributed to the integration of RXMD and PyTorch packages and helped to scale up NNQMD simulations to full use of Polaris.

Contact
Aiichiro Nakano
Collaboratory for Advanced Computing & Simulations (CACS)
University of Southern California
Email: anakano@usc.edu

Publication
Ibayashi, H., T.M. Razakh, L. Yang, T. Linker, M. Olguin, S. Hattori, Y. Luo, R.K. Kalia, A. Nakano, K. Nomura, and P. Vashishta. (2023). "Allegro-Legato: Scalable, Fast, and Robust Neural-Network Quantum Molecular Dynamics via Sharpness-Aware Minimization." In: Bhatele, A., Hammond, J., Baboulin, M., and Kruse, C. (eds.), *High Performance Computing*. ISC High Performance 2023. Lecture Notes in Computer Science, vol. 13948. Springer, Cham. DOI: https://doi.org/10.1007/978-3-031-32041-5_12

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: November 2023

8.1.2 First Principles Simulation of Hypersonic Flight

The Science

Hypersonic flight, the ability to fly at more than five times the speed of sound, has the potential to revolutionize technologies for national security, aviation, and space exploration. However, a fundamental understanding of its aerothermodynamics is needed to enable technological advances.

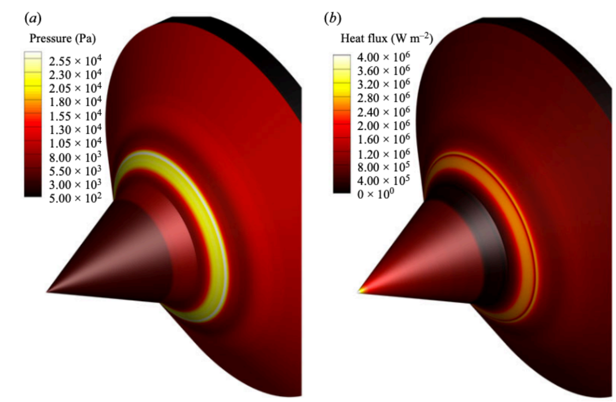
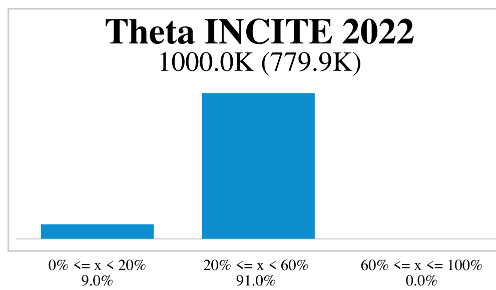


Figure 8.2 Three-dimensional view of the surface pressure and heat flux on the double-cone geometry.
Image courtesy of P. Valentini.

Strong shockwaves can cause the excitation of internal energy modes and chemical reactivity in the shock-heated gas. The rate of these phenomena is comparable to the fast, local flow time, causing the flow to be in thermal and chemical nonequilibrium. This nonequilibrium can even persist at the molecular level, with the shock-heated gas having a non-Boltzmann character [1,2]. The non-Boltzmann nature of the gas, in turn, affects the chemical reactivity of the flow in a nonlinear manner that even the most sophisticated computational fluid dynamics (CFD) models cannot resolve [2]. Therefore, a more fundamental approach is required to properly characterize the state of the gas in the shock layer. Accurate estimation of the thermal state and the chemical composition of the gas enveloping the vehicle is crucial to optimally designing the thermal protection system. Direct molecular simulation (DMS) [2,3] provides a unique opportunity to simulate fluid flows using only quantum mechanically guided molecular interactions. However, due to the cost of these simulations, DMS calculations had been relegated to use on small-scale systems, although ALCF INCITE resources have made large-scale simulations possible [2,3]. These simulations help researchers understand (1) how adequately reduced-order formulations, used in CFD codes, can capture the strong coupling between the fluid mechanics of the gas flow, the local gas-phase thermochemical non-equilibrium, and the transport properties of the high-temperature gas; and (2) how reduced-order formulations can be improved to obtain more reliable answers [1].



The Impact

The team used a modified version of the open-source SPARTA Direct Simulation Monte Carlo (DSMC) code on Theta to carry out DSMC and DMS of hypersonic flow experiments. Their goal was to conduct simulations relying solely on molecular-level interactions that are modeled using quantum mechanics, providing a fundamental comparison with experiments, and well-characterized solutions that can be used as benchmarks for reduced-order models. The team's work has demonstrated that their DSMC code can find solutions for these kinds of flows and, moreover, that it can seamlessly describe internal energy nonequilibrium for all modes.

Understanding the complex aerothermodynamics of hypersonic flight could provide insights to help inform the design of safer and more efficient technologies for space travel and defense.

Summary
The aerothermodynamics of hypersonic flight is extremely complex. A research team from the University of Dayton Research Institute and Air Force Research Laboratory is using ALCF supercomputers to study the complex thermal environment that vehicles in hypersonic flight encounter. They have developed large-scale computational tools that enable them to obtain entire flow fields solely from the fundamental interactions of atoms and molecules in the hypersonic, gaseous environment. The computations in this research produced the highest-fidelity CFD results obtained to date and will help to enable hypersonic transport and ready access to space.

In a new study published in the *Journal of Fluid Mechanics* [1], the team detailed a large-scale, fully resolved DSMC computation of a nonequilibrium, reactive flow of pure oxygen over a double cone (a canonical hypersonic test case) (Figure 8.2). The researchers used their highly accurate DMS method to obtain first-principles data to inform the parameters of the thermochemical and transport collision models. Their computations show good agreement with heat flux and pressure measured on the test article during the experiment. The computation also provided molecular-level insights regarding the nonequilibrium distribution of energy in the kinetic and vibrational modes in the shock layer. Such results show the importance of particle methods in verifying physical assumptions made by reduced-order models. Advancing our understanding of the complex aerothermodynamics of hypersonic flight could lead to the design of safer and more efficient technologies for space travel and defense.

ALCF Contribution: The ALCF catalyst provided cobalt scripts for submitting ensemble jobs; helped boost job priority to meet deadlines for conferences, workshops, and review meetings; and suggested affinity mappings to help manage workloads with increased memory requirements.

Contact
Maninder Grover
Air Force Research Laboratory (AFRL) and University of Dayton Research Institute (UDRI)
mgrover1@udayton.edu
ASCR Allocation PI (INCITE): Maninder Grover (AFRL & UDRI)

Publications
[1] Valentini, P., M.S. Grover, A.M. Verhoff, and N.J. Bisek. (2023). Near-Continuum, Hypersonic Oxygen Flow over a Double Cone Simulated by Direct Simulation Monte Carlo Informed from Quantum Chemistry. *Journal of Fluid Mechanics*.
[2] Grover, M.S., A.M. Verhoff, P. Valentini, and N.J. Bisek. (2023). First Principles Simulation of Reacting Hypersonic Flow over a Blunt Wedge. *Physics of Fluids*, 35(8). (Featured Article.)
[3] Grover, M.S., P., Valentini, N.J. Bisek, and A.M. Verhoff. (2023). First Principle Simulation of CUBRC Double Cone Experiments. In: *AIAA AVIATION 2023 Forum* (p. 3735).

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: October 2023

8.1.3 Assessment of Turbulent Prandtl Number for Heavy Liquid Metal Flow in a Bare Rod Bundle by Leveraging LES Simulation

The Science

Nuclear reactor vendor Westinghouse Electric Company, together with an international team, is developing its next-generation, high-capacity nuclear power plant based on lead-cooled fast reactor (LFR) technology. Accurate modeling and simulation of heat transfer and mixing in the heavy liquid metal (HLM) coolant is needed to help prepare the technology for licensing. HLM (e.g., lead) has different properties from other fluids, and the conventional choice of turbulent Prandtl number (Prt) in existing engineering turbulence models is not valid for HLM flow simulations. This campaign provides valuable high-fidelity reference data for comparison and calibration with less computationally expensive models.

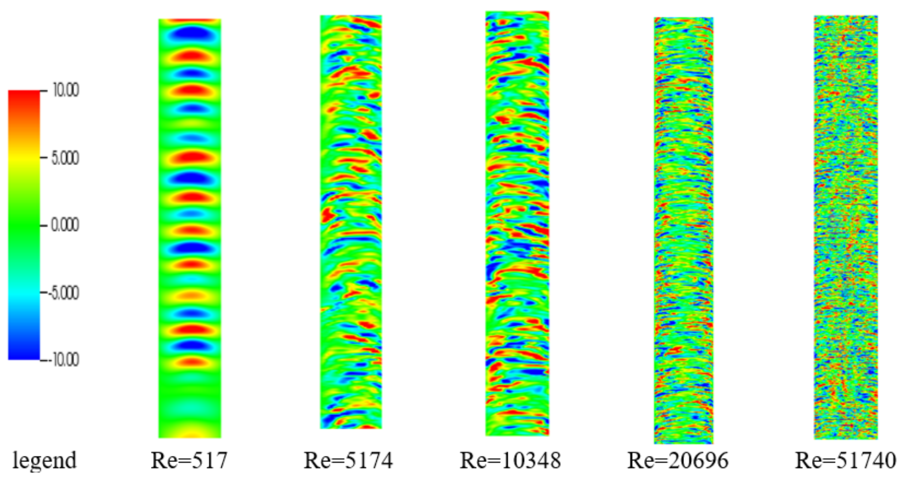
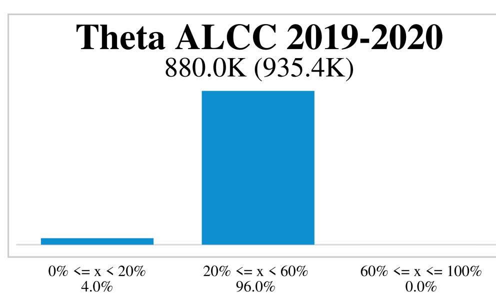


Figure 8.3 LES predictions of the turbulent structure (flow oscillations) in a lead-cooled rod bundle with varying Reynolds number (Re). As Re increases, the flow oscillations become smaller in scale and more chaotic. These simulations can be used as benchmarks for lower-fidelity calculations, which require tuning of the turbulent Prandtl number.

Image Courtesy of Yiqi Yu (Argonne).

The Impact



Large Eddy Simulations (LESs) using Nek5000 on Theta were applied to nuclear fuel rod bundles with HLM flow with the goal of providing detailed insight into flow physics and heat transfer mechanisms. LESs do not require Prt to model turbulence-driven heat transfer, and thus can be used as benchmarks for selecting Prt in a less computationally expensive Reynolds Averaged Navier Stokes (RANS) model that requires this parameter. This project provides detailed insight into the heat transfer mechanism in a nuclear fuel rod bundle with HLM flow. The assessment and selection of appropriate Prt can improve the accuracy of simulation for advanced nuclear reactors such as LFRs, which are potential candidates for next-generation nuclear reactors in the U.S. and abroad. Data are rarely reported or available for heat transfer of HLM in rod bundles with LES (Figure 8.3).

Summary

In HLM flow, the thermal boundary layer is much thicker than the viscous boundary layer — unlike in more conventional fluids like water. Analysis performed on a prototypical lead-cooled fast reactor assembly with different Prt values indicates that inappropriate Prt can introduce error in the Nusselt number (which is a measure of heat transfer) by up to 44%. Major over- or under-estimation of heat transfer directly follows from the error in the Nusselt number. Therefore, the selection of Prt contributes a significant impact to the accuracy of LFR simulations. While computationally expensive, LESs can provide detailed insight into the physics of the flow and the associated heat transfer in a rod bundle with HLM flows. Detailed temperature distributions obtained with less expensive RANS models and LES were compared to better understand the deviation introduced by the turbulence model. The analysis shows that the RANS model with Prt=1.5 shows the best agreement with LES on the prediction of local temperature distribution and global Nusselt number. The valuable reference data (temperature distribution and Nusselt number with different Reynolds numbers) appear in a published journal article. Other researchers can use these data to verify or calibrate less computationally expensive RANS models. This work could motivate the development of a more sophisticated local turbulent Prandtl number model.

ALCF Contribution: The Nek5000 code has received significant support from ALCF staff, including porting and optimizing the code for Theta.

Contact

Yiqi Yu
Argonne National laboratory
yyu@anl.gov
ASCR Allocation PI: Emily Shemon (ANL)

Publication

Yu, Y., E. Shemon, and E. Merzari. (2023). "LES Simulation on Heavy Liquid Metal Flow in a Bare Rod Bundle for Assessment of Turbulent Prandtl Number," *Nuclear Engineering and Design*, volume 404.

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: June 2023

8.1.4 A New Way to Predict Droughts

The Science

Extreme drought conditions in the United States are expected to increase with global warming; however, predicting potential changes can be challenging in regions with large variability in moisture levels across small regions, such as interconnected urban, suburban, and agricultural land regions. Unlike hurricanes, which have a rigorous classification scale, droughts are classified using various methodologies with diverse fidelity ranges. Based on analysis of a large community dataset and projections from three global climate models, a team has developed a new measure to help identify short-term droughts and predict future extreme droughts, including "flash drought" events that develop over a quick onset period of as short as a few weeks.

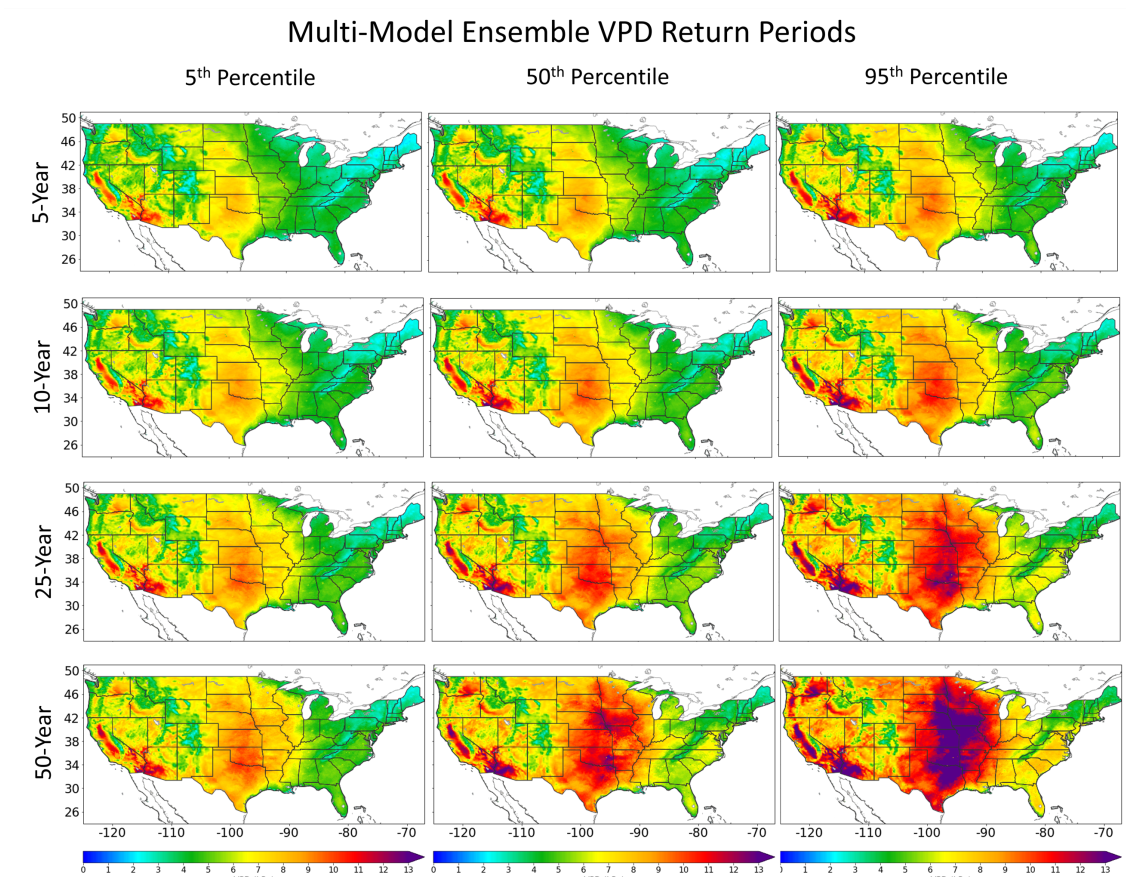
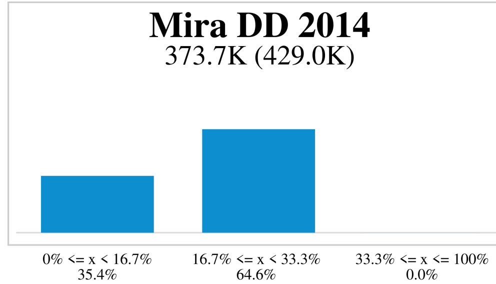


Figure 8.4 The 5-, 10-, 25-, and 50-year return periods in the United States for the 5th, median (50th), and 95th percentile of the sampled model ensemble for VPD return periods.

Image courtesy of Brandi L. Gamelin (Argonne).

The Impact



Using the new and uncomplicated methodology based on highly detailed regional climate models, the team found that droughts are likely to be exacerbated by global warming, particularly in regions like the Midwest, northwestern U.S., and California's Central Valley. The projected worsening of droughts because of global warming is likely to have significant consequences affecting water resources, wildfire activities, and crop loss. This technique, with its resolution of a single day, thus has more precise statistics, provides earlier drought detection, and can help researchers understand "flash drought" events that develop quickly and last only weeks.

Summary

Drought classification currently ranges from reports from farmers to noted precipitation deficits to assess which regions were experiencing a drought, which may not be adequate for areas that experience low precipitation to begin with. Typical drought monitoring relies on weekly or monthly data. Vapor pressure deficit (VPD), on the other hand, can be calculated based on a combination of the daily maximum temperature and minimum relative humidity and does not rely on precipitation deficit to identify drought. VPD is a measure of the difference between how much water vapor the air can hold when saturated and the total amount of water vapor available. Hotter air will typically have higher VPD than colder air because cold air retains less moisture. An extended period of higher-than-average VPD indicates that a drought is occurring.

A new drought index was also developed using VPD. The standardized VPD drought index (SVDI) was calculated with the NLDAS dataset. To validate the new index, it was compared to four other drought indices, and SVDI was shown to accurately identify the timing and magnitude of short-term droughts. VPD was also calculated from weather research and forecasting (WRF) simulations using three downscaled global climate models (WRF CCSM, WRF GFDI, and WRF HadGEM) for three separate timeframes (present, mid-century, and late-century) to identify return periods and the locations for potential drought extremes. A random sampling technique was applied to all three models to generate a multi-model ensemble and characterize uncertainties. It was projected that by the end of the century, the number of days with enhanced drought conditions increased by 10 along California's coastline, by 30–40 days in the Northwest and Midwest, and by 100 days in California's Central Valley.

ALCF Contribution: A large portion of these simulations have used ALCF resources dating back to 2012, including Intrepid. Cooley was used recently to analyze and visualize results that were backed up to HPSS, which has been critical to support research projects spanning several years. ALCF worked with the team to provide the allocations to support and complete this work.

Contact

Brandi L. Gamelin
Argonne National Laboratory
bgamelin@anl.gov
ASCR Allocation PI: Veerabhadra R. Kotamarthi, Argonne

Publication
Gamelin, B.L., et al., "Projected U.S. Drought Extremes through the Twenty-First Century with Vapor Pressure Deficit," *Scientific Reports* 12, 8615 (2022). DOI: 10.1038/s41598-022-12516-7

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: February 2023

8.1.5 High-Fidelity Simulations of Gas Turbine Combustor using Spectral Element Method

The Science
With a push toward decarbonizing the aviation sector, sustainable aviation fuels (SAFs) have gained prominence as a potential replacement for fossil fuels. The U.S. government's SAF Grand Challenge was recently formulated with a goal to achieve a minimum of 50% reduction in life cycle greenhouse gas emissions compared to conventional fuels and to supply sufficient SAFs to meet 100% of aviation fuel demand by 2050. To assess the viability of various SAFs, researchers must be able to understand and predict the complex flow, spray, and combustion processes taking place in the gas turbine combustors, as well as the influence of these processes on such events as lean blow out, high altitude relight, and cold start, which affect the performance of gas turbines. With recent advances in numerical methods and the availability of HPC facilities, computer simulations can provide unprecedented details about the underlying multi-physics processes; however, these simulations rely on the complex task of creating a detailed computational model of the gas turbine that is accurate and runs efficiently on modern computers.

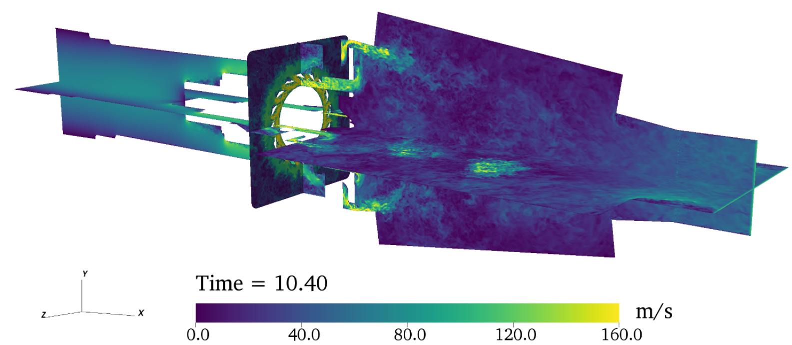
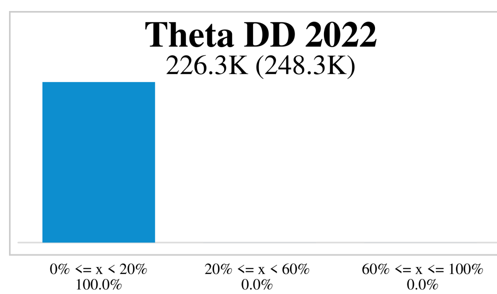


Figure 8.5 Snapshot of the instantaneous turbulent flow field in the ARC-M1 combustor. Seen in this image are multiple recirculation regions and complex interactions between the large- and small-scale turbulent flow structures.

Image courtesy of Sicong Wu.



The Impact
This project is developing the capabilities to perform fully resolved simulations of modern gas turbine combustors to enable an improved understanding of multi-physics processes for advancing the development of SAFs. In this study, wall-resolved large-eddy simulations (LESs) of the turbulent flow in the Army Research Laboratory's midsize combustor (ARC-M1) were performed using the high-order, spectral element fluid dynamics code Nek5000. The simulations were validated against experimental measurements and showed good agreement. These high-fidelity simulations that leverage the DOE leadership computing systems can help us understand

the combustion and heat transfer challenges introduced by using these low-carbon fuels. This project will help establish a high-fidelity, scalable, numerical framework that can be used for evaluating the effect of fuel properties on flow and flame dynamics in a practical gas turbine combustor.

Summary

This research has sought to develop capabilities for performing fully resolved simulations of modern gas turbine combustors using Nek5000 to enable improved understanding of the multi-physics processes for development of SAFs. Nek5000 is a high-order code based on the spectral element method (SEM), developed at Argonne and targeted for exascale systems. It shows excellent strong-scaling and transport properties, making it ideally suited for high-resolution, fully resolved, multi-scale simulations. It has the ability to handle complex geometries with body-fitted meshes, as well as to move boundaries using the Arbitrary Lagrangian Eulerian (ALE) method. The group's recent efforts have focused on implementing state-of-the-art spray, combustion, and heat transfer models in the code.

In the current project, the first-ever wall-resolved LESs of the turbulent flow and spray processes in the ARC-M1 combustor were performed. The simulations were validated using particle image velocimetry (PIV) measurements from a group at the University of Illinois at Urbana-Champaign and showed good agreement. The simulations demonstrated the presence of large and small recirculation regions generated due to mixing between the different flow streams. The accurate prediction of these recirculation regions is key in predicting the flame anchoring and dynamics for reacting simulations. Proper orthogonal decomposition (POD) of the turbulent flow field was performed to investigate the dynamics of the large- and small-scale turbulence in the combustor. Finally, simulations with fuel injection were used to determine the effect of fuel spray on the turbulent flow structures.

ALCF Contribution: ALCF staff helped with performance optimization on Mira and with debugging issues and job scheduling on both Mira and Theta. ALCF also provided excellent training opportunities for porting and performance tuning on both Mira and Theta. The training for Mira was essential for this work, and the training on Theta was useful for understanding the differences between Theta and Cori.

Contact

Muhsin Ameen
Transportation and Power Systems Division, Argonne National Laboratory
mameen@anl.gov
ASCR Allocation PI (DD): Sicong Wu (Argonne)
ASCR Allocation PI (INCITE): Sibendu Som (Argonne)

Publications

Wu, S., D. Dasgupta, M. Ameen, and S. Patel. (2022). LES Modeling of Gas Turbine Combustor using Nek5000. *Bulletin of the American Physical Society*.

Wu, S., D. Dasgupta, M. Ameen, and S. Patel. (2022). High-Fidelity Simulations of Gas Turbine Combustor using Spectral Element Method. *AIAA SciTech*.

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: August 2023

8.2 Research Activities / Vendor Engagements for Future Operations

8.2.1 Research Activity - Joint Laboratory for System Evaluation (JLSE)

Argonne's JLSE enables researchers to assess and improve next-generation computing platforms of interest to the DOE. Established by the computing divisions of Argonne's CELS Directorate (Data Science and Learning, Mathematics and Computer Science, Computational Science, and ALCF) and run by ALCF, JLSE centralizes Argonne's research activities aimed at evaluating future extreme-scale computing systems technologies, and capabilities.

JLSE users leverage existing infrastructure and next-generation hardware and software to explore systems-level experimental computer and computational science, including operating systems, messaging, compilers, benchmarking, power measurements, Input/Output (I/O), and new file systems. By providing access to leading-edge computing resources and fostering collaborative research, JLSE enables researchers to address Argonne's and DOE's needs in a variety of areas by:

- Improving science productivity on future hardware and software platforms.
- Providing an avenue for Argonne researchers to work collaboratively with HPC vendors on prototype technologies, including exascale and beyond.
- Investigating alternative approaches to current and future system deployments.
- Maintaining a range of hardware and software environments for testing research ideas.
- Helping to drive standards on benchmarks, programming models, programming languages, memory technologies, etc.

ALCF closely collaborates with Intel on Aurora. This includes accelerating their software roadmap for traditional HPC and for data and AI pillars to support the science workloads from ESP and ECP projects. JLSE testbeds and software used to prepare for Aurora include:

- **Florentia:** Six nodes with Intel Server Board (codenamed Denali Pass), four Intel Data Center GPU Max (codenamed Ponte Vecchio), and two Xeon CPU Max Processors with HBM2e (codenamed Sapphire Rapids) in each node. The lessons learned in bringing up the Florentia nodes have been especially helpful for Aurora and Sunspot.
- **Iris:** SuperMicro X11SSH-GF-1585 Server Motherboard with Intel Xeon E3-1585 v5 CPU and Iris Pro Graphics P580 GPU (Intel integrated Gen9 GPUs).
- **Presque:** Intel DAOS nodes (DCPMM and NVMe storage) with Intel DAOS file system software.

Other active JLSE testbeds include:

- NVIDIA GPUs:
  - DGX–1 (V100 GPUs)
  - Gigabyte NVIDIA A100 and A40 cluster
  - Supermicro NVIDIA H100, V100 and P100 cluster

- AMD GPUs:
    - AMD GPU MI250, MI100 and MI50 cluster
- Intel Xeon Clusters: Ice Lake, Skylake, Cascade Lake, and Cooper Lake
- ARM Clusters:
    - NVIDIA Grace Hopper GH200
    - HPE Apollo 80 – 8 node Fujitsu A64X CPU
    - NVIDIA ARM Dev Kit – Ampere Altra Q80-30 ARM CPU, NVIDIA A100 GPU
- IBM Power System AC922 (Power9 CPU, V100 GPU)
- NextSilicon Maverick-1
- Atos Quantum Learning Machine

In 2023, the JLSE supported over 500 users, spanning over 160 projects. These projects ranged from application portability to software development—including operating systems, compilers, deep learning frameworks, and performance tools. Teams within the ECP's Application Development and Software Technology projects used the JLSE Aurora testbeds and the Aurora SDK to develop applications and software for Aurora. Access to all major vendor GPUs has led to the development of performance portable APIs, profiling tools, and portable applications across all three platforms (AMD, Intel, and NVIDIA GPUs). It also helped measure and contrast performance against native programming models. Moreover, access to novel accelerators such as Next Silicon has led users to understand the performance of next-generation architectures such as data flow architectures. The following summaries represent a sampling of current JLSE projects:

- **Exascale Computing Project:** Projects from ECP had continued access to the Intel GPU nodes in Florentia and oneAPI software to port their applications and software to the platform.
- **ALCF Early Science for Aurora:** ESP project teams had access to the Florentia nodes to prepare applications for Aurora. They also had access to the NVIDIA and AMD GPUs to compare their applications' performance using different native programming models and different hardware.
- **THAPI:** stands for Tracing Heterogeneous APIs. It provides a lightweight profiling and debugging layer for commonly encountered heterogeneous APIs (OpenCL, Level-Zero, CUDA, OpenMP). Developers used JLSE as an experimental platform to develop their tools before making them available to ALCF users.
- **HEP-CCE:** High-energy physics (HEP) experiments have developed millions of lines of code optimized to run on traditional x86 systems. However, leadership-class computing facilities and traditional data centers use new accelerator architectures, such as GPUs. HEP experiments are faced with the untenable prospect of rewriting millions of lines of code. The architecture-specific languages and APIs promoted by the different vendors make this task more challenging. The Portable Parallelization Strategies team of the HEP Center for Computational Excellence is using JLSE GPU resources to investigate Kokkos, SYCL, Alpaka, etc., as potential portability solutions using representative use cases from DUNE, ATLAS, and CMS experiments.

- **Benchmarking Genetic Variant Calling Algorithms:** The project used JLSE resources to benchmark several algorithms for calling genetic variants (structural variants, SNPs, etc.) using approaches that use HPC and deep learning-based architectures. The project specifically benchmarked variant callers like Cue from Broad Institute and Parabricks tools from NVIDIA using publicly available genetics data that can be used without restrictions.
- **Using ML to configure QoS classes for Slingshot networks:** Slingshot networks provide quality-of-service (QoS) mechanisms to isolate different types of traffic in separate virtual networks, or traffic classes. The full benefits of QoS classes can only be realized if the different traffic flows are assigned to appropriately configured classes. ALCF developed machine learning methods to aid in configuring QoS classes for ALCF's Slingshot networks. Specifically, ALCF used the Skylake nodes to run CODES network simulations of Slingshot QoS and used the simulation data to develop machine learning models that can predict QoS class configurations to improve application performance.

8.2.2 Research Activity - ALCF AI Testbed

With an eye toward the future of scientific computing, ALCF has deployed an advanced AI platform testbed for the research community. This testbed enables the facility and its user community to help define the role of AI accelerators in next-generation scientific machine learning. It also helps shape the vendor's roadmap and development of AI accelerators for science. The testbed's innovative AI platforms complement Argonne's GPU-accelerated supercomputers, Polaris and Aurora, to provide a state-of-the-art computing environment that supports pioneering research at the intersection of AI and HPC.

The ALCF AI testbed consists of systems from Cerebras, Graphcore, Groq, Intel Habana, and SambaNova. ALCF actively worked with several AI accelerator systems and plans to include new systems as part of the testbed. The systems from Cerebras, Graphcore, Groq, and SambaNova are available via the Director's Discretionary allocation program to the open-science user community. In CY 2023, ALCF actively collaborated with new AI accelerator vendors, including Untether and Tenstorrent.

Active users of the systems span university, industry, and national labs and include applications in domains such as material science, cosmology, bioscience, imaging science, high-energy physics, and climate sciences. ALCF has conducted several user workshops on the AI testbed and held a tutorial at SC23 in collaboration with AI testbed partners to help the community leverage these systems for science. These systems were also covered as part of the ATPESC and the ALCF Introduction to AI for Science on Supercomputers training program.

The ALCF AI Testbed includes the following systems:

- The **Cerebras** CS-2 is a wafer-scale deep learning accelerator comprising 850,000 processing cores, each providing 48KB of dedicated SRAM memory for an on-chip total of 40 GB interconnected to optimize bandwidth and latency. To enable large models, the system has been scaled to two CS-2 wafer-scale engine nodes interconnected by the SwarmX fabric together with the MemoryX memory subsystem. Its software platform integrates popular machine learning frameworks such as TensorFlow and PyTorch.

- The **SambaNova** DataScale system is architected around the next-generation Reconfigurable Dataflow Unit (RDU) processor for optimal dataflow processing and acceleration. The system has been upgraded to a second-generation processor. The system consists of eight nodes, each of which features eight RDUs interconnected to enable model and data parallelism. SambaFlow, its software stack, extracts, optimizes and maps dataflow graphs to the RDUs from PyTorch machine learning frameworks.
- The **Graphcore** 22 petaflops Bow Pod64 system is the latest-generation accelerator from Graphcore. It is a one-rack system consisting of 64 Bow-class IPUs with a custom interconnect. The Graphcore software stack includes support for TensorFlow and PyTorch and the Poplar SDK used by machine learning frameworks.
- The **Habana** Gaudi processor features eight fully programmable VLIW SIMD tensor processor cores, integrating ten 100-GbE ports of RDMA over Converged Ethernet (RoCE) into each processor chip to efficiently scale training. The Gaudi system consists of two HLS-1H nodes, each with four Gaudi HL-205 cards. The software stack comprises the SynapseAI stack and provides support for TensorFlow and PyTorch.
- A **Groq** Tensor Streaming Processor (TSP) provides a scalable, programmable processing core and memory building block capable of achieving 250 TFlops in FP16 and 1 PetaOp/s in INT8 performance. The Groq system has recently been upgraded to a GroqRack with nine Groq nodes, each consisting of eight GroqChip v1.5 Tensor streaming processors (TSP) accelerators. GroqChip accelerators are interconnected via a proprietary chip-to-chip interconnect to enable larger models and data parallelism. The Groq system is highly optimized for inference.

Key activities of the testbed include:

- Maintaining a range of hardware and software environments for AI accelerators.
- Providing a platform to benchmark applications, programming models, and ML frameworks.
- Supporting science application teams in the porting and evaluation of their applications.
- Coordinating with vendors during their product development.

The AI Testbed effort supports remote access to the systems, collects feedback and use cases from users, develops online tutorials in conjunction with each vendor, and conducts in-person training and hackathon events.

Common Software Environment: ALCF worked with AI testbed vendors to use PBSPro to manage and schedule resources. This will enable better integration with the rest of ALCF resources.

Since the AI Testbed was established, there have been many opportunities for lessons learned on challenges faced and significant accomplishments achieved when utilizing state-of-the-art AI appliances for major campaigns, such as Gordon Bell runs.

Key lessons learned and accomplishments of the testbed include:

- AI accelerators support large memory capacities, which greatly benefit training large-language models with long context lengths. Given the inherent long sequence lengths of the data, long context lengths are critical in domains including biology, climate, and materials, among others. This has resulted in improved training performance results for science over what has been possible on GPUs. This has been key to accomplishing science for the GenSLM science runs with 10K–30K sequence lengths.
- To get the best performance on AI accelerators, one benefits by adapting the model's zoo implementations provided by the vendors to meet the needs of science rather than executing the model code one runs on the GPUs. This reflects the state of the software optimizations available. Efforts are being made by the AI accelerators to support various models 'out of the box,' and we expect improved support going forward.
- Given the long compilation times for models, strategies wherein we can trade-off the compilation times to the performance of generated code has proven to be critical during the model development cycle. This is akin to the various "O1/2/3" optimizations available with C/C++ compilers today. ALCF has worked closely with the AI vendors to incorporate these, and it has improved the productivity of developers.
- AI testbed architectures, including dataflow architectures, are very attractive for traditional HPC applications. The Monte Carlo XSBench Lookup has achieved >100X improvement in performance using a CS-2 over an A100 off the compilation times to the performance of generated code have proven critical during the model development cycle. This is akin to the various "O1/2/3" optimizations available with C/C++ compilers today. ALCF has worked closely with the AI vendors to incorporate these, and it has improved developers' productivity (see: https://doi.org/10.1016/j.cpc.2023.109072). Similarly, the Graphcore architecture has achieved improved performance for graph applications in comparison to GPUs (see: https://doi.org/10.1145/3624062.3624608). ALCF has ongoing efforts working with AI vendors, including libraries and tutorials, to enable porting of HPC applications to the testbeds.

8.2.3 Vendor Engagement – NextSilicon

Along with OCLF and PNNL, ALCF collaborated with Next Silicon in the development of their first-generation accelerator, Maverick-1. ALCF had two face-to-face workshops with Next Silicon on Maverick-1 on August 22 and September 19–20 to port and test applications onto Maverick-1 at JLSE. Applications studied included HACC and QMCPACK. Several ALCF staff members attended these hands-on workshops. The Maverick-1 is installed at JLSE and is available to any Argonne staff covered by the nondisclosure agreement (NDA) with Next Silicon.

8.2.4 Vendor Engagement – Altair and OpenPBS

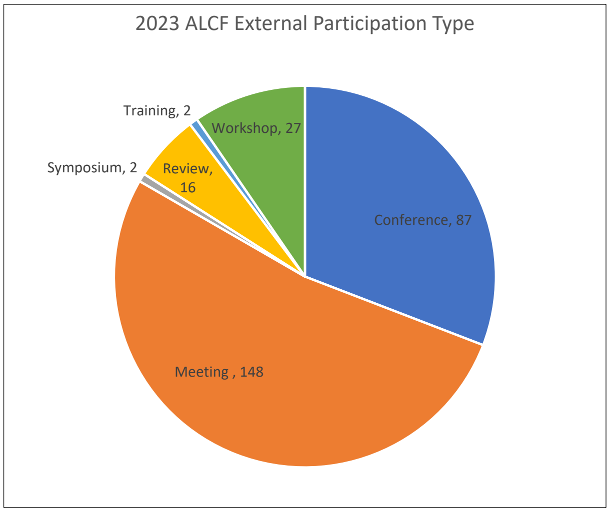
The ALCF team continued its collaboration with Altair Engineering and the OpenPBS community. In 2023, the primary focus was getting PBS working on the AI testbed systems. These systems often have unique scheduling constraints because of the nature of their hardware. For instance, on the Graphcore system, the Intelligence Processing Units (IPUs) can be visualized as being arranged in a binary tree. If a user asks for two IPUs, they must share a common parent, so if they are numbered from 0, 0 and 1 would be a valid allocation, but 1 and 2 would not. Trying to get these scheduling constraints enforced within the standard PBS scheduler in a performant way has been a challenge. The scheduler daemon is a separate component; and although it can be rewritten if needed, ALCF is trying to avoid that. ALCF has also been working with Altair on its new "Liquid Scheduling" by providing feedback on the design and installing and testing early versions. For ALCF, the two important features are (1) the ability to avoid qstat "denial of service attacks" since the liquid scheduling daemon now handles those directly, and (2) in the future (it isn't implemented yet) the ability to use it as a "meta scheduler" to schedule across multiple other schedulers. This capability could allow co-scheduling of jobs that need resources from multiple systems, such as Polaris and Samba Nova. ALCF also worked with Altair to fix a number of bugs, including one in particular that was only noticeable at scale on Aurora.

8.3 DOE Program Engagements / Requirements Gathering

To help ensure that the ALCF delivers on its mission of delivering breakthrough science, ALCF staff need to closely engage with domain science and keep a close eye on directions for supercomputing technologies. The ALCF provides a crucial balance of understanding how production science applications and computer science technologies can move into new and exciting machine architectures in the near term and the future. While training efforts bring people to the ALCF, below we cover the participation of ALCF staff in external events.

ALCF staff support a wide range of computer science and domain science projects and work in close collaboration with the project teams to advance their use of production resources and future resources alike. Additionally, staff members participate in HPC community and domain science activities, including conferences, workshops, reviews, and meetings. In 2023, staff participated in 282 events. Figure 8.6 breaks down these events by type and community. Staff members support DOE mission needs by serving on review committees and advisory boards and participating in and organizing DOE and broader community workshops. ALCF staff regularly participate in DOE and National Science Foundation (NSF) workshops and reviews. Staff are engaged in standards committees and boards for future and current software and hardware technologies. Not only do these activities maintain the expertise of the staff, but they show the respect the staff has in the community.

The ALCF works with other SC Facilities on daily basis. ALCF staff serve on committees for NERSC and OLCF projects. The ALCF has worked closely with APS as we engage how ALCF resources will support some of their needs with the APS Upgrade. The ALCF and OLCF jointly run the flagship allocation program, INCITE, which is a year-round collaboration.



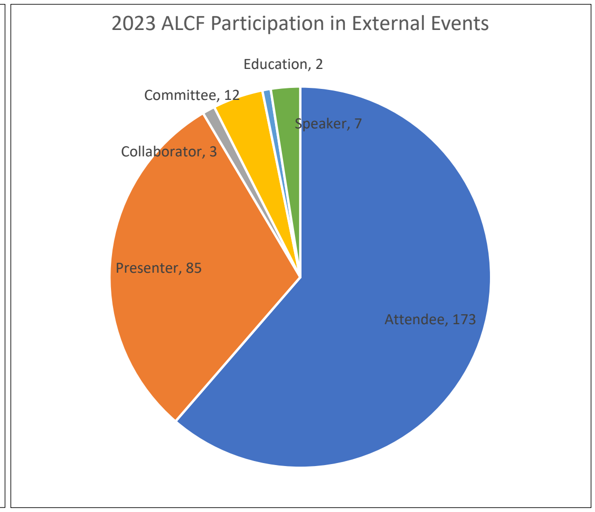


Figure 8.6 Breakdown of key activities by ALCF in CY 2023. The first pie chart (left) breaks down the 282 events by type, primarily based on how the event identified itself. The second pie chart (right) breaks down the same events by the staff member's role.

8.3.1 Engagement Highlights

Supercomputing 2023

SC is one of the key events in the field of supercomputing and covers every area of the field. Participation is one of the primary opportunities to document and share key knowledge. ALCF has significant participation in the event, as shown in Table 8.3.

Table 8.3 Summary of ALCF Participation in SC23

Program	Total
Tutorials	3
Workshops	16
Posters	1
Birds of a Feather	8
Booth Talks	9
Panels	1

Engagement in Standards and Community Groups

ALCF staff members remain actively involved in several HPC standards and community groups that help drive improvements in the usability and efficiency of scientific computing tools, technologies, and applications. Staff activities include contributions to the Better Scientific Software, C++ Standards Committee, Cray User Group, DAOS Foundation, Energy Efficient High-Performance Computing, HPC User Forum, HPSF High-Performance Software Foundation, Intel eXtreme Performance Users Group, Khronos OpenCL and SYCL Working Groups, LDMS User Group, UXL Foundation, MLCommons (HPC, Science, and Storage Working Groups), NITRD Middleware and Grid Infrastructure Team, OCHAMI, Open Fabrics Alliance, OpenMP Architecture Review Board, OpenMP Language Committee, OSTI ORCID

Consortium Membership, Open Scalable File Systems (OpenSFS) Board, and Standard Performance Evaluation Corporation (SPEC) HPG (High-Performance Group).

Performance, Portability, and Productivity in HPC (P3HPC) Forum

The Performance, Portability and Productivity in HPC workshop at SC23 was jointly organized by ALCF, LLNL, Intel, and NVIDIA to bring together developers and researchers with an interest in the development of performance-portable applications across current and future high-performance computers. The topic of performance, portability, and productivity focuses on enabling applications and libraries to run across multiple architectures without significant impact on achieved performance and with the goal of maintaining developer productivity. It is important that developers understand and enhance the best practices in this area to enable applications to run efficiently across the diverse hardware platforms that exist today.

8.3.2 Summary of Engagements with the Exascale Computing Project

Argonne is a core laboratory of the ECP, and several members of ALCF's leadership team are engaged in the ECP project. Susan Coghlan and David Martin are a part of the ECP leadership team: Coghlan is deputy director of Hardware and Integration (HI), and Martin is co-executive director of the ECP Industry and Agency Council. Haritha Siddabathuni Som is the level-3 lead for Facility Resource Utilization, and Scott Parker is the level-3 lead for Application Integration. Christopher Knight is a level-4 lead for the Aurora Application Integration area. Yasaman Ghadar is a level-4 lead for Training and Productivity. Other leadership team members participate in the various working groups, including Bill Allcock and Jini Ramprakash. ALCF Division Director Michael E. Papka regularly participates in teleconferences with the ECP project director and other facility directors. In addition, numerous other ALCF staff members have roles in the projects and working groups listed above.

In CY 2023, 25 ALCF staff members attended the ECP Annual Meeting held January 17–20, 2023, to participate in technical conversations, project discussions, and facility-specific breakouts. In addition, ALCF participated in several planning meetings with ECP and the other computing facilities (NERSC, OLCF) to augment and execute the ECP/Facilities engagement plan and worked with ECP's Training Lead to promote ECP training activities to ALCF users.

ECP-Funded Positions in ALCF

The ALCF's ECP Hardware & Integration effort made great strides in 2023, continuing the team's work in porting and testing ECP applications across many GPUs, including Intel GPUs. A total of 46 ALCF staff members were funded at various levels to work with ECP Application Development and Software Technology projects. Of these 46, several staff members focused specifically on training, and several more on Intel's Center of Excellence (COE) for Aurora. Additional staff members were funded to develop and deploy ECP continuous integration (CI) capabilities, support software technologies, and work with others within ECP on containers, whereas others were funded to aid ECP applications in porting and testing.

Continuous Integration (CI) Pipeline

In 2023, ALCF continued to support the growth of CI Pipelines using GitLab-CI, which is a crucial tool for projects to carry out regular automated testing on ALCF resources. Regular automated testing helps catch errors early in the development process, leading to increased development efficiency. Through GitLab-CI, users can set up CI pipelines on ALCF machines. In 2023, users ran 21,064 jobs using the ECP GitLab-CI infrastructure, with 86.6% of those jobs successfully completed. Post-ECP, in 2024, ALCF will continue to offer CI infrastructure on current and future ALCF systems, with a focus on enabling the use of CI on Aurora.

Communication between the ALCF and the ECP Resource Allocation Council

In 2018, the ECP ALCC allocation ended, and the DOE computing facilities switched to the Resource Allocations Council (RAC) to support ECP computing needs on the pre-exascale systems at OLCF and ALCF. The RAC, composed of representatives of the facilities and the ECP, met monthly to review project progress and to assess new project needs. To help automate how the RAC consumes system data, the ALCF sent allocation and usage data for Theta and Polaris as CSV (comma-separated values) files to the ECP each day. The files were uploaded to a Box folder accessible by the ECP, from where they were downloaded, processed, and merged into the data pipeline that fed into the ECP User Program dashboard.

The ESRAC (Exascale Systems Resource Allocations Council) was established in 2023 to review and respond to allocation needs on the Exascale systems, Aurora and Frontier, via the Exascale Systems User Program. This program did not have a proposal process; rather, access to all ECP teams was provided once the exascale systems were stable and ready for users. Management of utilization, prioritization of runs, monitoring the status of ECP Key Performance Parameter (KPP) runs, and facilitating dialogue between the focus areas and the facilities were handled through the ESRAC that met at least once a week and provided rapid turnaround on support to the Application Development and Software Technology teams. Similar to usage data on Theta and Polaris, data on Sunspot was also sent to the ECP as a CSV file each day.

Conclusion

The ALCF continued to enable scientific achievements, consistent with DOE's strategic goals of scientific breakthroughs and foundational science, through projects carried out on ALCF resources. In 2023, researchers using ALCF resources published 240 papers in high-quality conferences and journals. ALCF projects have had success in a variety of fields, using many different computational approaches. ALCF projects were able to reach their scientific goals and successfully use their allocations. Several of the projects and PIs subsequently received awards or were recognized as achieving significant accomplishments in their fields.

This page intentionally left blank.

Appendix A – Calculations

A.1 Scheduled Availability

Scheduled availability is the percentage of time a designated level of resource is available to users, excluding **scheduled outage** time for maintenance and upgrades. To be considered a scheduled outage, 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). Users will be notified of regularly scheduled maintenance in advance, on a schedule that provides sufficient notification, and no less than 72 hours prior to the event—and preferably as much as seven calendar days prior. If the regularly scheduled maintenance is not needed, users will be informed of the cancellation of the 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.

Formula:

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

Where

time in period = start time – end time
start time = end of last outage prior to reporting period
end time = start of first outage after reporting period (if available) or start of the last outage in the reporting period

A.2 Overall Availability

Overall availability is the percentage of time a system is available to users. Outage time reflects both scheduled and unscheduled outages.

***Formula*:**

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

A.3 System Mean Time to Interrupt (MTTI)

MTTI (Mean Time to Interrupt) is defined as time, on average, to any outage of the full system, whether unscheduled or scheduled. It is also known as MTBI (Mean Time Between Interrupts).

***Formula*:**

$$\mathbf{MTTI} = \frac{\textbf{time in period} - (\textbf{duration of scheduled outages} + \textbf{duration of unscheduled outages})}{\textbf{number of scheduled outages} + \textbf{number of unscheduled outages} + \mathbf{1}}$$

A.4 System Mean Time to Failure (MTTF)

MTTF (Mean Time to Failure) is defined as the time, on average, to an unscheduled outage of the full system.

Formula:

$$\mathbf{MTTF} = \frac{\textbf{time in period} - \textbf{duration of unscheduled outages}}{\textbf{number of unscheduled outages} + \mathbf{1}}$$

A.5 Total System Utilization

Total System **Utilization** is the percent of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user group, including staff and vendors. Jobs that ran during an outage are excluded.

***Formula*:**

$$\mathbf{Utilization} = \left(\frac{\textit{Node Hours used in period}}{\textit{Node Hours available in period}}\right) * \mathbf{100}$$

A.6 Capability

Capability is an attribute assigned to user jobs that meet the capability definition for a machine. **High Capability** is an attribute assigned to user jobs that meet the high capability definition for a machine.

Table A.1 shows the capability definitions for reportable machine Theta.

Table A.1 Capability Definitions for Theta

Theta				
Capability	**High Capability**	**Range**	**Minimum Nodes**	**Maximum Nodes**
No	No	0% <= x < 20.0%	1	799
Yes	No	20.0% <= x < 60.0%	800	2,399
Yes	Yes	60.0% <= x	2,400	See: A.7 Theta Nodes

Capability also refers to a calculation. The capability calculation is the percentage of node-hours of jobs with the capability attribute versus the total node-hours of all jobs. The calculation can be applied to a class of jobs. For example: Innovative and Novel Computational Impact on Theory and Experiment (INCITE) capability is the percentage of node-hours of INCITE jobs with the capability attribute versus the total node-hours of all INCITE jobs for a time period.

Formula:

$$\textbf{OVERALL CAPABILITY} = \left(\frac{\textbf{\textit{Capability Node Hours Consumed}}}{\textbf{Total Node Hours Consumed}}\right) * \mathbf{100}$$

$$\textbf{HIGH CAPABILITY} = \left(\frac{\textbf{\textit{High Capability Node Hours Consumed}}}{\textbf{Total Node Hours Consumed}}\right) * \mathbf{100}$$

A.7 Theta Nodes

The number of reportable nodes on Theta is fewer than the total number of nodes. The total node count for Theta changed during 2017, as shown in Table A.2.

Table A.2 Total and Reportable Nodes for Theta

Theta		
Data Range	Total Nodes	Reportable Nodes
07/01/2017 – 12/12/2017	3,624	3,240
12/13/2017 – 12/31/2017	4,392	3,240
01/01/2018	4,392	4,008

The reportable node count is used in the following calculations:

- Scheduled Availability: Affects the scheduled outage and unscheduled outage calculations when the node count in the outage was fewer than the total number of nodes.
- Overall Availability: Affects the scheduled outage and unscheduled outage calculations when the node count in the outage was fewer than the total number of nodes.
- Utilization: The calculation capped the daily utilization at 100 percent of reportable nodes. The number of node-hours for each day was calculated as the minimum of the node-hours used and the node-hours possible.
- Overall Capability: 20 percent of the reportable nodes.
- High Capability: 60 percent of the reportable nodes.

This page intentionally left blank.

Appendix B – ALCF Director's Discretionary Projects

January 1, 2023–December 31, 2023

Director's Discretionary (DD) Projects on Polaris (by Project Name)

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
3DWholeGenome	Jie Liang	University of Illinois at Chicago	High-Resolution Ensemble 3D Structures of Genome across Tissues	Biological Sciences	391
ACO2RDS	John J Low	Argonne National Laboratory	Adsorptive CO2 Removal from Dilute Sources	Materials Science	4,000
AI-based-NDI-Spirit	Rajkumar Kettimuthu	Argonne National Laboratory	Framework and Tool for Artificial Intelligence & Machine Learning Enabled Automated Non-Destructive Inspection of Composite Aerostructures Manufacturing	Engineering	1,272
AI4NMR	Eric Michael Jonas	The University of Chicago (UChicago)	Structure Elucidation for Nuclear Magnetic Resonance via Structured Prediction	Chemistry	13,762
AIAstroEC	Eugenio Culurciello	Purdue University	AI and Astronomy: training an artificial astronomer assistant	Physics	3,628
AIHPC4Edu	Ravi Kiran Madduri	Argonne National Laboratory	AI and HPC for Science and Education	Biological Sciences	3,803
ALCFNULLPROJECT	UNKNOWN	Argonne National Laboratory	If a job runs and there are no allocations anymore to charge to then the job will get charged to ALCFNULLPROJECT allocation.	UNKNOWN	115
ALCF_for_DUNE	Aleena Rafique	Argonne National Laboratory	Enabling precise measurement of neutrino interactions using advanced computing in DUNE	Physics	4,618
alcf_training	Yasaman Ghadar	Argonne National Laboratory	ALCF Training	Training	62
ALEXIS	Jorge Rene Padial	Vanderbilt University Medical Center	Automatically Labelled EUV and XRay Incedent SolarFlare catalog	Physics	3,662
amr-direct	Damyn Maxwell Chipman	Boise State University	Scalable CPU/GPU Direct Solver for Adaptive Mesh Refinements	Mathematics	1,183
AMRdetonations	Venkatramanan Raman	University of Michigan	Adaptive Simulation of Detonations	Engineering	3,203
APSDataAnalysis	Rafael Vescovi	Argonne National Laboratory	APS Beamline Data Processing and Analysis	Computer Science	19,281
APSDataProcessing	Nicholas Schwarz	Argonne National Laboratory	Advanced Photon Source (APS) Data Processing	Computer Science	2,988
APS_ILLUMINE	Nicholas Schwarz	Argonne National Laboratory	APS ILLUMINE Project Space	Physics	9

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
ArgonneGPUAccess	Craig Stacey	Argonne National Laboratory	GPU access for Argonne researchers	Internal	5,000
argonne_tpc	Venkatram Vishwanath	Argonne National Laboratory	Trillion Parameter Consortium	Computer Science	6,000
ArtISD	Tuhin Sahai	SRI International	Artificial Intelligence for Scientific Discovery	Computer Science	413
astroPic	Gregory R Werner	University of Colorado-Boulder	PIC simulation of astrophysical plasmas	Physics	1,000
AS_EN_TZ	Italo Moletto	AgroSpace	Water and Food Security for Tanzania using Satellite Imagery + ML/AI	Earth Science	199
atlas_aesp	Walter Howard Hopkins	Argonne National Laboratory	Simulating and Learning in the ATLAS detector at the Exascale	Physics	6,873
ATPESC2023	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program for Extreme-Scale Computing 2023	Training	15,000
ATPESC_Instructors	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing Instructors	Training	1,000
Auriga	Christine Mary Simpson	Argonne National Laboratory	The Auriga Project	Physics	300
autopology_alcf	Rafael Gomez-Bombarelli	Massachusetts Institute of Technology (MIT)	End to end classical force field parametrization for polymer electrolytes using machine learning	Materials Science	1,276
BACH1	Marsha Rosner	The University of Chicago (UChicago)	Structural basis of BACH1 proline hydroxylation regulation of the hypoxia response and metastasis in Triple Negative Breast Cancer	Biological Sciences	6,489
BACTWA	Philippe Regis-Guy Piot	Northern Illinois University (NIU)	Beam Acceleration and Control in Structure Wakefield Accelerator	Physics	3,836
BenchmarkUNM	Amanda Bienz	University of New Mexico	Benchmarking and Optimizing Data Movement on Emerging Heterogeneous Architectures	Computer Science	2,549
BFTrainer	Rajkumar Kettimuthu	Argonne National Laboratory	Rescaling DNN Training Tasks to Fit Dynamically Changing Holes in Supercomputer Schedule	Computer Science	3,822
BIP167	Philip Kurian	Howard University	Computing superradiance and van der Waals many-body dispersion effects from MD simulations of biomolecular complexes	Physics	1,999
BirdAudio	Nicola Joy Ferrier	Argonne National Laboratory	Machine Learning for Classification of Birdsong	Computer Science	3,982

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
BPC	Christopher Michael Graziul	The University of Chicago (UChicago)	Optimization of audio processing pipeline for broadcast police communications	Computer Science	9,494
BRAIN	Getnet Dubale Betrie	National Renewable Energy Laboratory (NREL)	Scalable Brain Simulator for Extreme Computing	Biological Sciences	5,934
BrainImagingML	Thomas David Uram	Argonne National Laboratory	Large-scale Brain Imaging and Reconstruction	Biological Sciences	4,032
C-Star	Matthew Long	[C]Worthy, LLC	Computational Systems for Tracking Ocean Carbon	Earth Science	582
cac	Chang Liu	Princeton Plasma Physics Laboratory (PPPL)	Center for Advanced Computation: Disruption Avoidance	Fusion Energy	888
CampSwift	Barry Schneider	National Institute of Standards and Technology (NIST)	Physics Based Modeling for Forced Fuel Management	Earth Science	600,875
candle_aesp	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	Virtual Drug Response Prediction	Biological Sciences	27,495
catalysis_aesp	David Hamilton Bross	Argonne National Laboratory	Exascale Computational Catalysis	Chemistry	13,747
Catalyst	Katherine M Riley, Christopher James Knight, James Clifton Osborn, Timothy Joe Williams	Argonne National Laboratory	Catalyst	Internal	69,398
cfdml_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Data Analytics and Machine Learning for Exascale CFD	Engineering	13,747
CFS_UX_TEST	Haritha Siddabathuni Som	Argonne National Laboratory	TESTING CFS	Support	40
CHAMPS_aero	Roberto Paoli	Polytechnique Montreal	Aerodynamics simulations with Chapel programming language	Engineering	2,877
CharmRTS	Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia	University of Illinois at Urbana-Champaign	Charm++ and its applications	Computer Science	5,156
climate_severe	Vittorio Angelo Gensini	Northern Illinois University (NIU)	Anticipating Severe Weather Events via Dynamical Downscaling	Earth Science	450
cloud-avoid	Yufeng Luo	University of Wyoming	Cloud avoidance algorithm for astronomical observations	Physics	2,000
CloudAnalysis	Michael E. Papka	Argonne National Laboratory	CloudAnalysis	Computer Science	1,000
co2-ads-mof	Jonathan Rutherford Owens	General Electric Company (GE)	Understanding CO2 Adsorption on Metal-Organic-Frameworks	Chemistry	2,931

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CobaltDevel	Paul Michael Rich, William Edward Allcock	Argonne National Laboratory	Cobalt Development	Internal	36
connectomics_aesp	Nicola Joy Ferrier, Thomas David Uram	Argonne National Laboratory	Enabling Connectomics at Exascale to Facilitate Discoveries in Neuroscience	Biological Sciences	687
covid-ct	Ravi Kiran Madduri	Argonne National Laboratory	Medical Imaging Domain-Expertise Machine Learning for Interrogation of COVID	Computer Science	5,572
Cray	Torrance Ivan Leggett, Mark Richard Fahey, Susan Marie Coghlan, Timothy Joe Williams, William Edward Allcock	Hewlett Packard Enterprise	Cray Installation	Internal	57,652
CSC249ADCD01	Ian Foster	Argonne National Laboratory	2.2.6.03 ADCD01-CODAR	Computer Science	4,000
CSC249ADCD02	Susan Marie Mniszewski, Timothy C. Germann	Los Alamos National Laboratory (LANL)	2.2.6.04 ADCD02-COPA: Co-Design Center for Particle Applications	Physics	4,000
CSC249ADCD04	Tzanio Valentinov Kolev, Misun Min, Paul Frederick Fischer	Lawrence Livermore National Laboratory (LLNL)	2.2.6.06 CEED: Center for Efficient Exascale Discretizations	Computer Science	17,000
CSC249ADCD05	Mahantesh Halappanavar	Pacific Northwest National Laboratory (PNNL)	2.2.6.07 ADCD05-ExaGraph	Computer Science	13,000
CSC249ADCD08	Francis Joseph Alexander	Brookhaven National Laboratory (BNL)	2.2.6.08 ADCD08-ExaLearn	Physics	4,000
CSC249ADCD09	John Bell	Lawrence Berkeley National Laboratory (LBNL)	2.2.6.05 ADCD03-AMREX: Block-Structured AMR Co-Design Center	Mathematics	3,100
CSC249ADCD502	Kenneth John Roche	Pacific Northwest National Laboratory (PNNL)	2.2.6.02 ADCD502 Application Assessment	Computer Science	4,000
CSC249ADCD504	Jeanine Cook, Shirley Victoria Moore	Lawrence Livermore National Laboratory (LLNL)	2.2.6.01 ADCD504-Proxy Applications	Computer Science	4,000
CSC249ADOA01	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	2.2.4.03 ADOA01 CANDLE: Exascale Deep Learning Enabled Precision Medicine for Cancer	Biological Sciences	13,000
CSC249ADSE01	Salman Habib, Katrin Heitmann	Argonne National Laboratory	2.2.3.02 ADSE01-ExaSky	Physics	15,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC249ADSE03	Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie	Fermi National Accelerator Laboratory (Fermilab)	2.2.1.01 ADSE03-LatticeQCD: Exascale Lattice Gauge Theory Opportunities/Reqmts for Nuclear & High Energy Physics	Physics	22,000
CSC249ADSE04	Danny Perez	Los Alamos National Laboratory (LANL)	2.2.1.04 ADSE04-EXAALT - Molecular dynamics at the exascale	Nuclear Energy	4,000
CSC249ADSE05	David Paul Trebotich	Lawrence Berkeley National Laboratory (LBNL)	2.2.3.04 ADSE05-Subsurface	Earth Science	4,000
CSC249ADSE06	Jean-Luc Yves Vay	Lawrence Berkeley National Laboratory (LBNL)	2.2.2.06 ADSE06-WarpX	Physics	4,000
CSC249ADSE08	Steven Hamilton, Paul Kollath Romano	Oak Ridge National Laboratory (ORNL)	2.2.2.03 ADSE08 ExaSMR	Nuclear Energy	4,000
CSC249ADSE09	Paul Richard Charles Kent, Anouar Benali	Oak Ridge National Laboratory (ORNL)	2.2.1.06 QMCPACK: Predictive and Improvable Quantum-mechanics Based Simulations	Materials Science	4,000
CSC249ADSE10	Matthew Thomas Bement	Oak Ridge National Laboratory (ORNL)	2.2.1.05 ADSE10-ExaAM	Engineering	3,000
CSC249ADSE11	Theresa Windus	University of Washington	2.2.1.02 ADSE11-NWChemEx: Tackling Chemical, Materials, & Biomolecular Challenges in Exascale	Chemistry	4,000
CSC249ADSE12	Amitava Bhattacharjee	Princeton Plasma Physics Laboratory (PPPL)	2.2.2.05 ADSE12 WDMAPP	Computer Science	4,000
CSC249ADSE14	Jacqueline Chen	Sandia National Laboratories, California	2.2.2.02 ADSE14-Combustion-Pele: Transforming Combustion Science & Technology with Exascale Simulations	Engineering	4,000
CSC249ADSE15	Mark Alan Taylor	Sandia National Laboratories, New Mexico	2.2.3.05 ADSE15-E3SM-MMF	Earth Science	3,100
CSC249ADSE16	Mark S Gordon	Ames Laboratory	2.2.1.03 ADSE16-GAMESS	Chemistry	44,600
CSC249ADSE18	Daniel Kasen	Lawrence Berkeley National Laboratory (LBNL)	2.2.3.01 ADSE18 Exastar	Physics	4,000
CSC249ADSE20	Katherine Yelick, Leonid Oliker	Lawrence Berkeley National Laboratory (LBNL)	2.2.4.04 ADSE20-ExaBiome: Exascale Solutions for Microbiome Analysis	Biological Sciences	1,000
CSC249ADSE22	Christopher Stephen Oehmen, Andres Marquez, Zhenyu Huang	Pacific Northwest National Laboratory (PNNL)	2.2.4.02 ADSE22-ExaSGD	Energy Technologies	4,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC249ADSE23	Jordan Michael Musser	National Energy Technology Laboratory (NETL)	2.2.2.04: MFIX-Exa: Perf Prediction of Multiphase Energy Conversion Device	Energy Technologies	4,000
CSC249ADTR01	Daniel Edward Laney	Lawrence Livermore National Laboratory (LLNL)	2.3.5.10 ADTR01-ExaWorks	Computer Science	4,000
CSC249ADTR02	Osni Marques	Oak Ridge National Laboratory (ORNL)	2.4.6.02 ADTR02- Productivity	Computer Science	4,000
CSC250STDA05	Kenneth Dean Moreland	Oak Ridge National Laboratory (ORNL)	2.3.4.13 STDA05-ECP/VTK-m	Computer Science	4,000
CSC250STDM10	Surendra Byna, Venkatram Vishwanath	Lawrence Berkeley National Laboratory (LBNL)	2.3.4.15 ExaIO - Delivering Efficient Parallel I/O on Exascale Computing Systems with HDF5 and Unify	Computer Science	40,000
CSC250STDM11	Scott Klasky, Norbert Podhorszki	Oak Ridge National Laboratory (ORNL)	2.3.4.09 STDM11-ADIOS Framework for Scientific Data on Exascale Systems	Computer Science	4,000
CSC250STDM12	Robert B. Ross, Robert J Latham	Argonne National Laboratory	2.3.4.10 STDM12-DataLib: Data Libraries and Services Enabling Exascale Science	Computer Science	4,000
CSC250STDM14	Franck Cappello	Argonne National Laboratory	2.3.4.14 STDM14 - VeloC-SZ: Very Low Overhead Transparent Multilevel Checkpoint/Restart/SZ: Fast, Effective, Parallel Error-bounded Exascale Loss....	Computer Science	4,000
CSC250STDM16	James Paul Ahrens, Terece Louise Turton	Los Alamos National Laboratory (LANL)	2.3.4.16 STDM16-ALPINE/ZFP	Computer Science	4,000
CSC250STDT10	Jeffrey S Vetter	Oak Ridge National Laboratory (ORNL)	2.3.2.10 STDT10 PPROTEAS-TUNE	Computer Science	4,000
CSC250STDT11	Sunita Chandrasekaran, Doss ay Oryspayev	Stony Brook University	2.3.2.11 SOLLVE: Scaling OpenMP with LLVm for Exascale	Computer Science	4,000
CSC250STDV01	Patrick O'Leary	Kitware Inc.	2.3.4.01 Data and Visualization Software Development Kit	Computer Science	4,000
CSC250STML12	Carol Woodward	Lawrence Livermore National Laboratory (LLNL)	2.3.3.12 Enabling Exascale Simulations with SUNDIALS and hypre	Mathematics	4,000
CSC250STML13	Hartwig Andreas Anzt	The University of Tennessee at Knoxville	2.3.3.13 STML13 - CLOVER	Computer Science	3,000
CSC250STML14	Andrey Valeryevich Prokopenko	Oak Ridge National Laboratory (ORNL)	2.3.3.14 ALExa: Accelerated Libraries for Exascale/ForTrilinos: Sustainable Production Fortran Interoperability with Trilinos Libraries	Mathematics	3,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC250STML15	Siva Rajamanickam	Sandia National Laboratories, New Mexico	2.3.3.15 STML-Sake	Computer Science	5,000
CSC250STMS05	Ulrike Meier Yang, Satish Balay	Argonne National Laboratory	2.3.3.01 STMS05-Extreme-scale Scientific xSDK for ECP	Mathematics	4,000
CSC250STMS07	Todd S. Munson, Hong Zhang, Richard Tran Mills, Satish Balay	Argonne National Laboratory	2.3.3.06 STMS07-PETSc/TAO for Exascale	Mathematics	4,000
CSC250STMS08	Xiaoye Sherry Li	Lawrence Berkeley National Laboratory (LBNL)	2.3.3.07 STMS08 STRUMPACK/SuperLU/FFTX: Factorization Based Sparse Solvers and Preconditioners for Exascale	Mathematics	2,100
CSC250STNS01	Michael Lang, Terece Louise Turton	Los Alamos National Laboratory (LANL)	2.3.6.01 - STNS01 -LANL ATDM ST Projects	Computer Science	4,000
CSC250STPM01	Sameer Suresh Shende	University of Oregon	2.3.1.01 Programming Models & Runtimes Software Development Kit	Computer Science	4,000
CSC250STPM08	Patrick McCormick	Los Alamos National Laboratory (LANL)	2.3.1.08 Legion: Enhancing and Hardening the Legion Programming System	Computer Science	2,100
CSC250STPM09	Yanfei Guo	Argonne National Laboratory	2.3.1.07 STPM09-Exascale MPI	Computer Science	4,000
CSC250STPM11	George Bosilca, Earl Luther Carr, Jack Dongarra, Thomas Herault	The University of Tennessee at Knoxville	2.3.1.09 STPM11 ParSEC: Distributed Tasking	Computer Science	4,000
CSC250STPM16	Scott Dov Pakin	Los Alamos National Laboratory (LANL)	2.3.1.16 SICM: Simplified Interface to Complex Memory	Computer Science	4,000
CSC250STPM17	Paul Hamilton Hargrove, Erich Strohmaier	Lawrence Berkeley National Laboratory (LBNL)	2.3.1.14 STPM17-UPC++ & GASNet	Computer Science	6,000
CSC250STPM18	Christian Trott	Sandia National Laboratories, California	2.3.1.18 RAJA/Kokkos	Computer Science	4,000
CSC250STPR19	Peter Hugh Beckman	Argonne National Laboratory	2.3.1.19 STPR19 Argo: Argo/Power Steering	Computer Science	4,000
CSC250STPR27	David Edward Bernholdt	Oak Ridge National Laboratory (ORNL)	2.3.1.17 STPR27-OMPI-X: Open MPI for Exascale	Materials Science	4,000
CSC250STTO09	Hartwig Andreas Anzt, Anthony Danalis, Earl Luther Carr, Heike Jagode	The University of Tennessee at Knoxville	2.3.2.06 STTO09 EXAPAPI	Computer Science	4,000
CSC250STTO11	John Michael Mellor-Crummey	Rice University	2.3.2.08 STTO11 HPCToolkit	Computer Science	4,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC251HIHE05	Scott Dov Pakin, Simon David Hammond	Los Alamos National Laboratory (LANL)	2.4.2.01 HIHE05-Analytical Modeling - Hardware Evaluation Working Groups	Computer Science	4,000
CSC251HISD01	Ryan Charles Prout	Los Alamos National Laboratory (LANL)	2.4.4.01 HISD01-Software Integration	Computer Science	4,000
CSCSTDT12345	Patrick McCormick	Los Alamos National Laboratory (LANL)	2.3.2.12 Flang: open-source Fortran front end for the LLVM infrastructure	Computer Science	4,000
CVD_CityCOVID	Jonathan Ozik	Argonne National Laboratory	Agent-based model called CityCOVID capable of tracking detailed COVID-19 transmission	Biological Sciences	2,032
darkskyml_aesp	Salman Habib	Argonne National Laboratory	Dark Sky Mining	Physics	687
datascience	Venkatram Vishwanath	Argonne National Laboratory	ALCF Data Science and Workflows Allocation	Internal	212,267
DataServicePrototype	Benoit Joseph Serge Cote	Argonne National Laboratory	Data Service Prototype	Computer Science	2,032
DAVSDK	Patrick O'Leary	Kitware Inc.	Data and Visualization SDK	Computer Science	309
dcp35	George Em Karniadakis	Brown University	Quantification of extreme weather events and their future changes using Physics-Informed DeepONet modeling and functional priors	Mathematics	2,000
DEAC_fermions	Adrian Giuseppe Del Maestro	The University of Tennessee at Knoxville	Analytic Continuation of Interacting Fermion Spectra	Physics	4,000
deepspeed_collab	Venkatram Vishwanath	Argonne National Laboratory	Deepspeed Collaboration with Microsoft for Large Language Models scaling	Computer Science	3,786
determined_eval	Venkatram Vishwanath	Argonne National Laboratory	Evaluation of Determined.AI HPO on Polaris	Computer Science	1,500
DFTBENCH	Anouar Benali	Argonne National Laboratory	DFT Benchmark for Exascale	Materials Science	4,000
diffDNS-turbulence	Jianxun Wang	University of Notre Dame	Differentiable DNS Simulations of Wall-bounded Turbulence with GPU Acceleration	Engineering	4,000
DiNTGPU	Yeonjun Jeong	Argonne National Laboratory	Towards Exascale Chemical Dynamics via GPU-Accelerated Polynomial Expansions	Chemistry	1,472
dist_relational_alg	Sidharth Kumar	The University of Alabama at Birmingham	Distributed relational algebra at scale	Computer Science	2,666
DL4VIS	Hanqi Guo	Argonne National Laboratory	Deep Learning for In Situ Analysis and Visualization	Computer Science	1,594

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
DLHMC	Sam Alfred Foreman	Argonne National Laboratory	Deep Learning HMC	Physics	3,855
DL_MODEX	MARUTI KUMAR MUDUNURU	Pacific Northwest National Laboratory (PNNL)	Towards a robust and scalable deep learning workflow for fast, accurate, and reliable calibration of watershed models	Earth Science	1,000
DMPE	Shizhong Han	Lieber Institute for Brain Development	Deep learning models for predicting enhancers in single cells	Biological Sciences	3,833
DNS3D	Ramesh Balakrishnan	Argonne National Laboratory	Direct Numerical Simulation of Three Dimensional Turbulence	Engineering	2,207
DNSVCMHD	Keith Daniel Brauss	Francis Marion University	DNS Simulations of Velocity-Current Magnetohydrodynamic Equations	Mathematics	853
Drag-Reduction	Paul Fischer	University of Illinois at Urbana-Champaign	DNS of Drag Reduction	Engineering	1,636
DSMC_NEQ	Jacqueline Chen	Sandia National Laboratories, California	Molecular level simulations of reacting flows under thermal and chemical non-equilibrium	Chemistry	1,508
DTLLM	Bo Li	The University of Chicago (UChicago)	Trustworthiness evaluation for Llama2	Computer Science	888
E2H	Michael E. Papka	Argonne National Laboratory	Edge to HPC	Computer Science	948
E3SM_RRM	Brandi Lee Gamelin	Argonne National Laboratory	Coupled high-resolution GPU enabled E3SM simulations for mid-century extreme events	Earth Science	371
earnest	Robert Clyde Jackson	Argonne National Laboratory	An Equitable, Affordable & Resilient Nationwide Energy System Transition (EARNEST)	Earth Science	2,000
EarthWorks	Richard Dana Loft	National Science Foundation	Preparing EarthWorks for GPU-Based Climate Simulations at Global Storm-Resolving Scales	Earth Science	493
ecc_pynta	Shin Ae Kim	Sandia National Laboratories, California	Exascale chemistry computing of heterogeneous catalytic reactions using automated workflow code "Pynta" on GPU support	Chemistry	1,483
ECP_SDK	Sameer Suresh Shende	University of Oregon	Deploying the ECP SDK software stack at ALCF	Computer Science	698
EE-ECP	Xingfu Wu, Valerie Taylor	Argonne National Laboratory	Energy efficient tradeoff among execution time and power of ECP applications	Computer Science	6,215
EfficientLLM	Mi Zhang	The Ohio State University	Efficient Large Language Model	Computer Science	2,345

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
efficios_thapi	Thomas Applencourt	Argonne National Laboratory	Improving Argonne Tracing Capabilities of Heterogeneous API	Computer Science	11,891
ElecDynComplexSys_DD	Yosuke Kanai	The University of North Carolina-Chapel Hill	Simulations of Electron Dynamics for Complex Systems	Chemistry	2,000
electrolyte-chibueze	Chibueze Vincent Amanchukwu	The University of Chicago (UChicago)	Molecular dynamics-driven discovery of novel liquid electrolytes	Energy Technologies	3,010
ELRDE	Praveen Kumar Ramaprabhu	The University of North Carolina-Chapel Hill	Euler-Lagrange simulations of liquid-fueled Rotating Detonation Engines	Engineering	888
EmergingTech2	Benjamin Alan Blakely	Argonne National Laboratory	Identification of Emerging Technologies from Publication Databases	Computer Science	2,935
EngineDNS	Christos Frouzakis	Eidgenössische Technische Hochschule Zürich (ETH Zurich)	Towards reactive DNS in complex internal combustion engine geometries	Engineering	1,285
ENL	Marc Berman	The University of Chicago (UChicago)	Environmental Neuroscience	Biological Sciences	3,497
epic-ecal	Sylvester Johannes Joosten	Argonne National Laboratory	Imaging Calorimeter Design & Optimization for the ePIC Detector	Physics	4,000
ESGF2	Ian Foster	Argonne National Laboratory	Earth System Grid Federation 2 Project	Earth Science	499
ESSAI_2023	Jeremy Allen Sauer	The National Center for Atmospheric Research (NCAR)	Exploring potential emerging hardware technologies for AI and Accelerated Earth System Science	Earth Science	2,385
EVITA	George K Thiruvathukal, Nicholas John Eliopoulos, Nicholas Michael Synovic	Loyola University Chicago	Energy-efficient VIsual Transformer Architecture: Transformer Models for Deployment on Embedded Systems	Computer Science	3,894
ExaNek_NEUP_IRP2	Saumil Sudhir Patel	Argonne National Laboratory	Exascale Simulations of Thermal-Hydraulics Phenomena In Advanced Reactors and Validation Using High Resolution Experimental Data	Engineering	538
ExtraNuc	Krishnan Raghavan	Argonne National Laboratory	Extrapolation of Nuclear Observables	Physics	396
ExtremeConvection	Mahendra Kumar Verma	Indian Institute of Technology Kanpur	Simulation of Turbulent Convection at Extreme Parameters	Physics	1,793
fallwkshp23	Yasaman Ghadar	Argonne National Laboratory	fall workshop 2023	Computer Science	3,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
fciqmc-julia	Mingrui Yang	Washington University-St. Louis	Full Configuration Interaction Quantum Monte Carlo in Julia	Physics	1,793
FD4GW	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Federated learning for gravitational wave astrophysics	Physics	278
filmcooling	Pinaki Pal	Argonne National Laboratory	Robust Film Cooling Under Manufacturing Uncertainty For Improved Jet Engine LifeCycle Energy Efficiency	Energy Technologies	666
FLUPS	Gilles Poncelet	Université Catholique de Louvain	3D distributed Fourier-based Poisson solver	Engineering	818
fm_electrolyte	Venkatasubramanian Viswanathan	Carnegie Mellon University	Foundation Models for Electrolyte Design	Materials Science	4,795
FourierHPO	Sri Hari Krishna Narayanan	Argonne National Laboratory	Large-scale Hyperparameter Optimization and Neural Architecture Search for Fourier Neural Operators in Ocean Modeling	Computer Science	807
fuelspray	Miaoqi Chu	Argonne National Laboratory	Ultrafast X-ray Vision of fuel injection and near-field spray	Energy Technologies	500
FUN3D_SYCL	Eric john Nielsen	National Aeronautics and Space Administration (NASA), Langley	Evaluation of FUN3D SYCL Implementation	Engineering	8,000
FURx	Yuri Alexeev	Argonne National Laboratory	Variational quantum circuit simulations	Computer Science	7,500
FusAblator	Ivan Oleynik	University of South Florida (USF)	Predictive Simulations of Inertial Confinement Fusion Ablator Materials	Fusion Energy	2,385
fusiondl_aesp	William Tang	Princeton University	Accelerated Deep Learning Discovery in Fusion Energy Science	Fusion Energy	687
GeomicVar	Ravi Kiran Madduri	Argonne National Laboratory	Scaling Genomic Variant callers to Leadership-class systems: A collaboration between VA-MVP and DOE	Biological Sciences	2,385
GPU-DG	Pinaki Pal	Argonne National Laboratory	GPU-enabled Discontinuous Galerkin simulations of complex fluid flows	Engineering	112
GPU-Trajectories	Maninder Singh Grover	University of Dayton	Molecular Trajectory Analysis on GPU Platform	Chemistry	4,000
GPUBenchDFT	Ganesh Sivaraman	Argonne National Laboratory	Benchmark of GPU based Real Space and Plane Wave DFT Codes	Chemistry	1,636
gpu_hack	Yasaman Ghadar, Raymond M. Loy	Argonne National Laboratory	GPU Hackathon	Training	40,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
GRACE	Sayan Ghosh	Pacific Northwest National Laboratory (PNNL)	Graph Analytics Codesign on GPUs	Computer Science	2,471
graphs_	Dossay Oryspayev	Brookhaven National Laboratory (BNL)	Exploration of parallelization of graph algorithms	Computer Science	2,000
gtcarbon	Chao Xu	Argonne National Laboratory	Computational modeling of cost-effective carbon capture technologies on industrial gas turbines to reduce CO2 emission	Energy Technologies	2,345
h-amr	Alexander Chekhovskoy	Northwestern University	Testing the scaling of GPU-accelerated code H-AMR for INCITE application	Physics	4,000
HACC_aesp	Katrin Heitmann	Argonne National Laboratory	Extreme-Scale Cosmological Hydrodynamics	Physics	687
HACC_P3HPC	Esteban Miguel Rangel	Argonne National Laboratory	P3HPC Performance Results for CRK-HACC	Computer Science	1,244
HEDM_Viz	Chihpin Chuang	Argonne National Laboratory	Visualization for High Energy Diffraction Microscopy	Materials Science	311
HighReyTurb_PostProc	Robert D. Moser, Myoungkyu Lee	The University of Alabama	Data analysis of turbulent Channel Flow at High Reynolds number	Engineering	954
HIV-DRUG	Ao Ma	University of Illinois Chicago	Understanding the rigorous molecular mechanism of drug resistance to HIV protease inhibitors	Biological Sciences	9,843
hp-ptycho	Tekin Bicer	Argonne National Laboratory	High Performance 3D Ptychographic Reconstruction and Image Enhancement	Materials Science	2,450
HPC-GUI	Wei Jiang	Argonne National Laboratory	Development of High Performance Computing - Graphical User Interface for ALCF	Computer Science	4,000
hpc-spectacle	Kevin Antoney Brown	Argonne National Laboratory	Evaluate and Optimize Data Movement Strategies in AI and Climate Science Workloads	Computer Science	2,242
hpcbdsm	Tanwi Mallick	Argonne National Laboratory	High-Performance Computing and Big Data Solutions for Mobility Design and Planning	Computer Science	3,816
hpe_dragon_collab	Venkatram Vishwanath	Argonne National Laboratory	Collaboration with HPE on the Dragon project evaluation and scaling	Computer Science	2,000
hydrosm	Jeremy A Feinstein	Argonne National Laboratory	Improving the predictability of hydrological systems with AI	Earth Science	5,493
hypersonic01	Chonglin Zhang	University of North Dakota	Kinetic Modeling of Hypersonic Nonequilibrium Flows	Engineering	280

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
hyper_bound_layer	Carlo Scalo	Purdue University	Passive control of hypersonic boundary layers via wall treatments	Engineering	3,266
IBM-GSS	Venkatram Vishwanath	Argonne National Laboratory	IBM GeoSpatial Software System	Earth Science	3,160
IMEXLBM	Saumil Sudhir Patel	Argonne National Laboratory	ECP ProxyApp Development for the Lattice Boltzmann Method	Computer Science	918
insitu	Silvio Humberto Rafael Rizzi	Argonne National Laboratory	In situ instrumentation of NekRS	Computer Science	2,000
Intel	Kalyan Kumaran, Scott Parker, Timothy Joe Williams, Venkatram Vishwanath	Argonne National Laboratory	Intel employees in support of Theta	Internal	3,351
introtohpc_faculty	Paige Carolyn Kinsley	Argonne National Laboratory	ALCF Faculty Curriculum Development Program	Computer Science	121
IQC	Murat Keceli	Argonne National Laboratory	Interactive Quantum Chemistry	Chemistry	4,000
IRIBeta	Thomas David Uram	Argonne National Laboratory	IRI explorations	Computer Science	1,000
LAR-EM	Zheng Zhang	The University of Chicago (UChicago)	Electron Tomography from Limited-Angular-Range Data	Materials Science	1,999
LASSCF_gpudev	Christopher James Knight	Argonne National Laboratory	GPU Development of LASSCF	Chemistry	2,613
LatticeQCD_aesp	Norman Howard Christ, Paul Mackenzie	Fermi National Accelerator Laboratory (Fermilab)	Lattice Quantum Chromodynamics Calculations for Particle and Nuclear Physics	Physics	687
lbpm_sycl	James Edward McClure	Virginia Polytechnic Institute and State University (Virginia Tech)	LBPM Performance Optimization	Earth Science	1,289
LeadChalco	Burak Guzelturk	Argonne National Laboratory	Understanding of local structure in lead chalcogenides and their nanocrystals	Physics	1,494
les120	George Em Karniadakis	Brown University	Learning the sub-grid model in Large Eddy Simulations using domain-decomposition based parallel physics-informed neural networks (PINNs)	Mathematics	2,000
libensemble	Stephen Tobias Paige Hudson	Argonne National Laboratory	libEnsemble: Project testing, development and scaling study	Computer Science	35

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
LibPressioTomo	Robert Raymond Underwood	Argonne National Laboratory	Science Preserving Data Approximation and Input Output Optimization Using Artificial Intelligence for Real-time High-resolution Tomography on Integrated Research	Computer Science	500
LLM4CS	Franck Cappello	Argonne National Laboratory	LLMs training with ACM Digital Library	Computer Science	285
LoopSynch	Shina Caroline Lynn Kamerlin	Georgia Institute of Technology (Georgia Tech)	Link between loop dynamics and turnover number across 30 extant Triosephosphate Isomerases	Chemistry	330
LQCDdev	James Clifton Osborn	Argonne National Laboratory	Lattice QCD development	Physics	2,611
lqcdml_aesp	William Detmold	Massachusetts Institute of Technology (MIT)	Machine Learning for Lattice Quantum Chromodynamics	Physics	18,902
LSSMEQ	Marco Govoni	Argonne National Laboratory	Large-scale simulations of materials for energy and quantum information science	Materials Science	143,000
LSWDD_SC	David Weihua He	University of Illinois at Chicago	Large Scale Wafer Defect Diagnosis using Supercomputer	Computer Science	180
Maintenance	William Edward Allcock, John Francis O'Connell, John Patrick Reddy, Ryan Milner, Torrance Ivan Leggett	Argonne National Laboratory	LCF Operations System Maintenance	Internal	32,501
matml_aesp	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials	Materials Science	16,497
Mayo	Nicholas Lee-Ping Chia	Argonne National Laboratory	Colorectal Cancer Evolution Project - Mayo & Argonne	Biological Sciences	557
mayopath	Nicholas Lee-Ping Chia	Mayo Clinic-Minnesota	Mayo Pathology	Biological Sciences	9,595
MBXOpt	Christopher James Knight	Argonne National Laboratory	Enabling Chemical Accuracy Through GPU-accelerated Large-Scale Many-body Molecular Dynamics Expansions	Chemistry	473
MFIX-Exa	William David Fullmer	National Energy Technology Laboratory (NETL)	MFIX-Exa: Performance Prediction of Multiphase Energy Conversion Device	Energy Technologies	1,597
MI2Dmaterials	Trevor David Rhone	Rensselaer Polytechnic Institute (RPI)	Materials informatics study of two-dimensional magnetic materials and their heterostructures	Materials Science	167

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
ML-Coupling	Emil Mihai Constantinescu	Argonne National Laboratory	Data-driven Coupling Methods for Atmospheric-Ocean Interactions	Earth Science	1,109
MLP4THERMO	Cem Sevik	University of Antwerp	Machine Learning Potentials for Thermal Properties of Two-Dimensional Materials	Materials Science	10,677
mm_protein	Andrew L Ferguson	The University of Chicago (UChicago)	Multimodal contrastive learning for natural language-prompted data-driven protein design	Biological Sciences	2,207
MOAB_App	Vijay Subramaniam Mahadevan	Argonne National Laboratory	MOAB Algorithmic Performance Portability	Mathematics	2,556
MPICH_MCS	Kenneth James Raffenetti, Pavan Balaji	Argonne National Laboratory	MPICH - A high performance and widely portable MPI implementation	Computer Science	163
MultiActiveAI	Dario Dematties	Northwestern Argonne Institute of Science and Engineering (NAISE)	Multimodal Intelligence for Federated Edge Computing Simulations	Computer Science	2,611
multiphysics_aesp	Amanda Randles	Duke University	Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations	Engineering	18,559
MVAPICH2	Dhabaleswar Kumar Panda	The Ohio State University	Optimizing and Tuning MVAPICH2-GDR Library and study Its Impact on HPC and AI Applications	Computer Science	4,886
N2M3O7	Iwnetim Iwnetu Abate	Massachusetts Institute of Technology (MIT)	Tunable catalysis in transition metal oxide: NaxMn3O7	Materials Science	309
nacatho1	Wenhua Zuo	Argonne National Laboratory	DeePMD simulations of the synthesis of layered oxide cathodes for advanced batteries	Materials Science	1,634
NAISE-NIH-NIAMS	Marta Garcia Martinez	Argonne National Laboratory	Automatic MRI segmentation for upper limb muscles for clinical applications	Biological Sciences	1,466
NAMD_aesp	Benoit Roux, James Christopher Phillips	The University of Chicago (UChicago)	Free Energy Landscapes of Membrane Transport Proteins	Biological Sciences	13,747
NAQMC_RMD_aesp	Aiichiro Nakano	University of Southern California (USC)	Metascalable Layered Materials Genome	Materials Science	687
nekrs-scaling	Pinaki Pal	Argonne National Laboratory	NekRS scalability studies for gas turbine film cooling high-fidelity simulations	Energy Technologies	3,834
nekrsCombustion	Benjamin William Keeton	Argonne National Laboratory	NekRS Combustion - Gas Turbines	Engineering	403

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
NeuralDE	Romit Maulik	The Pennsylvania State University (Penn State/PSU)	Embedding partial differential equation solvers within neural networks for in-situ scientific machine learning	Computer Science	1,606
neutrinoGPU	Corey J Adams	Argonne National Laboratory	GPU-Accelerated Neutrino Software	Physics	1,607
novacosmics	Alexander I Himmel	Fermi National Accelerator Laboratory (Fermilab)	NOvA Cosmic Rejection	Physics	466
NovelSemi	Feliciano Giustino	The University of Texas at Austin	Computational design of novel semiconductors for power and energy applications	Materials Science	2,385
NSLS2DataProcessing	Stuart Ian Campbell	Brookhaven National Laboratory (BNL)	National Synchrotron Light Source II (NSLS-II) Data Processing	Computer Science	1,000
NUCat_Micro-CT	Marta Garcia Martinez	Argonne National Laboratory	Large volume feline spinal cord microtomography	Biological Sciences	1,999
NucMatDyn	Aurel Bulgac	University of Washington	Nuclear Matter Dynamics in Real Time and the Heaviest Elements in Nature	Chemistry	57,395
NuQMC	Alessandro Lovato	Argonne National Laboratory	Nuclear quantum Monte Carlo	Physics	941
Nu_Novel	Zelimir Djurcic	Argonne National Laboratory	Novel LArTPC Simulation and Reconstruction	Physics	2,147
NWChemEx_aesp	Theresa Windus, Alvaro Vazquez Mayagoitia	Pacific Northwest National Laboratory (PNNL)	NWChemEx: Tackling Chemical, Materials & Biochemical Challenges in the Exascale Era	Chemistry	17,184
Oceananigans	Michel Schanen	Argonne National Laboratory	Large-scale fluid dynamics simulation in Julia using Oceananigans	Earth Science	2,000
OmniverseEval	Joseph A Insley	Argonne National Laboratory	NVIDIA Omniverse Evaluation	Computer Science	5,438
OpenCosmo	Katrin Heitmann	Argonne National Laboratory	HACC Simulation Data Portal	Physics	2,345
Operations	William Edward Allcock	Argonne National Laboratory	Systems administration tasks	Internal	69,573
OptADDN	Sandeep Madireddy	Argonne National Laboratory	Optimal Architecture discovery for deep probabilistic models and neuromorphic systems	Computer Science	4,373
PARTURB3D	Ramesh Balakrishnan	Argonne National Laboratory	Simulating turbulent particulate flows inside enclosures	Engineering	3,531
PCAD	Brigitte Wex	Lebanese American University	Utilization of Metal-Phthalocyanines toward Mitigation of Alzheimer's Disease	Chemistry	2,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
pCLAMD	Madhurima Vardhan	Argonne National Laboratory	pCLAMD - Parallelized Clinical Language model to Aid Medical Decision-making	Biological Sciences	1,303
PEDAL	Tom Peterka	Argonne National Laboratory	PEDAL: Parallel Extreme Scale Data Analysis	Computer Science	1,208
Performance	Scott Parker, Raymond M. Loy	Argonne National Laboratory	Performance	Internal	69,594
PHASTA_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Extreme Scale Unstructured Adaptive CFD: From Multiphase Flow to Aerodynamic Flow Control	Engineering	13,747
PHASTA_NCSU	Igor A Bolotnov	North Carolina State University (NCSU)	Multiphase Simulations of Nuclear Reactor Thermal Hydraulics	Engineering	309
PHTMD	Cem Sevik	University of Antwerp	Phase Transition Properties of Transition Metal Dichalcogenides	Physics	3,814
Polaris	Torrance Ivan Leggett	Hewlett Packard Enterprise	Polaris project for installation and related work for the vendors	Internal	738
PolyEFluc	Pamela Chenyang Cai	The University of Chicago (UChicago)	Theoretical prediction of dynamics from density fluctuations in polyelectrolyte complexes	Materials Science	203
PPI	Arjun Saha	University of Wisconsin-Milwaukee	Simulation of Protein Interfaces for Skin Cancer	Biological Sciences	2,549
Pred_AI	Zhen Xie	The State University of New York at Binghamton	A Performance Predictor to Select the Optimal AI Accelerator for DNN Training	Computer Science	1,129
prs-atlas	Ravi Kiran Madduri	Argonne National Laboratory	An Atlas of Multi-ancestry Polygenic Risk Scores for Heritable Human Disease	Biological Sciences	5,000
ptau	Christopher Robert Brue	University of Pennsylvania (UPenn)	Mechanistic Insights into the Role of the Proline Rich Region of Tau in Tubulin Binding and Polymerization	Biological Sciences	2,000
QE_dev	Ye Luo	Argonne National Laboratory	Quantum ESPRESSO and its downstream code development and benchmark	Materials Science	2,000
QMCPACK_aesp	Anouar Benali	Argonne National Laboratory	Extending Moore's Law computing with Quantum Monte Carlo	Materials Science	18,559
QSupremacy	Yuri Alexeev	Argonne National Laboratory	Quantum supremacy	Computer Science	2,345
QTensor	Yuri Alexeev	Argonne National Laboratory	Quantum circuit simulations	Computer Science	9,809
quantom_scidac	John Taylor Childers	Argonne National Laboratory	The QuantOm Event-Level Inference Framework	Physics	2,919

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
QuantumDS	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Quantum mechanics and Data Science	Chemistry	5,948
QuantumSupremacy	Yuri Alexeev	Argonne National Laboratory	Verification of quantum supremacy and advantage claims	Computer Science	4,000
radix-io	Philip Hutchinson Carns	Argonne National Laboratory	System software to enable data-intensive science	Computer Science	2,220
RAN	William Edward Allcock, Brian R. Toonen	Argonne National Laboratory	RAM Area Network	Computer Science	325
RAPINS	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Reproducible and Accelerated Physics-inspired Neural Networks	Physics	16,550
RaptorX	Jinbo Xu	Toyota Technological Institute at Chicago (TTIC)	Protein Folding through Deep Learning and Energy Minimization	Biological Sciences	3,809
RECUP	Line Catherine Pouchard	Argonne National Laboratory	RECUP - capture and management of in-depth provenance data from HPC workflows	Computer Science	4,939
ReForMerS	Sandeep Madireddy	Argonne National Laboratory	Robust and Reliable Foundation Models for Science	Computer Science	4,517
remote_offloading	Jose Manuel Monsalve Diaz	Argonne National Laboratory	Exploring collective operations with Remote Offloading	Computer Science	3,262
RL-fold	Arvind Ramanathan	Argonne National Laboratory	Targeting intrinsically disordered proteins using artificial intelligence driven molecular simulations	Biological Sciences	160,453
RNAtoImage	Ravi Kiran Madduri	Argonne National Laboratory	RNA-to-Image Synthesis project	Biological Sciences	5,895
S3D	Silvio Humberto Rafael Rizzi	Argonne National Laboratory	Porting S3D code to Polaris	Chemistry	1,434
safcomb	Marcus Steven Day	National Renewable Energy Laboratory (NREL)	Turbulent DNS of SAFs in an aero combustor	Chemistry	1,690
sbi-fair	Pete Beckman, Kamil Antoni Iskra	Argonne National Laboratory	FAIR Surrogate Benchmarks Supporting AI and Simulation Research	Computer Science	8,088
scalablepinns	Paris Perdikaris	University of Pennsylvania (UPenn)	Scalable PINNs	Engineering	4,279
SCPlasma	Ranganathan Gopalakrishnan	University of Memphis	Thermodynamics and Transport Models of Strongly Coupled Dusty Plasmas	Physics	3,656
SCREAM_Calib	Jiali Wang	Argonne National Laboratory	Towards Neighborhood Scale Climate Simulations using AI and Accelerated GPUs	Earth Science	5,253

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
SDR	Sheng Di	Argonne National Laboratory	Scalable Dynamic Scientific Data Reduction Framework (SDR)	Computer Science	1,309
SEEr-planning	Zhiling Lan	Illinois Institute of Technology (IIT)	Performance and Power Tradeoff Analysis of AI-Enabled Science on CPU-GPU System	Computer Science	6,079
SEEr-Polaris	Zhiling Lan	University of Illinois at Chicago	AI-enabled Benchmarking on Polaris	Computer Science	1,756
SENSEI	Silvio Humberto Rafael Rizzi, Joseph A Insley, Nicola Joy Ferrier, Venkatram Vishwanath	Argonne National Laboratory	Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery	Computer Science	3,731
SolarWindowsADSP	Jacqueline Manina Cole	University of Cambridge	Data-Driven Molecular Engineering of Solar-powered Windows	Materials Science	8,043
SOLLVE	Sunita Chandrasekaran	Brookhaven National Laboratory (BNL)	Scaling OpenMP With LLVm for Exascale Performance and Portability	Computer Science	2,000
spentFuel	Angela Di Fulvio	University of Illinois at Urbana-Champaign	Cask Mis-loads Evaluation Techniques	Nuclear Energy	2,000
SSLCos	Dario Dematties	Northwestern University	Self-supervised learning for cosmology	Physics	1,244
stabCNS	Ali Karakus	Middle East Technical University	GPU Accelerated High-order Compressible Flow Simulations	Engineering	1,946
StellTurbOpt	Walter Allen Guttenfelder	Type One Energy Group, Inc.	Optimizing multi-channel turbulent transport in stellarator fusion pilot plants	Fusion Energy	4,795
STlearn	Shinjae Yoo	Brookhaven National Laboratory (BNL)	Spatiotemporal learning for human neuroscience	Biological Sciences	9,706
Substrate-transport	Wonpil Im	Lehigh University	Characterizing Energy Landscape of Substrate Translocation in Bacterial Membrane	Biological Sciences	3,751
SuperBERT	Ian Foster	Argonne National Laboratory	Training of language models on large quantities of scientific text	Computer Science	30,612
Sustainable_RecSys	Yong Zheng	Illinois Institute of Technology (IIT)	Sustainable Deployment of Recommendation Models	Computer Science	4,500
swift-t-polaris	Justin Michael Wozniak	Argonne National Laboratory	Swift/T on Polaris	Computer Science	861
SYCLSupport	Kevin Harms	Argonne National Laboratory	SYCL Support on ALCF Systems	Computer Science	1,529

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
TernaryLBM	Chunheng Zhao	The City University of New York (CUNY)	Accelerating ternary flow simulations using the lattice Boltzmann method	Engineering	4,000
TFXcan	Ravi Kiran Madduri	Argonne National Laboratory	Predicting transcription factor binding and other epigenetic features to gain insight into the biology of diseases	Biological Sciences	4,000
ThroughFocal_DD	Jonathan Tyler Schwartz	University of Michigan	Aberration Corrected Through Focal Electron Tomography Through Focal_DD	Materials Science	5,153
TMEM_DEL	Diomedes Elias Logothetis	Northeastern University	Molecular Dynamics simulation on TMEM16A chloride channel	Biological Sciences	81
Tools	Scott Parker	Argonne National Laboratory	ALCF Performance Tools	Internal	2,836
Torantula	Benjamin Alan Blakely	Argonne National Laboratory	Topical Clustering of Web Data for Identification of Potential Material of Concern	Computer Science	2,000
tpc	Venkatram Vishwanath	Argonne National Laboratory	Trillion Parameter Consortium	Computer Science	4,000
training_polaris	Paige Carolyn Kinsley, Yasaman Ghadar	Argonne National Laboratory	Training Polaris	Training	1,000
transformer_eval	Rick Lyndon Stevens	Argonne National Laboratory	Transformers on AI accelerators	Computer Science	1,997
TRB	Parisa Mirbod	University of Illinois at Chicago	Turbulent Rayleigh-Benard Convection in Suspensions of Bubbles	Engineering	38,849
TRIDENT	Marta Galbiati	Politecnico di Milano	invesTigation of laseR-drIven raDiation sourcEs with advaNced Targets	Physics	267
TropicalMeteorology	Inna Polichtchouk	European Centre for Medium-Range Weather Forecasts	New Window into Tropical Meteorology with Global 1 km Atmosphere-Ocean Simulations	Earth Science	10,000
TURBSEP	Ricardo Vinuesa	KTH Royal Institute of Technology	Simulating the largest turbulent boundary layers subjected to pressure gradient, roughness, compressibility and separation	Engineering	8,000
uchicago-alcf-hub	Hakizumwami Birali Runesha	Argonne National Laboratory	University of Chicago Research Computing / ALCF Partnership Hub	Computer Science	878
UIC-CS494-Sp2023	Michael E. Papka	University of Illinois at Chicago	CS494 Introduction to High Performance Computing	Computer Science	6,000
UIC-HPC	Michael E. Papka, Zhiling Lan	Argonne National Laboratory	UIC High Performance Computing Research	Computer Science	517

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
UINTAH_aesp	Martin Berzins, John Andrew Schmidt	The University of Utah	Design and evaluation of high-efficiency boilers for energy production using a hierarchical V/UQ approach	Chemistry	687
UncertaintyDL	Maria Pantoja	Argonne National Laboratory	Quantifying Uncertainty in Deep Learning	Computer Science	1,450
User_Services	Haritha Siddabathuni Som	Argonne National Laboratory	User Services	Internal	2,498
vacuumms	Franklin Ted Willmore	Boise State University	Extending VACUUMMS for analysis of MD trajectories	Materials Science	740
variability_eht	Benjamin Scott Prather, Ben Prather	Los Alamos National Laboratory (LANL)	Variability Modeling of SgrA* for the EHT	Physics	4,000
VeloC	Bogdan Florin Nicolae	Argonne National Laboratory	VeloC: Very Low Overhead Checkpointing System	Computer Science	10,207
VenkatramVishwanath	Venkatram Vishwanath	Argonne National Laboratory	Venkatram Vishwanath	Computer Science	2,000
visualization	Joseph A Insley, Michael E. Papka	Argonne National Laboratory	Visualization and Analysis Research and Development for ALCF	Internal	69,586
Waggle	Pete Beckman	Argonne National Laboratory	AI at the edge for domain sciences	Computer Science	2,000
wall_turb_dd	Ramesh Balakrishnan	Argonne National Laboratory	Wall Resolved Simulations of Canonical Wall Bounded Flows	Engineering	3,731
WALSforAll	Christine Mary Simpson	Argonne National Laboratory	Workflows Across Labs	Computer Science	650
warpx_frc_stability	Roelof Erasmus Groenewald	TAE Technologies, Inc.	Global stability simulations of advanced FRC using ECP tools	Fusion Energy	41
wereszczynski	Jeffery Michael Wereszczynski	Illinois Institute of Technology (IIT)	MD Simulations of Chromatin Modification and Gene Regulation Mechanism	Biological Sciences	1,756
WMLES	Zhi Jian WANG	Kansas State University	Wall modeled large eddy simulation for turbomachinery applications	Engineering	1,756
WRF_col	Jose Manuel Monsalve Diaz	Argonne National Laboratory	Exploring weather forecast simulation in tropical regions	Earth Science	495
XGC_aesp	Choongseok Chang	Princeton Plasma Physics Laboratory (PPPL)	High fidelity simulation of fusion reactor boundary plasmas	Fusion Energy	13,747
XMultiImage	Phay J Ho	Argonne National Laboratory	Multimodal Imaging with Intense X-ray Pulses	Chemistry	61
YouDiffuse	Nicholas Lee-Ping Chia	Argonne National Laboratory	YouDiffuse: Training Latent Diffusion Models by Leveraging Educational Medical Content from YouTube	Computer Science	1,661
				Total DD	2,905,511

January 1, 2023–December 31, 2023
Director's Discretionary (DD) Projects on Theta (by Project Name)

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
3DChromatin	Jie Liang	University of Illinois at Chicago	Large Ensemble Model of Single-Cell 3D Genome Structures	Biological Sciences	64,000
ACO2RDS	John J Low	Argonne National Laboratory	Adsorptive CO2 Removal from Dilute Sources	Materials Science	45,000
ALCF_for_DUNE	Aleena Rafique	Argonne National Laboratory	Enabling precise measurement of neutrino interactions using advanced computing in DUNE	Physics	32,000
ALEXIS	Jorge Rene Padial	Vanderbilt University Medical Center	Automatically Labelled EUV and XRay Incedent SolarFlare catalog	Physics	5,000
Allinea	Raymond M. Loy, Kalyan Kumaran	Argonne National Laboratory	Allinea/Linaro Debug and Performance Tools Support	Internal	781
APSDataAnalysis	Rafael Vescovi	Argonne National Laboratory	APS Beamline Data Processing and Analysis	Computer Science	6,699
Aramco-PreChamber	Joohan Kim	Argonne National Laboratory	High-Fidelity LES of Turbulent Jet Combustion	Engineering	28,456
arfc-msr-ahtr	Madicken Munk	University of Illinois at Urbana-Champaign	Modeling of Molten Salt Reactor Design, Optimization, and Transient Behavior	Nuclear Energy	25,000
atlas_aesp	Walter Howard Hopkins	Argonne National Laboratory	Simulating and Learning in the ATLAS detector at the Exascale	Physics	19,500
ATPESC2023	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program for Extreme-Scale Computing 2023	Training	20,010
ATPESC_Instructors	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing Instructors	Training	1,000
Auriga	Christine Mary Simpson	Argonne National Laboratory	The Auriga Project	Physics	100
AutoBEM	Joshua Ryan New	Oak Ridge National Laboratory (ORNL)	Automatic Building Energy Modeling and analysis	Energy Technologies	8,716,068
BACTWA	Philippe Regis-Guy Piot	Northern Illinois University (NIU)	Beam Acceleration and Control in Structure Wakefield Accelerator	Physics	32,000
BIP167	Philip Kurian	Howard University	Computing superradiance and van der Waals many-body dispersion effects from MD simulations of biomolecular complexes	Physics	63,960
bloodflow_dd	Jifu Tan	Northern Illinois University (NIU)	Multiphysics modeling of clot mechanics	Engineering	5,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
BRAIN	Getnet Dubale Betrie	National Renewable Energy Laboratory (NREL)	Scalable Brain Simulator for Extreme Computing	Biological Sciences	16,000
BS-SOLCTRA	Esteban Meneses	Costa Rica National High Technology Center	Plasma Physics Simulations for SCR-1 Stellarator	Physics	5,918
candle_aesp	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	Virtual Drug Response Prediction	Biological Sciences	2,000
Carbon_composites	Hendrik Heinz	University of Colorado-Boulder	Designing Functional Nanostructures and Carbon-Based Composite Materials	Materials Science	13,090
catalysis_aesp	David Hamilton Bross	Argonne National Laboratory	Exascale Computational Catalysis	Chemistry	12,500
Catalyst	Katherine M Riley, Christopher James Knight, James Clifton Osborn, Timothy Joe Williams	Argonne National Laboratory	Catalyst	Internal	4,000
Cellulose-Simulation	Dewei Qi	Western Michigan University	Molecular dynamics simulation of nano-cellulose	Engineering	57,628
cfdml_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Data Analytics and Machine Learning for Exascale CFD	Engineering	8,000
CharmRTS	Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia	University of Illinois at Urbana-Champaign	Charm++ and its applications	Computer Science	8,000
climate_severe	Vittorio Angelo Gensini	Northern Illinois University (NIU)	Anticipating Severe Weather Events via Dynamical Downscaling	Earth Science	7,164
Clouds	Ian Foster	Argonne National Laboratory	Unsupervised analysis of satellite cloud imagery	Earth Science	12,080
CobaltDevel	Paul Michael Rich, William Edward Allcock	Argonne National Laboratory	Cobalt Development	Internal	5,657
CONUS-Carbon	Jinxun Liu	U.S. Geological Survey (USGS)	Terrestrial ecosystem carbon cycle of the conterminous U.S.	Earth Science	28,400
ConvReac_DD	Yiqi Yu	Argonne National Laboratory	Investigation of Flow and Heat Transfer Behavior in Involute Plate Research Reactor with Large Eddy Simulation to Support the Conversion of Research Reactors to Low	Nuclear Energy	32,000
covid-ct	Ravi Kiran Madduri	Argonne National Laboratory	Medical Imaging Domain-Expertise Machine Learning for Interrogation of COVID	Computer Science	1,636

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
Cray	Torrance Ivan Leggett, Mark Richard Fahey, Susan Marie Coghlan, Timothy Joe Williams, William Edward Allcock	Hewlett Packard Enterprise	Cray Installation	Internal	1,000
CSC249ADCD01	Ian Foster	Argonne National Laboratory	2.2.6.03 ADCD01-CODAR	Computer Science	4,000
CSC249ADCD02	Susan Marie Mniszewski,Timothy C. Germann	Los Alamos National Laboratory (LANL)	2.2.6.04 ADCD02-COPA: Co-Design Center for Particle Applications	Physics	4,000
CSC249ADCD04	Tzanio Valentinov Kolev, Misun Min, Paul Frederick Fischer	Lawrence Livermore National Laboratory (LLNL)	2.2.6.06 CEED: Center for Efficient Exascale Discretizations	Computer Science	11,392
CSC249ADCD05	Mahantesh Halappanavar	Pacific Northwest National Laboratory (PNNL)	2.2.6.07 ADCD05-ExaGraph	Computer Science	13,000
CSC249ADCD08	Francis Joseph Alexander	Brookhaven National Laboratory (BNL)	2.2.6.08 ADCD08-ExaLearn	Physics	8,000
CSC249ADCD09	John Bell	Lawrence Berkeley National Laboratory (LBNL)	2.2.6.05 ADCD03-AMREX: Block-Structured AMR Co-Design Center	Mathematics	3,100
CSC249ADCD502	Kenneth John Roche	Pacific Northwest National Laboratory (PNNL)	2.2.6.02 ADCD502 Application Assessment	Computer Science	4,000
CSC249ADCD504	Jeanine Cook, Shirley Victoria Moore	Lawrence Livermore National Laboratory (LLNL)	2.2.6.01 ADCD504-Proxy Applications	Computer Science	3,100
CSC249ADOA01	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	2.2.4.03 ADOA01 CANDLE: Exascale Deep Learning Enabled Precision Medicine for Cancer	Biological Sciences	4,000
CSC249ADSE01	Salman Habib, Katrin Heitmann	Argonne National Laboratory	2.2.3.02 ADSE01-ExaSky	Physics	1,000
CSC249ADSE02	David McCallen	Lawrence Berkeley National Laboratory (LBNL)	2.2.3.03 EQSIM: High Perf, Multidisciplinary Simulations for Regional Scale Earthquake Hazard/Risk Assmts	Earth Science	1,000
CSC249ADSE03	Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie	Fermi National Accelerator Laboratory (Fermilab)	2.2.1.01 ADSE03-LatticeQCD: Exascale Lattice Gauge Theory Opportunities/Reqmts for Nuclear & High Energy Physics	Physics	4,000
CSC249ADSE04	Danny Perez	Los Alamos National Laboratory (LANL)	2.2.1.04 ADSE04-EXAALT - Molecular dynamics at the exascale	Nuclear Energy	3,100

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC249ADSE05	David Paul Trebotich	Lawrence Berkeley National Laboratory (LBNL)	2.2.3.04 ADSE05-Subsurface	Earth Science	3,100
CSC249ADSE06	Jean-Luc Yves Vay	Lawrence Berkeley National Laboratory (LBNL)	2.2.2.06 ADSE06-WarpX	Physics	3,100
CSC249ADSE08	Steven Hamilton, Paul Kollath Romano	Oak Ridge National Laboratory (ORNL)	2.2.2.03 ADSE08 ExaSMR	Nuclear Energy	3,100
CSC249ADSE09	Paul Richard Charles Kent, Anouar Benali	Oak Ridge National Laboratory (ORNL)	2.2.1.06 QMCPACK: Predictive and Improvable Quantum-mechanics Based Simulations	Materials Science	4,000
CSC249ADSE10	Matthew Thomas Bement	Oak Ridge National Laboratory (ORNL)	2.2.1.05 ADSE10-ExaAM	Engineering	3,000
CSC249ADSE11	Theresa Windus	University of Washington	2.2.1.02 ADSE11-NWChemEx: Tackling Chemical, Materials, & Biomolecular Challenges in Exascale	Chemistry	3,100
CSC249ADSE12	Amitava Bhattacharjee	Princeton Plasma Physics Laboratory (PPPL)	2.2.2.05 ADSE12 WDMAPP	Computer Science	4,000
CSC249ADSE13	Michael Sprague	National Renewable Energy Laboratory (NREL)	2.2.2.01 ExaWind: Exascale Predictive Wind Plant Flow Physics Modeling	Energy Technologies	1,000
CSC249ADSE14	Jacqueline Chen	Sandia National Laboratories, California	2.2.2.02 ADSE14-Combustion-Pele: Transforming Combustion Science & Technology with Exascale Simulations	Engineering	3,100
CSC249ADSE15	Mark Alan Taylor	Sandia National Laboratories, New Mexico	2.2.3.05 ADSE15-E3SM-MMF	Earth Science	3,100
CSC249ADSE16	Mark S Gordon	Ames Laboratory	2.2.1.03 ADSE16-GAMESS	Chemistry	25,628
CSC249ADSE18	Daniel Kasen	Lawrence Berkeley National Laboratory (LBNL)	2.2.3.01 ADSE18 Exastar	Physics	3,100
CSC249ADSE20	Katherine Yelick, Leonid Oliker	Lawrence Berkeley National Laboratory (LBNL)	2.2.4.04 ADSE20-ExaBiome: Exascale Solutions for Microbiome Analysis	Biological Sciences	1,000
CSC249ADSE21	Amedeo Perazzo	SLAC National Accelerator Laboratory	2.2.4.05 ExaFEL: Data Analytics at the Exascale for Free Electron Lasers	Physics	1,000
CSC249ADSE22	Christopher Stephen Oehmen, Andres Marquez, Zhenyu Huang	Pacific Northwest National Laboratory (PNNL)	2.2.4.02 ADSE22-ExaSGD	Energy Technologies	3,100

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC249ADSE23	Jordan Michael Musser	National Energy Technology Laboratory (NETL)	2.2.2.04: MFIX-Exa: Perf Prediction of Multiphase Energy Conversion Device	Energy Technologies	3,100
CSC249ADTR01	Daniel Edward Laney	Lawrence Livermore National Laboratory (LLNL)	2.3.5.10 ADTR01-ExaWorks	Computer Science	1,300
CSC249ADTR02	Osni Marques	Oak Ridge National Laboratory (ORNL)	2.4.6.02 ADTR02-Productivity	Computer Science	1,300
CSC250STDA05	Kenneth Dean Moreland	Oak Ridge National Laboratory (ORNL)	2.3.4.13 STDA05-ECP/VTK-m	Computer Science	3,100
CSC250STDM10	Surendra Byna, Venkatram Vishwanath	Lawrence Berkeley National Laboratory (LBNL)	2.3.4.15 ExaIO - Delivering Efficient Parallel I/O on Exascale Computing Systems with HDF5 and Unify	Computer Science	17,000
CSC250STDM11	Scott Klasky, Norbert Podhorszki	Oak Ridge National Laboratory (ORNL)	2.3.4.09 STDM11-ADIOS Framework for Scientific Data on Exascale Systems	Computer Science	4,000
CSC250STDM12	Robert B. Ross, Robert J Latham	Argonne National Laboratory	2.3.4.10 STDM12-DataLib: Data Libraries and Services Enabling Exascale Science	Computer Science	4,000
CSC250STDM14	Franck Cappello	Argonne National Laboratory	2.3.4.14 STDM14 - VeloC-SZ: Very Low Overhead Transparent Multilevel Checkpoint/Restart/SZ: Fast, Effective, Parallel Error-bounded Exascale Loss....	Computer Science	4,000
CSC250STDM16	James Paul Ahrens, Terece Louise Turton	Los Alamos National Laboratory (LANL)	2.3.4.16 STDM16-ALPINE/ZFP	Computer Science	1,300
CSC250STDT10	Jeffrey S Vetter	Oak Ridge National Laboratory (ORNL)	2.3.2.10 STDT10 PPROTEAS-TUNE	Computer Science	3,100
CSC250STDT11	Sunita Chandrasekaran, Doss ay Oryspayev	Stony Brook University	2.3.2.11 SOLLVE: Scaling OpenMP with LLVm for Exascale	Computer Science	3,100
CSC250STDV01	Patrick O'Leary	Kitware Inc.	2.3.4.01 Data and Visualization Software Development Kit	Computer Science	4,000
CSC250STML12	Carol Woodward	Lawrence Livermore National Laboratory (LLNL)	2.3.3.12 Enabling Exascale Simulations with SUNDIALS and hypre	Mathematics	3,100
CSC250STML13	Hartwig Andreas Anzt	The University of Tennessee at Knoxville	2.3.3.13 STML13 - CLOVER	Computer Science	2,100

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC250STML14	Andrey Valeryevich Prokopenko	Oak Ridge National Laboratory (ORNL)	2.3.3.14 ALExa: Accelerated Libraries for Exascale/ForTrilinos: Sustainable Production Fortran Interoperability with Trilinos Libraries	Mathematics	1,000
CSC250STML15	Siva Rajamanickam	Sandia National Laboratories, New Mexico	2.3.3.15 STML-Sake	Computer Science	4,000
CSC250STMS05	Ulrike Meier Yang, Satish Balay	Argonne National Laboratory	2.3.3.01 STMS05-Extreme-scale Scientific xSDK for ECP	Mathematics	3,100
CSC250STMS07	Todd S. Munson, Hong Zhang, Richard Tran Mills, Satish Balay	Argonne National Laboratory	2.3.3.06 STMS07-PETSc/TAO for Exascale	Mathematics	4,000
CSC250STMS08	Xiaoye Sherry Li	Lawrence Berkeley National Laboratory (LBNL)	2.3.3.07 STMS08 STRUMPACK/SuperLU/FFTX: Factorization Based Sparse Solvers and Preconditioners for Exascale	Mathematics	2,100
CSC250STNS01	Michael Lang, Terece Louise Turton	Los Alamos National Laboratory (LANL)	2.3.6.01 - STNS01 -LANL ATDM ST Projects	Computer Science	1,300
CSC250STPM01	Sameer Suresh Shende	University of Oregon	2.3.1.01 Programming Models & Runtimes Software Development Kit	Computer Science	3,100
CSC250STPM08	Patrick McCormick	Los Alamos National Laboratory (LANL)	2.3.1.08 Legion: Enhancing and Hardening the Legion Programming System	Computer Science	1,000
CSC250STPM09	Yanfei Guo	Argonne National Laboratory	2.3.1.07 STPM09-Exascale MPI	Computer Science	3,100
CSC250STPM11	George Bosilca, Earl Luther Carr, Jack Dongarra, Thomas Herault	The University of Tennessee at Knoxville	2.3.1.09 STPM11 ParSEC: Distributed Tasking	Computer Science	3,100
CSC250STPM16	Scott Dov Pakin	Los Alamos National Laboratory (LANL)	2.3.1.16 SICM: Simplified Interface to Complex Memory	Computer Science	4,000
CSC250STPM17	Paul Hamilton Hargrove, Erich Strohmaier	Lawrence Berkeley National Laboratory (LBNL)	2.3.1.14 STPM17-UPC++ & GASNet	Computer Science	5,000
CSC250STPM18	Christian Trott	Sandia National Laboratories, California	2.3.1.18 RAJA/Kokkos	Computer Science	3,100
CSC250STPR19	Peter Hugh Beckman	Argonne National Laboratory	2.3.1.19 STPR19 Argo: Argo/Power Steering	Computer Science	3,100
CSC250STPR27	David Edward Bernholdt	Oak Ridge National Laboratory (ORNL)	2.3.1.17 STPR27-OMPI-X: Open MPI for Exascale	Materials Science	4,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CSC250STSE09	George Todd Gamblin	Lawrence Livermore National Laboratory (LLNL)	2.3.5.09 Software Packaging Technologies	Computer Science	100
CSC250STTO09	Hartwig Andreas Anzt, Anthony Danalis, Earl Luther Carr, Heike Jagode	The University of Tennessee at Knoxville	2.3.2.06 STTO09 EXAPAPI	Computer Science	4,000
CSC250STTO11	John Michael Mellor-Crummey	Rice University	2.3.2.08 STTO11 HPCToolkit	Computer Science	4,000
CSC251HIHE05	Scott Dov Pakin, Simon David Hammond	Los Alamos National Laboratory (LANL)	2.4.2.01 HIHE05-Analytical Modeling - Hardware Evaluation Working Groups	Computer Science	1,300
CSC251HISD01	Ryan Charles Prout	Los Alamos National Laboratory (LANL)	2.4.4.01 HISD01-Software Integration	Computer Science	1,300
CSCSTDT12345	Patrick McCormick	Los Alamos National Laboratory (LANL)	2.3.2.12 Flang: open-source Fortran front end for the LLVM infrastructure	Computer Science	3,100
CVD_CityCOVID	Jonathan Ozik	Argonne National Laboratory	Agent-based model called CityCOVID capable of tracking detailed COVID-19 transmission	Biological Sciences	105,373
darkskyml_aesp	Salman Habib	Argonne National Laboratory	Dark Sky Mining	Physics	8,000
datascience	Venkatram Vishwanath	Argonne National Laboratory	ALCF Data Science and Workflows Allocation	Internal	250,000
dist_relational_alg	Sidharth Kumar	The University of Alabama at Birmingham	Distributed relational algebra at scale	Computer Science	36,000
DL_MODEX	Maruti Kumar Mudunuru	Pacific Northwest National Laboratory (PNNL)	Towards a robust and scalable deep learning workflow for fast, accurate, and reliable calibration of watershed models	Earth Science	74,500
DNS3D	Ramesh Balakrishnan	Argonne National Laboratory	Direct Numerical Simulation of Three Dimensional Turbulence	Engineering	4,414
DNSVCMHD	Keith Daniel Brauss	Francis Marion University	DNS Simulations of Velocity-Current Magnetohydrodynamic Equations	Mathematics	4,265
dynstall_ss	Sarasija Sudharsan	Iowa State University (ISU)	Time-resolved Simulations of Unsteady, Separated Flows	Engineering	25,861
ECP_SDK	Sameer Suresh Shende	University of Oregon	Deploying the ECP SDK software stack at ALCF	Computer Science	1,500
EE-ECP	Xingfu Wu, Valerie Taylor	Argonne National Laboratory	Energy efficient tradeoff among execution time and power of ECP applications	Computer Science	16,125

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
EfficientLLM	Mi Zhang	The Ohio State University	Efficient Large Language Model	Computer Science	4,000
electrolyte-chibueze	Chibueze Vincent Amanchukwu	The University of Chicago (UChicago)	Molecular dynamics-driven discovery of novel liquid electrolytes	Energy Technologies	41,327
EmergingTech2	Benjamin Alan Blakely	Argonne National Laboratory	Identification of Emerging Technologies from Publication Databases	Computer Science	2,000
EngineDNS	Christos Frouzakis	Eidgenössische Technische Hochschule Zürich (ETH Zurich)	Towards reactive DNS in complex internal combustion engine geometries	Engineering	60,000
ENL	Marc Berman	The University of Chicago (UChicago)	Environmental Neuroscience	Biological Sciences	500
epic-ecal	Sylvester Johannes Joosten	Argonne National Laboratory	Imaging Calorimeter Design & Optimization for the ePIC Detector	Physics	20,000
fairscabio	Luiz M. R. Gadelha	Laboratorio Nacional de Computacao Cientifica (LNCC)	Advancing the Reproducibility and Scalability of Bioinformatics Workflows	Biological Sciences	2,000
fciqmc-julia	Mingrui Yang	Washington University-St. Louis	Full Configuration Interaction Quantum Monte Carlo in Julia	Physics	32,000
FRCPlasmas	Jaeyoung Park	TAE Technologies, Inc.	Particle-in-cell simulations of beam-driven, field-reversed configuration plasmas	Physics	100
fusiondl_aesp	William Tang	Princeton University	Accelerated Deep Learning Discovery in Fusion Energy Science	Fusion Energy	1,000
GeneModels	James Davis	Argonne National Laboratory	Development of an artificial intelligence-based framework for predicting gene content and function in microbial genomes	Biological Sciences	25,000
GrainBoundaries	Wissam A Saidi	University of Pittsburgh	Structure and Properties of Grain Boundaries in Materials for Energy Applications	Materials Science	14,400
GWrealtimeBSE	Jin Zhao	University of Science and Technology of China	Ab initio simulation for exciton dynamics in two-dimensional materials using GW + real-time BSE	Materials Science	15,005
HACC_aesp	Katrin Heitmann	Argonne National Laboratory	Extreme-Scale Cosmological Hydrodynamics	Physics	4,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
HEPCloud-FNAL	Burt Holzman	Fermi National Accelerator Laboratory (Fermilab)	High Energy Physics Computing for Fermilab experiments via HEPCloud	Physics	4,724
HEP_on_HPC	Jim B Kowalkowski, Marc Francis Paterno	Fermi National Accelerator Laboratory (Fermilab)	HEP analysis workflows on HPC	Physics	38,624
HighReyTurb_PostProc	Robert D. Moser, Myoungkyu Lee	The University of Alabama	Data analysis of turbulent Channel Flow at High Reynolds number	Engineering	64,000
HiPressMulticompflow	Hongyuan Zhang	University of Minnesota-Twin Cities	Physics-Based Modeling of Multicomponent Transcritical Phase Change and Spray Breakup in High-Pressure Liquid-Fueled Combustors	Engineering	4,090
HIV-DRUG	Ao Ma	University of Illinois Chicago	Understanding the rigorous molecular mechanism of drug resistance to HIV protease inhibitors	Biological Sciences	62,000
hpc-spectacle	Kevin Antoney Brown	Argonne National Laboratory	Evaluate and Optimize Data Movement Strategies in AI and Climate Science Workloads	Computer Science	16,709
Intel	Kalyan Kumaran, Scott Parker, Timothy Joe Williams, Venkatram Vishwanath	Argonne National Laboratory	Intel employees in support of Theta	Internal	4,000
IQC	Murat Keceli	Argonne National Laboratory	Interactive Quantum Chemistry	Chemistry	2,000
IRIBeta	Thomas David Uram	Argonne National Laboratory	IRI explorations	Computer Science	1,000
JCESR	Larry Curtiss, Anubhav Jain	Argonne National Laboratory	Development of High Throughput Methods	Materials Science	37,741
LatticeQCD_aesp	Norman Howard Christ, Paul Mackenzie	Fermi National Accelerator Laboratory (Fermilab)	Lattice Quantum Chromodynamics Calculations for Particle and Nuclear Physics	Physics	97,500
LESDNSHTECESHE2021	Lane Benjamin Carasik	Virginia Commonwealth University (VCU)	LES and DNS of Heat Transfer Enhancements in Clean Energy System Heat Exchangers	Engineering	3,802
lhtes	Kedar Prashant Shete	University of Massachusetts-Amherst	DNS of turbulent phase changing flows	Energy Technologies	7,604
LIGHTCONTROL	Sandra Gail Biedron, Trudy Beth Bolin	University of New Mexico	Light sources and their control using AI techniques	Physics	12,541

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
LoopSynch	Shina Caroline Lynn Kamerlin	Georgia Institute of Technology (Georgia Tech)	Link between loop dynamics and turnover number across 30 extant Triosephosphate Isomerases	Chemistry	23,174
LQCDdev	James Clifton Osborn	Argonne National Laboratory	Lattice QCD development	Physics	15,005
lqcdml_aesp	William Detmold	Massachusetts Institute of Technology (MIT)	Machine Learning for Lattice Quantum Chromodynamics	Physics	13,500
LSSMEQ	Marco Govoni	Argonne National Laboratory	Large-scale simulations of materials for energy and quantum information science	Materials Science	58,210
LSWDD_SC	David Weihua He	University of Illinois at Chicago	Large Scale Wafer Defect Diagnosis using Supercomputer	Computer Science	180
Maintenance	William Edward Allcock, John Francis O'Connell, John Patrick Reddy, Ryan Milner, Torrance Ivan Leggett	Argonne National Laboratory	LCF Operations System Maintenance	Internal	25,000
matml_aesp	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials	Materials Science	245,949
MEDDIAC	Ehud Strobach	Agricultural Research Organization	High resolution interactions of Mediterranean cyclone with ocean eddies	Earth Science	32,000
metastable	Subramanian Sankaranarayanan	Argonne National Laboratory	Metastable phase diagram of material	Materials Science	14,174
MI2Dmaterials	Trevor David Rhone	Rensselaer Polytechnic Institute (RPI)	Materials informatics study of two-dimensional magnetic materials and their heterostructures	Materials Science	85,892
MKM_catal	Wilfred T Tjalke Tysoe	University of Wisconsin-Milwaukee	Enantioselectivity in Heterogeneous Catalysts via the Addition of Chiral Modifiers	Chemistry	227,668
ML-Coupling	Emil Mihai Constantinescu	Argonne National Laboratory	Data-driven Coupling Methods for Atmospheric-Ocean Interactions	Earth Science	2,553
MOAB_App	Vijay Subramaniam Mahadevan	Argonne National Laboratory	MOAB Algorithmic Performance Portability	Mathematics	24,333
ModelingCoronaVirus	Zhangli Peng	University of Illinois at Chicago	Modeling Corona Virus	Biological Sciences	201,284

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
MPICH_MCS	Kenneth James Raffenetti, Pavan Balaji	Argonne National Laboratory	MPICH - A high performance and widely portable MPI implementation	Computer Science	9,736
multimode_comb	Pinaki Pal	Argonne National Laboratory	High-Fidelity CFD Simulations of Multi-Mode Combustion	Energy Technologies	18,557
multiphysics_aesp	Amanda Randles	Duke University	Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations	Engineering	10,000
N2M3O7	Iwnetim Iwnetu Abate	Massachusetts Institute of Technology (MIT)	Tunable catalysis in transition metal oxide: NaxMn3O7	Materials Science	128,000
NAMD_aesp	Benoit Roux, James Christopher Phillips	The University of Chicago (UChicago)	Free Energy Landscapes of Membrane Transport Proteins	Biological Sciences	19,500
NAQMC_RMD_aesp	Aiichiro Nakano	University of Southern California (USC)	Metascalable Layered Materials Genome	Materials Science	10,000
NekM1Comb	Sicong Wu	Argonne National Laboratory	Development of high-fidelity simulations of gas turbine combustors for sustainable aviation applications	Energy Technologies	153,954
Nek_Boost	Pinaki Pal	Argonne National Laboratory	Development of High-Fidelity and Efficient Modeling Capabilities for Enabling Co-Optimization of Fuels and Multi-Mode Engines	Energy Technologies	14,222
NitrateRemoval	Joshua Jodhimani Gabriel	Argonne National Laboratory	Selective Electrochemical Reduction of Nitrate to Value-Added Products using a Reactive Electrochemical Membrane System	Chemistry	17,041
novacosmics	Alexander I Himmel	Fermi National Accelerator Laboratory (Fermilab)	NOvA Cosmic Rejection	Physics	7,063
ns4surf	Robert Bruce Wexler	Washington University-St. Louis	Scratching the Surface Structure With Nested Sampling	Chemistry	114,000
NUCat_Micro-CT	Marta Garcia Martinez	Argonne National Laboratory	Large volume feline spinal cord microtomography	Biological Sciences	28,829
NucContainmentMix	Christopher Fred Boyd	U.S. Nuclear Regulatory Commission (NRC)	LES Simulations of Severe Accident Conditions in Nuclear Containment	Nuclear Energy	5,933

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
NWChemEx_aesp	Theresa Windus, Alvaro Vazquez Mayagoitia	Pacific Northwest National Laboratory (PNNL)	NWChemEx: Tackling Chemical, Materials & Biochemical Challenges in the Exascale Era	Chemistry	2,000
Operations	William Edward Allcock	Argonne National Laboratory	Systems administration tasks	Internal	78,125
OptADDN	Sandeep Madireddy	Argonne National Laboratory	Optimal Architecture discovery for deep probabilistic models and neuromorphic systems	Computer Science	10,000
PARTURB3D	Ramesh Balakrishnan	Argonne National Laboratory	Simulating turbulent particulate flows inside enclosures	Engineering	17,658
PCAD	Brigitte Wex	Lebanese American University	Utilization of Metal-Phthalocyanines toward Mitigation of Alzheimer's Disease	Chemistry	3,024
Performance	Scott Parker, Raymond M. Loy	Argonne National Laboratory	Performance	Internal	500,000
PHASTA_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Extreme Scale Unstructured Adaptive CFD: From Multiphase Flow to Aerodynamic Flow Control	Engineering	19,500
PHASTA_NCSU	Igor A Bolotnov	North Carolina State University (NCSU)	Multiphase Simulations of Nuclear Reactor Thermal Hydraulics	Engineering	15,000
PodPre	Romit Maulik	Argonne National Laboratory	Predictablity analysis of the ERA5 dataset	Earth Science	1,812
prs-atlas	Ravi Kiran Madduri	Argonne National Laboratory	An Atlas of Multi-ancestry Polygenic Risk Scores for Heritable Human Disease	Biological Sciences	5,000
ptau	Christopher Robert Brue	University of Pennsylvania (UPenn)	Mechanistic Insights into the Role of the Proline Rich Region of Tau in Tubulin Binding and Polymerization	Biological Sciences	2,152
PTLearnPhoto	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Materials for Organic Photovoltaics	Materials Science	206,596
QMCPACK_aesp	Anouar Benali	Argonne National Laboratory	Extending Moore's Law computing with Quantum Monte Carlo	Materials Science	19,500
QNAS	Yuri Alexeev	Argonne National Laboratory	Quantum neural architecture search	Computer Science	10,000
QuantumDS	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Quantum mechanics and Data Science	Chemistry	68,054

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
radix-io	Philip Hutchinson Carns	Argonne National Laboratory	System software to enable data-intensive science	Computer Science	43,649
RAPINS	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Reproducible and Accelerated Physics-inspired Neural Networks	Physics	9,554
RCM_4km	Jiali Wang	Argonne National Laboratory	Generation of a next level dataset for regional scale climate modeling: convective resolving spatial scales	Earth Science	11,627
RDD_U3Si2	Gyuchul Park	Argonne National Laboratory	Radiation-driven diffusion of U,Si, and Xe in amorphous U3Si2	Nuclear Energy	1,000
RNAtoImage	Ravi Kiran Madduri	Argonne National Laboratory	RNA-to-Image Synthesis project	Biological Sciences	8,000
rnn-robustness	Liam Benjamin Johnston	University of Wisconsin-Madison	Large-scale Factorial Experiment on RNN Robustness	Computer Science	63,475
RomanDESC	Katrin Heitmann	Argonne National Laboratory	Joint Roman-LSST DESC Time Domain Survey Simulations	Physics	848,030
RTI_DD	Tapan Kumar Sengupta	Indian Institute of Technology Dhanbad	Peta- and Exa-Scale Computing of Rayleigh-Taylor Instability	Engineering	29,252
sbi-fair	Pete Beckman, Kamil Antoni Iskra	Argonne National Laboratory	FAIR Surrogate Benchmarks Supporting AI and Simulation Research	Computer Science	5,000
SCREAM_Calib	Jiali Wang	Argonne National Laboratory	Towards Neighborhood Scale Climate Simulations using AI and Accelerated GPUs	Earth Science	383
SDR	Sheng Di	Argonne National Laboratory	Scalable Dynamic Scientific Data Reduction Framework (SDR)	Computer Science	1,000
SEEr-planning	Zhiling Lan	Illinois Institute of Technology (IIT)	Performance and Power Tradeoff Analysis of AI-Enabled Science on CPU-GPU System	Computer Science	16,841
SENSEI	Silvio Humberto Rafael Rizzi, Joseph A Insley, Nicola Joy Ferrier, Venkatram Vishwanath	Argonne National Laboratory	Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery	Computer Science	27,989
Shavalier_sims	Joshua Daniel Gezelter	University of Notre Dame	Heat Transport in Gold Interfaces Capped with thiolated-PEG, MTAB, and CTAB Using a Polarizable Force Field	Chemistry	24,965
shearlayers	Johan Larsson	University of Maryland	DNS of three-dimensional shear layers	Engineering	32,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
SolarWindowsADSP	Jacqueline Manina Cole	University of Cambridge	Data-Driven Molecular Engineering of Solar-powered Windows	Materials Science	100
spentFuel	Angela Di Fulvio	University of Illinois at Urbana-Champaign	Cask Mis-loads Evaluation Techniques	Nuclear Energy	47,824
SuperBERT	Ian Foster	Argonne National Laboratory	Training of language models on large quantities of scientific text	Computer Science	28,000
TDMD_thermostat	Charles Michael McCallum	University of the Pacific	Effect of the thermostat on the simulation of ESI processes	Chemistry	1,500
TIADE	Andre F. P. Ribeiro	Dassault Systemes BIOVIA	Aeroacoustic simulation of full wind turbine blade	Engineering	32,000
Tools	Scott Parker	Argonne National Laboratory	ALCF Performance Tools	Internal	5,000
TotalView	John Vincent DelSignore, Jr., Raymond M. Loy	Perforce Software, Inc.	TotalView Debugger ALCF Systems	Internal	1,000
TRB	Parisa Mirbod	University of Illinois at Chicago	Turbulent Rayleigh-Benard Convection in Suspensions of Bubbles	Engineering	706,965
TurbulentLiquidDrop	Arne Pearlstein	University of Illinois at Urbana-Champaign	Computation of Transitional and Turbulent Drop Flows for Liquid Carbon Dioxide Drops Rising in Seawater	Engineering	14,000
UIC-CS494-Sp2023	Michael E. Papka	University of Illinois at Chicago	CS494 Introduction to High Performance Computing	Computer Science	6,000
UINTAH_aesp	Martin Berzins, John Andrew Schmidt	The University of Utah	Design and evaluation of high-efficiency boilers for energy production using a hierarchical V/UQ approach	Chemistry	15,500
User_Services	Haritha Siddabathuni Som	Argonne National Laboratory	User Services	Internal	0
uso				UNKNOWN	10
VeloC	Bogdan Florin Nicolae	Argonne National Laboratory	VeloC: Very Low Overhead Checkpointing System	Computer Science	49,972
Vendor_Support	William Edward Allcock, Andrew J Cherry, Susan Marie Coghlan, Torrance Ivan Leggett, William R Scullin	Argonne National Laboratory	Vendor Support	Internal	68
vib_free_eng	Buu Q Pham	Ames Laboratory	Vibrational free energy from quantum-chemical calculations of large molecular systems	Chemistry	30,666

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
visualization	Joseph A Insley, Michael E. Papka	Argonne National Laboratory	Visualization and Analysis Research and Development for ALCF	Internal	25,000
Viz_Support	Joseph A Insley, William Edward Allcock	Argonne National Laboratory	Visualization Support	Computer Science	4,000
vsvb4cdx	Graham Donald Fletcher	Argonne National Laboratory	VSVB Study of Carbon Dot Excitons	Chemistry	32,000
WALSforAll	Christine Mary Simpson	Argonne National Laboratory	Workflows Across Labs	Computer Science	1,100
WaterHammer	Hong Zhang, Hong Zhang	Argonne National Laboratory	Water Hammer Simulation	Mathematics	5,020
WRF_LES	Haochen Tan	Argonne National Laboratory	Impact of sensible heat flux due to rainfall over Chicago area	Earth Science	64,000
XGC_aesp	Choongseok Chang	Princeton Plasma Physics Laboratory (PPPL)	High fidelity simulation of fusion reactor boundary plasmas	Fusion Energy	19,500
XMultiImage	Phay J Ho	Argonne National Laboratory	Multimodal Imaging with Intense X-ray Pulses	Chemistry	123,282
				Total DD	15,424,222

January 1, 2023–December 31, 2023
Director's Discretionary (DD) Projects on Theta Expansion (by Project Name)

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
AI-based-NDI-Spirit	Rajkumar Kettimuthu	Argonne National Laboratory	Framework and Tool for Artificial Intelligence & Machine Learning Enabled Automated Non-Destructive Inspection of Composite Aerostructures Manufacturing	Engineering	4,000
AI4NMR	Eric Michael Jonas	The University of Chicago (UChicago)	Structure Elucidation for Nuclear Magnetic Resonance via Structured Prediction	Chemistry	1,698
AIASMAAR	Rui Hu	Argonne National Laboratory	Artificial Intelligence Assisted Safety Modeling and Analysis of Advanced Reactors	Nuclear Energy	1,627
ALCF_for_DUNE	Aleena Rafique	Argonne National Laboratory	Enabling precise measurement of neutrino interactions using advanced computing in DUNE	Physics	200
APSDataAnalysis	Rafael Vescovi	Argonne National Laboratory	APS Beamline Data Processing and Analysis	Computer Science	969
ATPESC2023	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program for Extreme-Scale Computing 2023	Training	2,000
ATPESC_Instructors	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing Instructors	Training	500
Auriga	Christine Mary Simpson	Argonne National Laboratory	The Auriga Project	Physics	1,000
autopology_alcf	Rafael Gomez-Bombarelli	Massachusetts Institute of Technology (MIT)	End to end classical force field parametrization for polymer electrolytes using machine learning	Materials Science	1,276
biolearning	Chongle Pan	University of Oklahoma	Development of large-scale biomedical machine learning models	Biological Sciences	327
BIP167	Philip Kurian	Howard University	Computing superradiance and van der Waals many-body dispersion effects from MD simulations of biomolecular complexes	Physics	2,000
BirdAudio	Nicola Joy Ferrier	Argonne National Laboratory	Machine Learning for Classification of Birdsong	Computer Science	3,516
bloodflow_dd	Jifu Tan	Northern Illinois University (NIU)	Multiphysics modeling of clot mechanics	Engineering	500

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
BNN-Scale	Murali Krishna Emani	Argonne National Laboratory	Optimizing Bayesian Neural Networks for Scientific Machine Learning Applications	Computer Science	1,200
BPC	Christopher Michael Graziul	The University of Chicago (UChicago)	Optimization of audio processing pipeline for broadcast police communications	Computer Science	2,882
BRAIN	Getnet Dubale Betrie	National Renewable Energy Laboratory (NREL)	Scalable Brain Simulator for Extreme Computing	Biological Sciences	2,000
BS-SOLCTRA	Esteban Meneses	Costa Rica National High Technology Center	Plasma Physics Simulations for SCR-1 Stellarator	Physics	971
bubble-ai	Ben J. Blaiszik	Argonne National Laboratory	Discovery of Novel Fuel Cell Catalyst Materials via Development of High-Throughput AI-Guided Characterization Methods	Materials Science	1,875
candle_aesp	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	Virtual Drug Response Prediction	Biological Sciences	250
Carbon_composites	Hendrik Heinz	University of Colorado-Boulder	Designing Functional Nanostructures and Carbon-Based Composite Materials	Materials Science	409
Catalyst	Katherine M. Riley, Christopher James Knight, James Clifton Osborn, Timothy Joe Williams	Argonne National Laboratory	Catalyst	Internal	4,000
cfdml_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Data Analytics and Machine Learning for Exascale CFD	Engineering	25
CharmRTS	Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia	University of Illinois at Urbana-Champaign	Charm++ and its applications	Computer Science	4,000
Clouds	Ian Foster	Argonne National Laboratory	Unsupervised analysis of satellite cloud imagery	Earth Science	604
covid-ct	Ravi Kiran Madduri	Argonne National Laboratory	Medical Imaging Domain-Expertise Machine Learning for Interrogation of COVID	Computer Science	896
CSC249ADOA01	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	2.2.4.03 ADOA01 CANDLE: Exascale Deep Learning Enabled Precision Medicine for Cancer	Biological Sciences	3,095

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CVD_CityCOVID	Jonathan Ozik	Argonne National Laboratory	Agent-based model called CityCOVID capable of tracking detailed COVID-19 transmission	Biological Sciences	3,016
CyberAdvisor	Benjamin Alan Blakely	Argonne National Laboratory	INSURE: A Digital Cybersecurity Advisor for the Power Industry Built on Open Source Large Language Foundation Models	Computer Science	2,000
datascience	Venkatram Vishwanath	Argonne National Laboratory	ALCF Data Science and Workflows Allocation	Internal	12,500
dcp35	George Em Karniadakis	Brown University	Quantification of extreme weather events and their future changes using Physics-Informed DeepONet modeling and functional priors	Mathematics	4,000
DEAC_fermions	Adrian Giuseppe Del Maestro	The University of Tennessee at Knoxville	Analytic Continuation of Interacting Fermion Spectra	Physics	2,000
dist_relational_alg	Sidharth Kumar	The University of Alabama at Birmingham	Distributed relational algebra at scale	Computer Science	1,399
DL4VIS	Hanqi Guo	Argonne National Laboratory	Deep Learning for In Situ Analysis and Visualization	Computer Science	1,004
DLHMC	Sam Alfred Foreman	Argonne National Laboratory	Deep Learning HMC	Physics	927
DL_MODEX	Maruti Kumar Mudunuru	Pacific Northwest National Laboratory (PNNL)	Towards a robust and scalable deep learning workflow for fast, accurate, and reliable calibration of watershed models	Earth Science	1,000
Drug_FEP_Data	Wei Jiang	Argonne National Laboratory	Machine eLearning of drug binding and toxicity based on high throughput free energy computations	Biological Sciences	1,183
ECP_SDK	Sameer Suresh Shende	University of Oregon	Deploying the ECP SDK software stack at ALCF	Computer Science	1,000
EE-ECP	Xingfu Wu, Valerie Taylor	Argonne National Laboratory	Energy efficient tradeoff among execution time and power of ECP applications	Computer Science	871
electrolyte-chibueze	Chibueze Vincent Amanchukwu	The University of Chicago (UChicago)	Molecular dynamics-driven discovery of novel liquid electrolytes	Energy Technologies	2,000
EmergingTech2	Benjamin Alan Blakely	Argonne National Laboratory	Identification of Emerging Technologies from Publication Databases	Computer Science	2,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
Emerging_Tech_ML	Benjamin Alan Blakely	Argonne National Laboratory	Graph-based Bibliometric Analysis for Emerging Technology Discovery	Computer Science	1,487
EngineDNS	Christos Frouzakis	Eidgenössische Technische Hochschule Zürich (ETH Zurich)	Towards reactive DNS in complex internal combustion engine geometries	Engineering	4,000
ENL	Marc Berman	The University of Chicago (UChicago)	Environmental Neuroscience	Biological Sciences	3,000
epic-ecal	Sylvester Johannes Joosten	Argonne National Laboratory	Imaging Calorimeter Design & Optimization for the ePIC Detector	Physics	2,000
EvalDL	Natalia Sergeyevna Vasileva	Cerebras Systems Inc.	Evaluation of Deep Learning Models	Computer Science	11
EVITA	George K Thiruvathukal, Nicholas John Eliopoulos, Nicholas Michael Synovic	Loyola University Chicago	Energy-efficient VIsual Transformer Architecture: Transformer Models for Deployment on Embedded Systems	Computer Science	1,011
ExtremeConvection	Mahendra Kumar Verma	Indian Institute of Technology Kanpur	Simulation of Turbulent Convection at Extreme Parameters	Physics	2,000
gccy3	Marc Francis Paterno	Fermi National Accelerator Laboratory (Fermilab)	Cosmological Parameter Inference from Galaxy Clusters	Physics	2,000
GPU-DG	Pinaki Pal	Argonne National Laboratory	GPU-enabled Discontinuous Galerkin simulations of complex fluid flows	Engineering	1,004
GRACE	Sayan Ghosh	Pacific Northwest National Laboratory (PNNL)	Graph Analytics Codesign on GPUs	Computer Science	5,003
GrainBoundaries	Wissam A Saidi	University of Pittsburgh	Structure and Properties of Grain Boundaries in Materials for Energy Applications	Materials Science	474
graphs_	Dossay Oryspayev	Brookhaven National Laboratory (BNL)	Exploration of parallelization of graph algorithms	Computer Science	1,456
HighReyTurb_PostProc	Robert D. Moser, Myoungkyu Lee	The University of Alabama	Data analysis of turbulent Channel Flow at High Reynolds number	Engineering	1,000
hpc-spectacle	Kevin Antoney Brown	Argonne National Laboratory	Evaluate and Optimize Data Movement Strategies in AI and Climate Science Workloads	Computer Science	2,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
hpcbdsm	Tanwi Mallick	Argonne National Laboratory	High-Performance Computing and Big Data Solutions for Mobility Design and Planning	Computer Science	2,000
hydrosm	Jeremy A Feinstein	Argonne National Laboratory	Improving the predictability of hydrological systems with AI	Earth Science	2,000
IBM-GSS	Venkatram Vishwanath	Argonne National Laboratory	IBM GeoSpatial Software System	Earth Science	2,622
IMEXLBM	Saumil Sudhir Patel	Argonne National Laboratory	ECP ProxyApp Development for the Lattice Boltzmann Method	Computer Science	955
Intel	Kalyan Kumaran, Scott Parker, Timothy Joe Williams, Venkatram Vishwanath	Argonne National Laboratory	Intel employees in support of Theta	Internal	4,000
JCESR	Larry Curtiss, Anubhav Jain	Argonne National Laboratory	Development of High Throughput Methods	Materials Science	2,370
LAR-EM	Zheng Zhang	The University of Chicago (UChicago)	Electron Tomography from Limited-Angular-Range Data	Materials Science	1,670
les120	George Em Karniadakis	Brown University	Learning the sub-grid model in Large Eddy Simulations using domain-decomposition based parallel physics-informed neural networks (PINNs)	Mathematics	1,765
lhtes	Kedar Prashant Shete	University of Massachusetts-Amherst	DNS of turbulent phase changing flows	Energy Technologies	2,000
LIGHTCONTROL	Sandra Gail Biedron, Trudy Beth Bolin	University of New Mexico	Light sources and their control using AI techniques	Physics	2,472
LoopSynch	Shina Caroline Lynn Kamerlin	Georgia Institute of Technology (Georgia Tech)	Link between loop dynamics and turnover number across 30 extant Triosephosphate Isomerases	Chemistry	1,910
LQCD-ML	Wai Nim Alfred Tang	Phd Tutor Hub	LQCD parametric regression by neural networks	Physics	288
LQCDdev	James Clifton Osborn	Argonne National Laboratory	Lattice QCD development	Physics	500
lqcdml_aesp	William Detmold	Massachusetts Institute of Technology (MIT)	Machine Learning for Lattice Quantum Chromodynamics	Physics	250
LSWDD_SC	David Weihua He	University of Illinois at Chicago	Large Scale Wafer Defect Diagnosis using Supercomputer	Computer Science	180

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
Maintenance	William Edward Allcock, John Francis O'Connell, John Patrick Reddy, Ryan Milner, Torrance Ivan Leggett	Argonne National Laboratory	LCF Operations System Maintenance	Internal	6,008
matml_aesp	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials	Materials Science	1,670
metastable	Subramanian Sankaranarayanan	Argonne National Laboratory	Metastable phase diagram of material	Materials Science	1,543
MI2Dmaterials	Trevor David Rhone	Rensselaer Polytechnic Institute (RPI)	Materials informatics study of two-dimensional magnetic materials and their heterostructures	Materials Science	2,678
MICCoM	Justin Michael Wozniak	Argonne National Laboratory	Midwest Integrated Center for Computational Materials	Materials Science	33,000
ML-Coupling	Emil Mihai Constantinescu	Argonne National Laboratory	Data-driven Coupling Methods for Atmospheric-Ocean Interactions	Earth Science	1,000
mlrom_uas	Rohit Kameshwara Sampath Sai Vuppala	Oklahoma State University	ML based Reduced Order Models	Engineering	1,000
mpi_partitioned	Ahmad Afsahi	Queen's University	GPU-Initiated MPI Partitioned Point-to-Point Communication Over NVSHMEM	Computer Science	732
MultiActiveAI	Dario Dematties	Northwestern Argonne Institute of Science and Engineering (NAISE)	Multimodal Intelligence for Federated Edge Computing Simulations	Computer Science	1,970
multiphysics_aesp	Amanda Randles	Duke University	Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations	Engineering	552
MVAPICH2	Dhabaleswar Kumar Panda	The Ohio State University	Optimizing and Tuning MVAPICH2-GDR Library and study Its Impact on HPC and AI Applications	Computer Science	2,947
N2M3O7	Iwnetim Iwnetu Abate	Massachusetts Institute of Technology (MIT)	Tunable catalysis in transition metal oxide: NaxMn3O7	Materials Science	2,000
NAISE-NIH-NIAMS	Marta Garcia Martinez	Argonne National Laboratory	Automatic MRI segmentation for upper limb muscles for clinical applications	Biological Sciences	2,000
novacosmics	Alexander I Himmel	Fermi National Accelerator Laboratory (Fermilab)	NOvA Cosmic Rejection	Physics	6,648

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
NUCat_Micro-CT	Marta Garcia Martinez	Argonne National Laboratory	Large volume feline spinal cord microtomography	Biological Sciences	2,000
NuQMC	Alessandro Lovato	Argonne National Laboratory	Nuclear quantum Monte Carlo	Physics	4,250
Operations	William Edward Allcock	Argonne National Laboratory	Systems administration tasks	Internal	1,000
OptADDN	Sandeep Madireddy	Argonne National Laboratory	Optimal Architecture discovery for deep probabilistic models and neuromorphic systems	Computer Science	2,000
PARTURB3D	Ramesh Balakrishnan	Argonne National Laboratory	Simulating turbulent particulate flows inside enclosures	Engineering	1,765
PDE_ML	Ramin Baghgar Bostanabad	University of California-Irvine	Self-supervised Coupling of Deep Operator Surrogates for Scalable and Transferable Learning	Engineering	1,601
Performance	Scott Parker, Raymond M. Loy	Argonne National Laboratory	Performance	Internal	1,000
PUR-IRL	Nicholas Lee-Ping Chia	Mayo Clinic-Minnesota	Inferring the Reward Function of Cancer	Biological Sciences	1,016
QuantumDS	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Quantum mechanics and Data Science	Chemistry	2,405
radix-io	Philip Hutchinson Carns	Argonne National Laboratory	System software to enable data-intensive science	Computer Science	46
RAN	William Edward Allcock, Brian R. Toonen	Argonne National Laboratory	RAM Area Network	Computer Science	5,000
RAPINS	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Reproducible and Accelerated Physics-inspired Neural Networks	Physics	5,828
ReForMerS	Sandeep Madireddy	Argonne National Laboratory	Robust and Reliable Foundation Models for Science	Computer Science	2,000
remote_offloading	Jose Manuel Monsalve Diaz	Argonne National Laboratory	Exploring collective operations with Remote Offloading	Computer Science	1,354
RL-fold	Arvind Ramanathan	Argonne National Laboratory	Targeting intrinsically disordered proteins using artificial intelligence driven molecular simulations	Biological Sciences	31,210
RNAtoImage	Ravi Kiran Madduri	Argonne National Laboratory	RNA-to-Image Synthesis project	Biological Sciences	2,000
sbi-fair	Pete Beckman, Kamil Antoni Iskra	Argonne National Laboratory	FAIR Surrogate Benchmarks Supporting AI and Simulation Research	Computer Science	4,345

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
SCREAM_Calib	Jiali Wang	Argonne National Laboratory	Towards Neighborhood Scale Climate Simulations using AI and Accelerated GPUs	Earth Science	2,970
SDR	Sheng Di	Argonne National Laboratory	Scalable Dynamic Scientific Data Reduction Framework (SDR)	Computer Science	3,100
SEEr-planning	Zhiling Lan	Illinois Institute of Technology (IIT)	Performance and Power Tradeoff Analysis of AI-Enabled Science on CPU-GPU System	Computer Science	2,621
SENSEI	Silvio Humberto Rafael Rizzi, Joseph A Insley, Nicola Joy Ferrier, Venkatram Vishwanath	Argonne National Laboratory	Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery	Computer Science	932
SolarWindowsADSP	Jacqueline Manina Cole	University of Cambridge	Data-Driven Molecular Engineering of Solar-powered Windows	Materials Science	165
SOLLVE	Sunita Chandrasekaran	Brookhaven National Laboratory (BNL)	Scaling OpenMP With LLVm for Exascale Performance and Portability	Computer Science	1,693
spentFuel	Angela Di Fulvio	University of Illinois at Urbana-Champaign	Cask Mis-loads Evaluation Techniques	Nuclear Energy	989
SSLCos	Dario Dematties	Northwestern University	Self-supervised learning for cosmology	Physics	2,000
stabCNS	Ali Karakus	Middle East Technical University	GPU Accelerated High-order Compressible Flow Simulations	Engineering	2,000
SuperBERT	Ian Foster	Argonne National Laboratory	Training of language models on large quantities of scientific text	Computer Science	6,849
SuperRes	Jiali Wang	Argonne National Laboratory	Observation-informed AI downscaling for winds	Earth Science	2,000
TDMD_thermostat	Charles Michael McCallum	University of the Pacific	Effect of the thermostat on the simulation of ESI processes	Chemistry	750
TIADE	Andre F. P. Ribeiro	Dassault Systemes BIOVIA	Aeroacoustic simulation of full wind turbine blade	Engineering	2,000
TMEM_DEL	Diomedes Elias Logothetis	Northeastern University	Molecular Dynamics simulation on TMEM16A chloride channel	Biological Sciences	7,258
Tools	Scott Parker	Argonne National Laboratory	ALCF Performance Tools	Internal	500
User_Services	Haritha Siddabathuni Som	Argonne National Laboratory	User Services	Internal	504

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
VeloC	Bogdan Florin Nicolae	Argonne National Laboratory	VeloC: Very Low Overhead Checkpointing System	Computer Science	4,000
visualization	Joseph A Insley, Michael E. Papka	Argonne National Laboratory	Visualization and Analysis Research and Development for ALCF	Internal	1,000
WaggleAISummer2023	Sean A Shahkarami	Northwestern Argonne Institute of Science and Engineering (NAISE)	Waggle AI Summer 2023	Computer Science	2,000
wall_turb_dd	Ramesh Balakrishnan	Argonne National Laboratory	Wall Resolved Simulations of Canonical Wall Bounded Flows	Engineering	905
wereszczynski	Jeffery Michael Wereszczynski	Illinois Institute of Technology (IIT)	MD Simulations of Chromatin Modification and Gene Regulation Mechanism	Biological Sciences	878
WRFGPU	Veerabhadra Rao Kotamarthi	Argonne National Laboratory	WRF GPU testing	Earth Science	927
XMultiImage	Phay J Ho	Argonne National Laboratory	Multimodal Imaging with Intense X-ray Pulses	Chemistry	1,470
				Total DD	308,227

This page intentionally left blank.

Appendix C – ALCF CY2023 Science Highlights

The following Science Highlights were submitted to ASCR for the 2023 OAR performance period.

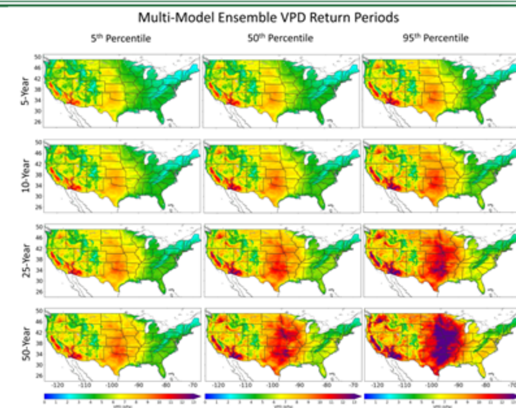
A New Way To Predict Droughts

The Science

Extreme drought conditions in the United States are expected to increase with global warming but predicting potential changes can be challenging in regions with large variability in moisture across small regions, such as interconnected urban, suburban, and agricultural land regions. Unlike hurricanes, which have a rigorous classification scale, various methodologies with diverse fidelity are currently used to classify droughts. Based on analysis of a large community dataset and projections from three Global Climate Models, a team has developed a new measure to help identify short-term droughts and predict future extreme droughts, including "flash drought" events with a quick onset period that could be as short as few weeks.

The Impact

Using the new uncomplicated methodology, based on highly detailed regional climate models, the team found that droughts are likely to be exacerbated by global warming, particularly in regions like the Midwest, Northwestern U.S., and California's Central Valley. The projected worsening of droughts because of global warming are likely to have significant consequences affecting water resources, wildfire activities, and crop loss. This technique with the resolution of a single day, and thus more precise statistics, provides earlier drought detection and can help researchers to understand "flash drought" events that quickly occur and last only weeks.



The 5, 10, 25, and 50-year return periods in the United States for the 5th, median (50th), and 95th percentile of the sampled model ensemble for VPD return periods. (Image courtesy of Brandi L. Gamelin of Argonne National Laboratory)

Contact PI: Brandi L. Gamelin (ANL)
ASCR Allocation PI: Veerabhadra R. Kotamarthi (ANL)
ASCR Program/Facility: DD/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: February 6, 2023
Publication(s) for this work: B.L. Gamelin et al., *Scientific Reports* ***12***, 8615 (2022)

Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

AT&T and ANL

Investigation of Cyclic Variations in Internal Combustion Engines

Using proper orthogonal decomposition to design better engines

The Science

A key scientific challenge for internal combustion engines (ICE) is the understanding, modeling, and control of cycle-to-cycle variability (CCV) in engine performance, which can contribute to unevenness in the running of the engine, excessive engine noise and emissions, and potentially damaging engine knock. Mitigating CCV requires an in-depth understanding of the stochastic in-cylinder processes. This starts with characterizing the behavior of the highly turbulent fluid flow inside the combustion chamber. It has been shown that the large-scale coherent structures in the ICEs have significant influence on the air-fuel mixing and transport during the intake and compression strokes. In the current study, high-fidelity multi-cycle simulations of the flow inside an internal combustion engine were performed and proper orthogonal decomposition (POD) was used to characterize the dominant and coherent flow structures.

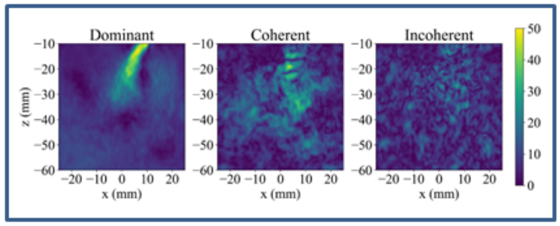


Image (courtesy of Sicong Wu): Comparison of the velocity magnitude (m/s) of the dominant, coherent and incoherent structures in the tumble plane obtained using the triple proper orthogonal decomposition (POD) applied to an experimental engine.

The Impact

This was the first ever use of a POD technique on a wall-resolved multi-cycle LES of a flow inside an internal combustion engine. The simulations provided unprecedent insights into the distribution of energy across the different scales within the engine and their effect on cyclic variability. This information can be useful for engine designers to modify the engine components to control the coherent scales of the in-cylinder flow, thus reducing the cyclic variability and enabling stable combustion.

Contact PI: Muhsin Ameen (ANL)
ASCR Allocation PI: Muhsin Ameen
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: March, 2023
Publication(s) for this work: Sicong Wu, Saumil Patel, and Muhsin Ameen, *Flow, Turbulence and Combustion* (2022).

Transportation and Power Systems Division (ANL), Computational Science Division (ANL)

U.S. DEPARTMENT OF ENERGY | Office of Science

Funded by DOE-Vehicles Technologies Office

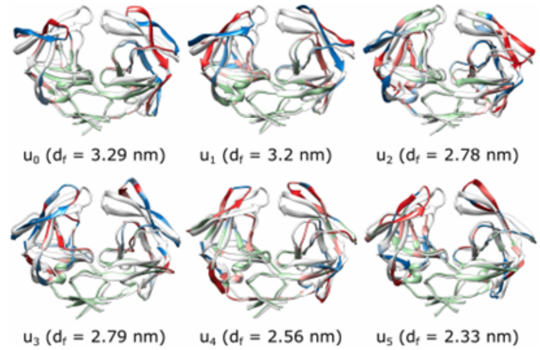
Identifying Reaction Coordinates for Complex Protein Dynamics

The Science

The primary goal of protein science is to understand how proteins function, which requires understanding the functional dynamics responsible for transitions between different conformations. If the exact reaction coordinates (RCs) were known, then one could easily determine the transition likelihood for any protein configuration. Despite intensive efforts, identifying the exact RCs in complex macromolecules remains a formidable challenge. Using the recently developed generalized work functional (GWF), a biomedical engineering group at the University of Illinois Chicago reported the discovery of the exact RCs for a large-scale functional process of an important protein system. The RCs were leveraged to explore the protein reaction mechanism and provide an efficient means for accelerating traversal of conformational space in computer simulations.

The Impact

Most protein functions, such as drug binding and enzymatic catalysis, involve significant conformational dynamics, for which RCs are of central importance in exploring reaction mechanisms and guiding design of computer simulations. Researchers demonstrated the robust identification of RCs for the HIV-1 protease flap opening process. Success in determining the RCs enabled acceleration of this important process 10^3-10^4 fold over that observed in traditional simulations. Identifying for the first time the exact RCs of a large functional protein has the potential to dramatically accelerate understanding of protein conformational dynamics benefiting biomedical research and drug discovery.



Representative open structures from accelerated trajectories along the identified six reaction coordinates. The reference semi-open structure is colored white and atoms are colored green, red, and blue according to its weight in the RC. (Image courtesy of Ao Ma (UIC); reproduced from Figure 6 of *Proc. Natl. Acad. Sci. U.S.A* **119**, e2214906119 (2022).)

Contact PI: Ao Ma (UIC, Dept. Biomed. Eng.)
ASCR Allocation PI: Ao Ma (UIC, Dept. Biomed. Eng.)
ASCR Program/Facility: DD/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: April 10, 2023
Publication(s) for this work: S. Wu, et al., *Proc. Natl. Acad. Sci. U.S.A.* ***119***, e2214906119 (2022)

University of Illinois Chicago

U.S. DEPARTMENT OF ENERGY | Office of Science

NIH & NSF

Multiphysics modeling of blood flow with cell suspensions in patient-specific capillary network

The Science

Modeling capillary flow accurately is challenging due to the complex structure with various vessel branches and loops, and moving cell suspensions whose size is comparable to vessel diameters. Simulations of blood flow with deformable red blood cells were performed for the first time in a patient-specific retina vascular network examining the impact of blockages on flow rate and cell transport dynamics. Large three-dimensional (3D) vascular networks, such as this, are typically represented by simplified one-dimensional (1D) models at a much lower computational cost; however, these reduced order models may not accurately describe the flow dynamics. The accuracy of 1D models with four effective viscosity models were directly compared to 3D simulations with cell suspensions.

The Impact

Cardiovascular disease, including heart attack and stroke, is the leading cause of death in the United States. Particulate cells and blockages in vascular networks can result in complex flow patterns. Cell transport along different branches of split vessels showed complicated behavior dependent on many factors, such as vessel diameter size, branching angles, local cell concentration and flow rate. Inclusion of larger white blood cells was found to significantly increase the transit time of red blood cells through vessels. The simulation of flow under partial vessel blockage (e.g., stenosis) with cells showed that cells could oscillate and be trapped in an adjacent vessel due to the fluctuating flow. The best performing 1D reduced order model still resulted in large errors in both the number of red blood cells and flow rate for short vessels, and such models may be more suitable for networks with larger vessels.

A snapshot of complex flow with cell suspensions in the retina network showing the transport of cells in different regions of the network. The blue/green coloring indicates the flow velocity within the network and red blood cells are colored red. (Image courtesy of Kacper Ostalowski (NIU), Jifu Tan (NIU) and Joseph A. Insley (ALCF).)

Contact PI: Jifu Tan (NIU, Dept. Mech. Eng.)
ASCR Allocation PI: Jifu Tan (NIU, Dept. Mech. Eng.)
ASCR Program/Facility: DD/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: May 15, 2023
Publication(s) for this work: K. Ostalowski and J. Tan., Phys. Fluids. ***34***, 041912 (2022)

Northern Illinois University

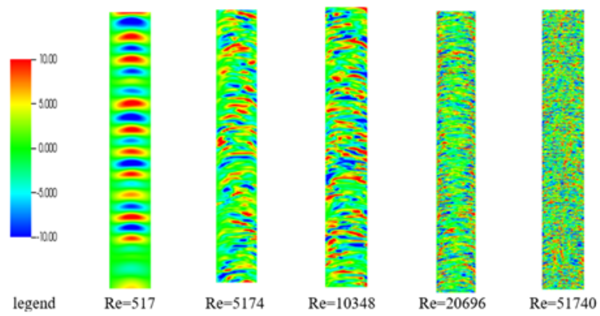
U.S. DEPARTMENT OF ENERGY | Office of Science

NIU

Assessment of Turbulent Prandtl Number for Heavy Liquid Metal Flow in a Bare Rod Bundle by Leveraging LES simulation

The Science

Nuclear reactor vendor Westinghouse Electric Company, together with an international team, is developing its next-generation high-capacity nuclear power plant based on lead-cooled fast reactor technology. Accurate modeling and simulation of heat transfer and mixing in the heavy liquid metal (HLM) coolant is needed to help prepare the technology for licensing. HLM (e.g. lead) has different properties from other fluids, and the conventional choice of turbulent Prandtl number (Pr_t) in existing engineering turbulence models is not valid for HLM flow simulations. This campaign provides valuable high-fidelity reference data to the nuclear reactor community, with which less computationally expensive models can be compared and calibrated.



LES predictions of the turbulent structure (flow oscillations) in a lead-cooled rod bundle with varying Reynolds numbers. As Re increases, the flow oscillations become smaller scale and more chaotic. These simulations can be used as benchmarks for lower fidelity calculations which require tuning of the turbulent Prandtl number. (Courtesy Yiqi Yu)

The Impact

Large Eddy Simulations (LES) using Nek5000 on Theta were applied to nuclear fuel rod bundles with HLM flow with the goal of providing detailed insight into flow physics and heat transfer mechanisms. LES simulations do not require Pr_t to model turbulence-driven heat transfer. They thus can be used as benchmarks for selecting Prt in a less computationally expensive Reynolds Averaged Navier Stokes (RANS) model, which requires this parameter. This project provides detailed insight into the heat transfer mechanism in a nuclear fuel rod bundle with HLM flow. The assessment and selection of appropriate Pr_t can improve simulation accuracy for advanced nuclear reactors, such as lead-cooled fast reactors (LFR), which are potential candidates for next-generation nuclear reactors in the U.S. and abroad. Data are rarely reported or available for heat transfer of HLM in rod bundles with LES.

Contact PI: Yiqi Yu (ANL)
ASCR Allocation PI: Emily Shemon (ANL)
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: June 2023
Publication(s) for this work: Yiqi Yu, Emily Shemon, Elia Merzari, Nuclear Engineering and Design, Volume 404, 2023.

U.S. DEPARTMENT OF ENERGY | Office of Science

Funding Agency: DOE-NEAMS program

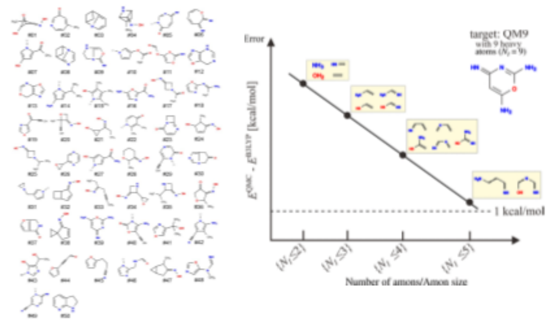
Towards DMC Accuracy Across Chemical Space with Scalable Δ-QML

The Science

With the development of AI/ML methods and the great amount of data generated using supercomputers, highly accurate data is becoming increasingly important. This is particularly true in the field of quantum chemistry, where accurate data is essential for predicting the properties of molecules and materials. Quantum Monte Carlo (QMC) has been used for years to predict the energetics and properties of molecules and solids by numerically solving the electronic many-body Schrödinger equation. However, the computational cost of QMC has been considered too expensive for very large molecules. The code development efforts of the last decades combined with more powerful high-performance computing (HPC) made it possible to study larger systems but limiting high throughput calculations to smaller systems.

The Impact

The combination of machine learning and quantum Monte Carlo (QMC) with delta learning to predict the energetics of larger molecules at chemical accuracy is a significant advancement in the field of quantum chemistry. This new method will allow researchers to study larger systems and more accurately predict the properties of molecules and materials. Moreover, the amons-based Δ-QML framework, in combination with costly DMC-referenced datasets of small molecules, holds great promise for robust yet efficient exploration campaigns of molecules and materials throughout the Compound Chemical Space (CCS). Using Δ–Quantum Machine Learning, scientists could predict the energetics of larger molecules at a fraction of the cost while maintaining chemical accuracy. This new method could significantly impact fields such as drug discovery, materials science, and energy research.



Figures from left to right show the structures of 50 test molecules made up of nine heavy atoms and the increase of accuracy as the size of the molecular fragment (Amon) is increased to reach chemical accuracy. (Image courtesy of A. Benali (ANL), J.T. Krogel (ORNL), B. Huang (U. Vienna), and o.A. von Lilienfeld (U of T)).

Contact PI: Anouar Benali (ANL)
ASCR Allocation PI: Paul R. C. Kent (ORNL)
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: July 10, 2023
Publication(s) for this work: B. Huang, O. A. von Lilienfeld, J. T. Krogel and A. Benali, *J. Chem. Theory Comput.* **19**, 1711 (2023).

ANL, ORNL, University of Toronto, University of Vienna

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE-BES, NCCR MARVEL, Swiss National Science Foundation

High-Fidelity Simulations of Gas Turbine Combustor

Using Nek5000 to enhance gas turbine efficiency

The Science

With a push towards decarbonizing the aviation sector, sustainable aviation fuels (SAFs) have gained prominence as a potential replacement for fossil fuels. The US government's SAF Grand Challenge was recently formulated with a goal to achieve a minimum of 50% reduction in life cycle greenhouse gas emissions compared to conventional fuel and to supply sufficient SAF to meet 100% of aviation fuel demand by 2050. To assess the viability of various SAFs, researchers must be able to understand and predict the complex flow, spray, and combustion processes taking place in the gas turbine combustors, and their influence on events such as lean blow out, high altitude relight and cold start, that affect the performance of gas turbines. With recent advances in numerical methods and the availability of high-performance computing facilities, computer simulations can provide unprecedented details of the underlying multi-physics processes, but they rely on the complex task of creating a detailed computational model of the gas turbine that is accurate and runs efficiently on modern computers.

The Impact

This project is developing the capabilities to perform fully-resolved simulations of modern gas turbine combustors to enable improved understanding of the multi-physics processes in the context of advancing the development of sustainable aviation fuels. In this study, wall-resolved large-eddy simulations of the turbulent flow in the Army Research Laboratory's midsize combustor (ARC-M1) were performed using Nek5000, a massively-parallel computational fluid dynamics code. The simulations were validated against experimental measurements and showed good agreement. These simulations leverage DOE leadership computing systems to help understand the combustion and heat transfer challenges introduced by using low-carbon fuels. This project will help establish a high-fidelity, scalable, numerical framework that can be used for evaluating the effect of fuel properties on flow and flame dynamics in a practical gas turbine combustor.

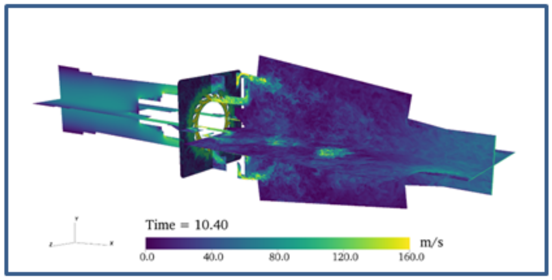


Image (Courtesy of Sicong Wu): Snapshot of the instantaneous turbulent flow field in the ARC-M1 combustor. Seen in this image are multiple recirculation regions and complex interactions between the large- and small-scale turbulent flow structures.

Contact PI: Muhsin Ameen (ANL)
ASCR Allocation PI: Sicong Wu (ANL)
ASCR Program/Facility: DD/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: August, 2023
Publication(s) for this work: Sicong Wu, Debolina Dasgupta, Saumil Patel, and Muhsin Ameen, AIAA SciTech 2023 Forum.

Transportation and Power Systems Division (ANL), Computational Science Division (ANL)

U.S. DEPARTMENT OF ENERGY | Office of Science

Funded by DOE-Vehicles Technologies Office

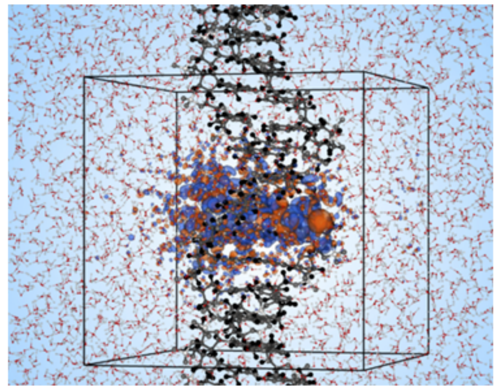
Electronic Excitation Response of DNA to High-Energy Proton Radiation in Water

The Science

Studying the mechanism of electronic excitation response of DNA with high energy proton radiation can help to elucidate damage during proton radiotherapy, or under extreme conditions such as those experienced by astronauts in space. Proton beam therapy has become a promising cancer therapy enabling more precise targeting of tumor cells reducing the damage to surrounding healthy tissue. The ultrafast nature of the excitations makes experimental investigation difficult. Quantum mechanical methods help to understand the mechanism of DNA damage, wherein nonequilibrium simulations can shed light on the details of the intricate energy transfer process of high-energy protons damaging DNA.

The Impact

Quantum mechanical simulations indicate that high-energy protons transfer significantly more energy to sugar-phosphate side chains than the nucleobases of DNA, generating highly energetic electron holes on the side chains as a key source of damage. The electronic excitation response of DNA was found to be highly localized near the path of the irradiating proton. Results from these simulations help fill the knowledge void in understanding detailed mechanisms for extensive DNA strand break lesions observed and build increasingly more sophisticated multi-scale models in medical physics to assist with customizing patient treatments and improve efficacy by reducing side effects.



Simulation cell for one full turn of a DNA double-helix solvated in water. Blue (orange) isosurfaces represent decreases (increases) in electron density in response to an irradiating proton having passed through. (Image courtesy of Yosuke Kanai (UNC))

Contact PI: Yosuke Kanao (UNC)
ASCR Allocation PI: Youke Kanai (UNC)
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: Sep. 11, 2023
Publication(s) for this work: C. Shepard, D.C. Yost, and Y. Kanai, *Phys. Rev. Lett.*, **130**, 118401 (2023).

University of North Carolina at Chapel Hill

U.S. DEPARTMENT OF ENERGY | Office of Science

National Science Foundation

First principles simulation of hypersonic flight

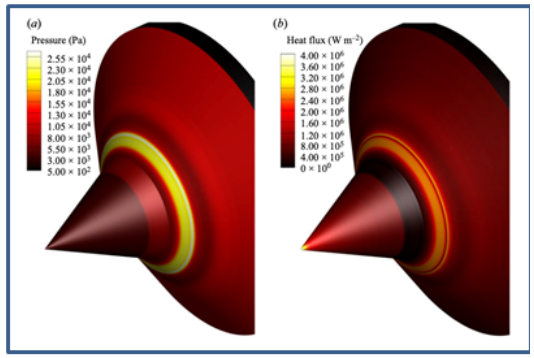
Improving our understanding of aerothermodynamics using quantum mechanical interactions

The Science

Hypersonic flight, the ability to fly at more than five times the speed of sound, has the potential to revolutionize technologies for national security, aviation, and space exploration. However, a fundamental understanding of the aerothermodynamics of hypersonic flight is needed to enable technological advances in this field. Strong shockwaves can cause the excitation of internal energy modes and chemical reactivity in the shock-heated gas. The rate of these phenomena is comparable to the fast, local flow time, causing the flow to be in thermal and chemical nonequilibrium. This nonequilibrium can even persist at the molecular level, with the shock heated gas having a non-Boltzmann character. The non-Boltzmann nature of the gas in tum affects the chemical reactivity of the flow in a non-linear manner, that even the most sophisticated computational fluid dynamics models cannot resolve. Therefore, a more fundamental approach is required to properly characterize the state of the gas in the shock layer. Accurate estimation of the thermal state and the chemical composition of the of the gas enveloping the vehicle is crucial for an optimal design for the thermal protection system. This work uses the Direct Molecular Simulation (DMS) method to provide benchmarks. DMS method uses quantum mechanically guided molecular interactions as the only input to simulate a flow field, providing a fundamental solution to this highly coupled problem.

The Impact

The team used the open-source SPARTA Direct Simulation Monte Carlo (DSMC) code on Theta to carry out direct molecular simulations (DMS) of hypersonic flow experiments. Their goal is to conduct simulations that rely solely on molecular level interactions modeled using quantum mechanics, providing a fundamental comparison with experiments, and well-characterized solutions that can be used as benchmarks for reduced-order models. The team's work demonstrates that their DSMC code has the ability to find solutions for these kinds of flows and, moreover, that it can seamlessly describe internal energy non-equilibrium for all modes. This research is helping to advance our understanding of the complex aerothermodynamics of hypersonic flight, providing insights that could help inform the design of safer and more efficient technologies for space travel and defense.



Three-dimensional view of the surface pressure and heat flux on the double cone geometry. (Image courtesy of P. Valentini)

Contact PI: Maninder Grover (U. of Dayton Research Institute)
ASCR Allocation PI: Maninder Grover (U. of Dayton Research Institute)
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: October, 2023
Publication(s) for this work: Valentini, P., et al., Journal of Fluid Mechanics (July 2023).

Air Force Research Laboratory & University of Dayton Research Institute

U.S. DEPARTMENT OF ENERGY | Office of Science

Funded by: Air Force Office of Scientific Research

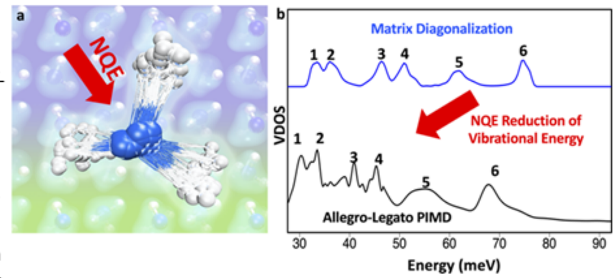
Allegro-Legato: Scalable, Fast, and Robust Neural-Network Quantum Molecular Dynamics via Sharpness-Aware Minimization

The Science

Neural-network quantum molecular dynamics (NNQMD) simulations based on machine learning are revolutionizing atomistic simulations of materials by providing quantum-mechanical accuracy but orders-of-magnitude faster. The state-of-art accuracy has now been combined with a record speed based on spatially localized descriptors in the latest NNQMD model named Allegro. The resulting Allegro-Legato (meaning fast and "smooth") model was more stable while maintaining the same inference speed and accuracy. In simulations of ammonia, the expected softening of high-energy vibrational modes was observed at finite temperature with inclusion of nuclear quantum effects and found to be consistent with high-end neutron experiments performed by the team.

The Impact

Allegro-Legato allows much larger spatio-temporal scale NNQMD simulations than are otherwise possible. Unlike typical MD methods for "effective" long-time sampling (e.g., for protein folding), which can skew dynamical information, Allegro-Legato enables "true" long-time dynamics that can be directly compared with high-resolution spectroscopic experiments. The prohibitive computational demand of accounting for subtle nuclear quantum effects (NQEs) simulations is no longer a problem, which is essential for accurately calculating thermodynamic properties. The synergy between the most advanced neutron experiment and leadership-scale NNQMD simulations lays a foundation for green ammonia-based fuel technology to achieve a sustainable society.



Computed vibrational spectra of ammonia: (a) representative image of ammonia molecule in path-integral molecular dynamics (PIMD) simulation and (b) vibrational spectrum computed at (top) zero temperature without nuclear quantum effects (NQE) and (bottom) at finite temperature with the Allegro-Legato PIMD simulation. (Image courtesy of Thomas Linker (USC))

Contact PI: Aiichiro Nakano (USC)
ASCR Allocation PI: Aiichiro Nakano (USC)
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: Nov. 6, 2023
Publication(s) for this work: H. Ibayashi, et al., In: A. Bhatele, J. Hammond, M. Baboulin, C. Kruse (eds) High Performance Computing. ISC High Performance 2023 Lecture Notes in Computer Science, vol 13948. Springer, Cham.

USC, Sony, ANL

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE BES & NSF

This page intentionally left blank.

Appendix D – Acronyms and Abbreviations

ACL	access-control list
ACT-SO	Afro-Academic, Cultural, Technological, and Scientific Olympics (NAACP)
AI	artificial intelligence
ALCC	ASCR Leadership Computing Challenge
ALCF	Argonne Leadership Computing Facility
API	application programming interface
APS	Advanced Photon Source (Argonne)
Argonne	Argonne National Laboratory
ASCR	Advanced Scientific Computing Research
ATPESC	Argonne Training Program on Extreme-Scale Computing
ATO	Authority To Operate
BIS	Business and Information Services Division (Argonne)
CCIO	custom collective I/O
CELS	Computing, Environment, and Life Sciences (Argonne)
CFD	computational fluid dynamics
CI	continuous integration
COE	Center of Excellence (Intel)
CPU	central processing unit
CSPO	Cyber Security Program Office (Argonne)
CY	calendar year
DAOS	Distributed Asynchronous Object Storage
DD	Director's Discretionary
DDPM	Denoising Diffusion Probabilistic Model
DFT	density-functional theory
DL	deep learning
DOE	U.S. Department of Energy
DSAV	Division Site Assist Visit
ECP	Exascale Computing Project
ESH	Environmental, Safety and Health (Argonne)
ESP	Early Science Program
FY	fiscal year
GAN	generative adversarial network
GenSLM	genome-scale language model
GNN	graph neural network
GPU	graphics processing unit
HA	high availability, Hamiltonian Annealing
HACC	Hardware/Hybrid Accelerated Cosmology Code

HDF5	Hierarchical Data Format version 5
HIP	Heterogeneous Interface for Portability
HPC	high-performance computing
HPE	Hewlett Packard Enterprise
HPCM	HPE Performance Cluster Manager
HPSS	high-performance storage system
IEEE	Institute of Electrical and Electronics Engineers
I/O	input/output
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
ISM	Integrated Safety Management (Argonne)
ISS	Internet Information Services
ISSF	Interim Supercomputing Support Facility (formerly of Argonne)
IT	information technology
JHQ	job hazard questionnaire
JLSE	Joint Laboratory for System Evaluation
JSA	Job Safety Analysis
LCF	Leadership Computing Facility
LDRD	Laboratory-Directed Research and Development (Argonne)
LES	large eddy simulation
LLM	large-language model
MD	molecular dynamics
MIT	Massachusetts Institute of Technology
ML	machine learning
MTBI	Mean Time Between Interrupt
MTTF	Mean Time to Failure
MTTI	Mean Time to Interrupt
NDA	nondisclosure agreement
NERSC	National Energy Research Scientific Computing Center
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NIU	Northern Illinois University
NRE	Non-recurring Engineering
NSF	National Science Foundation
Oak Ridge	Oak Ridge National Laboratory
OAR	Operational Assessment Report
OLCF	Oak Ridge Leadership Computing Facility
P3HPC	International Workshop on Performance, Portability and Productivity in HPC
PI	principal investigator
PM	preventative maintenance

PY	previous year
QCD	Quantum Chromodynamics
QEW	Qualified Electrical Worker
QMC	Quantum Monte Carlo
RAC	Resource Allocation Council (ALCF)
RDU	Reconfigurable Dataflow Unit
RMP	risk management plan
RTR	Response to Recommendation
SC	Office of Science (DOE)
SC23	2023 International Conference for High Performance Computing, Networking, Storage and Analysis (annual supercomputing conference)
SECAC	Secure ASCR Facilities
SME	subject matter expert
SSL	Secure Sockets Layer
STEM	science, technology, engineering and math
SULI	Science Undergraduate Laboratory Internships (DOE)
THAPI	Tracing Heterogeneous API
UAC	User Advisory Council (ALCF)
UB3	Userbase 3
UC,UChicago	University of Chicago
UIC	University of Illinois at Chicago
UTK	University of Tennessee, Knoxville
VAE	variational autoencoder
WCD	Work Control Documents
WPC	work planning and control
WRF	Weather Research and Forecasting

This page intentionally left blank.

Argonne Leadership Computing Facility
Argonne National Laboratory
9700 South Cass Avenue, Bldg. 240
Lemont, IL 60439-4832

www.anl.gov

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