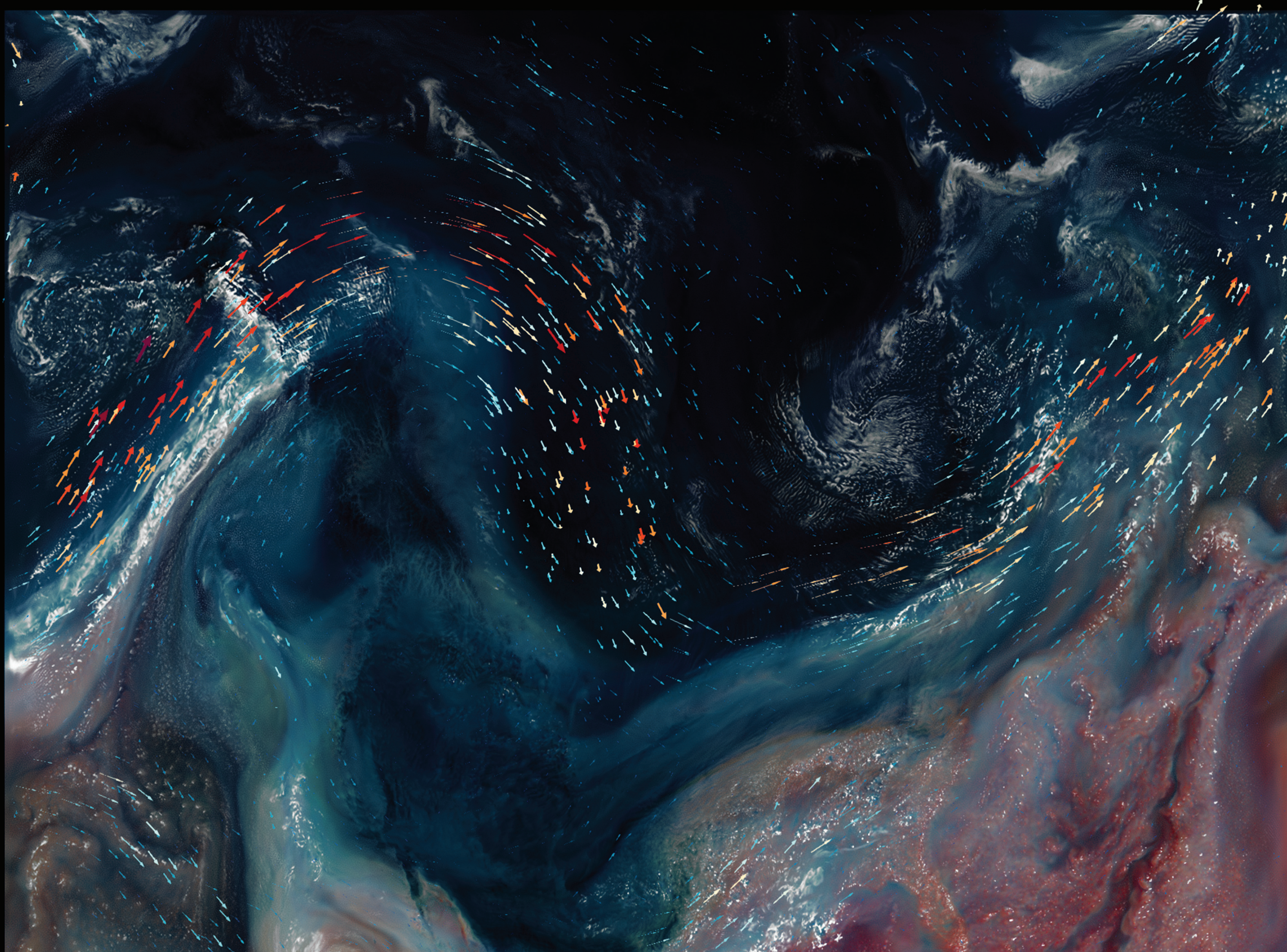
Argonne NATIONAL LABORATORY | Argonne Leadership Computing Facility

ANL/ALCF-25/1

2024 Operational Assessment Report



On the cover: *Researchers used the ALCF's Polaris supercomputer to perform GPU-enabled weather simulations at cloud-resolving (1 km) spatial resolution for the month of September 2017. Covering the entire North American continent and approximately 3 billion grid cells, this visualization shows water vapor (red-blue) and cloud water (white-grayscale) fields, with wind vectors highlighting the jet stream pattern.*

Image credit: *ALCF Visualization and Data Analytics Team; Rao Kotamarthi and Gökhan Sever, Argonne National Laboratory.*

Contents

Background ... x
Executive Summary ... ES-1
Response to Recommendation from the PY OAR ... RTR-1
Section 1. User Support Results ... 1-1
ALCF Response ... 1-1
Survey Approach ... 1-2
Likert Scale and Numeric Mapping ... 1-2
1.1 User Support Metrics ... 1-3
1.2 Problem Resolution Metrics ... 1-4
1.3 User Support and User Engagement ... 1-5
1.3.1 Tier 1 Support ... 1-5
1.3.1.1 Phone/Teams, Slack, and E-mail Support ... 1-5
1.3.1.2 User Support (UX) Accomplishments ... 1-6
NAIRR Projects ... 1-6
Slack Support during Training Events ... 1-6
Updated/Enhanced User Notifications ... 1-7
Miscellaneous ... 1-7
1.3.1.3 Userbase/Portal Improvements ... 1-7
UB3 ... 1-7
Ni/sbank ... 1-7
1.3.2 Application Support for Individual Projects ... 1-8
1.3.2.1 Projected long-term effects of colorectal cancer screening disruptions following the COVID-19 pandemic ... 1-8
1.3.2.2 Simulations of turbulence in black hole coronae ... 1-8
1.3.3 User Engagement ... 1-9
1.3.3.1 General Outreach ... 1-9
Facility Tours ... 1-9
User Advisory Council ... 1-9
1.3.3.2 Training Activities ... 1-9
Intro to AI-driven Science on Supercomputers: A Student Training Series (virtual) ... 1-9
ALCF INCITE GPU Hackathon (in-person) ... 1-10
2024 INCITE Proposal Writing Webinars (virtual) ... 1-10
Argonne Training Program on Extreme-Scale Computing (ATPESC) (in-person) ... 1-10
ALCF Hands-on HPC Workshop (in-person) ... 1-10
ALCF Webinars (virtual) ... 1-10
AI Testbed Workshops ... 1-11
1.3.3.3 Community Outreach ... 1-11
STEM Activities ... 1-11
1.3.4 Communications ... 1-13
1.3.4.1 Continuous Improvement ... 1-13

Contents (Cont.)

1.3.4.2 Consistent Cadence of ALCF Impact on Exascale and AI Efforts 1-13
1.3.4.3 Communicating Scientific Impact 1-14
1.3.4.4 Messaging for Users and Community 1-15
Conclusion 1-16

Section 2. Operational Performance 2-1

ALCF Response 2-1
2.1 ALCF Production Resources Overview 2-3
2.2 Definitions 2-3
2.3 Polaris 2-4
2.3.1 Scheduled and Overall Availability 2-4
2.3.1.1 Explanation of Significant Availability Losses 2-4
2.3.2 System Mean Time to Interrupt (MTTI) and System Mean Time to Failure (MTTF) 2-6
2.3.2.1 MTTI and MTTF Summary 2-6
2.3.3 Resource Utilization 2-6
2.3.3.1 System Utilization 2-6
2.3.3.2 Capability Utilization 2-8
2.4 Storage 2-11
2.4.1 Agile File System 2-11
2.4.1.1 Scheduled and Overall Availability 2-11
Agile – Explanation of Significant Availability Losses 2-11
2.4.1.2 MTTI and MTTF 2-12
Agile – MTTI and MTTF Summary 2-12
2.4.2 Swift File System 2-13
2.4.2.1 Scheduled and Overall Availability 2-13
Swift – Explanation of Significant Availability Losses 2-13
2.4.2.2 MTTI and MTTF 2-15
Swift – MTTI and MTTF Summary 2-15
2.4.3 Grand and Eagle Lustre File Systems 2-15
2.4.3.1 Grand Scheduled and Overall Availability 2-15
Grand – Explanation of Significant Availability Losses 2-16
Grand – MTTI and MTTF Summary 2-17
2.4.3.2 Eagle – Scheduled and Overall Availability 2-17
Eagle – Explanation of Significant Availability Losses 2-17
Eagle – MTTI and MTTF Summary 2-18
2.4.4 Tape Storage 2-19
2.4.4.1 Scheduled and Overall Availability 2-19
HPSS – Explanation of Significant Availability Losses 2-20
2.4.4.2 MTTI and MTTF 2-20
HPSS – MTTI and MTTF Summary 2-20
2.5 Center-Wide Operational Highlights 2-21
2.5.1 Polaris / Sirius 2-21
2.5.2 AI Testbed 2-21

Contents (Cont.)

2.5.3 Aurora / Sunspot ... 2-21
2.5.4 Crux ... 2-22
2.5.5 Sophia ... 2-22
2.5.6 Storage ... 2-22
2.5.7 Other Facility Notes ... 2-23
2.5.8 Decommissioned Systems ... 2-23
Conclusion ... 2-23

Section 3. Allocation of Resources ... 3-1

ALCF Response ... 3-1
3.1 Usage of the INCITE and ALCC Hours ... 3-1
3.2 Facility Director's Discretionary Reserve Time ... 3-4
3.3 Allocation Utilization ... 3-6
3.4 Other Large-Scale Managed Resources ... 3-6
Conclusion ... 3-7

Section 4. Operational Innovation ... 4-1

ALCF Response ... 4-1
4.1 Operational Innovation – Technical ... 4-1
4.1.1 MyALCF Portal ... 4-1
4.1.2 SimAI-Bench: Performance Benchmarks for Coupled Simulation and AI Workflows on HPC Systems ... 4-2
4.1.3 Enabling Integrated Research Infrastructure ... 4-3
4.1.4 Automation of Drug Discovery Workflows at Scale ... 4-4
4.2 Operational Innovation – Management/Workforce ... 4-5
4.2.1 ALCF Lighthouse Initiative ... 4-5
4.2.2 Sprinterns ... 4-5
4.2.3 Event Planning with Indico ... 4-6
4.2.4 Budget Planning Best Practices ... 4-6
4.2.5 Streamlining Editorial and Social Media Workflows ... 4-6
4.3 Postdoctoral Fellows ... 4-7

Section 5. Risk Management ... 5-1

ALCF Response ... 5-1
5.1 Risk Management Process Overview ... 5-1
5.1.1 Risk Review Board ... 5-3
5.1.2 Risk Management in Day-to-Day Operations ... 5-3
5.1.3 Continuation of the ALCF-3 Project ... 5-4
5.2 Major Risks Tracked ... 5-4
5.3 Risks Encountered in the Review Year ... 5-5
5.4 Retired Risks ... 5-5
5.5 New and Recharacterized Risks ... 5-6
5.6 Top Operating Risks Monitored Closely for the Next Year ... 5-6
Conclusion ... 5-7

Contents (Cont.)

Section 6. Environment, Safety, and Health 6-1

ALCF Response 6-1
6.1 Work Planning 6-1
6.1.1 Work Control Documents 6-1
6.1.2 Contractor Job Safety Analysis 6-2
6.1.3 Training 6-2
6.2 Safety Performance 6-3
6.2.1 Safety Shares 6-3
6.2.2 PDSA and Management Assessments 6-4
6.2.3 Operational Improvements 6-4
6.2.4 Lessons Learned 6-5
6.2.5 Ergonomics 6-5
6.2.6 Aurora 6-6
6.3 Incident Reports 6-6

Section 7. Security 7-1

ALCF Response 7-1
7.1 Demonstrating Effective Cybersecurity Practices 7-1
7.2 Cybersecurity Plan 7-5
7.3 Foreign Visitor Access and Export Controls 7-5

Section 8. Mission Impact, Strategic Planning, and Strategic Engagements 8-1

ALCF Response 8-1
8.1 Science Highlights and Accomplishments 8-1
Science Highlights 8-2
8.1.1 Demonstrating Cross-Facility Data Processing at Scale with Laue Microdiffraction 8-3
8.1.2 Efficient Algorithms for Monte Carlo Particle Transport on AI Accelerator Hardware 8-4
8.1.3 Computer Simulation-Aided Design of High-Performance Electrolytes for Lithium Batteries 8-6
8.1.4 A Theory for Neutron Star and Black Hole Kicks and Induced Spins 8-8
8.1.5 Three-Dimensional Shock Boundary Layer Interactions Over Flexible Walls 8-10
8.2 Research Activities / Vendor Engagements for Future Operations 8-12
8.2.1 Research Activity - Joint Laboratory for System Evaluation (JLSE) 8-12
8.2.2 Research Activity - ALCF AI Testbed 8-13
8.2.3 Vendor Engagements for Future Systems 8-16
8.3 DOE Program Engagements / Requirements Gathering 8-16
8.3.1 Engagement Highlights 8-17
Supercomputing 2024 8-17
Summary of Engagements with the Exascale Computing Project 8-17
Engagement in Standards and Community Groups 8-18
Performance, Portability, and Productivity in HPC (P3HPC) Forum 8-18

Contents (Cont.)

8.4 Integrated Research Infrastructure Activities 8-19
Nexus 8-19
Pathfinding Projects 8-19
DIII-D 8-19
APS 8-19
Advanced Light Source 8-20
Technical Subcommittees 8-20
Leadership Group 8-20
Conclusion 8-20

Appendix A – Calculations A-1

A.1 Scheduled Availability A-1
A.2 Overall Availability A-1
A.3 System Mean Time to Interrupt (MTTI) A-2
A.4 System Mean Time to Failure (MTTF) A-2
A.5 Total System Utilization A-2
A.6 Capability A-2
A.7 Polaris Nodes A-3

Appendix B – ALCF Director's Discretionary Projects B-1

Appendix C – ALCF CY2024 Science Highlights C-1

Appendix D – Acronyms and Abbreviations D-1

Figures

2.1 Polaris Weekly Availability for CY 2024 2-5
2.2 Polaris System Utilization over Time by Program 2-7
2.3 Polaris INCITE Overall Capability 2-9
2.4 Polaris Capability Node-Hours by Program 2-10
2.5 Polaris Job Usage by Size 2-10
2.6 Agile Weekly Availability for CY 2024 2-12
2.7 Swift Weekly Availability for CY 2024 2-14
2.8 Grand Weekly Availability for CY 2024 2-16
2.9 Eagle Weekly Availability for CY 2024 2-18
2.10 HPSS Weekly Availability for CY 2024 2-20
3.1 Polaris INCITE 2024 Allocation Usage (Note: Projects are randomly ordered.) 3-2
3.2 Polaris ALCC 2023–2024 Allocation Usage (Note: Projects are randomly ordered.) 3-3

Figures (Cont.)

3.3 Polaris ALCC 2024–2025 Allocation Usage as of December 31, 2024 (Note: Projects are randomly ordered.) ... 3-3

3.4 Polaris CY 2024 DD Allocations by Domain ... 3-5

4.1 Screenshot of the MyALCF dashboard, which displays project allocation and machine status information (left) and an enlarged view of Compute Days Needed (right). This view illustrates a PI's ability to fully utilize the allocation. If time is limited, as shown in the example project, the PI can seek assistance from the ALCF Catalyst to reallocate nodes for project completion. ... 4-2

5.1 Illustration of the Iterative Risk Management Process from ALCF's Risk Management Plan ... 5-2

8.1 Data from the new coded aperture at an APS beamline (top left) is automatically transferred to the Polaris supercomputer (top right), where it is reconstructed on demand in real time (bottom). ... 8-3

8.2 Speedup vs. serial CPU execution for macroscopic cross-section lookup kernel (adapted from XSBench). ... 8-4

8.3 (a–c) Simulation snapshots of electrolyte distribution on NMC62 and Li electrodes: cyan-IL cation; white-H on the PMpyr+ backbone; purple-F on the PMpyrf+; yellow-FSI–; green-Li+. (d) A representative snapshot of Li+ ion solvation structure with TfTHF that results in better mobility. ... 8-6

8.4 An isosurface (10%) of the 56Ni abundance in the explosion of a 9 M⊙ (left) and 23 M⊙ (right) progenitor model at ~1.95 and ~6.2 seconds, respectively, after the bounce, painted by Mach number. Red/purple is positive and white/blue is negative, indicating infall. ... 8-8

8.5 The computational setup of the full-span 3D SBLI simulations. By oscillating the wedge, the strength of the SBLI can be increased to study the effects of flow separation and that of an oscillating separation bubble coupled with the fluid-structure interaction. Simulation details compared well with measurements from flexible-panel experiments. ... 8-10

8.6 Breakdown of key activities by the ALCF in CY 2024. The first pie chart (left) breaks down the 161 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. ... 8-17

Tables

ES.1 Summary of the Target and Actual Data for the Previous Year (2024) Metrics ... ES-2

1.1 All 2024 User Support Metrics and Results ... 1-1

1.2 User Survey Results in 2024 grouped by Allocation Program ... 1-4

1.3 User Support Metrics ... 1-4

Tables (Cont.)

1.4 Problem Resolution Metrics 1-5
1.5 Ticket Categorization for 2023 and 2024 1-6
1.6 Publications Designed for Print 1-15
1.7 Primary Communication Channels 1-15
1.8 Target Audiences 1-16
2.1 Summary of All Metrics Reported in the Operational Performance Section 2-1
2.2 Summary of Operational Performance of HPC Storage Systems 2-2
2.3 Availability Results 2-4
2.4 MTTI and MTTF Results 2-6
2.5 System Utilization Results 2-7
2.6 Node-Hours Allocated and Used by Program 2-8
2.7 Capability Results 2-9
2.8 Availability Results 2-11
2.9 MTTI and MTTF Results - Agile 2-13
2.10 Availability Results - Swift 2-13
2.11 MTTI and MTTF Results - Swift 2-15
2.12 Availability Results - Grand 2-16
2.13 MTTI and MTTF Results - Grand 2-17
2.14 Availability Results - Eagle 2-17
2.15 MTTI and MTTF Results - Eagle 2-19
2.16 Availability Results - HPSS 2-19
2.17 MTTI and MTTF Results - HPSS 2-21
3.1 INCITE 2024 Time Allocated and Used on Polaris 3-1
3.2 ALCC Time Allocated and Used on Polaris in CY 2024 3-4
3.3 DD Time Allocated and Used on Polaris in CY 2024 3-5
5.1 Major Risks Tracked for CY 2024 5-5
5.2 Top Operating Risks Monitored for CY 2025 5-6
6.1 ALCF Injury and Illness Cases FY 2020 thru FY 2024 6-6
8.1 Summary of Users' Peer-Reviewed Publications in CY 2024 8-2
8.2 Summary of Users' Peer-Reviewed Publications for 5-year Moving Window 8-2
8.3 Summary of ALCF Participation in SC24 8-17
A.1 Capability Definitions for Polaris A-2
A.2 Total and Reportable Nodes for Polaris A-3

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-ranging areas such as energy assurance, environmental 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 cost to the DOE Office of Science and the U.S. Congress.

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) 2024.

In CY 2024, the ALCF operated Polaris, an AMD and NVIDIA-based Hewlett Packard Enterprise (HPE) Apollo 6500 Gen10+ system that provides a powerful platform for breakthrough science and prepares applications and workloads for Aurora, Argonne National Laboratory's Intel-HPE exascale computer.

Since going into production in August 2022, Polaris has supported research teams from the DOE Exascale Computing Project (ECP), which concluded in 2024, and from ALCF's Aurora Early Science Program (ESP). Polaris delivered a total of 2555.7K node-hours to 17 Innovative and Novel Computational Impact on Theory and Experiment (INCITE) 2024 projects and 601.6K node-hours to ASCR Leadership Computing Challenge (ALCC) projects (12 of which were awarded during the 2023–2024 ALCC year and 11 of which were awarded during the 2024–2025 ALCC year), as well as providing substantial time for Director's Discretionary (DD) projects (966.2K node-hours).

As Table ES.1 shows, Polaris performed exceptionally well in terms of overall availability (97.7 percent), scheduled availability (99.6 percent), and utilization (94.7 percent; Table 2.1). In CY 2024, ALCF supported more than 2,000 users. As of March 18, 2025, ALCF's users have published 254 papers in peer-reviewed journals and technical proceedings. Other CY 2024 highlights include:

- ALCF conducted 22 training activities, reaching more than 1,600 participants in events where participation was tracked.
- In April, Argonne and RIKEN, Japan's national research and development agency, signed a memorandum of understanding to establish a cooperative relationship in support of artificial intelligence (AI) computing projects. This includes a joint effort to train large language models on scientific data, exchange researchers, staff, and students, and share datasets and other scientific and technical research materials.
- The ALCF AI Testbed expanded its infrastructure and reach, adding new systems and supporting three of the initial project awards from the National Artificial Intelligence Research Resource (NAIRR) Pilot.
- In September, David Baker, a longtime ALCF user from the University of Washington, was awarded the 2024 Nobel Prize in Chemistry for his pioneering work in computational protein design.
- In November, an Argonne-based research team's AI-driven, multimodal protein design framework was selected as a finalist for the prestigious Gordon Bell Prize at the 2024 International Conference for High Performance Computing, Networking, Storage and Analysis (SC24). Two other ALCF teams received Best Paper awards at SC24. Both papers focused on in-situ visualization research and application performance on the Intel graphical processing units (GPUs) powering Aurora.

Table ES.1 Summary of the Target and Actual Data for the Previous Year (2024) Metrics

Area	Metric	2024 Target	2024 Actual
User Results	User Survey – **Overall Satisfaction**	3.5/5.0	4.4/5.0
	User Survey – **User Support**	3.5/5.0	4.5/5.0
	User Survey – **Problem Resolution**	3.5/5.0	4.4/5.0
	User Survey – **Response Rate**	25.0%	39.6%
	% User Problems Addressed within Three Working Days	80.0%	93.1%
Business Results	Polaris Overall Availability	90.0%	97.7%
	Polaris Scheduled Availability	90.0%	99.6%
	% of INCITE node hours from jobs run on 20.0% or more of Polaris (99–496 nodes)	20.0%	61.4%
	% of INCITE node hours from jobs run on 60.0% or more of Polaris (297–496 nodes)	N/A [a]	11.3%
	Grand File System Overall Availability	90.0%	97.5%
	Grand File System Scheduled Availability	90.0%	100.0%
	Eagle File System Overall Availability	90.0%	97.9%
	Eagle File System Scheduled Availability	90.0%	100.0%
	Agile File System Overall Availability	90.0%	99.7%
	Agile File System Scheduled Availability	90.0%	100.0%
	Swift File System Overall Availability	90.0%	97.4%
	Swift File System Scheduled Availability	90.0%	100.0%
	HPSS [b] Archive Overall Availability	90.0%	97.8%
	HPSS Archive Scheduled Availability	90.0%	100.0%

[a] N/A = not applicable.

[b] HPSS = high-performance storage system.

In early 2025, Argonne will fully open Aurora to researchers around the world, marking a new era in computing-driven discoveries. In addition to running a variety of AI and computational science applications, Aurora will be used to train a large language model with over 1 trillion parameters specifically for scientific research, further transforming the computing landscape. Looking ahead to 2025 and beyond, the Argonne Leadership Computing Facility will continue to engage in strategic initiatives that push the limits of computational science and engineering.

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 received no 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

The ALCF has processes in place to effectively support its users, resolve user problems, and conduct outreach. The 2024 annual user survey measured overall satisfaction, user support, and problem resolution—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) 2024.

Table 1.1 All 2024 User Support Metrics and Results [a]

Metrics		2023 Target	2023 Actual	2024 Target	2024 Actual
Number Surveyed		N/A [c]	1,495	N/A [c]	1,852
Number of Respondents (Response Rate)		25.0%	562 (37.6%)	25.0%	733 (39.6%)
Overall Satisfaction	Mean	3.5	4.5	3.5	4.4
	Variance	N/A	0.6	N/A	0.7
	Standard Deviation	N/A	0.8 [d]	N/A	0.8
Problem Resolution	Mean	3.5	4.5	3.5	4.4
	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
Totals		**2023 Target**	**2023 Actual**	**2024 Target**	**2024 Actual**
% of Problems Addressed Within Three Working Days [b]		80.0%	93.6%	80.0%	93.1%

[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 into 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 the note above.

[c] N/A = not applicable.

[d] This was incorrectly reported as 0.6 in the CY 2023 OAR and is corrected here.

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 OAR, and making all other questions optional. The facility surveys workshop attendees separately.

The 2024 user survey closely resembled the 2023 user survey, with minor modifications to operations, infrastructure, performance, and performance tools/services questions. Questions regarding Theta/ThetaGPU were removed, as both systems retired in 2023.

The 2024 survey was administered through a contract with Statistical Consulting Services at the Department of Statistics and Actuarial Science at Northern Illinois University. It consisted of 6 required questions and 26 optional questions. The survey and associated e-mail campaign ran from November 18, 2024, through January 10, 2025. Each reminder e-mail included a user-specific link to the 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. This method also allows for the 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

Some of the non-metric survey questions capture sentiments about various aspects of ALCF's user services that used the options below:

Select all that apply.
Praise
Suggestions for Improvement
Problem Experienced
Complaint

1.1 User Support Metrics

Everyone who met the definition of a facility user during the fiscal year (FY) 2024—that is, users who would be counted under the ALCF's annual submission of user statistics to the Office of Science (SC)—was asked to complete the survey, 2,013 individuals in total. Of those individuals, 64 did not receive the email due to undeliverable messages and 97 chose to opt out of the survey. Of the 1,852 remaining users, 733 responded, for a 39.6 percent response rate.

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

Table 1.2 User Survey Results in 2024 grouped by Allocation Program

Metrics by Program		INCITE	ALCC	INCITE + ALCC	DD	All
Number Surveyed		302	102	404	1,448	1,852
Number of Respondents		117	45	162	571	733
Response Rate		38.7%	44.1%	40.1%	39.4%	39.6%
Overall Satisfaction	Mean	4.4	4.2	4.3	4.5	4.4
	Variance	0.9	1.1	0.9	0.6	0.7
	Standard Deviation	0.9	1.0	1.0	0.8	0.8
Problem Resolution	Mean	4.5	4.4	4.5	4.4	4.4
	Variance	0.6	0.6	0.6	0.5	0.5
	Standard Deviation	0.8	0.8	0.8	0.7	0.7
User Support	Mean	4.5	4.4	4.5	4.5	4.5
	Variance	0.6	0.6	0.6	0.5	0.5
	Standard Deviation	0.8	0.8	0.8	0.7	0.7
All Questions	Mean	4.5	4.4	4.5	4.5	4.5
	Variance	0.6	0.7	0.6	0.5	0.6
	Standard Deviation	0.8	0.8	0.8	0.7	0.7

As Table 1.3 shows, in CY 2024, the ALCF again exceeded the targets for overall satisfaction and user support.

Table 1.3 User Support Metrics

Survey Area	2023 Target	2023 Actual	2024 Target	2024 Actual
Overall Satisfaction Rating	3.5/5.0	4.5/5.0	3.5/5.0	4.4/5.0
Average 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 and actual for the percentage of user problem tickets or "trouble tickets" addressed in three days or less. The ALCF exceeded the target. 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

	2023 Target	2023 Actual	2024 Target	2024 Actual
% of Problems Addressed Within Three Working Days [a]	80.0%	93.6%	80.0%	93.1%
Avg. of Problem Resolution Ratings	3.5/5.0	4.5/5.0	3.5/5.0	4.4/5.0

[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/Teams, Slack, and E-mail Support

In CY 2024, ALCF addressed and categorized 6,651 user support tickets generated by email and calls, reflecting a three percent decrease from CY 2023. The most significant decreases were observed in the Accounts/Access and Systems categories, while the largest increase occurred in the Scheduling category (Table 1.5).

In CY 2024, ALCF introduced five new sub-categories for tickets classified as Accounts to provide a clearer understanding of the issues within this category.

Phone, Teams, or Slack support primarily addressed account-related queries, such as unlocking user tokens or verifying a user identity to allow the user to submit a reactivation request. Slack was used for screen sharing to troubleshoot portal access issues. See the section below for details about Slack support during ALCF training events.

Table 1.5 Ticket Categorization for 2023 and 2024

Category	2023	2024
Access	1,285 (19%)	1,133 (17%)
Accounts [a]	2,841 (42%)	2,635 (40%)
Allocations	1,196 (17%)	1,155 (17%)
Applications Software	271 (4%)	329 (5%)
Bounced Emails	29 (0%)	77 (1%)
Compilers	32 (0%)	29 (0%)
Data Transfer	58 (1%)	57 (1%)
Debuggers	8 (0%)	3 (0%)
File System	87 (1%)	74 (1%)
HPSS [b]	17 (0%)	19 (0%)
I/O and Storage	176 (3%)	190 (3%)
Libraries	25 (0%)	33 (0%)
Miscellaneous	157 (2%)	118 (2%)
Network	11 (0%)	35 (1%)
Performance and Performance Tools	27 (0%)	6 (0%)
Reports	169 (2%)	113 (2%)
Scheduling	319 (5%)	547 (8%)
System	129 (2%)	92 (1%)
Visualization	1 (0%)	6 (0%)
TOTAL TICKETS	6,838 (100%)	6,651 (100%)

[a] In Q4 of CY 2024, we introduced 5 new categories to capture additional details for account-related tickets – Login (8%), New (14%), Renew/Reactivate (48%), Tokens (24%), and Misc. (6%).

[b] HPSS = high-performance storage system; I/O = input/output.

1.3.1.2 User Support (UX) Accomplishments

NAIRR Projects

All four ALCF AI Testbed systems were included in the National Artificial Intelligence Research Resource (NAIRR) Pilot allocation program. Ten NAIRR projects were awarded time at the ALCF. Collaboration with the NAIRR team led to the development of a workflow that notified ALCF Support when NAIRR project tickets were submitted to their Jira system, allowing the ALCF to track the requests in the ALCF ticketing system.

Slack Support during Training Events

It is now a standard practice for the ALCF to provide real-time account support during training events via Slack. A dedicated Slack workspace is created for participants and support team members to collaborate. The support covers many of the same areas for which users typically submit support tickets, such as access issues, debugging user code, job submissions, and filesystem queries. Using Slack is particularly effective for short-lived events with a small user base, as they require a quicker turnaround time, usually within an hour.

Updated/Enhanced User Notifications

The ALCF reformatted the welcome emails for each system notification list to include links to user guides, offering system-specific guidance to new users during the onboarding process. Additionally, the ALCF converted the user notifications from text-only to HTML format to improve readability and engagement.

Miscellaneous

Community sharing capability has been extended to all projects on the Eagle file system, including those migrated from the Grand file system hardware. Previously, it was only available to projects created on Eagle, and PIs had to submit a request to Support. In 2024, this capability was automatically enabled for all active ALCF projects, eliminating the need for PIs to send a request. This change has made it easier for project teams to share their data with the scientific community.

1.3.1.3 Userbase/Portal Improvements

UB3

In November, the MyALCF user portal was launched to help users manage their awards and teams. The portal also provides system status, information on training and events, and facility updates. Userbase3 (UB3), ALCF’s accounts and project management website, was integrated into the MyALCF user portal. For more information on MyALCF, refer to Section 4.1.1. Before the launch of the MyALCF portal, several enhancements were made to UB3. Principal Investigators and Proxies can now access a project’s resource allocations on the Project Management screen. Additionally, the Director’s Discretionary Allocation Request screen was updated to allow users to request time on Sophia and Crux.

Ni/sbank

Infrastructure changes were made in Ni (sbank), the ALCF’s resource usage accounting system, to support integration with the Applications Integration Group web framework. This development enabled the ALCF to incorporate Ni (sbank) into the MyALCF user portal, providing a graphical user interface (GUI) for sbank. The interface provides an easy-to-use alternative to the command line interface, as well as on-demand help. This feature saves the user time and accelerates the learning of sbank functionality. Additionally, Vault, a secure, centralized secret management tool, was integrated into the Ni workflow to store and distribute authentication tokens for all machines using Ni. Previously, the tokens were stored and distributed via a system-administered directory.

1.3.2 Application Support for Individual Projects

1.3.2.1 Projected long-term effects of colorectal cancer screening disruptions following the COVID-19 pandemic

Colorectal cancer (CRC) remains the second-leading cause of cancer deaths in the United States, and there is clear evidence that screening has a major impact on reducing the burden of CRC and is cost-effective. While CRC screening rates have increased in the past two decades, they are still below established population-level targets, and these low rates were further exacerbated by the COVID-19 pandemic, as elective procedures were postponed during the initial months of the pandemic, which led to sharp declines in CRC screening exams during that period. Two independently developed microsimulation models of CRC were used to estimate the effects of pandemic-induced disruptions in colonoscopy screening for several population cohorts and analyze variations of two commonly used screening strategies. The impact on disruptions was investigated for three scenarios: delays in screening, regimen switching, and screening discontinuation. ALCF staff worked with a team at the ALCF INCITE Hackathon 2023 to port a simplified version of the CRC model SimCRC from R to Python to run on GPUs. In one benchmark, speedups of ~2x were observed on a single central processing unit (CPU) with the Python port, and further speedups of ~36x were observed using all four GPUs on a single Polaris node. While short-term delays in screening of 3–18 months are predicted to result in minor decreases in life expectancy, discontinuing screening is predicted to cause a much more significant decline. The worst-case scenario considered was that of 50-year-olds who postponed screening until the age of 65 when they became Medicare-eligible. These findings demonstrate that unequal recovery of screening following the pandemic can further widen disparities and highlight the importance of ensuring the resumption of screening following the pandemic.

1.3.2.2 Simulations of turbulence in black hole coronae

Matter accreting onto black holes makes up some of the brightest X-ray sources in the universe. Yet, since the early days of X-ray astronomy in the 1960s, researchers have been struggling to explain the origins of the observed emission. A new theoretical study by the project team proposes that the X-rays are produced in a strongly turbulent medium threaded by intense magnetic fields near the black hole. Their model was validated by first-of-a-kind computer simulations of electromagnetic turbulence performed on Theta, which tracked the interactions between billions of charged particles and photons. The researchers compared their simulation results to the black hole Cygnus X-1 in the Milky Way — one of the brightest persistent sources of X-rays in the sky — and demonstrated good agreement with the observed emissions. The project's catalyst met with the PI several times early in the project's first year to help create job submission scripts, advise on job queueing strategies, and help find the best mode and affinity settings for Theta's processors for efficient execution of the project's code. The catalyst also helped the PI apply for and successfully use overburn time to achieve the project's first-year milestones.

1.3.3 User Engagement

1.3.3.1 General Outreach

Facility Tours

Visitors to Argonne can request a tour of the ALCF. The ALCF welcomes a diverse array of visitors, including student groups, Congressional representatives, government officials, industry researchers, summer research students, visiting researchers, and journalists. Different staff members lead tours that may include the main machine room, the Visualization Lab, and the Aurora supercomputer.

In 2024, staff members hosted over 400 groups of visitors. The leadership team at ALCF welcomed DOE Deputy Assistant Secretary for Buildings and Industry Dr. Carolyn Snyder on January 23, 2024; former DOE Under Secretary for Infrastructure David Crane on February 9, 2024, former Treasurer of the United States Lynn Malerba on April 30, 2024, U.S. Representative Raja Krishnamoorthi (D-IL) on August 27, 2024; and former Principal Deputy Director of National Intelligence, Dr. Stacey Dixon on September 24, 2024, among other VIP visitors.

User Advisory Council

The 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 challenges and requirements of specific applications or domain areas. The current council consists of 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), Vikram Mulligan (Flatiron Institute), and Aiichiro Nakano (University of Southern California). ALCF leadership suspended UAC meetings in 2024 and will restore the council and resume meetings in 2025.

1.3.3.2 Training Activities

The ALCF offers workshops and webinars on various tools, systems, and frameworks. These intensive 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 2024, ALCF conducted 22 training activities, reaching more than 1,600 participants in events where participation was tracked. Below is a list of ALCF 2024 training activities:

Intro to AI-driven Science on Supercomputers: A Student Training Series (virtual)

This 7–8 week introductory webinar series targeted undergraduate and graduate students enrolled at U.S. universities and community colleges. It aimed to attract a new generation of AI users by maintaining a low entry barrier; attendees only needed basic experience with the Python programming language as a prerequisite. In 2024, the training series ran twice, in spring and in fall. Both series focused on introducing participants to foundational concepts in AI and machine learning (ML) and then dove into the fundamentals of large language models (LLMs). ALCF

computer scientists led the weekly sessions and hands-on exercises along with talks by Argonne scientists who use AI in their own research. The programs trained more than 500 attendees from over 100 universities and included undergraduate students, graduate students, postdocs, and faculty from fields spanning computer science to materials science to biology.
(Dates: February 6–March 26 and October 1–November 12, 2024)

ALCF INCITE GPU Hackathon (in-person)
ALCF organized an in-person hackathon to help researchers prepare for the 2024 INCITE Call for Proposals. In CY 2024, ALCF received 40 applications, which is twice as many compared to previous years. Twenty teams were selected to work closely with ALCF and vendor experts to accelerate their AI or HPC research 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. Out of 20 teams, 11 were part of ongoing INCITE projects, and out of the nine remaining teams, six applied for 2025 INCITE awards. (Dates: May 7–8, 2024: Team-Expert Virtual Introductory Meeting, May 21–23, 2024: In-Person Hackathon at Argonne National Laboratory)

2024 INCITE Proposal Writing Webinars (virtual)
The INCITE program, the ALCF, and the Oak Ridge Leadership Computing Facility (OLCF) jointly hosted two webinars on effective strategies for writing an INCITE proposal.
(Dates: April 23, 2024, and May 7, 2024)

Argonne Training Program on Extreme-Scale Computing (ATPESC) (in-person)
ATPESC, initially founded by Argonne and later integrated into the Exascale Computing Project, is an intensive, two-week program that focuses on essential skills and tools for using supercomputers in scientific research. The program features talks from leading computer scientists and HPC experts and hands-on training using the DOE leadership-class systems at the ALCF, the OLCF, and the National Energy Research Scientific Computing Center (NERSC). In 2024, ATPESC accepted 75 attendees from 54 different institutions worldwide. Video recordings of ATPESC sessions are available on ALCF’s YouTube channel. (Dates: July 28–Aug. 9, 2024)

ALCF Hands-on HPC Workshop (in-person)
The ALCF hosted a three-day hands-on HPC Workshop at Argonne, attracting over 170 registrants with an average of 90 participating each day. In 2024, the ALCF offered two distinct tracks: one focused on programming models, performance optimization, and debugging tools, and the other centered on AI and ML frameworks. Participants had the opportunity for hands-on experience on Polaris and the AI Testbed, focusing on porting applications to heterogeneous architectures (CPU + GPU), improving code performance, and exploring AI/ML application development on ALCF systems. (Dates: October 29–31, 2024)

ALCF Webinars (virtual)
In 2024, the ALCF webinar program focused on both Polaris and getting ready for Aurora. All talks are posted to the ALCF’s YouTube channel, and the associated training materials can be found on the ALCF Events website. The ALCF collaborated with the OLCF and NERSC to provide the Portability Webinar series. The ALCF also participates in community events, the IDEAS productivity project webinar series, and Intel webinars. The 2024 webinar program series included:

- Introducing Visualization on Polaris, January 2024
- ATPESC 2024 Application Webinar, February 2024
- Overview of Aurora Exascale Compute Blades, February 2024
- Overview of Integrated Research Infrastructure, March 2024
- QMCPACK: Journey to Exascale on Aurora, April 2024
- Deep Learning Frameworks on Aurora, May 2024
- Portable SYCL Code Using oneMKL on AMD, Intel, and NVIDIA GPUs, June 2024
- Getting Started Bootcamp on Polaris, July 2024
- chipStar: a Heterogeneous Interface for Portability (HIP) implementation for Aurora, August 2024
- Intel Performance Profiling Tools on Aurora, September 2024
- Overview of DAOS and Best Practices, October 2024
- Remote Workflows at ALCF, December 2024

AI Testbed Workshops

ALCF offered hands-on training for all four AI Testbed systems that were 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 employ next-generation AI for science.

The training sessions consisted of:

- SambaNova AI Training Workshop (April 30–May 1, 2024)
- Cerebras AI Training Workshop (May 7–8, 2024)
- Groq AI Training Workshop (June 4–5, 2024)
- Graphcore AI Training Workshop (June 11–12, 2024)

1.3.3.3 Community Outreach

In CY 2024, the ALCF hosted or participated in seven community outreach events, reaching 1,500+ students, both in their classrooms and on-site, alongside over 400 tours of the ALCF. These activities ranged from giving tours to industry groups and DOE leadership to participating in science, technology, engineering, and math (STEM) events and K-12 classroom visits. The ALCF supported several summer coding camps, and ALCF staff participated in computer science education events such as the Hour of Code. Additionally, the ALCF's annual summer student program allowed undergraduate and graduate students to work side-by-side with staff members on real-world research projects on ALCF resources, working in areas like computational science, system administration, and data science.

STEM Activities

Break Through Tech Sprinternship Program

The ALCF and the University of Illinois Chicago ran a pilot program to introduce undergraduate students to HPC earlier in their computing studies. The program offered short, immersive micro-internships to a team of five computer science undergraduates over three weeks, during which they developed a program aimed at tracking allocation awards and their impact on scientific publications more effectively. All students were granted an extension for their sprinternship until the end of summer 2024, and four out of the five students accepted the extension.
(Dates: May 6–24, 2024)

ALCF Lighthouse Initiative

In 2024, the ALCF launched the Lighthouse Initiative, which aims to establish sustainable partnerships with university research computing centers and foster connections with the next generation of computing professionals. As part of this initiative, university partners collaborate with ALCF staff to onboard their computing center personnel to ALCF systems. This onboarding process includes a half-day, in-person workshop, hosted at Argonne or the partner institution. During the workshop, university computing center staff are guided through tutorials that introduce them to ALCF machines and their capabilities. In 2024, the ALCF held three onboarding events:

- The University of Chicago Research Computing Center (Date: February 19, 2024)
- University of Illinois Chicago Advanced Cyberinfrastructure for Education and Research (Date: June 28, 2024)
- University of Wyoming Advanced Research Computing Center (Date: November 6, 2024)

Summer 2024 Research Internships

The ALCF hosted 41 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 researchers worked closely with their mentors from the 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. (Dates: June–August 2024)

2024 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 15, 2024)

2024 Science Careers in Search of Women

Argonne's 2024 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 190 girls and their teachers participated in the event. (Date: April 10, 2024)

2024 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 promoted 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 the ALCF, hosted 27 students.
(Dates: June 24–28, 2024)

2024 CodeGirls@Argonne Camp

Argonne held its annual CodeGirls@Argonne summer camp, a weeklong STEM course for 6th- and 7th-grade girls taught by the Argonne Learning Center and ALCF staff. Twenty-four 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: July 8–12, 2024)

2024 Big Data Camp

Argonne's sixth annual Big Data Camp introduced high school juniors and seniors to the advanced tools used by professional data scientists. Campers were required to have coding experience and learned techniques for probing and analyzing massive scientific datasets, such as the dataset from the Array of Things (AoT) urban sensor project. Participants were also introduced to the foundational concepts underlying artificial intelligence and machine learning and why supercomputers are so important to use these tools effectively. This camp was organized by Argonne's Educational Programs and Outreach staff and taught by ALCF scientists and visualization experts, hosting 12 participants in 2024. (Dates: July 15–19, 2024)

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 2024, 31 Argonne staff volunteers, including 6 from the ALCF, visiting 36 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 9–13, 2024)

1.3.4 Communications

1.3.4.1 Continuous Improvement

The ALCF website, GitHub repository, and YouTube channel serve as primary repositories for diverse resources, including onboarding guides, community announcements, and video training tutorials. These platforms, managed by the ALCF communications team, underwent regular internal content reviews by technical staff, ensuring continuous updates with new content, features, and redesigned web pages.

In CY 2024, ALCF adopted new tools to align with Argonne's transition to Federal Risk and Authorization Management Program (FedRAMP)-compliant products. FedRAMP is a government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services, ensuring effective and repeatable cloud security for federal agencies. GovDelivery was implemented to distribute monthly newsletters and promote training events, while Microsoft Planner became the team's primary project management tool.

1.3.4.2 Consistent Cadence of ALCF Impact on Exascale and AI Efforts

ALCF staff wrote many articles during CY 2024 to communicate the value that ALCF's expert staff and computing resources bring to users' science campaigns.

The ALCF leveraged long-standing partnerships with Intel, Hewlett Packard Enterprise (HPE), Argonne's central communications, and several AI accelerator vendors to facilitate the broad dissemination of ALCF success stories. These partnerships supported joint communication plans for Aurora and the ALCF AI Testbed.

For the International Supercomputing Conference (ISC) 2024, the ALCF communications team coordinated a press event with HPE and Intel to showcase Aurora's progress prior to the Top500 announcement. This event generated favorable media coverage for Argonne and highlighted Aurora. Complementary videos featuring Aurora Early Science Program researchers were shared on social media throughout the year, maintaining focus on Aurora's scientific impact. Additionally, the team collaborated with engineers and computational scientists to create a "Case Studies" section on the ALCF website, providing developer-focused insights into code porting and performance optimization.

In promoting the ALCF AI Testbed, the communications team worked with vendors to host intensive two-day workshops on systems from SambaNova, Cerebras, Graphcore, and Groq. Recorded sessions were converted into training videos and uploaded to YouTube, reaching potential testbed users and participants in the NAIRR Pilot program.

During the 2024 International Conference for High Performance Computing, Networking, Storage and Analysis (SC24), the team collaborated with SambaNova to announce the expansion of its cluster solution, enhancing ALCF's AI capabilities. The team also produced videos featuring Argonne staff for social media and the SC24 exhibit hall, while coordinating with Cerebras on potential recognition for a Gordon Bell Award.

1.3.4.3 Communicating Scientific Impact

The ALCF produced science stories and articles and promoted HPC training opportunities throughout CY 2024, including coverage of Aurora's performance benchmark results presented at the ISC 2024 and SC24 conferences.

The stories were picked up by several media outlets and resulted in some of the ALCF's and Argonne's most popular social media posts of the year, further solidifying Argonne's reputation as a leader in the supercomputing field. Furthermore, the ALCF planned marketing campaigns around major annual HPC conferences and events such as the ISC 2024, Exascale Day, and SC24.

In CY 2024, ALCF placed 58 original science stories in various news outlets in coordination with Argonne's Communications and Public Affairs (CPA) Division and other ALCF direct relationships. The ALCF tracked media hits through its media monitoring service, Meltwater. In CY 2024, Meltwater reported 2,969 unique ALCF media hits (an increase of 1,761 from 2023), 145 of which were chronicled on the ALCF website, with an audience reach of 6.9 billion. 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.

In CY 2024, the ALCF streamlined its social media strategy by shifting its focus from Twitter to LinkedIn. This targeted approach led to significant growth on LinkedIn, with an increase of

1,938 followers, bringing the total audience to approximately 5,300. The ALCF also used Instagram to showcase science visualizations and promote training opportunities to students.

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. For CY 2024, the team published a digital version of the ALCF Annual Report.

Table 1.6 Publications Designed for Print

Publication	Frequency	When
Press and Visitor Packets	As Needed	As Needed
Industry Brochures	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

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

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

In CY 2024, ALCF's monthly e-newsletter, *Newsbytes*, highlighted ALCF-supported research advances, promoted training events and allocation program announcements, and linked to relevant news stories. Special announcements about specific training opportunities and fellowships were sent throughout the year as needed.

Conclusion

Maximizing the use of ALCF resources is at the forefront of all improvements to the ALCF's customer service. In CY 2024, the ALCF continued to support project teams in adapting their codes to new architectures and preparing others to apply for major project allocations. The Exascale Computing Project (ECP) and Early Science Program (ESP) users continued their work on Aurora and Sunspot. The ALCF partnered with other national laboratories and the ECP team to present work in premier scientific journals and at top professional meetings and conferences. Community sharing capability was extended by default to all users of the Eagle file system, including projects that were previously on the Grand file system. The AI Testbed systems were included in the NAIRR Pilot program. The ALCF launched the MyALCF user portal, where users can manage their awards and teams while also receiving various facility updates. Multiple improvements were made to the account and project management software, which was integrated into the user portal.

In 2024, ALCF members hosted over 400 groups of visitors, including many VIPs. In CY 2024, the ALCF conducted 22 training activities (in-person and virtual), reaching more than 1,600 participants in events where participation was tracked. In addition, the ALCF hosted or participated in seven community outreach events, reaching over 1,500 students, both in their classrooms and at Argonne. The ALCF supported several summer coding camps, student programs, and STEM events. The ALCF utilized its website, GitHub, and YouTube as key platforms for sharing information such as onboarding guides and video tutorials. Transitioning to FedRAMP-compliant tools, the ALCF adopted GovDelivery for its newsletters. In 2024, the ALCF published 58 science stories, achieving 2,969 unique media hits and a 6.9 billion audience reach, tracked via Meltwater. Shifting focus from Twitter to LinkedIn, the ALCF significantly grew its LinkedIn audience and adopted Instagram for visual content. The facility maintained diverse communication channels, including newsletters and social media, to engage its audience and highlight research achievements.

Section 2. Operational Performance

Did the facility's operational performance meet established targets?

ALCF Response

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

	Polaris (HPE Apollo 6500 Gen 10+) 560-node, 17920-core 2240 NVIDIA A100 GPU 287 TB DDR4 90 TB HBM2			
	CY 2023		CY 2024	
	Target	Actual	Target	Actual
Scheduled Availability	90.0%	99.2%	90.0%	99.6%
Overall Availability	90.0%	93.8%	90.0%	97.7%
System MTTI	N/A [a]	13.72 days	N/A	29.83 days
System MTTF	N/A	91.03 days	N/A	121.92 days
INCITE Usage	2000.0K	1764.3K	2500.0K	2628.7K
Total Usage	N/A	3680.7K	N/A	4196.6K
System Utilization	N/A	81.1%	N/A	94.7%
INCITE Overall Capability [b]	20.0%	56.7%	20.0%	61.4%
INCITE High Capability [c]	N/A	13.7%	N/A	11.3%

[a] N/A = not applicable.

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

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

Table 2.2 Summary of Operational Performance of HPC Storage Systems

Agile File System (09/10/2024 through 12/31/2024) DDN AI400X with 243TB of capacity				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
Scheduled Availability	N/A	N/A	90.0%	100.0%
Overall Availability	N/A	N/A	90.0%	99.7%
System MTTI	N/A	N/A	N/A	56.32 days
System MTTF	N/A	N/A	N/A	113.00 days
Swift File System (01/01/2024 through 09/10/2024) DDN AI400X with 120TB of capacity				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
Scheduled Availability	N/A	N/A	90.0%	100.0%
Overall Availability	N/A	N/A	90.0%	97.4%
System MTTI	N/A	N/A	N/A	27.29 days
System MTTF	N/A	N/A	N/A	253.00 days
Grand File System (01/01/2024 through 02/07/2024) HPE ClusterStor E1000 with 100 PB of storage at 650 GB/s				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
Scheduled Availability	90.0%	99.5%	90.0%	100.0%
Overall Availability	90.0%	93.2%	90.0%	97.5%
System MTTI	N/A	13.08 days	N/A	12.02 days
System MTTF	N/A	72.68 days	N/A	37.00 days
Eagle File System HPE ClusterStor E1000 with 100 PB of storage at 650 GB/s				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
Scheduled Availability	90.0%	99.9%	90.0%	100.0%
Overall Availability	90.0%	94.2%	90.0%	97.9%
System MTTI	N/A	14.94 days	N/A	35.82 days
System MTTF	N/A	182.30 days	N/A	366.00 days
HPSS Archive LTO8 tape drives and tape with 305 PB of storage capacity				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
Scheduled Availability	90.0%	99.6%	90.0%	100.0%
Overall Availability	90.0%	93.8%	90.0%	97.8%
System MTTI	N/A	12.69 days	N/A	27.52 days
System MTTF	N/A	60.58 days	N/A	366.00 days

[a] N/A = not applicable.

2.1 ALCF Production Resources Overview

During CY 2024, 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 during CY 2024.
- Agile File System: DDN AI400X with 243TB of capacity. It is mounted on Polaris and replaced the Swift file system on September 10, 2024.
- Swift File System: DDN AI400X with 120TB of capacity. It was mounted on Polaris through September 9, 2024, and was replaced by the Agile File System.
- Grand and Eagle are both 100PB Lustre file systems located on an HPE ClusterStor E1000 platform. Previously, they were mounted facility wide. However, to support Aurora during the bring-up and acceptance process, the ALCF migrated all user data from Grand to Eagle, after which the Grand hardware was exclusively mounted as Flare on Aurora. Eagle still has Globus sharing enabled as it did in the past, but Flare does not.
- 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 (for emergency fixes). Users will be notified of regularly scheduled maintenance in advance, with 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 past 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 effectiveness 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 Interrupts).
- *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 computing nodes were running 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, the 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 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				
Years	CY 2023		CY 2024	
Metrics	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	90.0	99.2	90.0	99.6
Overall Availability	90.0	93.8	90.0	97.7

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, 2024, through December 31, 2024, as shown in Figure 2.1.

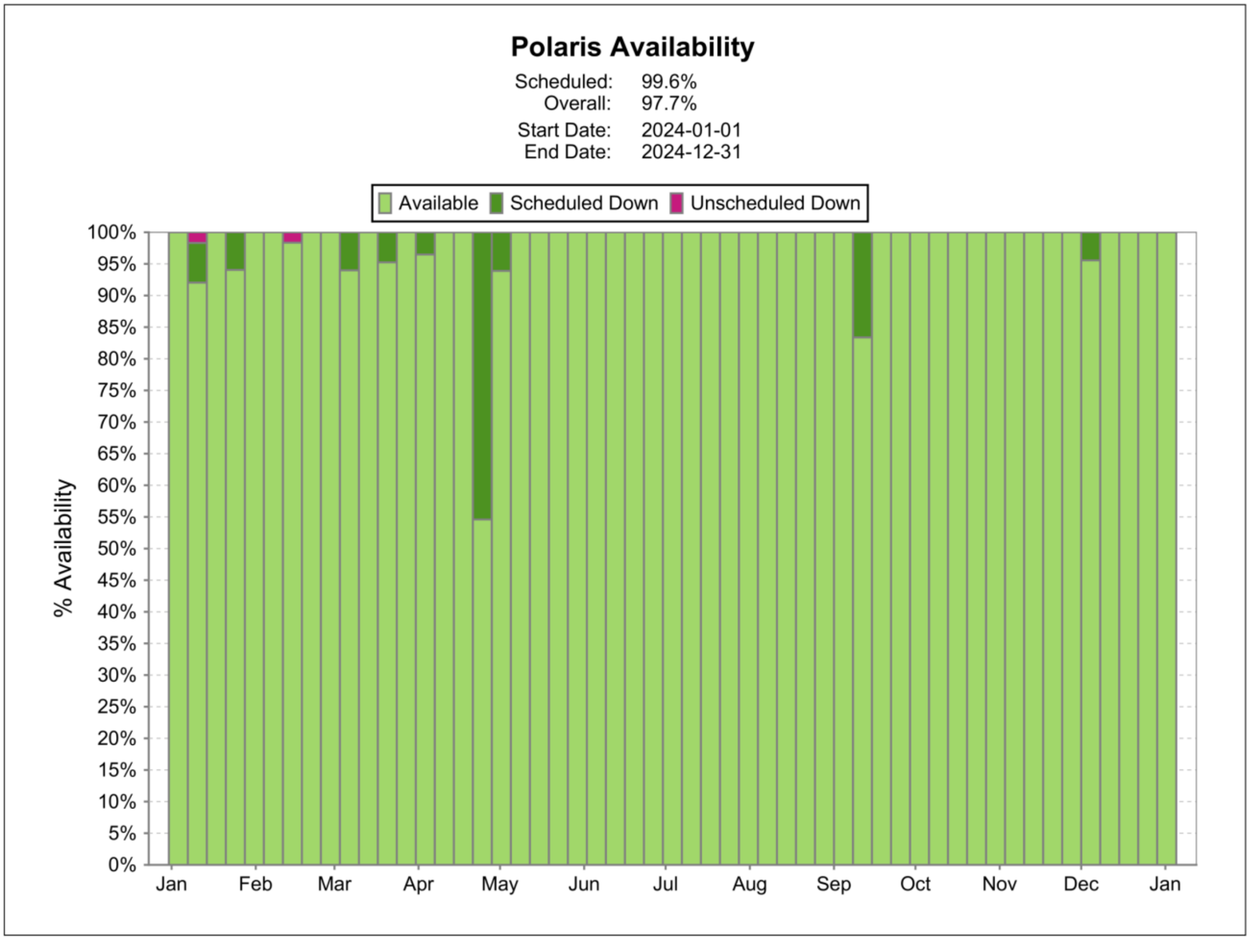


Figure 2.1 Polaris Weekly Availability for CY 2024

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.

January 9, 2024: Unscheduled outage – Image Issue
Polaris was shut down to address recent software image issues. Jobs were failing to load Conda modules.

February 12, 2024: Unscheduled outage – Grand File System removal
An image update was needed to remove the Grand File System from Polaris.

April 23, 2024 – Apr 25, 2024: Scheduled outage – HPE Performance Cluster Manager (HPCM) Upgrade
The management software on Polaris was successfully upgraded to HPCM 1.10.

September 9, 2024 – September 10, 2024: Scheduled outage – Swift to Agile
During the scheduled maintenance the Agile File System replaced the Swift File System on Polaris.

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				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
System MTTI	N/A	13.72 days	N/A	29.83 days
System MTTF	N/A	91.03 days	N/A	121.92 days

[a] N/A = not applicable.

Polaris has occasional maintenance outages scheduled to perform upgrades, hardware replacements, OS upgrades, etc. The 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.

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 the 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				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
System Utilization	N/A	81.1%	N/A	94.7%

[a] N/A = not applicable.

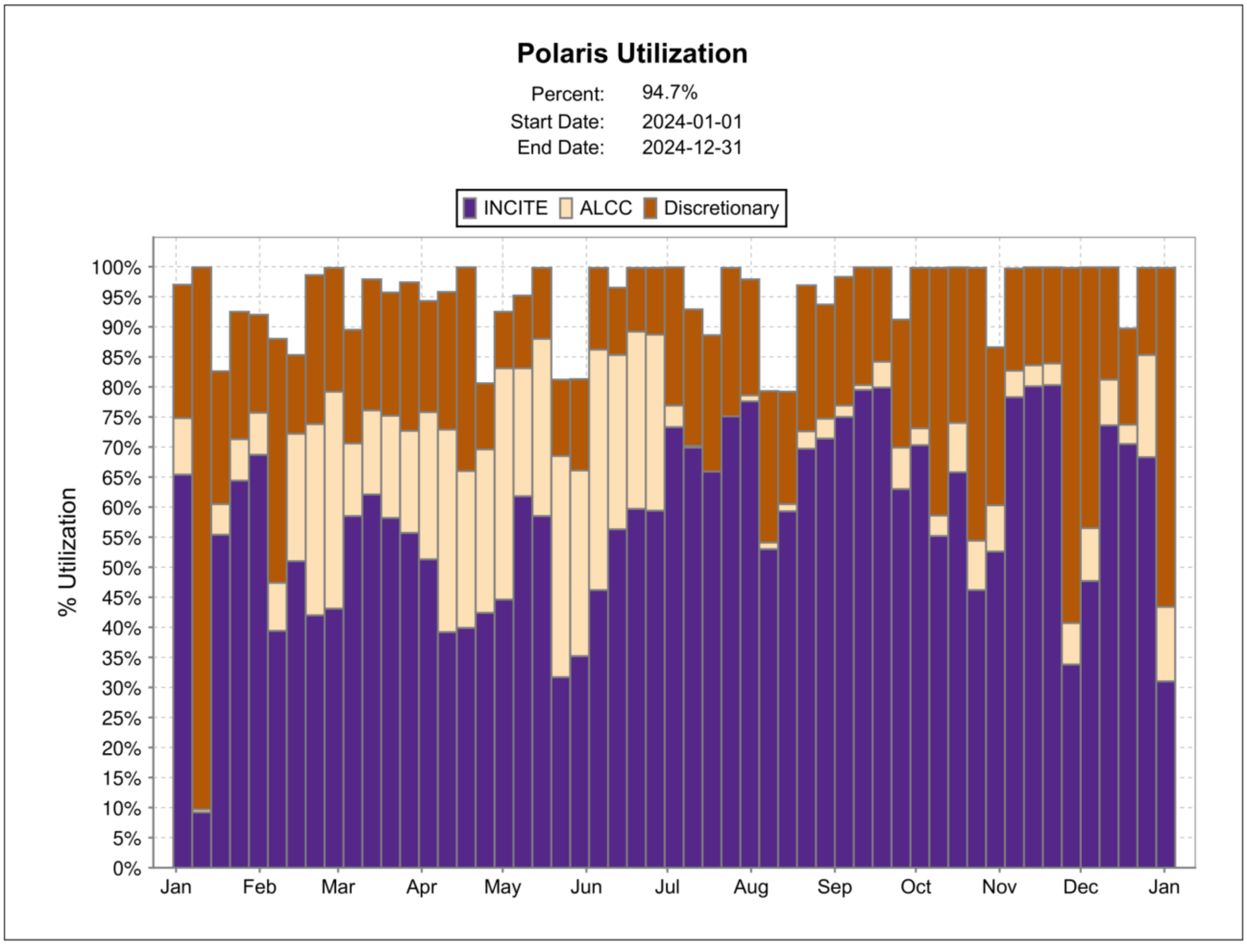


Figure 2.2 Polaris System Utilization over Time by Program

The system utilization for Polaris was 94.7 percent for its CY 2024 production period of January 1, 2024, through December 31, 2024.

Table 2.6 shows how Polaris's system hours were allocated and used by allocation program. The INCITE program awarded 2500.0K node-hours to 17 projects on Polaris in 2024. The ALCF augmented this with an additional 900.0K node-hours over the course of calendar year 2024 to a subset of those projects to ensure that they had sufficient computational time to achieve their science milestones and to ensure that time delivered to INCITE met the target. The DD allocation is greater than the actual node-hours used. Most DD projects are exploratory investigations, so the time allocated is often not used in full. DD allocations are discussed in detail in Section 3.2.

In CY 2024, Polaris delivered a total of 4,196,561 node-hours. The Used node-hours in Table 2.6 for each award program is the total node-hours used for the calendar year. The ALCC Used node-hours includes ALCC 2023 and ALCC 2024 for jobs that ran during CY 2024.

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						
Years	CY 2023			CY 2024		
Metrics	Allocated	Used		Allocated	Used	
Allocation Programs	Node-hours	Node-hours	% of Total Used	Node-hours	Node-hours	% of Total Used
INCITE	2000.0K	1764.3K	47.9	2500.0K	2628.7K [a]	62.6
ALCC	943.4K	724.3K	19.7	1304.1K	601.6K	14.4
DD	2895.8K	1192.0K	32.4	2855.0K	966.2K	23.0
Total	5839.2K	3680.7K	100.0	6659.1K	4196.6K	100.0

[a] The INCITE Used node-hours includes jobs that ran in CY 2024. This includes hours used by INCITE 2023 projects that received a 13th month extension into January 2024.

Summary: For CY 2024, the system usage for INCITE exceeded targets. System utilization values aligned with expectations. The calculation for system utilization is described in Appendix A.

2.3.3.2 Capability Utilization

Polaris has a total of 560 nodes, of which 56 were purchased for use as on-demand nodes (used for Integrated Research Infrastructure (IRI) activities, see Section 8.4), and 8 are dedicated to debugging. Therefore, ALCF used a count of 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 details on the capability calculation. Table 2.7 and Figure 2.4 show that the ALCF has exceeded 20 percent capability usage 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 17920-core 2240 NVIDIA A100 GPU 287 TB DDR4 90 TB HBM2						
Metrics	CY 2023			CY 2024		
Capability Usage	Total Hours	Capability Hours	Percent Capability	Total Hours	Capability Hours	Percent Capability
INCITE Overall [a]	1764.3K	999.6K	56.7	2628.7K	1613.6K	61.4
INCITE High [b]	1764.3K	241.4K	13.7	2628.7K	295.9K	11.3
ALCC Overall	724.3K	443.6K	61.2	601.6K	327.9K	54.5
ALCC High	724.3K	275.3K	38.0	601.6K	77.9K	13.0
Director's Discretionary Overall	1192.0K	493.1K	41.4	966.2K	365.1K	37.8
Director's Discretionary High	1192.0K	118.2K	9.9	966.2K	119.5K	12.4
TOTAL Overall	3680.7K	1936.3K	52.6	4196.6K	2306.6K	55.0
TOTAL High	3680.7K	634.9K	17.2	4196.6K	493.4K	11.8

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

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

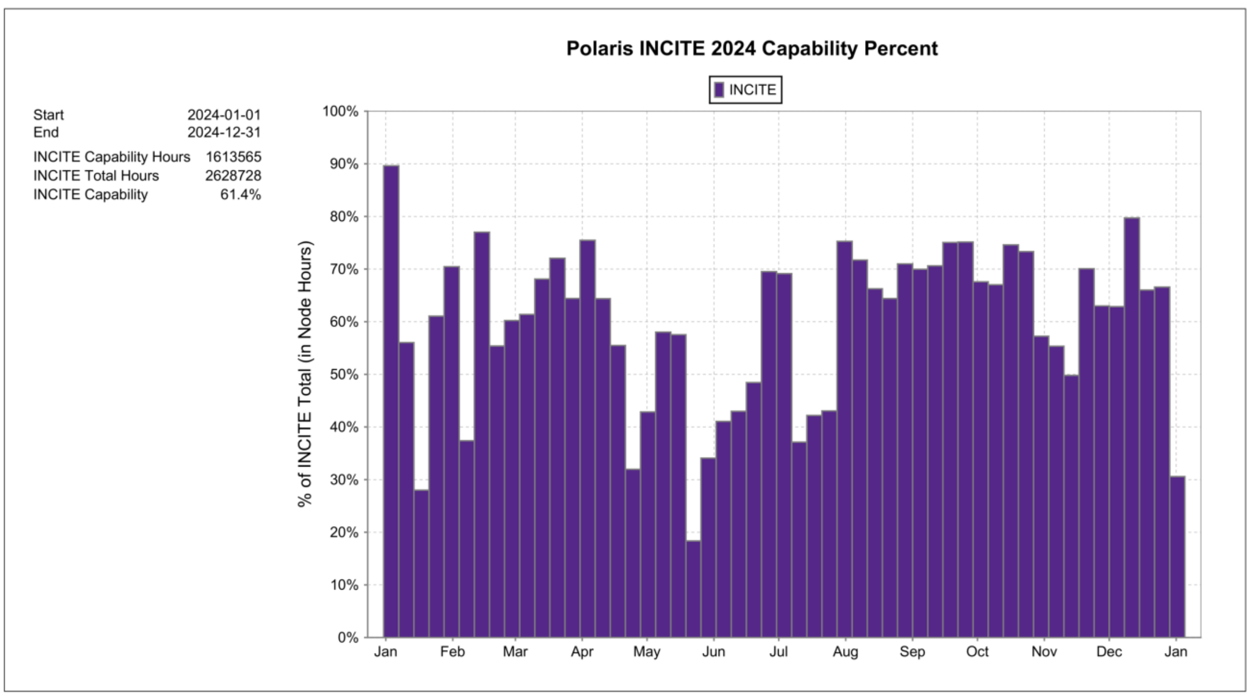


Figure 2.3 Polaris INCITE Overall Capability

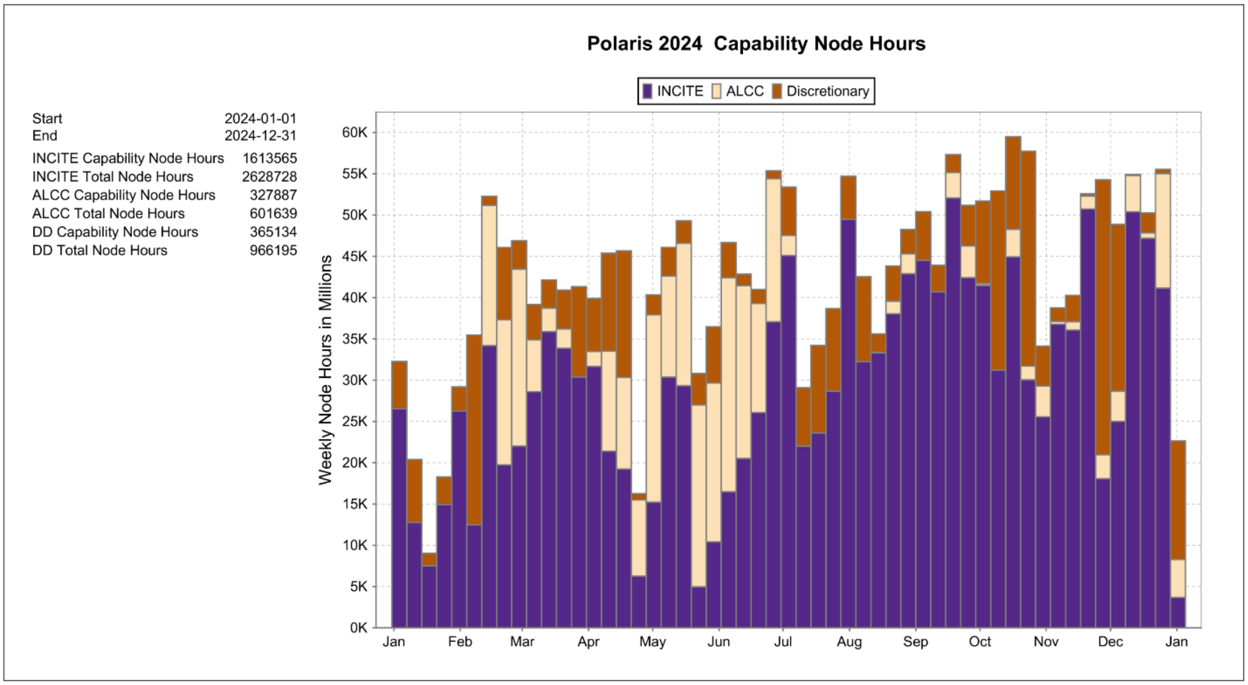


Figure 2.4 Polaris Capability Node-Hours by Program

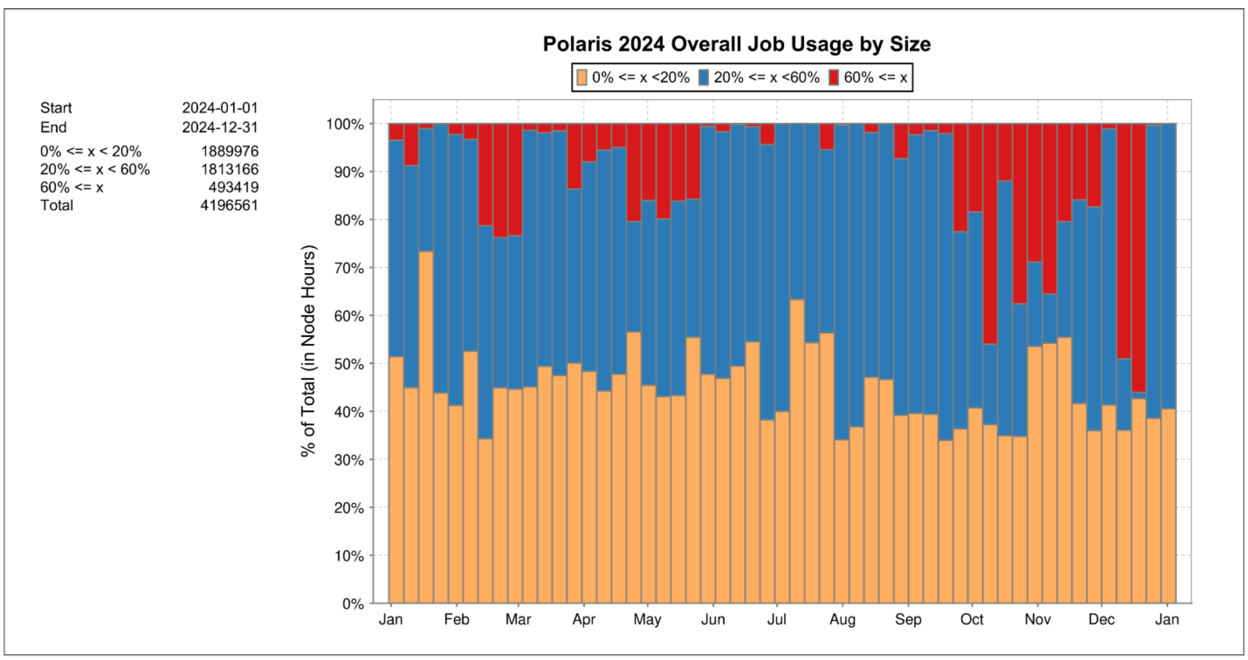


Figure 2.5 Polaris Job Usage by Size

2.4 Storage

This section covers availability, MTTI metrics, and MTTF metrics for production storage resources.

2.4.1 Agile File System

Agile is a DDN AI400X with 243TB of capacity. It is mounted by Polaris. It replaced the Swift File System. Agile entered production on September 10, 2024.

2.4.1.1 Scheduled and Overall Availability

The ALCF uses the target of 90 percent overall availability and 90 percent scheduled availability. 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 2.8 summarizes the availability results.

Table 2.8 Availability Results

Agile File System DDN AI400X with 243TB of capacity		
Year	CY 2024	
Metrics	Target (%)	Actual (%)
Scheduled Availability	90.0	100.0
Overall Availability	90.0	99.7

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.

Agile – Explanation of Significant Availability Losses

There were no significant outages of the Agile File System for the period of September 10, 2024, through December 31, 2024, as annotated in Figure 2.6.

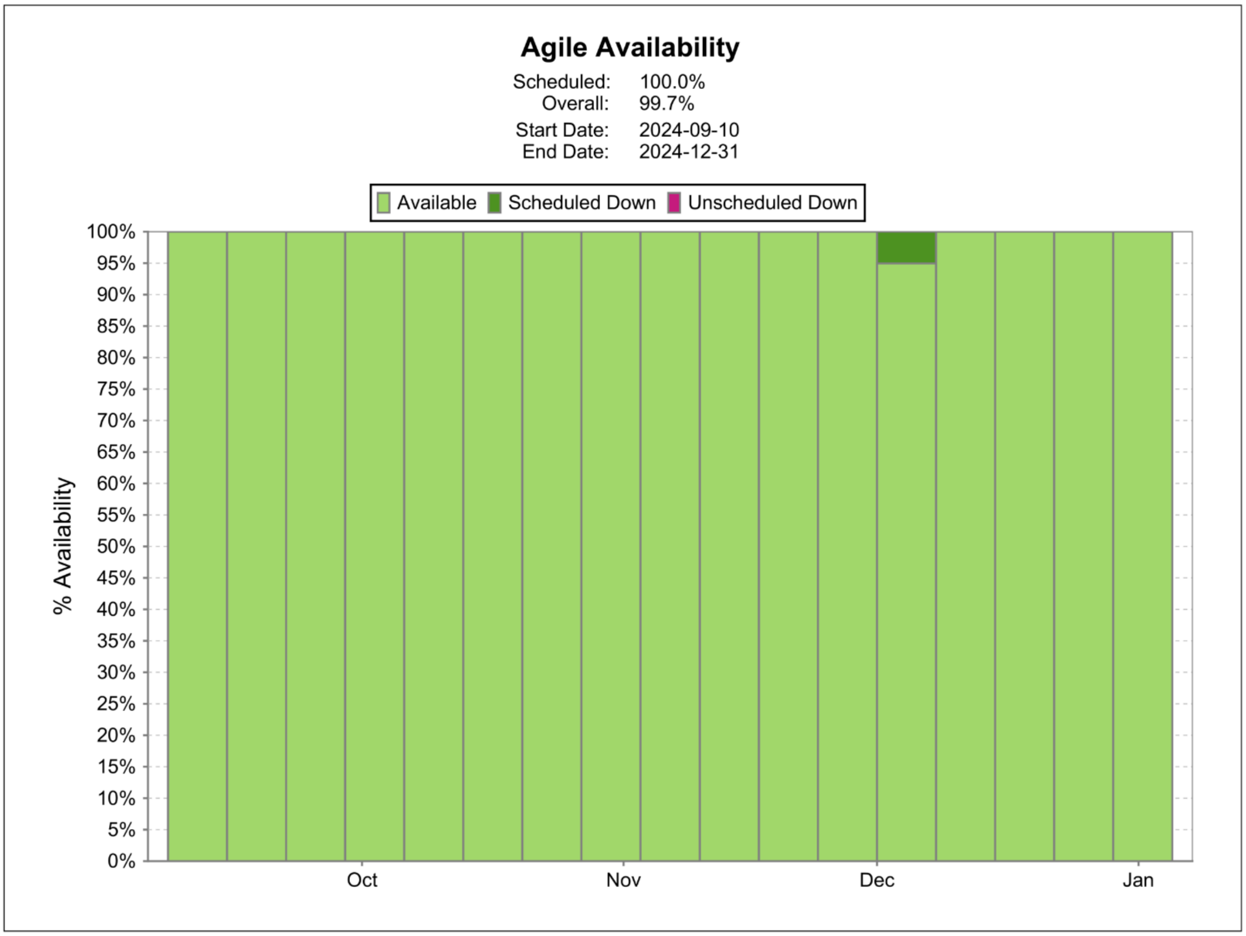


Figure 2.6 Agile Weekly Availability for CY 2024

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.

2.4.1.2 MTTI and MTTF

Agile – MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.9 summarizes the current MTTI and MTTF values.

Table 2.9 MTTI and MTTF Results - Agile

Agile File System DDN AI400X with 243TB of capacity		
Year	CY 2024	
Metrics	Target	Actual
System MTTI	N/A [a]	56.32 days
System MTTF	N/A	113.00 days

[a] N/A = not applicable.

The Agile File System currently follows the Polaris maintenance schedule. The PMs are often used to apply upgrades and patches.

2.4.2 Swift File System

Swift is a DDN AI400X with 120TB of capacity. It was mounted by Polaris through September 9, 2024, and was replaced by the Agile File System on September 10, 2024. The ALCF began tracking Swift's availability on January 1, 2024.

2.4.2.1 Scheduled and Overall Availability

The ALCF uses the target of 90 percent overall availability and 90 percent scheduled availability. 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 2.10 summarizes the availability results.

Table 2.10 Availability Results - Swift

Swift File System DDN AI400X with 120TB of capacity		
Year	CY 2024	CY 2024
Metrics	Target (%)	Actual (%)
Scheduled Availability	90.0	100.0
Overall Availability	90.0	97.4

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.

Swift – Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability of the Swift File System for the period of January 1, 2024, through September 9, 2024, as annotated in Figure 2.7.

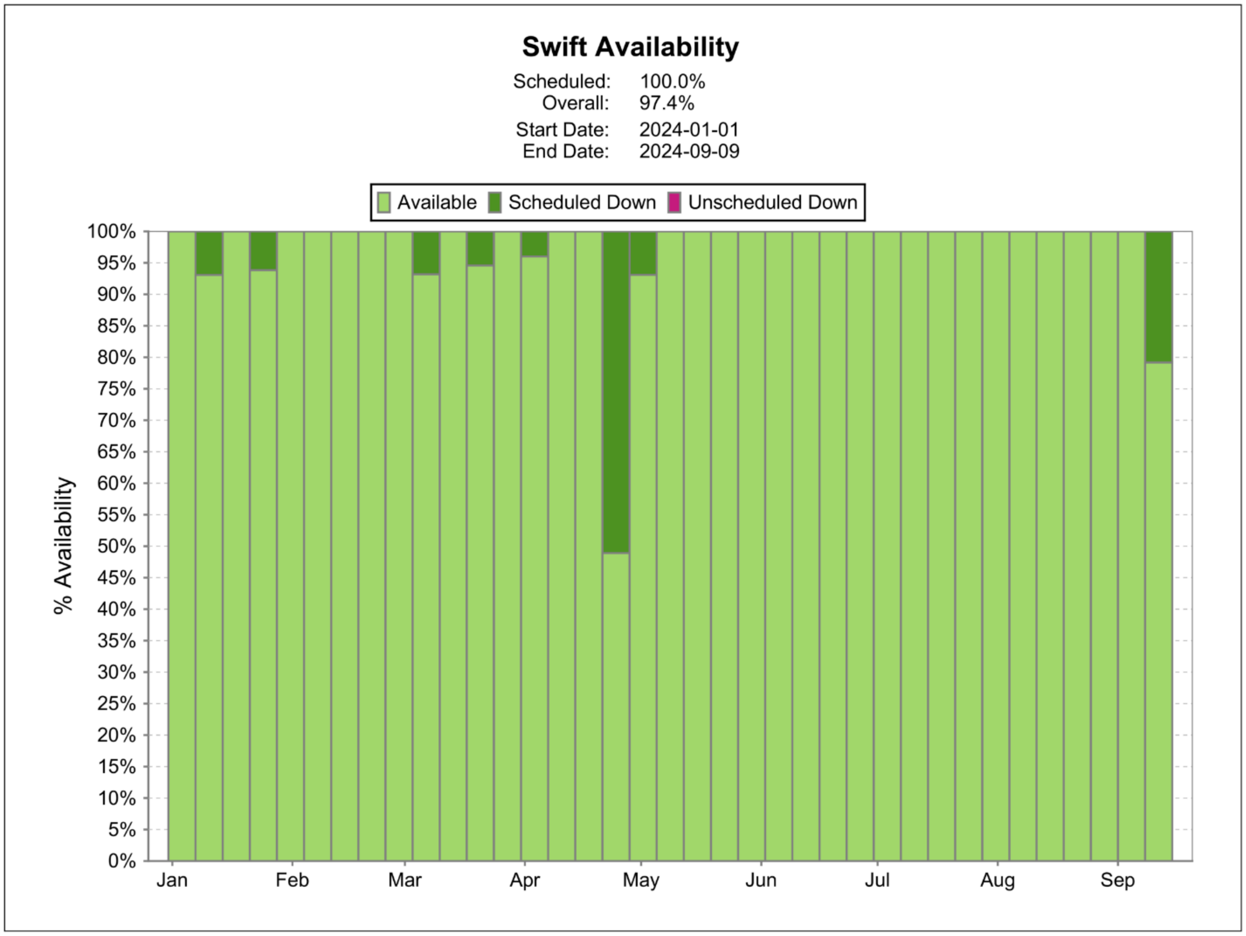


Figure 2.7 Swift Weekly Availability for CY 2024

Graph Description: Each bar in Figure 2.7 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.

April 23, 2024 – April 25, 2024: Scheduled outage – HPCM Upgrade
The management software on Polaris was successfully upgraded to HPCM 1.10. During this time, the Swift File System was unavailable.

September 9, 2024 – September 10, 2024: Scheduled outage – Swift to Agile
During the scheduled maintenance, the Agile File System replaced the Swift File System on Polaris.

2.4.2.2 MTTI and MTTF

Swift – MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.11 summarizes the current MTTI and MTTF values.

Table 2.11 MTTI and MTTF Results - Swift

Swift File System DDN AI400X with 243TB of capacity		
Year	CY 2024	
Metrics	**Target**	**Actual**
System MTTI	N/A	27.29 days
System MTTF	N/A	253.00 days

[a] N/A = not applicable.

The Swift File System currently follows the Polaris maintenance schedule. PMs are often used to apply upgrades and patches.

2.4.3 Grand and Eagle Lustre File Systems

The ALCF installed a new set of Lustre file systems in 2020 based on HPE ClusterStor E1000 storage solutions. Grand and Eagle each offer 100 PB of storage at 650 GB/s and provides availability protection if one fails. Additionally, the file systems have the capability of sharing via Globus, which is a move towards providing a community file system. The file systems went into production in January 2021. The Grand file system changed over to the Flare file system on February 7, 2024. The Flare file system became an Aurora file system and was not reportable in 2024. It will be reportable in 2025.

ALCF uses the target of 90 percent overall availability and 90 percent scheduled availability. 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.

2.4.3.1 Grand Scheduled and Overall Availability

Table 2.12 summarizes the availability results for the Grand File System.

Table 2.12 Availability Results - Grand

Grand File System HPE ClusterStor E1000 with 100 PB of storage at 650 GB/s				
Years	CY 2023		CY 2024	
Metrics	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	90.0	99.5	90.0	100.0
Overall Availability	90.0	93.2	90.0	97.5

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

There were no significant outages for the period of January 1, 2024, through February 7, 2024, as annotated in Figure 2.8.

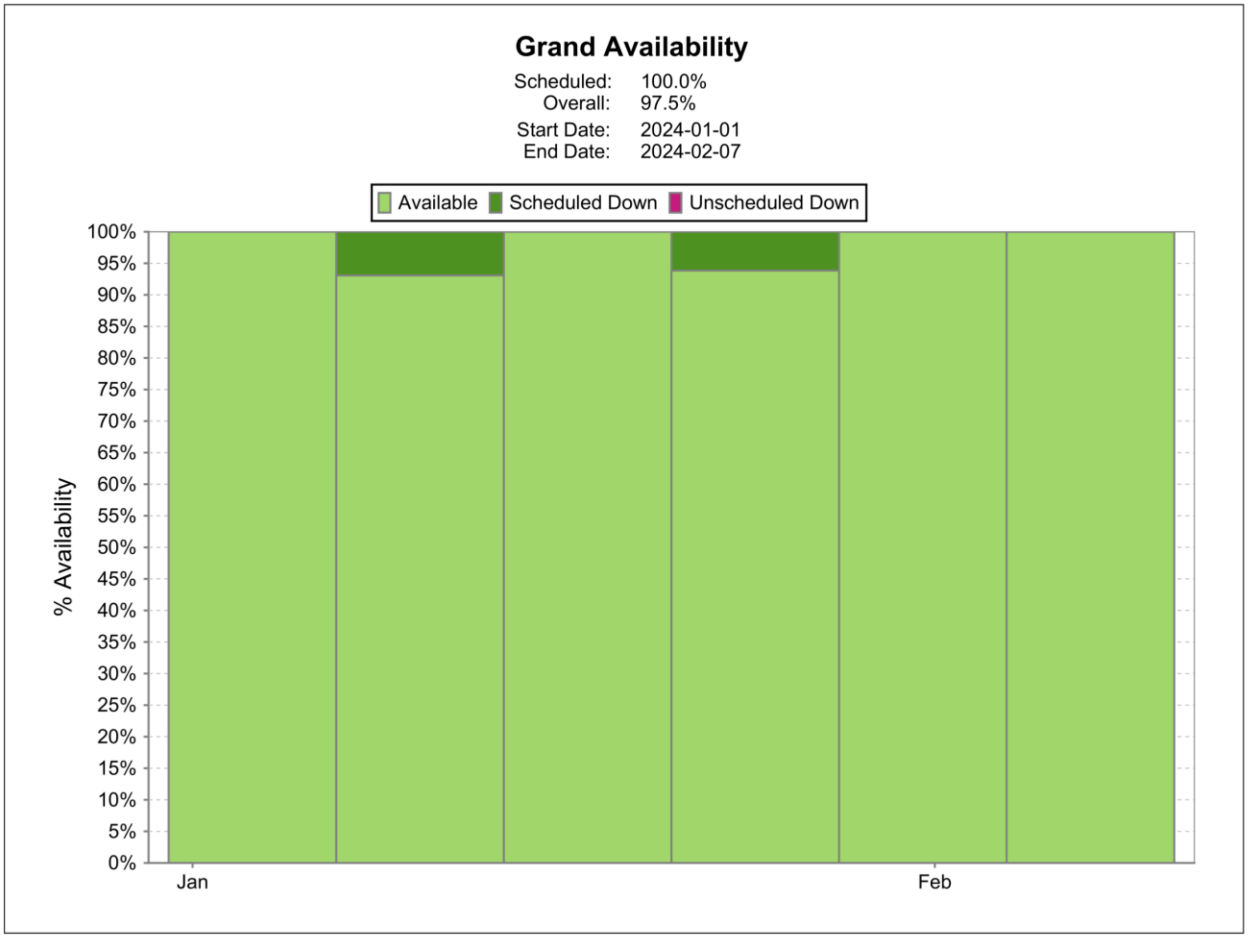


Figure 2.8 Grand Weekly Availability for CY 2024

Graph Description: Each bar in Figure 2.8 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.

Grand – MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.13 summarizes the current MTTI and MTTF values.

Table 2.13 MTTI and MTTF Results - Grand

Grand File System HPE ClusterStor E1000 with 100 PB of storage at 650 GB/s				
	CY 2023		CY 2024	
	Target	Actual	Target	Actual
System MTTI	N/A [a]	13.08 days	N/A	12.02 days
System MTTF	N/A	72.68 days	N/A	37.00 days

[a] N/A = not applicable.

Grand generally follows Polaris's maintenance schedule. Maintenance windows are often used to apply upgrades and patches.

2.4.3.2 Eagle – Scheduled and Overall Availability

Table 2.14 summarizes the availability results for the Eagle File System.

Table 2.14 Availability Results - Eagle

Eagle File System HPE ClusterStor E1000 with 100 PB of storage at 650 GB/s				
Years	CY 2023		CY 2024	
Metrics	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	90.0	99.9	90.0	100.0
Overall Availability	90.0	94.2	90.0	97.9

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.

Eagle – Explanation of Significant Availability Losses

This section briefly describes the causes of major losses in Eagle's availability from January 1, 2024, through December 31, 2024, as shown in Figure 2.9.

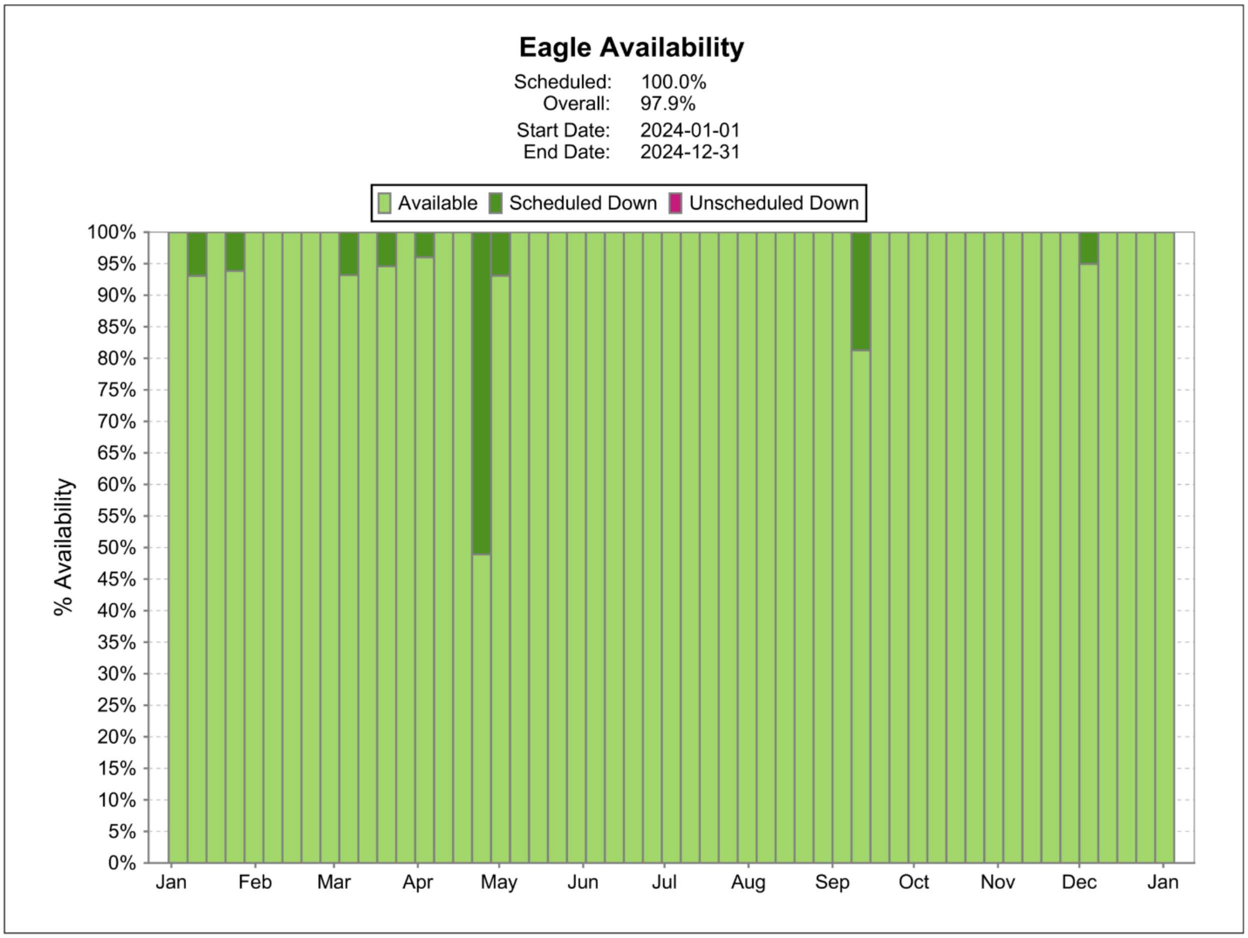


Figure 2.9 Eagle Weekly Availability for CY 2024

Graph Description: Each bar in Figure 2.9 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.

April 23, 2024 – April 25, 2024: Scheduled outage – HPCM Upgrade
The management software on Polaris was successfully upgraded to HPCM 1.10.

September 9, 2024 – September 10, 2024: Scheduled outage – Swift to Agile
During scheduled maintenance, the Agile File System replaced Swift on Polaris.

Eagle – MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.15 summarizes the current MTTI and MTTF values.

Table 2.15 MTTI and MTTF Results - Eagle

Eagle File System HPE ClusterStor E1000 with 100 PB of storage at 650 GB/s				
	CY 2023		CY 2024	
	Target	Actual	Target	Actual
System MTTI	N/A	14.94 days	N/A	35.82 days
System MTTF	N/A	182.30 days	N/A	366.00 days

Eagle generally followed Polaris's maintenance schedule. Maintenance windows were often used to apply upgrades and patches.

2.4.4 Tape Storage

The facility-wide HPSS (high-performance storage system) tape archive was available to all ALCF users on all compute resources in CY 2024, as 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 at Argonne 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.

2.4.4.1 Scheduled and Overall Availability

ALCF uses the target metrics of 90 percent overall availability and 90 percent scheduled availability. Table 2.16 summarizes the availability results.

Table 2.16 Availability Results - HPSS

HPSS Archive LTO8 tape drives and tape with 305 PB storage capacity				
	CY 2023		CY 2024	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	90.0	99.6	90.0	100.0
Overall Availability	90.0	93.8	90.0	97.8

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

HPSS – Explanation of Significant Availability Losses

This section briefly describes the causes of major availability losses of HPSS for the period of January 1, 2024, through December 31, 2024, as shown in Figure 2.10.

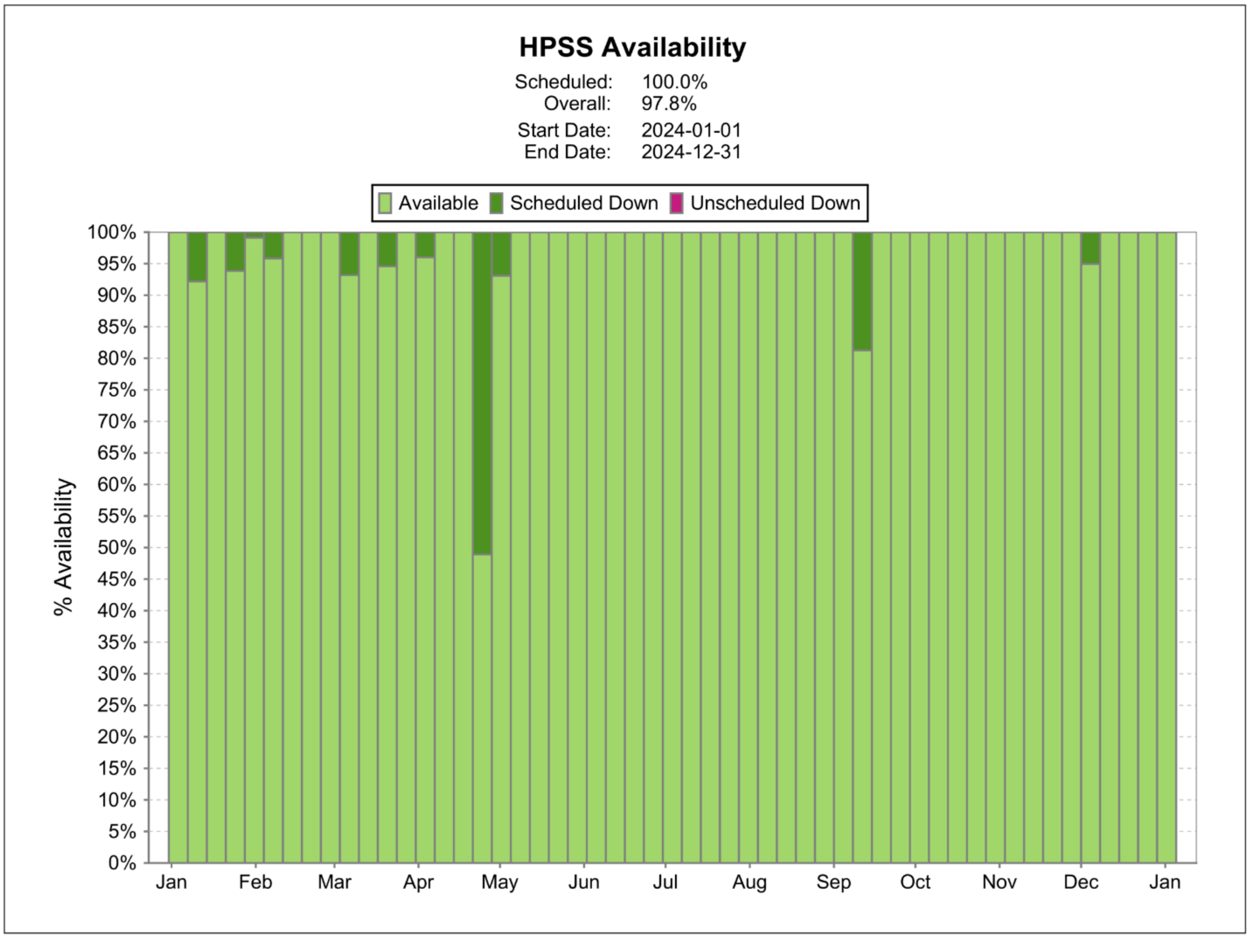


Figure 2.10 HPSS Weekly Availability for CY 2024

April 23, 2024 – April 25, 2024: Scheduled outage – HPCM Upgrade
The management software on Polaris was successfully upgraded to HPCM 1.10.

September 9, 2024 – September 10, 2024: Scheduled outage – Swift to Agile
During scheduled maintenance, the Agile File System replaced Swift on Polaris.

2.4.4.2 MTTI and MTTF

HPSS – MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.17 summarizes the current MTTI and MTTF values.

Table 2.17 MTTI and MTTF Results - HPSS

HPSS Archive LTO8 tape drives and tape with 305 PB storage capacity				
Years	CY 2023		CY 2024	
Metrics	Target	Actual	Target	Actual
System MTTI	N/A	12.69 days	N/A	27.52 days
System MTTF	N/A	60.58 days	N/A	366.00 days

[a] N/A = not applicable.

HPSS maintenance is infrequent, but is typically aligned with Polaris's maintenance schedule. HPSS is often available even though other resources may be in preventative maintenance.

2.5 Center-Wide Operational Highlights

2.5.1 Polaris / Sirius

In CY 2024, Polaris was the ALCF workhorse for delivering compute cycles for science. Sirius functioned as the test and development system (TDS) for Polaris and was accessible only to ALCF staff. Polaris delivered nearly 4.2M node-hours to science, achieving an overall availability of 97.7% and a scheduled availability of 99.6%.

The ALCF Operations Team has been implementing a series of infrastructure and process improvements to facilitate more "hot maintenance" and minimize the number of scheduled outages on Polaris. These efforts came to fruition in 2024, allowing the ALCF to go four months without any maintenance on Polaris. Moving forward, the goal is to perform only three scheduled PMs per year.

The ALCF has been collaborating with several scientific facilities to enable "on-demand" computing. The Advanced Photon Source (APS) was the first to start using Polaris in its production runs. The 8-ID-I beamline now regularly uses Polaris, and the ALCF anticipates that more beamlines and other facilities will begin adopting Polaris for on-demand computing.

2.5.2 AI Testbed

The AI Testbed remains highly useful and popular among users and is available through the DD allocation program. In 2024, there were no major hardware expansions; however, numerous software upgrades were implemented, and workshops were held. Additionally, resources in the AI Testbed were made available via the National Artificial Intelligence Research Resource (NAIRR) Pilot.

2.5.3 Aurora / Sunspot

Getting Aurora through acceptance testing and ready for users was a major focus for the Operations Team in 2024. The ALCF successfully completed acceptance testing in December. To manage the acceptance testing, the ALCF utilized a tool called ReFrame to implement the tests, along with Jenkins, which was used for continuous integration and continuous delivery

(CI/CD). The ALCF introduced a new feature that involved using the GitLab application programming interface (API) to create an issue for every failed test, including information on how to rerun the test, which nodes it ran on, and the locations of the various aggregated logs. This improvement made debugging and tracking the root causes of failures much easier, and labels on the issues allowed aggregation by root cause. It performed very well, and the ALCF will continue to use this approach in the future.

When working on a system the size of Aurora, seemingly simple tasks can present technical challenges. For example, if the ALCF filtered nothing and passed every log line from every device and service needed to operate Aurora, it would generate well over 1 trillion log lines per day. While the ALCF does filter data at the source, enormous amounts of information still need to be processed. Before Aurora, ETL (Extract, Transform, Load) operations went directly into the ALCF's DB2 database for the business intelligence (BI) system. There are currently over 1 trillion rows in that database. However, with Aurora, everything is funneled through Kafka, which allows easy access to much more detailed information, such as syslog on all the compute nodes. This is not an optimal use of DB2, so the ALCF explored alternatives and implemented a "data lake." This is a file-based system with additional metadata that enables it to be treated similarly to a database in certain respects. In CY 2024, there were about 238 million entries, and the query-like capabilities allow operations to generate "deep job logs" that aggregate all data from all sources for a job into a single directory, facilitating the debugging of job failures.

2.5.4 Crux

The ALCF brought online a relatively modest all-CPU system, at about 1 Petaflop (PF). Initially, this was a test system used to evaluate system management software stacks before Aurora. That evaluation is now complete, so the ALCF has made it available to users who have not yet transitioned to GPUs and for workloads that are not suitable for GPUs.

2.5.5 Sophia

When the ALCF shut down Theta at the end of 2023, the DGX systems known as ThetaGPU still had some useful life left. The ALCF separated the systems, added the required management and ancillary hardware, and then released it back to users in July 2024 with a new hostname, Sophia.

2.5.6 Storage

The storage team had a busy year. The deployment of a Distributed Asynchronous Object Storage (DAOS) system was a significant effort that involved coordination between the development team and system administration, along with the development of tools, processes, and documentation. The ALCF also engaged in a good bit of "musical data," migrating data from one hardware platform to another. The ALCF performed this with the home file system to increase capacity and enable offline upgrades, reducing risk. Additionally, the ALCF migrated all the data in the Grand file system to free up Grand hardware for mounting on Aurora during acceptance. At first glance, one might wonder, "How hard is it to copy a bunch of files?" However, when dealing with roughly 10 billion files, ensuring that soft and hard links are managed correctly, sparse files remain sparse, and sufficient hardware is allocated to complete the transition in an acceptable timeframe, it turns into a very non-trivial task.

2.5.7 Other Facility Notes

The ALCF implemented a significant upgrade to the ALCF web applications, called MyALCF. The ALCF added important functionality, including tools for PIs to visualize their progress on allocations, a web front end for the accounting system, and a tool to generate a command line for those who wish to script their web actions for regular reporting. The ALCF will continue to expand the system's capabilities, aiming to make it the "single pane of glass" for interacting with the ALCF. For more details on MyALCF, refer to Section 4.1.1.

2.5.8 Decommissioned Systems

The ALCF did not retire any systems in 2024.

Conclusion

The ALCF maximized the use of its HPC systems and other resources consistent with its mission. The ALCF has exceeded the metrics of system availability and capability hours delivered. For the reportable areas—MTTI, MTTF, and utilization—the ALCF is on par with the OLCF and NERSC, and the values reported are reasonable. These measures are summarized in Table 2.1.

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 Polaris, the breakdown of time allocated to INCITE, ALCC, and DD is 60 percent, 30 percent, and 10 percent, respectively. 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 the ALCF's mechanisms to promote user allocation utilization, including the 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 in Section 8 show, these are reasonable allocations of resources. Below are a few areas that ALCF considers when analyzing usage statistics for the allocation programs.

3.1 Usage of the INCITE and ALCC Hours

The 2024 INCITE program had a target allocation of 2500.0K node-hours on Polaris, and a total of 2555.7K node-hours were delivered to 17 INCITE 2024 projects on Polaris (Table 3.1). (The ALCF augmented this with an additional 900.0K node-hours over the course of CY 2024 to a subset of those projects to ensure that they had sufficient computational time to meet their science milestones and to ensure that time delivered to INCITE met the target.) Figure 3.1 shows that 6 projects were able to use at least 85 percent of their allocation, including 3 that used more than 100 percent of their initial allocation. Of the other 11 projects, 10 used between 32 and 73 percent of their allocation, and the remaining project used 18 percent of its allocation.

Table 3.1 INCITE 2024 Time Allocated and Used on Polaris

Projects	Polaris
Targeted Node-Hours Allocation	2500.0K
Used Node-Hours [a]	2555.7K

[a] Polaris INCITE usage includes usage from jobs of projects with an extension into January 2025.

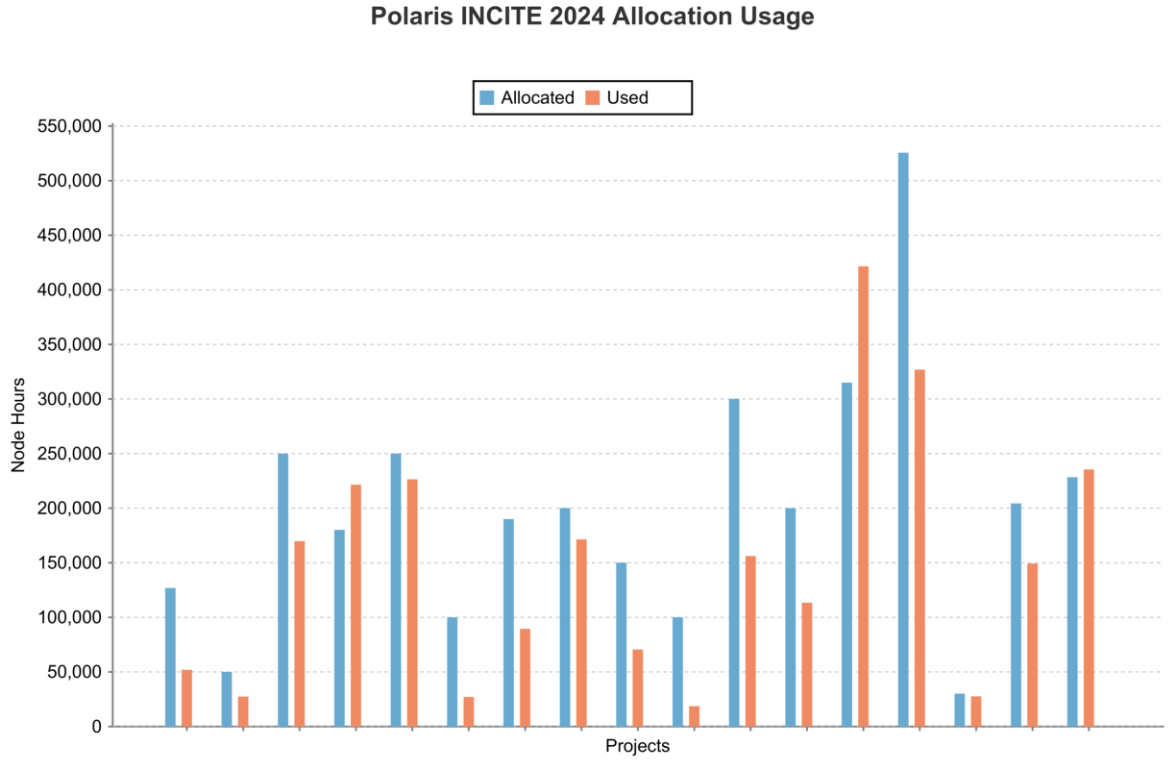


Figure 3.1 Polaris INCITE 2024 Allocation Usage (Note: Projects are randomly ordered.)

For the 2023–2024 ALCC year, 12 projects had allocations on Polaris with a total of 500.5K node-hours used. Figure 3.2 shows that 3 projects used more than 98 percent of their allocation, including 2 that used more than 100 percent of their initial allocation. Of the other 9 projects, 5 used between 22 and 71 percent of their allocation, and the remaining 4 projects used less than 6 percent of their allocation.

The 2024–2025 ALCC year was approximately halfway through its allocation cycle at the end of CY 2024. At that point, 10 projects received allocations totaling 1,085.9K node-hours on Polaris. The 2024–2025 ALCC projects on Polaris used a total of 101.2K node-hours from July 1, 2024, through December 31, 2024, and their allocation usage is shown in Figure 3.3.

The ALCF worked with ASCR to address the issue of underutilization in ALCC projects. Several factors contributed to project underperformance, including distractions from other compute resources, a lack of readiness that could not be resolved during the award period, and changes in project priorities. Although ALCC initiated pullback on larger underutilized projects, this was not reflected in the final allocations reported. Furthermore, subsequent awards did not rectify this discrepancy. One mid-cycle award aimed to increase the amount of an original award that was insufficient for the PI to use effectively. Unfortunately, this time extension came after the PI had moved on. A very large, mid-cycle effort made progress, but it only used 35 percent of the allocated time due to the limited amount of time the allocation was available.

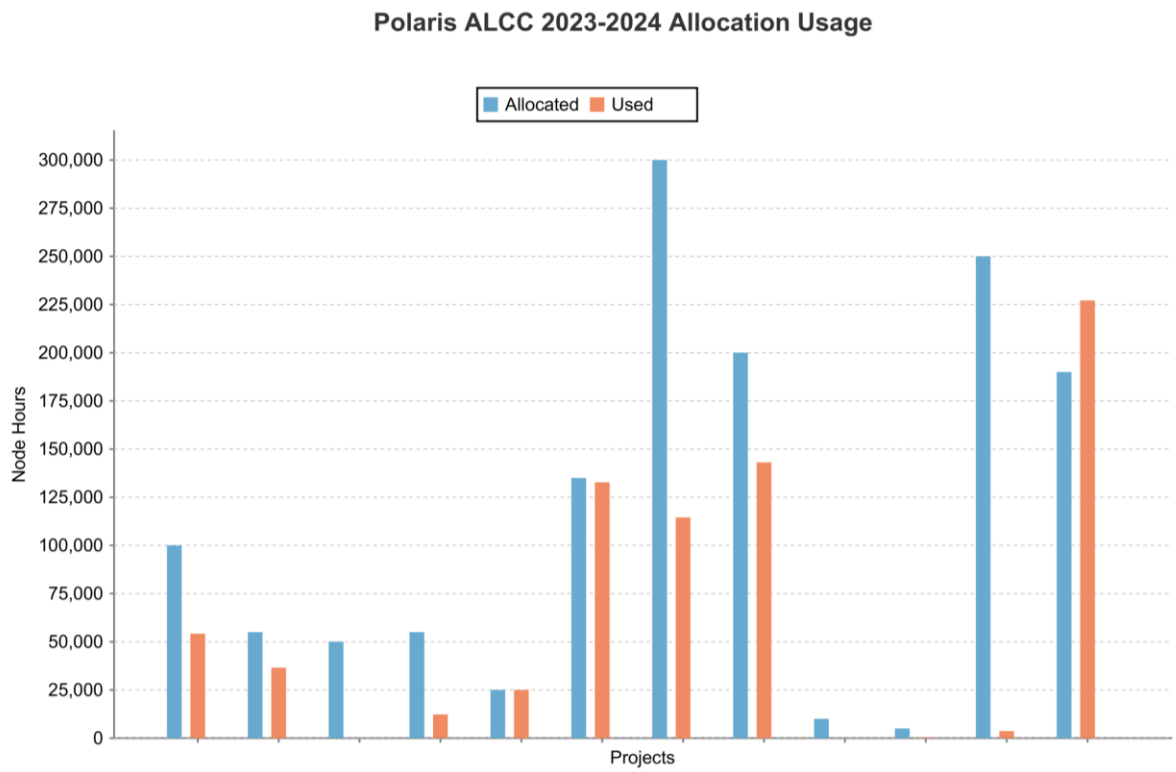


Figure 3.2 Polaris ALCC 2023–2024 Allocation Usage (Note: Projects are randomly ordered.)

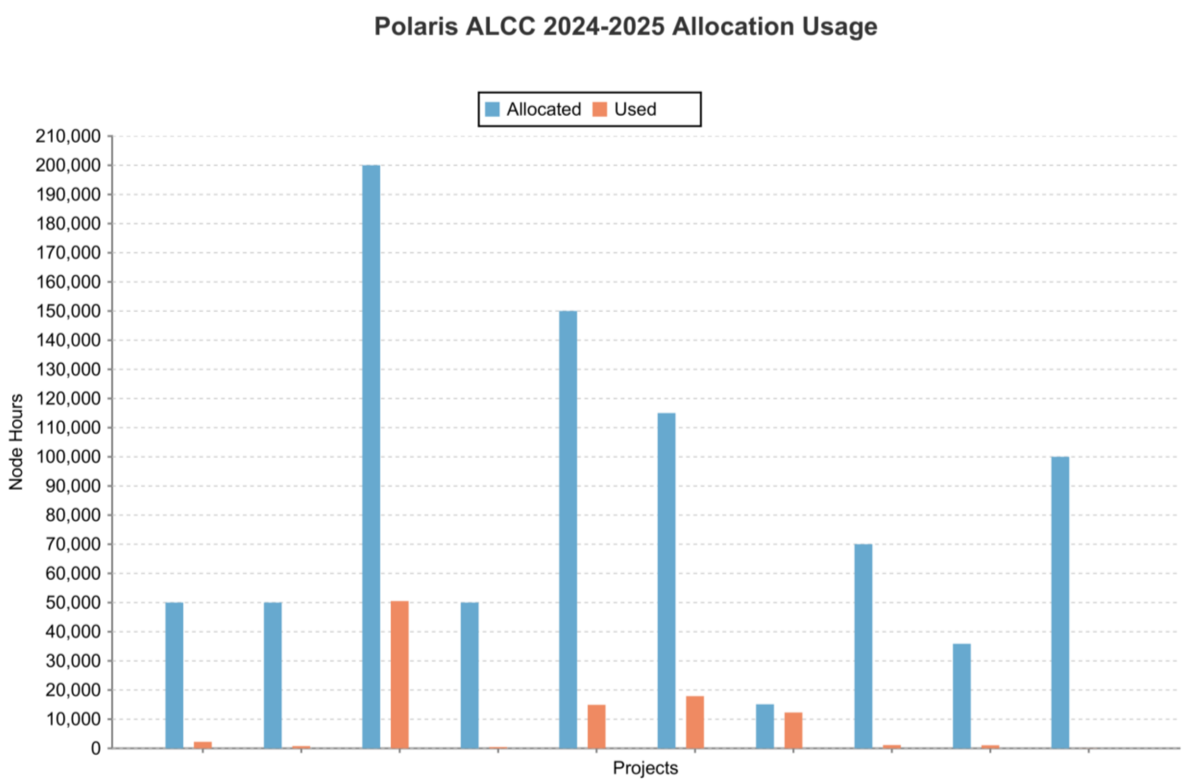


Figure 3.3 Polaris ALCC 2024–2025 Allocation Usage as of December 31, 2024 (Note: Projects are randomly ordered.)

Table 3.2 summarizes the ALCC node-hours allocated and used on Polaris in CY 2024. The 1304.1K ALCC node hours allocated are calculated by adjusting the 2023–2024 and 2024–2025 Polaris ALCC year allocations by the percentage of their award cycle occurring in CY 2024. The total 601.6K ALCC node-hours used is the sum of all node-hours used by any ALCC project on Polaris in CY 2024.

Table 3.2 ALCC Time Allocated and Used on Polaris in CY 2024

Projects	Polaris
Allocated Node-Hours	1304.1K
Used Node-Hours	601.6K

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) Integrated Research Infrastructure (IRI) support

INCITE and ALCC proposal preparation allocations are offered to 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 like the Gordon Bell Prize. Projects in this category have been responsible for bringing new capabilities to the ALCF.

The 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. The 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.

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 team). 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.

Table 3.3 shows the total time allocated and used in the DD program on Polaris during 2024. By its very nature, the DD program is amenable to overallocation since time often remains 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 certain internal projects that facilitate the acceptance of new hardware or assist users, aligning with the ALCF's core mission.

Table 3.3 DD Time Allocated and Used on Polaris in CY 2024

Projects	Polaris
Allocated Node-Hours	2855.0K
Used Node-Hours	966.2K

Lists of the CY 2024 DD projects on Polaris are provided in Appendix B. Figure 3.4 offers breakdowns for the CY 2024 DD allocations by domain for Polaris.

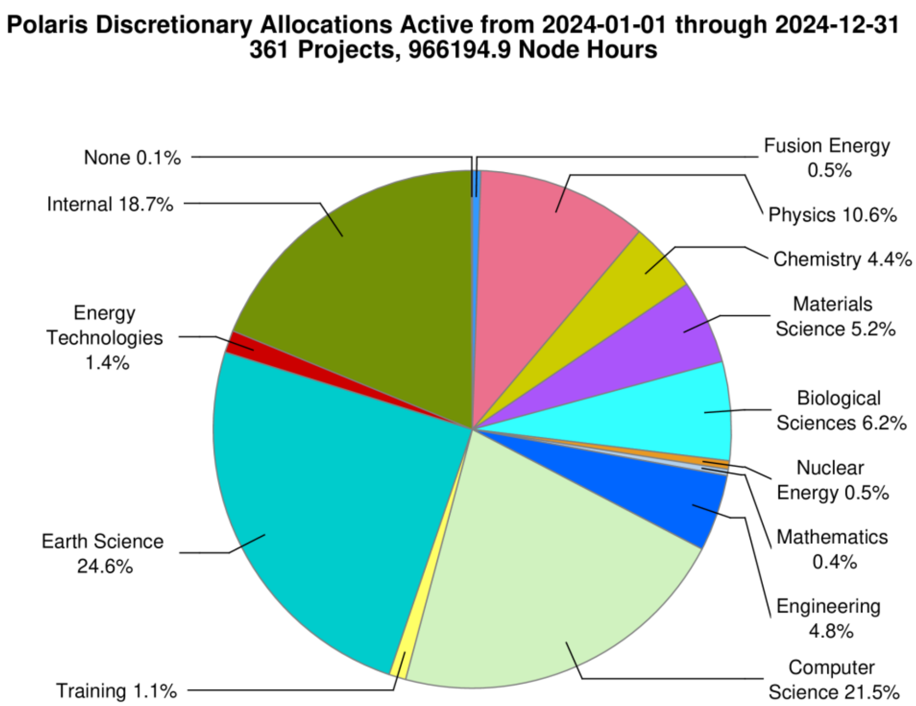


Figure 3.4 Polaris CY 2024 DD Allocations by Domain

3.3 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 brief description of what the project intends to achieve with the extra hours, emphasizing specific goals or milestones and the computing time required to accomplish them. The scheduling committee, allocations committee, and ALCF management will review these requests. Non-capability jobs from projects that have exhausted their allocation will be placed in the backfill queue.

This overburn policy does not guarantee extra time, and the ALCF reserves the right to prioritize jobs submitted by projects that have not yet used 100 percent of their allocation. The earlier an INCITE or ALCC project exhausts its allocation, the more likely it is to fully benefit from the overburn policy.

Underutilization earlier in the allocation year is primarily managed through personal contact with the projects to understand issues and assist in resolving any problems. Examples of these challenges can include data movement, scheduling conflicts, porting issues, and bugs. Most of the time, this approach is effective, but if significant underutilization persists later in the allocation year, then the ALCF may implement its stated policies for pulling back time from INCITE or ALCC projects that can be redistributed among other projects in the program. For INCITE, utilization is assessed, and a time pullback 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, the ALCF supports its user base with home and project file systems, which are mounted on all production systems. The AI Testbed systems are in a different network enclave and do not mount these file systems. In 2024, the home file systems were migrated from a system called Swift to one called Agile. The project file systems are Grand and Eagle. 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 these file systems can be found in Section 2.5.

As part of its mission to support collaborative and data-driven scientific discoveries, the ALCF provides researchers with services to securely share and reliably transfer data using Globus. In 2024, the ALCF supported 16 data-only projects using these services. Globus sharing is a key component in ongoing efforts to automate the process of taking data from scientific instruments like the APS or DIII-D, processing that data with on-demand computing resources on Polaris, and sharing the processed data with collaborators and the scientific community.

Conclusion

The ALCF delivered the following node hours to the allocation programs in CY 2024: 2.6 million to INCITE, 0.6 million to ALCC, and 1.0 million to DD. The DD program was used not only to develop INCITE and ALCC proposals but also to conduct real science of strategic importance and to drive the development and scaling of potential INCITE and ALCC projects. Excellent ALCF support and reliable, high-performing ALCF resources enabled INCITE and ALCC projects to run jobs efficiently and to achieve scientific goals that may not have been attainable otherwise.

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

Below are the innovations and best practices carried out at the ALCF during CY 2024. These innovations and best practices have helped prepare for future systems, enabled more efficient operations, and strengthened collaboration and engagement across ASCR facilities and beyond. As outlined in Section 4.3, the ALCF is effectively utilizing its postdoctoral fellows.

4.1 Operational Innovation – Technical

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

4.1.1 MyALCF Portal

Challenge: To provide a central access point for PIs and project teams to access project and allocation information.

Approach: The ALCF released a significant upgrade to the account and project management website in November 2024 with the launch of MyALCF, which features a new graphical dashboard that displays project allocations in real time. As a "one-stop shop" for users, it assists PIs and project teams in managing their projects and resources, enabling them to see all of their project's allocations (Figure 4.1). At a glance, the dashboard displays several graphs showing allocation usage, allocation on-track trends, daily/14-day/monthly job activity, and node usage throughout a project. MyALCF also expands the ALCF's web tool offerings, including a tool to help new users learn Ni (sbank) allocation management. The Ni (sbank) tool is also useful to help users learn of options in sbank that they may not have seen or used previously.

Impact/Status: With the introduction of the portal, the ALCF developed the ability to run multiple servers to support and integrate new applications within the software framework, which allows ALCF's software to scale up and will seamlessly integrate with other applications in the future. Users still have the functionality (to request an allocation, manage projects and Unix groups, etc.) as in the project management system in Userbase, but with a cleaner user interface (UI). The ALCF plans to continue expanding MyALCF with an improved web UI for job management and a user notification framework. The MyALCF portal has also realized the impacts, made earlier in the year on UserBase3, of improved cybersecurity and automated testing through: (1) improvements to security on release images, which are now placed in Harbor and scanned for software vulnerabilities before public release; (2) nightly running of automated tests, which has resulted in lowering test failure rate since mid-March 2024 from 4.3% to 0.5%, with results stored automatically on an internal Wiki; and (3) providing of automated PI information for foreign national users to the Argonne security team.

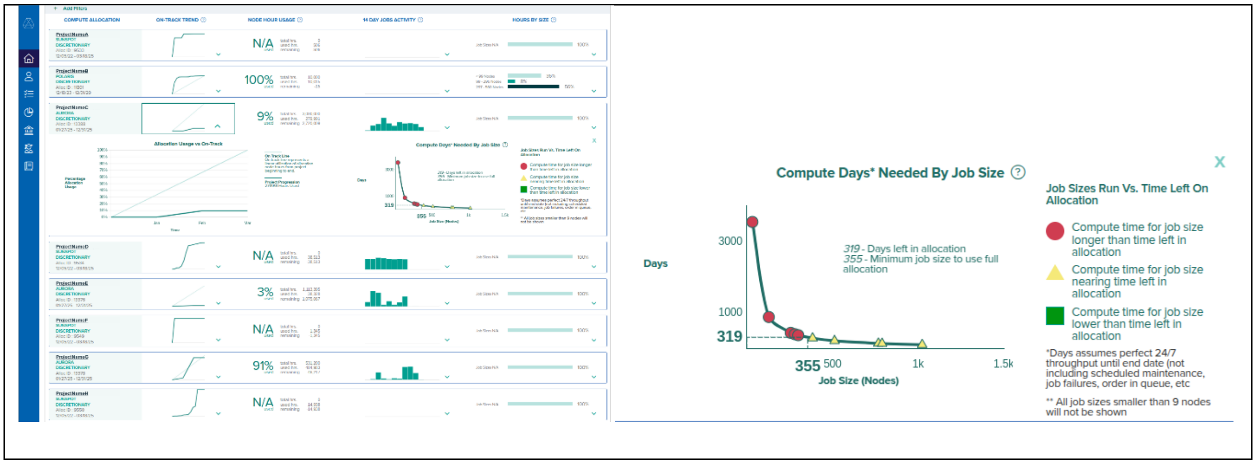


Figure 4.1 Screenshot of the MyALCF dashboard, which displays project allocation and machine status information (left) and an enlarged view of Compute Days Needed (right). This view illustrates a PI's ability to fully utilize the allocation. If time is limited, as shown in the example project, the PI can seek assistance from the ALCF Catalyst to reallocate nodes for project completion.

4.1.2 SimAI-Bench: Performance Benchmarks for Coupled Simulation and AI Workflows on HPC Systems

Challenge: With the ever-increasing adoption of AI/ML methods to accelerate or steer scientific campaigns, users are running new types of applications on ALCF systems. These applications couple traditional HPC simulation with AI training and inference tasks, creating complex workflows with intricate data dependencies and large memory requirements. Just as the AI/ML and HPC landscapes contain multiple programming models and libraries, multiple tools are available for implementing coupled simulation and AI workflows on ALCF systems. However, these tools' relative advantages and disadvantages for various workflow motifs and data transfer patterns, especially their performance at very large scales, are not well known and rarely reported in the literature. This gap in knowledge poses a challenge to ALCF in providing high-quality support to users aiming to run such workflows on ALCF systems.

Approach: To better understand the landscape of coupled simulation and AI workflows on HPC systems, the ALCF developed a first-of-its-kind, open-source suite of performance benchmarks called SimAI-Bench. The goal of SimAI-Bench is to host a series of micro, mini, and full benchmarks for various coupled simulation and AI/ML workflow motifs designed for leadership HPC systems. These benchmarks may be used to evaluate the performance and scalability of different workflow libraries or compare the same workflow motif across different hardware. These scalable mini-apps are inspired by real scientific applications run by ALCF users, such as NekRS-ML; however, they simplify the software stack while maintaining accurate data transfer patterns and dependencies. Similar to how a traditional mini-app is benchmarked by implementing it with different programming models or backends (e.g., OpenMP, Kokkos, SYCL), SimAI-bench implements the same mini-app with different workflow backends (e.g., SmartSim, Dragon, ADIOS-2, and the parallel file system), thus enabling accurate comparisons of various libraries and implementations. Moreover, SimAI-Bench supports various deployment strategies on HPC systems, allowing further understanding of possible bottlenecks and scalable solutions. Finally, SimAI-Bench is designed to be portable across all DOE systems, making this an effective tool for the broad user base. **SimAI-Bench:** https://github.com/argonne-lcf/SimAI-Bench

Impact/Status: SimAI-Bench will significantly impact the ALCF and other DOE user facilities in the coming years. By developing a deep understanding of the coupled simulation and AI workflow landscape and tools, the ALCF will be able to provide users running similar applications with valuable insight and support. SimAI-Bench is also being used to evaluate the performance and suitability of current hardware for these types of workflows and provide insights into possible new hardware solutions. Moreover, SimAI-Bench has expanded to a multi-facility collaboration, including scientists from ORNL and PPPL. SimAI-Bench currently implements a single workflow motif inspired by the NekRS-ML applications, in which simulation data is streamed to train a graph neural network (GNN) surrogate simultaneously. This mini-app is implemented in SmartSim, Dragon, and ADIOS-2, and has been tested on Polaris and Frontier. Results from this mini-app were presented at the 2024 Platform for Advanced Scientific Computing (PASC) conference. (Note: see Section 4.1.4 for our involvement in co-development of Dragon.)

4.1.3 Enabling Integrated Research Infrastructure

Challenge: Modern experimental science increasingly depends on near–real-time data or experiment-time analysis to inform subsequent decisions. Two exemplary cases are the DIII-D fusion experiments and the recently-upgraded APS. At DIII-D, a DOE User Facility focused on fusion energy, the challenge was to process experimental data within a tight 20-minute cycle to enable rapid decision-making in the Tokamak control room. Similarly, as the APS emerges from its upgrade with dramatically enhanced beamline brightness and data rates, its users — while lacking direct access to ALCF resources — required a robust, integrated computing infrastructure that automates data analysis at experiment time.

Approach: To address these challenges, the ALCF has developed a flexible, remote computing infrastructure that adapts to the diverse needs of DOE user facilities, collaborating closely with DIII-D scientists to integrate the facility's experiments with Polaris. ALCF staff reconfigured this workflow to leverage an on-demand queue for rapid processing, integrate with the PBS scheduler, and utilize service accounts to enable autonomous operation. Programmable interfaces and GPU acceleration were integrated into the workflow to further enhance the efficiency of data analysis, ensuring that the computed results are returned promptly to inform the next experimental configuration in the Tokamak control room. Next, because APS users do not possess direct ALCF accounts, dedicated service accounts were established within the APS data management system to trigger automated analysis workflows. Then, in collaboration with APS and the Globus team, the ALCF extended its remote computing capabilities to beamline experiments.

Impact/Status: This workflow redesign now operates where data arriving at the beamline triggers a Globus Flow that coordinates the secure transfer of data to the ALCF, where jobs are executed via the demand queue to meet the experiment-time requirements of the experiments.

Data collected during each 20-minute experimental cycle is streamed to Polaris, where a Globus Compute endpoint manages near–real-time job submissions. Once processed, results are transferred back to APS users. More than ten beamlines have already adopted this seamless integration, ensuring that the increased data rates from the APS upgrade are met with commensurate computational support.

Both implementations demonstrate the versatility and scalability of the ALCF's remote computing innovations. The DIII-D system has transitioned to production experiments, with its capabilities showcased at SC24 and extended to other platforms such as NERSC's Perlmutter system. At the APS, the infrastructure not only meets the immediate needs of users on enhanced beamlines but also sets a new standard for experiment-time computing, positioning the experimental user facility to realize analysis capabilities hundreds of times greater than before. Collectively, these efforts provide a robust model for integrating near real-time computational analysis across diverse DOE user facilities, driving scientific discovery in both fusion energy and photon science.

4.1.4 Automation of Drug Discovery Workflows at Scale

Challenge: The use of AI surrogate models in drug discovery workflows has accelerated the task of screening millions of molecules to identify the best candidates for binding to specific proteins. Nevertheless, these workflows are in many cases still composed of distinct tasks that run sequentially on HPC clusters to iteratively screen through the list of compounds, identify top candidates, perform molecular dynamics simulations to produce accurate training data, and fine-tune the AI surrogate. The sequential nature of the workflow results in multiple job submissions with long queue times and heavy use of the parallel file system. These bottlenecks can be addressed by running all workflow components simultaneously within the same job, storing intermediate data in memory instead of the file system, and continuously screening compounds and refining the AI surrogate. Still, the effective management and movement of terabytes' worth of data remains a challenge in scaling such workflows to HPC systems like Polaris and Aurora.

Approach: By leveraging a novel, open-source software tool called Dragon, ALCF staff partnered with HPE on co-designing a drug discovery workflow that automates the execution of compound screening, sorting of the top candidates, simulation of protein binding affinity, and fine-tuning of the model in one unified workflow. The workflow can make use of both CPUs and GPUs by carefully distributing the components on all available compute resources. Moreover, the workflow limits the use of the file system to a single read operation at the beginning of the run that loads the list of compounds to screen into memory, while performing all other data-sharing operations through local memory or fast Remote Direct Memory Access (RDMA) transfers through the system's interconnect. For enhanced efficiency and scalability, ALCF staff have also tested running workflow components concurrently to maximize the use of heterogeneous hardware with promising initial results.

Impact/Status: ALCF staff scaled the automated drug discovery workflow on Polaris and presented results at the 2024 Platform for Advanced Scientific Computing (PASC) conference. The automated workflow communicates simulation results, trained model weights, and inference results as they are produced across Polaris's interconnect with fast RDMA transfers between nodes in memory. This approach to coupling data between components produced significant speed-ups relative to the previous implementation that stored data on a Lustre filesystem. In comparing this workflow to the SimAI-Bench efforts, ALCF staff have identified fast data movement as a significant challenge in coupling AI/ML models to traditional simulations. ALCF staff plan to scale the drug discovery workflow on Aurora in 2025 and use it to explore different data movement patterns and workflow tools for AI/ML-simulation coupled workflows.

4.2 Operational Innovation – Management/Workforce

The ALCF works to prepare for next-generation systems through collaboration with vendors and other DOE facilities. The ALCF reports on participation in research projects in Sections 1.3.2 and 8.1, and professional community activities in Section 8.3.1.

4.2.1 ALCF Lighthouse Initiative

Challenge: As the use of HPC and AI grows across fields, the need to expand the ALCF user base and cultivate connections with the next generation of computing professionals increases alongside it.

Approach: The ALCF Lighthouse Initiative was created with the mission of building sustainable partnerships between the ALCF and university research computing centers. Through the initiative, computing center staff receive access to ALCF machines, tailored training, and a dedicated ALCF point of contact, which prepares them to support university researchers as they transition from local university clusters to the ALCF's leadership-class systems. As researchers become familiar with ALCF machines, they progress to their own ALCF allocations to continue scaling their research, aiming to submit proposals for ALCC or INCITE projects.

Impact/Status: In 2024, three Lighthouse partners—a private university and two state universities—were onboarded to the initiative. Collaborating with these pilot partners, the ALCF has iterated and enhanced the onboarding process for partners by customizing training for university computing center staff according to partner requirements. Among the three current Lighthouse partners, two introduced university researchers to ALCF systems.

4.2.2 Sprinterns

Challenge: To increase the pipeline of students entering the national laboratory/ASCR user facility workforce.

Approach: Break Through Tech is a national program that trains and connects students with potential employers. The Sprinternship program through Break Through Tech focuses on providing real-world experience through micro-internships lasting three weeks. Students join the tech workforce by participating in a "challenge project" as part of a team of five students. The ALCF hosted one such team, comprised of students majoring or minoring in computing, information technology, network engineering, or a related discipline.

Impact/Status: The "challenge project" for the Sprinterns involved modifying a piece of code used to generate Digital Object Identifiers (DOIs) for new allocation awards to the INCITE program to include awards that were renewed for the year. The students improved their skills in

Python, Git, data manipulation, and documentation. Additionally, four out of the five students returned to the ALCF for summer internships in 2024 to participate in other ALCF projects.

4.2.3 Event Planning with Indico

Challenge: To better manage ALCF events (meetings, workshops, and conferences) with differing complexities across numerous meeting sites.

Approach: The ALCF adopted Indico, an open-source tool for event organization, communications, registration, and collaboration. Indico supports easy upload and retrieval of presentations and other documents, links to room booking and management, and provides permanent archives of all event material, among other features.

Impact/Status: The previous event planning and management tool (CVENT) had associated costs and required external laboratory staff to build and maintain the event web pages. Since switching to Indico, the ALCF administrative support team has been better equipped to execute event planning using this robust and scalable event management tool.

4.2.4 Budget Planning Best Practices

Challenge: ALCF staff assisted their ESnet colleagues with ideas for improving budget planning.

Approach: ALCF staff leveraged ten years of refining the budget planning process by sharing best practices with ESnet. The former financial business manager of ALCF and the current divisional operations lead met with ESnet to guide them on conducting one-dive and deep-dive budget exercises.

Impact/Status: The ALCF's best practices for budget planning were shared with the DOE User Facility community.

4.2.5 Streamlining Editorial and Social Media Workflows

Challenge: The ALCF Communications Team tracks articles and approval workflows and coordinates social media campaigns across multiple channels. With various staff managing and executing our editorial calendar and social postings, the team needed FedRamp-approved software that supported an asynchronous planning process for everyone.

Approach: After weeks of experimentation, the team selected Microsoft Planner as their go-to platform for task organization and collaboration. The project manager begins by creating a dedicated plan for each project or campaign, establishing a centralized hub for all activities. They define task categories to break the project into manageable segments, ensuring clarity and focus. Once the plan is structured, the project manager assigns tasks to team members, sets deadlines, and ensures clear responsibilities. Throughout the project, the manager oversees progress using Planner's interface, fostering team communication, maintaining organization, and driving the project toward successful completion. A team member can view all their assignments and monitor their progress, which aids in managing workloads effectively and staying current on tasks.

Impact/Status: Switching to the new, asynchronous process was seamless, enabling team members to track articles from drafts through approvals and schedule and coordinate social media posts for Facebook, LinkedIn, and Instagram more efficiently. Team members have a clear view into tracking and can adjust deadlines in a centralized hub, as needed, and can share images and copies for posts. Organizing their work this way has eliminated the need for team status meetings because the staff can self-manage their projects, since the interface enables them to have visibility into all project-related activity.

4.3 Postdoctoral Fellows

The ALCF supports a steady-state postdoctoral fellowship program. Within this program, the ALCF supports one named postdoctoral fellow, the Margaret Butler Fellowship in Computational Science Fellow. Postdocs are granted a one-year term with an option to renew for an additional year (typically the case), along with a similar option for a third year. The program's primary goal is to convert 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 manages applications for postdoctoral positions. In 2024, ALCF supported **24** postdoctoral researchers (fellows), including **6** new hires, representing various scientific domains. Throughout the year, **5** postdocs transitioned to ALCF staff, **1** postdoc moved to CPS staff, and **2** departed from ALCF—unfortunately, one postdoc, Raymundo Hernandez-Esparza, is now deceased; his accomplishments are detailed below; the other, Brendan Wallace, left for a position in industry. **Three** postdocs were part of ESP projects, and **21** were part of the 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 postdocs weekly and engaged in the postdocs' research efforts. The supervisor evaluated the progress and completed a yearly standardized review submitted to ALCF management for consideration in appointment renewals. The mentor was responsible for meeting with the postdoc to discuss career development milestones and personal goals. This interaction occurred 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, which skills or experience would be most beneficial for 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.

The ALCF supported the following 24 postdoctoral researchers in CY 2024:

Riccardo Balin (Ph.D., Aerospace Engineering, University of Colorado Boulder). **Hired:** January 2021. **Transitioned to ALCF staff:** 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.

Reet Barik (Ph.D., Computer Science, Washington State University). **Hired:** November 2024. **Research area:** Performance portability for science applications; scaling AI workflows on Aurora. **Projects:** (1) working with the High Energy Physics Center for Computational Excellence (HEP-CCE) to develop the next generation of particle physics simulation codes to help predict research outcomes at the Large Hadron Collider at CERN; and (2) will work with two SciDAC teams, in nuclear physics (NP) and HEP. As a member of the NP SciDAC team, Reet is providing HPC expertise to run a complex AI workflow on Polaris and eventually Aurora. This workflow aims to invert the process of doing science at accelerator facilities. Instead of collecting data for long periods of time, then performing an analysis after the fact, this SciDAC team is building an AI driving pipeline to essentially *fit* the data in real-time as it is recorded by the experiment. Over time, the pipeline will *fit* the parameters of the theoretical model to the data. **Accomplishments:** None thus far, as he started his postdoc in November 2024.

Solomon Abera Bekele (Ph.D., Computer Engineering, Indian Institute of Technology Delhi). **Hired:** February 2023. **Research area:** Computer architectures, performance-energy tradeoffs in heterogeneous HPC. **Projects:** (1) developing a performance and power tracing framework for heterogeneous HPC systems; and (2) researching power and energy optimization strategies in heterogeneous HPC. **Accomplishments:** Expanded the tracing framework Tracing Heterogeneous API (THAPI) to encompass the comprehensive collection of device telemetry, including fabric traffic and memory metrics such as memory bandwidth and allocation size; presented THAPI at the International Workshop on Performance, Portability and Productivity in HPC (P3HPC) tools session at SC24; and evaluated and reviewed the Aurora nonrecurring engineering (NRE) deliverable milestones of GEOPM, which explores power and energy optimizations on heterogeneous platforms.

Fakhrul Hasan Bhuiyan (Ph.D. Mechanical Engineering, University of California, Merced). **Hired:** September 2024. **Research area:** Computational materials, Molten salts, Machine Learning. Developing ML-based interatomic force fields for highly accurate atomistic simulations of molten salt-based electrolyte materials. **Project:** ALCF, CSTEEL, Laboratory-Directed Research and Development (LDRD) project: "Highly Efficient and Low-Cost Electrocatalysts for Converting Carbon Dioxide to Chemicals Containing Two or More Carbons." **Accomplishments:** Code improvements/contributions: (1) created Balsam configurations for ALCF Sophia and Crux; (2) created Balsam apps for LAMMPS and VASP to facilitate the automation of forcefield development on ALCF systems; (3) created a Github repository with instructions to install LAMMPS software with MACE forcefield package on ALCF Polaris, Sophia, and Crux; and (4) created an online app to explore the National Institute of Standards and Technology (NIST) SRD 46 database, which is an important data source for electrochemical experimentalists.

Krishna Teja Chitty-Venkata (Ph.D., Computer Engineering, Iowa State University). **Hired:** August 2023. **Research area:** AI efficiency, AI accelerators, benchmarking. **Projects:** (1) *WActiGrad*, a structured pruning algorithm called Weight Activation and Gradient (WActiGrad) to obtain smaller LLMs from large pretrained models; (2) *LLM-Inference-Bench*, a comprehensive benchmarking suite to evaluate the hardware inference performance of LLMs; (3) *LLaMA-Finetune-Bench*, a framework for fine-tuning and deploying LLMs and vision-language models (VLMs); (4) *Paged Compression*, a novel block-wise structured key value (KV) cache eviction method to enhance the efficiency of virtual large language model's (vLLM's) PagedAttention mechanism; and (5) *LangVision-LoRA-NAS*, a novel framework to integrate neural architecture search (NAS) with low-rank adaptation (LoRA) to optimize LLMs and VLMs for variable-rant adaptation. **Accomplishments:** (1) published a paper in the proceedings of SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis, (2) published a paper in the proceedings of the European Conference on Parallel Processing, (3) made code contributions to Inference-Bench and Finetune-Bench, (4) gave talks at EuroPar, SC24-W, GE Vernova, and ALCF's Fall HPC Workshop, and presented a poster at the Monterey Data Conference, (5) volunteered as a scribe for multiple breakout sessions for the Argonne-hosted National Artificial Intelligence Research Resource (NAIRR) software workshop, and (6) assisted the Aurora GPT team to run different LLMs using several Inference frameworks across NVIDIA and Intel GPUs, and benchmarked several LLMs and efficient Inference methods.

Farah Ferdaus (Ph.D., Electrical and Computer Engineering, Florida International University). **Hired:** February 2023. **Research area:** Power/energy profiling of AI accelerators. **Project:** Evaluating AI accelerators in the AI testbed; investigating the impact of hardware and software libraries on the energy consumption of the deep learning (DL) frameworks for both training and inference using different benchmarks. **Accomplishments:** (1) developed a high-level power analysis subroutine using *NVML, PYHLML, and gcipuinfo* APIs to record power from NVIDIA GPU, Habana Gaudi, and Graphcore IPU, respectively; (2) developing a similar power analysis subroutine for AMD and Intel GPUs; (2) published a paper in the proceedings of the 2024 Institute of Electrical and Electronics Engineers (IEEE)/ACM International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS).

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 plane-wave density functional theory (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 the “Frameworks” module and continuous testing and validation of the AI/ML software stack as part of the Aurora NRE project; (2) scaling out deep learning applications on Aurora, specifically CosmicTagger (an application originating in neutrino physics, part of the Aurora NRE project); (3) continuous validation of large-scale communication methods using different collective communication libraries, such as OneCCL, MPI, etc.; (4) testing and validation of profiling strategies and tools for AI/ML users on Aurora (also educating users on these tools); and (5) large-scale AI/ML experiments on both Aurora and OLCF’s Frontier. **Accomplishments:** (1) submitted two papers under review and co-wrote a workshop talk at SC24; (2) gave a talk about AI/ML profiling strategies at the ALCF Hands-on HPC Workshop 2024; and (3) participated in the ATPESC and INCITE technical review committees for 2024.

Junoh Jung (Ph.D., Mechanical Engineering, University of Michigan). **Hired:** November 2024. **Research area:** scientific machine learning to enhance simulations of dynamical systems environments (e.g., computational fluid dynamics and flow estimation and control) in HPC, particularly using PETSc. **Project:** (1) developing PDE-based simulations using scientific machine learning for exascale platforms. **Accomplishments:** None thus far, as he started his ALCF postdoc in late 2024.

Bharat Kale (Ph.D., Computer Science, Northern Illinois University). **Hired:** September 2023. **Research area:** Visual AI. **Projects:** (1) genome-scale language models; (2) visual analytics for AI; and (3) LLM Inference Bench. **Accomplishments:** a finalist co-author for the 2024 Gordon Bell prize for a paper published in the proceedings of SC24: International Conference for High Performance Computing, Networking, Storage and Analysis and SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis; also, a co-author contributing a workshop paper at SC24 (also published in the conference proceedings).

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, and optimized the Kokkos version of HARVEY by offloading the cell communications to GPUs; (2) tested the optimized code and analyzed the performance levels on Aurora and other platforms; (3) integrated Ascent-based in-situ visualization into the optimized Kokkos version of HARVEY; (4) contributed to the final report of the ESP project; and (5) co-authored an SC24 workshop paper.

Nathan Nichols (Ph.D., Material Sciences, University of Vermont). **Hired:** September 2021. **Transitioned to ALCF staff:** July 2024. **Research area:** Aurora ESP Application Performance, Aurora LLM Quantization. **Projects:** (1) leads the ATLAS ESP project, which enables simulations of proton-proton collisions in the Large Hadron Collider at CERN and involved in porting the leading-order MadGraph event generator to SYCL to compare performance across multiple architectures (Intel CPUs, NVIDIA GPUs, AMD GPUs, and Intel GPUs), enabling the application to scale up to 2,125 nodes on Aurora—one of the first applications to meet the ESP scaling criteria; (2) primary developer for the DEAC algorithm within the SmoQyDEAC.jl project, which has enhanced DEAC’s capability to process both fermionic and bosonic data (critical for precise spectral function analysis in quantum many-body simulations); its advances have been benchmarked against established methods, confirming DEAC's strong performance and robustness in computational physics research; (3) contributed to optimizing AI and machine learning performance on Intel hardware by porting and testing 8-bit quantized linear layers and optimizers to SYCL, enabling efficient execution on Intel GPUs; and (4) fostered new collaborations within the ALCF community (e.g., Adrian Del Maestro's group from the University of Tennessee, Knoxville). **Accomplishments:** (1) publications in *Phys. Rev. B* and *Proceedings of the 12th International Workshop on OpenCL and SYCL*; and (2) presented the MadGraph-to-SYCL results at leading high-energy physics conferences.

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:** (1) analyzing how job characteristics impact wait times on Theta and Polaris and using machine learning to predict job wait times; (2) analyzing the relationship between workloads and hardware failures; and (3) benchmarking distributed machine learning algorithms from the cuML library on Polaris. **Accomplishments:** (1) Submitted one paper for review and a second is in preparation; (2) presented a poster at the 2024 Argonne Postdoctoral Symposium; and (3) created and presented a new tutorial at the 2024 ALCF Hands-on HPC Workshop on scikit-learn and Dask.

Thang Pham (Ph.D., Chemical and Biological Engineering, Northwestern University). **Hired:** December 2024. **Research area:** Advancing the application of LLMs as intelligent agents in computational chemistry, with a focus on automating complex workflows and enhancing efficiency in material design and discovery. **Project:** LLM agents for autonomous computational chemistry workflows. **Accomplishments:** Implemented and tested LLM agents to automate geometry optimization and simulation input generation.

Siddhisanket (Sid) Raskar (Ph.D., Electrical and Computer Engineering, University of Delaware). **Hired:** June 2021. **Transitioned to ALCF staff:** March 2024. **Research areas:** Dataflow models and architectures, HPC, ML, AI accelerators, performance modeling. **Projects:** (1) LLM inference and finetune benchmarking; (2) LLM inference optimizations; (3) and porting and optimization of HPC workloads on AI accelerators. **Accomplishments:** (1) published in *Applied Sciences* and in the proceedings of ACM MACS 2024, IEEE IPDPSW 2024, and IEEE/ACM PMBS 2024; and (2) gave invited talks at the 2024 ALCF Hands-on HPC Workshop, two tutorials at SC24, and a talk at the University of Michigan Summer School (SciFM 2024).

Thilina Ratnayaka (Ph.D., Computer Science, University of Illinois Urbana-Champaign). **Hired:** September 2024. **Research area:** High-performance computing, scientific computing, GPU programming models. **Projects:** (1) Has worked on packaging OCCA, a portable and vendor neutral framework for parallel programming on heterogeneous platforms, and Conda: an OS-agnostic, system-level binary package and environment manager, using the high-performance package manager Spack; (2) added CTest-based testing for NekRS, a fast and scalable CFD solver using OCCA to run on GPUs; (3) is evaluating SYCL in NVIDIA hardware by comparing it to CUDA and high-performance, high-level APIs like CUTLASS ; and (4) added Zoltan2 as an optional partitioner to Nek5000, which is used by NekRS (also refactoring existing partitioner logic in Nek5000). **Accomplishments:** (1) Ran the entire NekRS test suite on Aurora, helping to fix multiple NekRS bugs; (2) prepared detailed documentation to help people learn SYCL; and (3) helped fix an integer overflow bug causing runtime errors in the Nekbone proxy app used for Aurora acceptance, liaising with the Intel engineer responsible for the Nekbone application.

Shilpika (Ph.D., Computer Science, University of California, Davis). **Hired:** September 2023. **Research area:** Explainable AI, Log Analysis and Visual Analytics. **Projects:** Recent work spans three primary areas: (1) anomaly detection supercomputers; (2) explainable AI; and (3) the visualization of protein-protein interaction (PPI) graphs. **Accomplishments:** (1) built a holistic visual analytics frontend consisting of a Jupyter notebook integration and D3-based visualizations for error log analysis. This environment integrates diverse logs and monitoring data, enabling more efficient identification of issues. In addition, we developed an incremental approach to manage high-frequency time series data (e.g., environmental readings), facilitating faster data processing and storage. Currently, we are focusing on sequence mining using an automaton to quickly extract error message patterns from text-based log messages, thereby improving the accuracy and speed of anomaly detection; (2) efforts have centered on diagnosing and enhancing LLMs. Investigated bottlenecks in a Smiles-based SST transformer model and introduced a probabilistic attribution method to pinpoint key tokens in text inputs for next-word prediction tasks. Ongoing work aims to extend these methods by computing multi-token attribution scores and examining the impact of various tokenization schemes on trained LLMs; (3) created a 3js and D3-based visualization for dense, interactive PPI graphs and are now extending this tool to enable the comparison of two PPI networks: one derived from a trained LLM and another from empirical research. This comparative approach will provide deeper insights into model accuracy and the biological relationships captured by advanced AI systems; (4) publications in proceedings of the 2024 IEEE 24th International Symposium on Cluster, Cloud and Internet Computing (CCGrid) and SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis; and (5) placed third at the Argonne Postdoctoral Slam; presented at Argonne's public "Argonne Outloud" event; and at Argonne's Board of Governors' meeting.

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) high-level heterogeneous programming model research (studying efficient ways to utilize intra-node resources using parallel 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, Intel compiler teams, and ORNL; Flash-X has demonstrated linear scaling across clusters and has been proven effective up to 2K nodes; (2) published articles in the proceedings of the 2024 International Workshop on OpenMP (IWOMP-24) and Cray User Group (CUG-24) conference; and (3) was invited to speak at the SC-24 Intel booth.

Umesh Unnikrishnan (Ph.D., Aerospace Engineering, Georgia Institute of Technology). **Hired:** November 2021. **Research area:** HPC, performance engineering, software portability, CFD. **Current projects:** *Hardware integration and optimization on LCF resources:* (1) scaling and performance benchmarking of CNS application on Polaris (up to 128 nodes) and Aurora (up to 512 nodes) with GPU-aware MPI; (2) scaling and testing the CEED-PHASTA Early Science application on Aurora (up to 4096 nodes); (3) power measurement for the CNS application on PVC and A100 GPUs. *Application development to attract usage on LCF resources:* (1) implementation and testing of artificial dissipation scheme for shock capturing, evaluating the robustness of the scheme in test cases involving supersonic flow over aerodynamic bodies; (2) developing a runtime compilation framework for the CEED-PHASTA application. **Accomplishments:** (1) Presented work at the Cray User Group (CUG-24) conference; the In Situ Infrastructures for Enabling Extreme-scale Analysis and Visualization (ISAV) Workshop; and the SC24-P3HPC Workshop; and (2) completed numerous code improvements and Github contribution for the CEED-PHASTA application.

Madhurima Vardhan (Ph.D., Biomedical Engineering, Duke University). **Hired:** September 2022. **Transitioned to ALCF staff:** September 2024. **Award:** Margaret Butler Fellow. **Research area:** Biomedical engineering. **Projects:** (1) evaluating LLMs for generation of synthetic health records against traditional DL and state-of-the-art diffusion models on multidimensional axes of accuracy, utility, privacy-preserving, and augmentation axes; (2) enabling LLMs to function as personalized health coach via In-context learning; (3) developing a personalized and automated healthcare assistant by fine-tuning LLMs with deidentified structured and unstructured clinical datasets; and (4) evaluating the performance of generative models on HPC resources, Polaris and specialized AI accelerators, Cerebras and SambaNova. **Accomplishments:** (1) published two conference papers (AI Health 2024, IEEE CAI 2024), published two journal papers (*Structural Heart*, *International Journal of Molecular Sciences*); (2) gave two invited lectures; and (3) made code improvements including implementing an LLM code on Polaris using structured electronic health records data; evaluating other generative models such as VAE, GAN, and DDPM on Polaris to compare against LLMs; and implementing LLM inference code using free text clinical data.

Archit Vasan (Ph.D., Biophysics, University of Illinois Urbana-Champaign). **Hired:** February 2023. **Research area:** Computational biophysics/drug discovery workflows. **Projects:** (1) small molecule drug-screening platforms on Aurora, (2) peptide design workflow development, and (3) interfacing direct preference optimization for small molecule/large binder refinement. **Accomplishments:** (1) presented a workshop paper at the IEEE International Workshop on High Performance Computational Biology; (2) co-author on a Gordon Bell Finalist paper at SC24 and on an IPDPS 2025 paper; and (3) lectured on LLMs at the ALCF AI student training series.

Kaushik Velusamy (Ph.D., Computer Science, University of Maryland, Baltimore). **Hired:** January 2022. **Transitioned to ALCF staff:** July 2024. **Research area:** data management, storage, and I/O. **Projects:** DAOS and Copper. **Accomplishments:** (1) Evaluated various I/O patterns and benchmarks on DAOS, the major filesystem on Aurora; assisted our users in integrating their applications to use DAOS for I/O and checkpointing such that HACC, XGC, Phasta, NekRS, Candle, E3SM, and Aurora Acceptance Test applications like thunderSVM, Resnet50, BERT, and UNO are now actively using DAOS for their full system-wide runs (which was not the case at the beginning of FY 2024); also delivered hands-on sessions and a paper on DAOS at SC-24; (2) developed Copper, which uses cooperative caching, with an overlay message passing network, on an n-ary tree of compute nodes (without MPI) using the Mochi Thallium and Argobot framework, which has significantly reduced the stress on our Lustre subsystems and thus on network interference and scaled easily to the full system; (3) contributed to the acceptance of the Aurora NRE frameworks and improved the scaling performance of collective communication in Intel MPICH, OneCCL, and several Python frameworks along with profilers like Unitrace and Thapi; also assisted in the ALCF training and webinars, reviewed INCITE computational readiness and conference/workshop papers, and guided tours at the ALCF data center.

Brendan Wallace (Ph.D., Atmospheric Science, SUNY-Albany). **Hired:** August 2022. **Left the lab:** March 2024 for employment at Leidos. **Research area:** atmospheric science/climate dynamics. **Projects:** (1) Large-scale MCS environments in a dynamically downscaled regional climate model (supported by the National Science Foundation [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 four publications appearing in *Climate Dynamics* (August 2023), *Earth and Space* (December 2023), *Springer Climate Dynamics* (February 2024), and *Journal of Climate* (December 2024); co-author on papers published in *Climate Change* (June 2024), and *Climate and Atmospheric Science* (August 2024).

Azton Wells (Ph.D., Physics, University of California, San Diego). **Hired:** September 2022. **Switched ALCF research focus:** April 2024. **Transitioned to CPS staff:** August 2024. **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 modeling both peptide and genome sequences; (3) AI for Science: (a) Foundation models in cosmology, (b) Agentic models for cosmological data analysis and workflows, (c) Generative models of galaxy images from observational or simulation catalogs. **Accomplishments:** (1) authored or co-authored several publications submitted after postdoc concluded, based on postdoctoral research (EAIRA, HiPerRAG, AstroMLab 3, AstroMLab 1, Galaxies and Their Environment at z ~ 10—I); (2) gave two invited lectures; and (3) led a project leading to a presentation at Science Understanding through Data Science (SUDS) Conference, Caltech, 2024.

Section 5. Risk Management

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

ALCF Response

The details of the risk management process that the ALCF follows (and laid out in Section 5.1) demonstrate that the ALCF successfully managed operational risks in 2024. Guided by 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. The ALCF currently has 38 open risks, with two risks with a post-mitigated score in the severe range: Funding/Budget Shortfalls, which is managed by careful planning with the DOE program manager and the implementation of austerity measures as necessary, and Staff Recruitment Challenges, which is managed by ongoing recruiting and re-tasking of current staff as needed. The major risks tracked for 2024 are listed in Section 5.2, along with the details of these risks in Table 5.1. There were no risks encountered in 2024. Section 5.6 and Table 5.2 provide details on the major risks that will be tracked in 2025.

5.1 Risk Management Process Overview

The 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 reviewed annually and updated as needed throughout the year to reflect changes, incorporate new risk management techniques as they are adopted, and incorporate best practices from other facilities. Risk management is part of the ALCF's culture, and the RMP processes are part of normal operations and all projects.

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

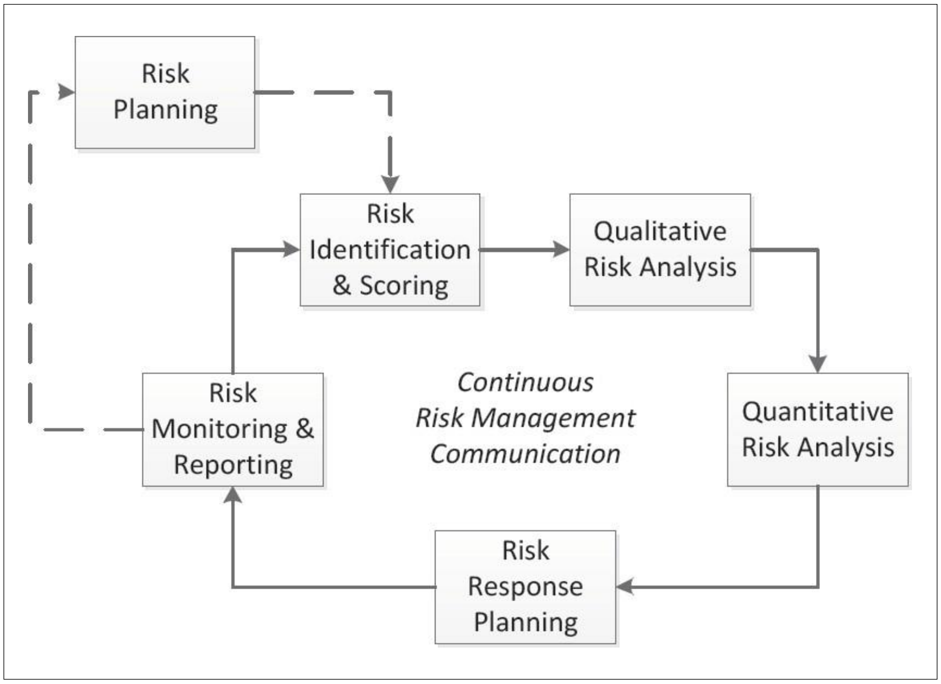


Figure 5.1 Illustration of the Iterative Risk Management Process from ALCF's Risk Management Plan

The ALCF risk management process consists of the following steps, which are performed on a continual basis in all routine operations and 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 funds 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's 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 the 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 the ALCF.

5.1.2 Risk Management in Day-to-Day Operations

The ALCF currently has 38 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 ALCF. Risks are owned by group leads or managers. In ALCF Operations, subject matter experts estimate risk mitigation costs for management consideration. In CY 2024, the Risk Review Board did not convene because no new risks were proposed, and no risks were re-characterized during the year. The ALCF holds informal individual risk discussions as often as needed. Risk owners submit monthly reports on the risks within their respective areas throughout the calendar year.

Risks are identified and evaluated, and mitigation actions are developed for all changes that occur at the 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 register. 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 request.

Other tools besides the risk register are used for managing risks in day-to-day operational risk. One 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 such as evaluating and managing risks around INCITE and ALCC proposals (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 the ALCF's supercomputer Aurora, known as ALCF-3, continued in 2024. Project Risk Register managers continue to maintain a separate project risk register and track a set of detailed risks. Risk mitigation costs are developed using bottom-up cost analysis and then entered in 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 Polaris during CY 2024 and planned the growth of the staff and the budget to bring the facility to full strength. Because of this, the ALCF continues to monitor a number of major risks for the facility. No risks were retired in 2024.

Five major operations risks were tracked for 2024, one with a risk rating of High, one Moderate, and three 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.

Table 5.1 Major Risks Tracked for CY 2024

ID	Title	Encountered	Rating	Notes
1059	Funding/Budget Shortfalls	No	High	The ALCF regularly worked with the program office to plan a budget for handling the impact of a Continuing Resolution in FY 2024 and FY2025, new hires, and changes in the laboratory indirect expense rate. This risk remains a major concern as the facility moves forward with Polaris and Aurora in CY 2025.
1049	Staff Retention	No	Moderate	The ALCF added eleven new regular staff members in CY 2024: Eight of those were external new hires, two of those were converted from postdocs to staff, and another was converted from a contractor to the staff. The ALCF lost only one staff member in CY 2024. The 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. 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	The ALCF had significant increased data center activity in CY 2024, but zero incidents given the ALCF's safety culture.
1019	Users Perceive Scheduling Performance as Poor	No	Low	The 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.
1078	System Stability Issues	No	Low	With Aurora becoming operational in CY 2024 and completing its transition from the ALCF-3 project into ALCF Operations in CY 2025, ALCF is closely monitoring the system stability of this new machine.

5.3 Risks Encountered in the Review Year

The ALCF did not encounter any risks in CY 2024.

5.4 Retired Risks

No risks were retired in CY 2024.

5.5 New and Recharacterized Risks

No risks were recharacterized in CY 2024. While no new risks were added to the register in CY 2024; in April 2025, the ALCF introduced a new risk related to cybersecurity incidents to address the possibility of unauthorized individuals accessing the facility.

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 2025, and the current risk rating and management strategies for each risk. The first risk is one that experience has shown is likely to be encountered in any fiscal year. The ALCF removed Staff Retention and Injury to Workers/Overall Safety of the Division from this list for CY 2025. The ALCF has been effective with staff retention in the past year and is confident in its ability to retain staff for the upcoming year. Given that the active installation of Aurora is now complete, and the system is moving into production, the ALCF has removed Injury to Workers/Overall Safety of the Division from the list of top risks being monitored for the year.

For CY 2025, we are closely monitoring two additional risks: ***Compute Performance Issues*** and ***Incorrect answers come out of the system***. With Aurora becoming operational in CY 2024 and completing its transition from the ALCF-3 project into ALCF Operations in CY 2025, ALCF is closely monitoring the performance, correctness, and system stability of this new machine, as reflected in risks 1079, 1038, and 1078.

Table 5.2 Top Operating Risks Monitored for CY 2025

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.
1079	Compute Performance Issues	Low	Conduct regression testing to monitor system health. Monitor and analyze system failure data. Evaluate initially on test system 2 weeks prior to production.
1038	Incorrect answers come out of the system	Low	Run validation on the production codes. Use any tests developed from codes that have been previously identified as triggering wrong answers due to hardware malfunctions (e.g., the CPS code test). Build a regression test with benchmarks and applications. Engage pertinent vendor(s).
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

The 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 the ALCF's culture and is used by all staff, from senior management to summer students. A formal risk assessment is performed for every major activity within the 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 the ALCF, particularly in the area of safety. The ALCF's effective risk management plan has contributed to the successful management of all significant risks monitored in CY 2024.

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

The ALCF has consistently demonstrated a strong commitment to environment, safety, and health (ES&H) practices, fostering a culture of continuous improvement. In line with this commitment, the ALCF upholds the principles of the laboratory's Integrated Safety Management System (ISM) to ensure the well-being of its staff, users, the public, and the environment while complying with 10 CFR 851.

As part of the ISM, the ALCF defines the scope of work, which includes planning and executing operations and projects to meet scientific and research objectives while ensuring the highest safety and environmental performance. The ALCF utilizes its Work Planning and Control (WPC) Manual (LCF-PROC-1) to establish clear worker roles and responsibilities. ALCF utilizes the Laboratory's training management system (TMS) to identify the necessary competencies to perform its tasks safely and effectively. LCF-PROC-1 details the ALCF's authorization process, which includes reviewing and approving work plans, conducting hazard assessments, and implementing control measures before work begins. Regular monitoring and auditing activities are carried out to assess compliance with safety protocols and evaluate the effectiveness of controls and procedures. Any necessary adjustments are made to improve safety performance and ensure continual improvement.

The ALCF also uses a graded approach to incorporate controls for both quality and hazard risks related to the specific work being performed. It implements a hierarchy of controls to eliminate or mitigate potential hazards. Through this approach, risks are systematically identified, analyzed, and reduced to acceptable levels. The ALCF places great emphasis on the appropriate training and competence of its staff and users to ensure they possess the knowledge and skills necessary to carry out their work safely.

6.1 Work Planning

6.1.1 Work Control Documents

Work planning at the ALCF begins during the earliest stages of a project. It is guided by the Laboratory's safety and quality policies, alongside the more detailed WPC Manual, LCF-PROC-1, and the Computing, Environment, and Life Sciences (CELS) Directorate's Quality Manual CLS-MNL-1. The ALCF emphasizes a systematic approach that identifies and manages risks through hazard identification and mitigation. This process starts by conducting work planning meetings with stakeholders, including workers, subject matter experts (SMEs), line management, engineers, and contractors, to establish the scope of work and start the hazard identification process. The ALCF works closely with the work planning stakeholders to develop acceptable hazard mitigation and a plan for executing work. This process is documented and reviewed by all involved in the work planning using Work Control Documents (WCDs) and Job Safety Analyses

(JSAs). This planning, reviewing, and documenting process ensures that all potential hazards and necessary controls are recognized and addressed within the data center environment.

All ALCF work activities outside of office duties are assessed and documented using WCDs. Several types of WCDs address different types of work that happen in the data center. ALCF's primary WCD is a Skill of the Worker WCD, Operating Skill of the Worker Standard Tasks. This document addresses low-risk standard work that occurs daily in the data center. More complex work involving higher risk or requiring specific skills, training, and approval is generally covered under task-based WCDs. For example, the ALCF has a dedicated WCD for creating and removing power connections associated with power distribution units (PDUs). All WCDs are subject to annual reviews and updates to account for changes or new hazards that may arise. This ongoing evaluation aids in maintaining a thorough and current understanding of the risks associated with the work performed by ALCF staff.

6.1.2 Contractor Job Safety Analysis

Additionally, Job Safety Analyses (JSAs) are explicitly used for contractor work within the facility. These analyses require collaboration between ALCF personnel and contractors, helping to identify hazards and ensuring that all work complies with established scope limits and mitigation measures. Planning for contractor work starts during the procurement process. ALCF personnel procuring contractor services collaborate with their Environment, Safety, and Health (ESH) Coordinator to assess the risks associated with any service work performed in the data center. The 'Risk Level Determination Checklist' form records the identified risks. This form is utilized to outline the correct terms and conditions, which include safety requirements the contractor must agree to before work can commence. Once a risk level is established, ALCF personnel initiate a more detailed work planning process documented in the JSA. This process guarantees that established safety expectations between the contractor and the ALCF are documented and met.

ALCF personnel oversee all work conducted by their contractors and are accountable for ensuring compliance with ALCF safety work practices. Before ALCF personnel can procure service contractors, they must qualify as a Technical Representative. This qualification requirement is confirmed in the Argonne procurement system and by the ALCF ESH Coordinator. Technical Representatives receive extensive training regarding their responsibilities for overseeing contractors to ensure they understand their role and have the necessary resources. In 2024, ALCF developed and managed the work of 5 high-risk JSAs, 18 moderate-risk JSAs, and 3 low-risk JSAs, all without incident.

6.1.3 Training

The ALCF is committed to ensuring staff have access to adequate training, so they are qualified to perform their required tasks. The Training Management System (TMS) is an application used to track the completion of required training. The Job Hazard Questionnaire (JHQ) is an exhaustive tool used to evaluate all tasks and hazards that ALCF staff may encounter while performing their work. A worker's TMS profile will generate required training based on the hazards and tasks selected on the JHQ. The JHQ is reviewed with ALCF staff when they begin employment, change roles, and at least once a year with their supervisor. It is then reviewed and approved by their ESH Coordinator. Some roles require documented qualifications, like being a Technical Representative. These roles are identified in the JHQ and require an SME to certify and document an individual's competency once it has been demonstrated. Some roles require more hands-on training or on-the-job training (OJT). Several tasks and roles specific to the ALCF data center necessitate documented OJT, such as operating lift equipment and becoming a Qualified Electrical Worker (QEW). Completed OJT is recorded in the TMS, and any additional OJT not tracked in TMS is recorded in a WCD.

All personnel working in the data center must complete the CELS100 and CELS101 training. CELS100 reviews CLS-PROC-9, "Environmental Health and Safety Plan for Working in the Building 240 Data Center," in detail. CLS-PROC-9 is a procedure developed and maintained by the data center manager that houses important health and safety information about the data center. This includes information about access requirements, emergency egress and response, personal protective equipment (PPE) requirements, training requirements, and confined space information and contains the current year's noise reports. CLS-PROC-9 is reviewed and approved by ALCF management annually and is updated at least every 36 months. CELS101 is the Server Lift and Material Handling training. This training is required for anyone working inside the data center and is taught in person by the ALCF ESH Coordinator. CELS101 covers the safe use of server lifts and the safe principles for handling material. After completing the initial CELS101, ALCF staff working in the data center are required to take a CELS101 refresher every 36 months.

Contractors working under a JSA must complete Contractor Safety Orientation (CSO) training annually. Additionally, for any training requirements identified in the JSA, such as NFPA 70(e), the contractor must provide records of completion before starting work. The ALCF also requires all contractors working in the data center to complete CELS100 training and building orientations. Records for both are documented in the completed JSA.

6.2 Safety Performance

The ALCF fosters a strong safety culture that emphasizes a safe working environment. The ALCF actively promotes a proactive safety culture by implementing safety pauses and recognizing the importance of continuous improvement through feedback mechanisms.

In CY 2024, the ALCF participated in the laboratory-wide safety pause, and each team supervisor led a feedback session to share observations and suggestions. Feedback was explicitly provided to the data center regarding clarification on the required PPE. The Aurora Floor Manager addressed this feedback by providing clear visual indicators when entering an area with more restrictive PPE requirements.

The Aurora Install Team was recognized in 2024 with the UChicago Argonne Board of Governors James B. Porter, Jr. Team Award for Outstanding Safety.

6.2.1 Safety Shares

The ALCF's ESH Coordinator participates in the bi-weekly ALCF Ops All-Hands Meeting to share relevant safety updates with the Operations Team. These meetings also enable staff to give safety feedback and promote open communication with the ALCF ESH Coordinator. Safety shares consist of useful information, tips, or stories related to safety, aiming to keep safety a priority for ALCF staff. Anyone in the ALCF can contribute safety shares, which are intended to encourage discussion and facilitate the identification of best practices. Staff members can present a safety share of their choice for informal discussion during the operations meeting. Additionally, they can propose a topic to the ESH Coordinator, who will turn it into a safety share using a laboratory-provided template. This will then be reviewed by the laboratory safety committee and communicated throughout the laboratory. For example, during a safety share about the services provided by the Argonne Fire Department (AFD), ALCF staff expressed interest in hands-on training for automated external defibrillators (AEDs) for data center workers. A hands-on AED training was organized in the data center building, and multiple drop-in sessions were offered. The AFD conducted this training, which saw strong attendance.

6.2.2 PDSA and Management Assessments

The ALCF adheres to the Plan-Do-Study-Act (PDSA) philosophy, integrating ISM functions within its operations. The facility conducts biannual health and safety inspections, led by the ALCF's Division Director and ESH Coordinator, to evaluate work practices and identify areas for improvement. Any inspection findings are tracked and documented in the laboratory application Cority. Data from Cority is then used to help identify risk opportunities that are addressed using management assessments. Any health and safety issues or issues requiring corrective action are documented and tracked within the laboratory performance management system, Prism.

Management assessments are also performed according to guidance in the laboratory's performance management system and are documented in Prism. The CELS Directorate Operating Officer and ALCF management meet to review the required assessment topics and identify additional topics based on risk. The assessment topics that are developed yearly serve as a feedback mechanism to drive continuous improvement efforts. The assessments are then entered into Prism to be tracked. Ad hoc assessments can be added at any time should the need arise. A management assessment completed for the data center in 2024 identified a need to improve noise guidelines for visitors. Based on the findings, the ALCF ESH Coordinator worked with the noise SME to develop controls for data center tours that protect visitors and remain compliant with DOE orders.

The DOE ANL Site Office participates in the facility health and safety inspections of all ALCF-occupied spaces on the ANL site, providing recommendations that are documented and tracked in Cority. The CELS Argonne Site Officer (ASO) regularly participates as a team member in ALCF management assessments and independent management assessments. The CELS ASO is also included in regularly held facility operations meetings and work planning and control reviews to ensure the ALCF operates and executes work consistent with the Site Office's requirements. The Site Office does not offer safety guidance to the ALCF.

6.2.3 Operational Improvements

In 2024, a particular focus was placed on enhancing electrical safety practices within the data center. The ALCF worked closely with Argonne's Electrical Safety Program (ESP) Manager and

Electrical SMEs to address outdated electrical policies that hindered work processes and lacked adequate hazard control measures. By providing feedback during an ESP assessment and assisting in the subsequent working groups to update these laboratory policies, the ALCF improved its ability to conduct work safely and mitigate risks associated with electrical hazards in the data center setting. The ESP now includes more laboratory stakeholders, including those from the ALCF, on its Electrical Safety Committee.

6.2.4 Lessons Learned

The ALCF actively promotes a culture of learning through lessons learned (LL). The facility examines incidents and near misses, conducting causal analyses to develop action plans and follow-up procedures. These lessons are shared with the laboratory to enhance organizational knowledge and prevent future accidents. One LL shared with the ALCF occurred when a small network device was plugged into a 208V receptacle, but the device was only rated for 120V. This caused a small spark and rendered the network device inoperable. This LL was shared with all workers in the data center, and more emphasis was placed on explaining the reasoning behind the rule that it is prohibited to plug in any device to an in-rack PDU without prior data center manager approval. The ALCF also added a map indicating all ground fault circuit interrupter (GFCI) 120V receptacles to its CELS100 document, which is reviewed with all workers in the data center, including contractors.

6.2.5 Ergonomics

The ALCF maintains a robust ergonomic program for office and data center environments. A dedicated Ergonomics Coordinator works with the Physical Therapy (PT) Department to conduct ergonomic consultations and evaluations for on-site and remote work. Based on these evaluations, the facility loans out ergonomic task chairs, keyboards, mice, and fatigue mats, providing employees with personalized support. All equipment that the ALCF ergo coordinator recommends or loans out of the ergo room is first vetted by the laboratory PT to ensure that it meets their guidelines. Ergonomic evaluations are promoted by including them as an onboarding activity during a new hire’s second week. The ALCF also has the ergo coordinator present at all-hands meetings and sends out reminder information detailing the process of requesting an evaluation. The ergo coordinator also provides weekly open office hours during which ALCF staff can visit the ergo room to try out the equipment and ask questions. In 2024, ergonomic evaluations were conducted for 21 staff members, 4 of which were remote home office evaluations.

Specific focus is given to material handling programs in the data center, ensuring safe practices for workers. The ALCF’s ESH Coordinator worked with the Argonne PT group to develop the compressive material handling training for data center workers, CELS101. CELS101 is reviewed by the Argonne PT group when any updates are made or as necessary. CELS101 is conducted in-person and reviews material handling principles to reduce ergonomic risk when performing physical moves in the data center. The Argonne PT group assists with concerns outside the ALCF ESH Coordinator’s skill set.

6.2.6 Aurora

During the Aurora installation in 2024, the ALCF effectively managed the delivery and placement of additional compute racks. The Aurora Floor Manager effectively planned and oversaw the subcontractors' daily activities, including offloading and loading daily deliveries, blade maintenance, installation, and material handling. The floor manager also worked closely with the contractors and consultants to document and proceduralize Aurora-specific tasks to ensure a smooth transition to the ALCF Operations Team. Additionally, the floor manager worked closely with the laboratory's confined space SME and building management to ensure updates were made to confined space procedures to prioritize safety during the installation. This included upgrading the visual indicators above and below the raised floor to indicate the location of the confined spaces.

6.3 Incident Reports

In CY 2024, the ALCF experienced zero Days Away Restricted or Transferred (DART) cases and one first aid case, resulting in zero Total Recordable Cases (TRCs). The first aid case involved a strained elbow that occurred when a postdoc lifted a heavy laptop computer to place it in a backpack. The recorded cases over the past five years are summarized in Table 6.1.

Table 6.1 ALCF Injury and Illness Cases FY 2020 thru FY 2024

Years	DART	First Aid	TRCs
2020	0	1	0
2021	0	0	0
2022	0	0	1
2023	0	0	0
2024	0	1	0

The ALCF continues to prioritize ES&H practices, implement appropriate measures, and foster a safety-first culture throughout the facility. These efforts ensure the well-being of staff, users, the public, and the environment.

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

The ALCF demonstrates effective cybersecurity practices through a commitment to enhancing these practices and expanding staff involvement in cybersecurity initiatives. This is evident in the ALCF's efforts to formalize a cybersecurity program within the ALCF and expand cybersecurity working groups to incorporate staff from other groups. As a result, the ALCF has recorded zero security incidents this year and effectively mitigated security risks. Additionally, the ALCF possesses a valid Authority to Operate (ATO) and maintains a robust process for compliance with export controls and foreign visitor access.

7.1 Demonstrating Effective Cybersecurity Practices

The ALCF maintains a strong cybersecurity posture and is vigilant in monitoring evolving threats, vulnerabilities, and potential security incidents. The ALCF's Security Team has ongoing partnerships with the OLCF, NERSC, and other facilities to jointly assess the risks of security threats and incidents unique to HPC facilities. Working together, the facilities can quickly identify, coordinate, and understand mitigations used to protect HPC systems against new vulnerabilities. The Secure ASCR Facilities (SECAF) working group continues to be a strong partnership for information sharing and collaboration between facility security staff.

In 2024, the ALCF continued the process of formalizing groups and practices. The ALCF continued moving toward a centralized cyber security team at the directorate level. This team enables security staff members at the CELS level to work more closely with ALCF's security staff as a single team. The team currently contains three full-time equivalents (FTEs) with the flexibility to grow and expand to meet the future needs of the facility. These changes do not affect the day-to-day work of ALCF's security staff as they are integrated into the facility operations work. This shift is intended to give more coverage and spread out tasks to allow cybersecurity engineers to focus better on their areas of expertise. The Security Team will continue working to formalize cyber security documentation, policies, procedures, and practices to help guide the facility in its decision-making processes. The CELS Security Team leadership holds regular meetings with the Cybersecurity Program Office (CSPO) Leadership to collaborate on ongoing security efforts.

A large security effort in 2024 was working with CSPO in preparations for the Enterprise Assessment Audit that took place in October 2024. This involved lab-wide efforts to reduce existing vulnerabilities and develop strategies for reducing vulnerabilities long term with the hygiene improvement project. In addition to the vulnerability remediations, the ALCF also engaged a third-party contractor to complete a full evaluation of the Major Application Certification Package, which included the NIST 800-18, NIST 800-30, NIST 800-34, and NIST 800-53, and migrate these from NIST Rev 4 to NIST Rev 5. This was an extensive project that required approximately 2,000 hours of work from the contracting team over an 8-week period. During this time, ALCF security worked through documented policies, procedures, and practices with the team to ensure that security documentation was thorough and would meet the high standards of the Enterprise Assessment Team.

In CY 2024, the ALCF continued to improve its cybersecurity practices to ensure an effective cybersecurity program. It also gained access to improved tools, both locally and at the laboratory level, to increase cybersecurity capabilities. The 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 the ALCF's effective cyber security practices.

In CY 2024, there were zero cybersecurity incidents on ALCF systems. The ALCF's Cybersecurity Team adopts a proactive approach to problem management. Examples of efforts in CY 2024 include:

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 with internal discussions/reviews and external training courses.
5. Integrate security auditing into ALCF 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. Manage 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 recommended 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 the ALCF's scheduler software 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. Collaborate with the ALCF's system administration teams on configuration management tooling and configurations of various systems.

Additionally, the 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 the ALCF with additional protection from attacks on Argonne networks as they are detected.
3. CSPO provides Cloudflare, a powerful web application firewall (WAF), which protects the ALCF's public sites from known attacks and ensures that encryption requirements are met.
4. CSPO provides a Tenable security center (Tenable.sc) instance that the ALCF can tie Nessus scanners into to leverage 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 the 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 to help reduce vulnerabilities in the environment.

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 2024, 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 logs showed no unauthorized access. The ALCF changed the way passwords were stored and changed all related passwords as a precaution against potential exploitation.
2. **Issue:** Secret values were committed into a Git repository.
   Mitigation: Evaluation was performed to determine that no unauthorized users were able to access the credentials. The ALCF changed all secrets and worked with individuals to emphasize the importance of reporting these issues and to thank them for their diligence in reporting the event.
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, the ALCF was able to identify the person behind these attempts, and they were informed that their actions were not acceptable.
4. **Issue:** ALCF staff identified critical security fixes for upstream services by monitoring security feeds.
   Mitigation: ALCF staff members worked with system administrators to ensure that critical patches were deployed in reasonable timeframes.

The ALCF will continue to proactively investigate security issues and monitor and respond to all vulnerabilities. Plans for improving the security of ALCF's 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.

The 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. Continuing the ALCF-specific cybersecurity working group to bring all operational teams together to ensure cybersecurity coverage in all aspects of ALCF operations.
2. Creating a modernized incident response plan specific to the ALCF that will align with the Argonne-wide incident response plan.
3. Developing a framework for a cybersecurity-specific risk registry for the ALCF that will use risk modeling frameworks to quantify the level of risk.
4. Expanding in-house security expertise by hiring two roles within the CELS directorate. The first is filling the Cybersecurity Lead at the CELS level who will assist the ALCF. The second is a new Security Engineer position that will be able to expand into a variety of areas.

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 2024, a DSAV was not completed, as the lab's security team was in direct and constant contact with the divisions and directorates to ensure everyone was prepared for the Enterprise Assessment Audit. This looked different for each division based on what their individual needs were. For ALCF, this was largely covered under the NIST Major Application Security Documentation, which involved mapping all existing NIST 800-53 Rev 4 controls to NIST 800-53 Rev 5 controls and updating them appropriately. This also involved adding new NIST 800-53 Rev 5 controls to ensure that we are in line with the latest NIST standards. In addition to reviewing the NIST 800-53 controls, the contracting team also worked on updates to NIST 800-18, NIST 800-30, and NIST 800-34.

ALCF looks forward to partnering with CSPO to continue the DSAV process in 2025. 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 July 19, 2024. It remains valid as long as Argonne maintains robust, continuous monitoring of its Cyber Security Program, as provided in the ATO letter, which is included at the end of this section.

7.3 Foreign Visitor Access and Export Controls

The 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. The 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 approved and active 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 completes a secure web form in the ALCF Account and Project Management system, Userbase 3 (UB3), providing details such as their legal name, a valid email address, work address, phone number, and country of citizenship. They also specify the ALCF project with which they are affiliated. Additionally, all foreign nationals (non-U.S. citizens) must 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 email is sent to the user's project PI for approval. If the user is a foreign national, their details are electronically attached to an ANL-593, the Foreign National Access Request form, and submitted to the Foreign National Access Program (FNAP) Office for review. The Argonne FNAP Office is responsible for overseeing compliance with laboratory adherence to DOE directives related to foreign national access to our site and systems.

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 process for handling foreign visitors and assignments integrates with the DOE Foreign Access Central Tracking System (FACTS), which documents and tracks access control records for international visits, assignments, and employment at DOE facilities and contractor sites. Once the user's ANL-593 is approved, the UB3 database is automatically updated with the user's ANL-593 start and end dates. The ALCF User Experience Team then creates the user account and notifies the user. Any changes to the ANL-593 dates are automatically reflected in UB3. Accounts are suspended when the user's ANL-593 expires.

The ALCF permits only a limited subset of export control data on its systems. It collaborates closely with Argonne's Export Control Office to develop a detailed security plan outlining the permitted export control classifications and the security measures required for each instance of export-controlled data. If the ALCF wishes to introduce new classifications of export control data on its systems, a new security plan must be created and approved by both 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

July 19, 2024

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

Reference: Letter, R. Snyder to P. Kerns, dated April 26, 2023, 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 2023, 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 NIST SP 800-53 Revision 5 security controls annually on a rotating basis as part of the self-assessment program.

This has demonstrated that the Laboratory's IT Infrastructure is operating at an acceptable level of risk, and I am, therefore, as the Authorizing Official, renewing the Authority to Operate for the General Computing - Low enclave at FIPS-199 Low and General Computing - Moderate enclave at FIPS-199 Moderate. 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)
- Argonne Biomedical Learning Enclave (ABLE)

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 report the status of any open Plan of Action and Milestones associated with this ATO during the continuous monitoring meetings. Failure to do so or complete actions timely may result in revocation of this authorization. The Laboratory should retain a copy of this letter with its 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 (331) 444-0595.

Sincerely,

Whitney Begner-Romozzi
Digitally signed by Whitney Begner-Romozzi
Date: 2024.07.19 15:05:40 -05'00'

Whitney S. Begner-Romozzi
Manager

cc: M. Wisniewski, DOE-OIM
F. Healy, DOE-OIM
S. Hannay, ANL
M. Skwarek, ANL
R. Denney, 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 scientific accomplishments of the INCITE, ALCC, and DD projects demonstrate ALCF's impact in facilitating scientific breakthroughs. In CY 2024, ALCF staff members effectively collaborated with individual project teams to adapt their simulation, data analytics, and machine learning codes to run efficiently in the ALCF environment, leading to scientific achievements that would not have been possible otherwise.

In this section, the ALCF reports:

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

8.1 Science Highlights and Accomplishments

The ALCF employs various methods and processes to monitor scientific accomplishments. Monthly scientific highlights, primarily from the Catalyst Team, are collected and documented in a quarterly report. The determination and coordination of these highlights are managed by ALCF's Applications Team, which consists of members from the Catalyst Team, the Data Science Team, and the Performance Engineering Team, in consultation with ALCF's Director of Science. Other sources of scientific highlights include technical communications between ALCF staff and a project PI or co-PI, significant findings reported in high-impact publications or conference presentations, and a Catalyst's involvement in a publication.

The ALCF tracks and annually reports the number of peer-reviewed publications that result (in whole or in part) from the use of ALCF's resources. Tracking occurs over a five-year period following the completion of the project. This may include publications that are in press or accepted for publication, but does not include papers that have been submitted or are in preparation. The count is a reported figure rather than a metric. The facility may report additional publications where appropriate. Methods for gathering publication data include requesting users to verify or update the ALCF's online publications database and ALCF staff conducting Google Scholar and Crossref searches.

Table 8.1 presents a breakdown of refereed publications that are based, in whole or in part, on the use of ALCF resources and highlights those published in major journals and proceedings as of March 18, 2025. This includes one publication in *Nature*, four in *Nature Communications*, one in *npj Climate and Atmospheric Science*, two in *npj Computational Materials*, two in *Scientific Data*, one in *Scientific Reports*, one in *Nature Biomedical Engineering*, one in *Nature Nanotechnology*, and one in *Communications Chemistry* (grouped in the *Nature* journals category in the table below); two in *Science Advances* (listed under the *Science* journals category in the table below); three in the *Proceedings of the National Academy of Sciences (PNAS)*; eight in *Physical Review Letters (PRL)*; one in the *39th International Conference, ISC High Performance 2024 (ISC)*; and four in the proceedings of the *2024 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*. Table 8.2 reflects updated publication counts from prior years based on new information received after the previous year's OAR deadline.

Table 8.1 Summary of Users' Peer-Reviewed Publications in CY 2024

Nature Journals	Science	PNAS	Physical Review Letters	ApJL	Nano Letters	ISC	SC	Total 2024 Publications
14	2	3	8	3	1	1	4	254

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

OAR Year	CY 2020	CY 2021	CY 2022	CY 2023	CY 2024
Total Publications	257	249	213	240	254

Science Highlights

Scientific highlights are short narratives that illustrate the ALCF'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 total allocation, 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 Demonstrating Cross-Facility Data Processing at Scale with Laue Microdiffraction

The Science

Laue microdiffraction is a technique used at synchrotron facilities, such as the APS, to examine the local structure of materials with crystalline formations. This method demands extensive computation: a single CPU workstation would take over 250 days to analyze data from a 12-hour scan. With the recent upgrade at APS providing X-ray beams that are 500 times more powerful than before, the volume of experimental data produced will increase correspondingly. A team from the APS, the ALCF, and Globus has developed high-performance software tools and data infrastructure to integrate APS data processing and analysis with ALCF computing resources in near-real-time using GPU-based software.

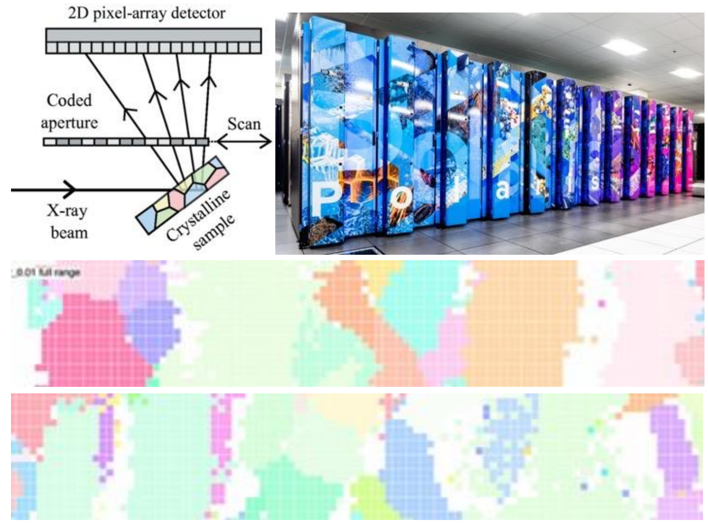
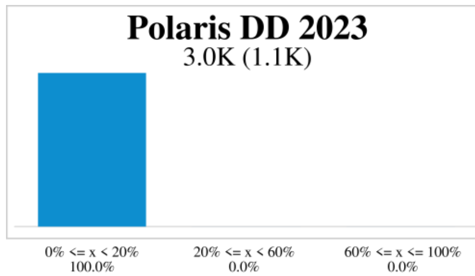


Figure 8.1 Data from the new coded aperture at an APS beamline (top left) is automatically transferred to the Polaris supercomputer (top right), where it is reconstructed on demand in real time (bottom).

Image courtesy of Michael Prince et al. (Argonne).



The Impact

The team demonstrated the on-demand use of ALCF resources to reconstruct data obtained from the APS beamline, returning reconstructed scans to the APS within 15 minutes of transfer to the ALCF (Figure 8.1). With ALCF resources, the APS is able to reconstruct coded aperture Laue datasets automatically in near-real-time, allowing users to collect data approximately an order of magnitude faster than is possible today, accelerating time to science. Building a common infrastructure such as this enables the APS, and potentially other facilities, to use ALCF supercomputing resources.

Summary

As a focused X-ray beam passes through a material in the beamline, individual crystallites along the path of the beam diffract along different angles depending on their orientation. Obtaining a full 3D map of the structure requires resolving the angle and position of each diffracted beam. Instead of using a very time-intensive scan to complete the analysis, APS users employ coded aperture Laue microdiffraction. Decoding each beam's sequence can be performed in parallel as a post-processing step, where each of the ~400 analysis points (where each point consists of 300–400 images) is collected every 60–90 seconds and queued as a separate job. While suitable for completion on large-scale computational resources, the post-processing adds computational expense.

Near-ideal scaling for the workflow has been measured with up to 100 nodes. However, during continuous use of up to 50 nodes on Polaris, the system using the new GPU-based algorithm managed to process the scans as they came in every 1–2 minutes throughout 6- to 12-hour runs. Before the APS shutdown and after implementing some code optimizations, the team demonstrated the need for 40–45 of the 56 available nodes to meet the demand of the beamline. The team is currently developing user-facing web portals to help beamline scientists and APS users manage their data and launch reconstructions more easily.

ALCF Contribution: ALCF staff worked closely with the team enabling the infrastructure (e.g., PBS on-demand queues) to allow on-demand use of Polaris resources.

Contact
Nicholas Schwarz
Argonne National Laboratory
Email: nschwarz@anl.gov
ASCR Allocation PI: Nicholas Schwarz, Argonne National Laboratory

Publication
[1] M. Prince, D. Gürsoy, D. Sheyfer, R. Chard, B. Côté, H. Parraga, B. Frosik, J. Tischler, and N. Schwarz, “Demonstrating Cross-Facility Data Processing at Scale with Laue Microdiffraction,” *Proceedings of the SC-W 2023*, 2133, Nov. 12–17, 2023, Denver, CO, USA. DOI: 10.1145/3624062.2634613.

8.1.2 Efficient Algorithms for Monte Carlo Particle Transport on AI Accelerator Hardware

The Science
AI accelerators, such as the Cerebras Wafer-Scale Engine 2 (WSE-2), are custom-made to efficiently perform AI training tasks at high speed. While not intended for traditional modeling and simulation workloads, these accelerators are nevertheless attractive for use with some simulation algorithms. A team developed new algorithms and performance optimization strategies to enable a key Monte Carlo (MC) particle transport simulation kernel to use the device effectively. Speedups of 182× over a single GPU’s performance were observed for the algorithm using a production-quality workload typical of a nuclear reactor simulation.

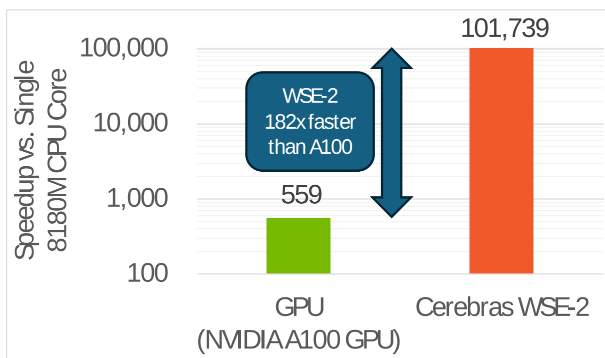


Figure 8.2 Speedup vs. serial CPU execution for macroscopic cross-section lookup kernel (adapted from XSBench).
Image courtesy of John Tramm (Argonne).

The Impact
Significant speed and power advantages over highly optimized CPU and GPU implementations were identified, indicating that acceleration of a full MC particle transport code on WSE-2 is

feasible. The team developed and assessed new algorithms to minimize communication costs and manage load balancing that could facilitate the creation of optimized algorithms for related simulation methods. AI accelerators, such as the WSE-2, can provide substantial benefits to traditional simulation workloads, and the advancement of higher-level programming models to enhance software development and exploration could greatly impact HPC simulations.

Summary

MC particle transport is a method for simulating the behavior of particles as they move through (and interact with) a medium. In MC neutron particle transport, the cross-section lookup kernel (Figure 8.2) accounts for the majority of the runtime. This lookup kernel assembles data in a number of tables depending on several properties of the particle and material: energy of the neutron, isotopic composition of the material the particle is traveling through, and the temperature of the material. The XSBench mini-app, which includes a simplified implementation of the macroscopic cross-section lookup kernel, was ported to different architectures. An optimized binary search acceleration method for the Cerebras architecture was also developed to improve the performance of the lookup kernel.

A new hyper-domain decomposition technique more effectively spread simulation data across >700K processing elements of the WSE-2 and minimized communication costs for moving particle data. Another important aspect of the WSE-2 hardware is the amount of low-latency memory available compared to typical CPUs and GPU architectures. With these changes, the WSE-2 was found to deliver speedups of 182× for the cross-section lookup algorithms compared to a highly optimized CUDA version and an overall 3.2× power efficiency gain over the GPU.

ALCF Contribution: The team worked closely with ALCF staff to get started on the Cerebras system and on developing code with the CSL Cerebras SDK.

Contact

John Tramm
Computational Science Division, Argonne National Laboratory
jtramm@anl.gov
ASCR Allocation PI: John Tramm (Argonne)

Publications

[1] J. Tramm, B. Allen, K. Yoshii, A. Siegel, and L. Wilson, “Efficient algorithms for Monte Carlo particle transport on AI accelerator hardware,” *Comput. Physics, Commun*., 298, 109072 (2024).

[2] J. Tramm, B. Allen, K. Yoshii, and A. Siegel, “Monte Carlo with Single-Cycle Latency: Optimization of a Continuous Energy Cross Section Lookup Kernel for AI Accelerator Hardware,” in: PHYSOR 2024 – International Conference on Physics of Reactors, San Francisco, USA.

8.1.3 Computer Simulation-Aided Design of High-Performance Electrolytes for Lithium Batteries

The Science

For scientists working on future lithium-battery chemistries with high energy densities, designing a robust protective “interphase” formed by sacrificial electrolytes becomes the driver for developing better electrolytes. Experiments point to electrolyte fluorination as a promising solution to create such an interphase. However, designing a fluorination strategy without relying on demanding lab implementations remains the challenge. Researchers have unlocked the fundamental mechanism of fluorination effects on battery stability, lifetime, and capacity through extensive molecular dynamics (MD) simulations under realistic experimental conditions. This study opens a broad design space for next-generation electrolytes for these coveted battery chemistries.

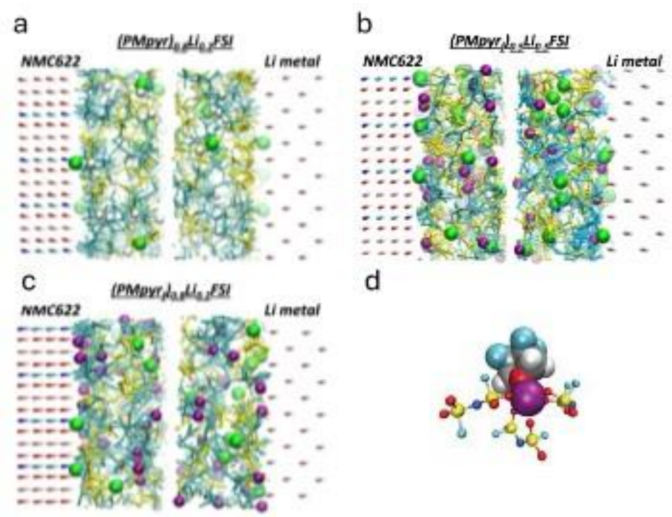
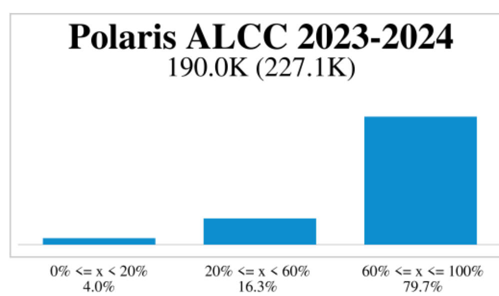


Figure 8.3 **(a–c) Simulation snapshots of electrolyte distribution on NMC62 and Li electrodes: cyan-IL cation; white-H on the PMpyr+ backbone; purple-F on the PMpyrf+; yellow-FSI–; green-Li+. (d) A representative snapshot of Li+ ion solvation structure with TfTHF that results in better mobility.**

Image courtesy of Wei Jiang (Argonne).

The Impact



Fundamental knowledge of the prevailing fluorination process, particularly at the atomic scale, will speed up the discovery of new electrolytes without extensive wet lab synthesis and validations, and minimize related hazards. Developing a high-fidelity, all-atom simulation, and modeling methodology that can predict fluorination effects on the interphase formation and charge conduction, as well as physicochemical mechanisms within a realistic electrolyte-electrode matrix, represents a key step toward the rational design of fluorinated electrolytes (Figure 8.3). The novel all-atom models for the cathode and anode and the ensemble-enhanced sampling algorithm open the door to the virtual characterization of fluorination effects on interphase formation and lithium-ion transport properties.

Summary

Novel all-atom models for high-capacity Li-metal anodes and Mn-Ni-O cathodes are adopted to obtain atom-scale characterization of interface-related structural changes induced by fluorination. Simulations with enhanced sampling enabled studying the dynamics of viscous electrolytes that require large timescales beyond brute-force simulations. Multiple copies (replicas) of each simulated system were generated with different initial configurations, and each replica annealed the interactions of the electrode-electrolyte complexes within a predesigned range of effective temperatures. The MD simulations, featuring various combinations of fluorinated/nonfluorinated electrolytes and diluents, accurately identified where, how, and why fluorinations facilitate the formation of protective interphase and boost the mobility of charge carriers. Strong electrostatic attractions of fluorine and electrodes that induce tight electrolyte interface structure are regarded as the prelude to robust interphase, whereas looser interface structure without fluorination or with suboptimal fluorination position risks interphase instability. The statistical analysis revealed the need to apply fluorinations to specific positions of an electrolyte molecule and/or specific molecular structures of diluents to achieve the best performance gains — offering a path toward rational design of fluorinated electrolytes for future lithium battery fabrication.

ALCF Contribution: ALCF staff maintained the NAMD code and sampling methods and ensured optimal usage of the compute resources by coordinating production runs; implemented modeling and parameterization of the electrodes for MD simulations; managed the massive simulation data; and advised on the analysis of MD trajectory data.

Contact

Zhengcheng Zhang
Chemical Sciences and Engineering Division, Argonne National Laboratory
zzhang@anl.gov
ASCR Allocation PI: Wei Jiang (Argonne)

Publications

[1] Q. Liu, J. Xu, W. Jiang, J. Gim, A.P. Tornheim, R. Pathak, Q. Zhu, P. Zuo, Z. Yang, K.Z. Pupek, E. Lee, C. Wang, C. Liu, J.R. Croy, K. Xu, and Z. Zhang, “High-Energy LiNiO2 Li Metal Batteries Enabled by Hybrid Electrolyte Consisting of Ionic Liquid and Weakly Solvating Fluorinated Ether,” *Adv. Sci*., 2409662 (2024).

[2] Q. Liu, W. Jiang, J. Xu, Y. Xu, Z. Yang, D.-J. Yoo, K.Z. Pupek, C. Wang, C. Liu, K. Xu, and Z. Zhang, “A fluorinated cation introduces new interphasial chemistries to enable high-voltage lithium metal batteries,” *Nat. Commun*., 14, 3678 (2023).

8.1.4 A Theory for Neutron Star and Black Hole Kicks and Induced Spins

The Science

Supernovae shape our universe, yet how a massive star explodes as a supernova remains a mystery. NASA's James Webb Space Telescope provides unprecedented views of supernovae remnants, but scientists need to peer deep inside massive stars to understand the mechanisms behind these distant explosions. The project team generated one of the largest collections of 3D supernova simulations, aiming to illuminate how and why a massive star goes supernova.

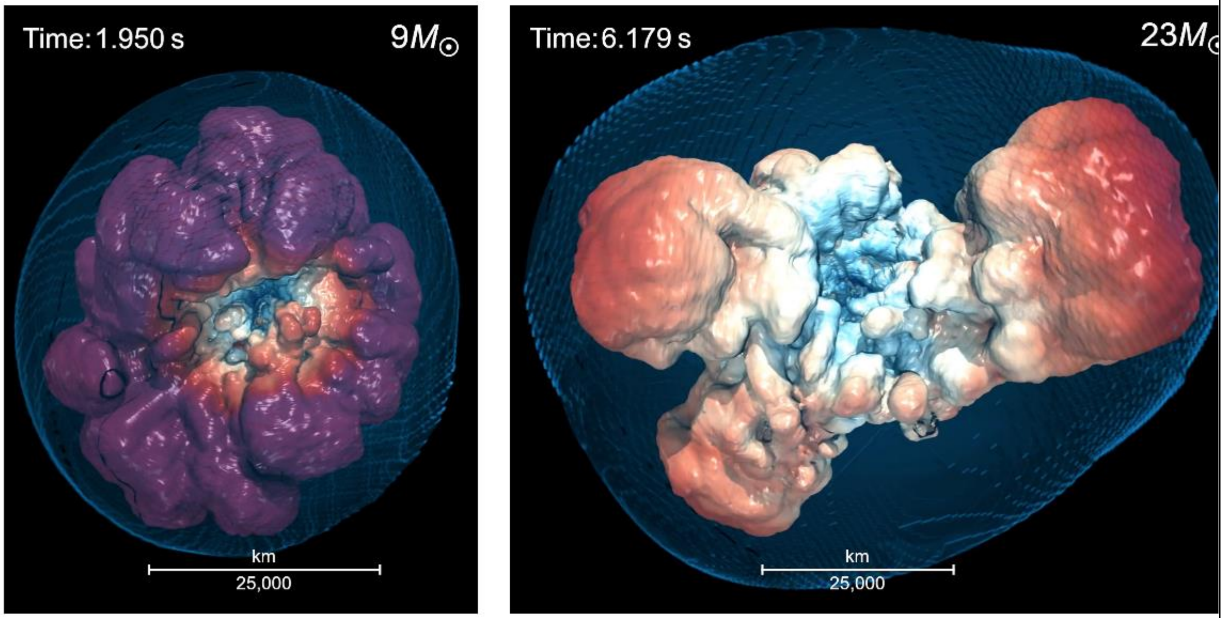
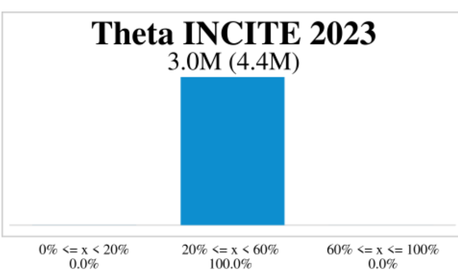
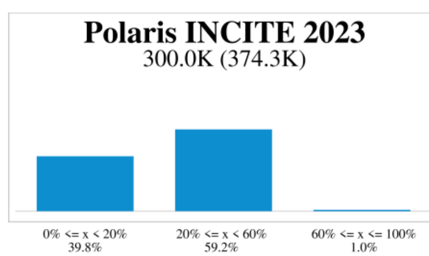


Figure 8.4 An isosurface (10%) of the 56Ni abundance in the explosion of a 9 M⊙ (left) and 23 M⊙ (right) progenitor model at ~1.95 and ~6.2 seconds, respectively, after the bounce, painted by Mach number. Red/purple is positive and white/blue is negative, indicating infall.

Image courtesy of Joseph Insley (Argonne).





The Impact

Using 20 long-term 3D core-collapse supernova simulations, the team found that progenitors with lower compactness that explode quasi-spherically due to the short explosion delay experience smaller neutron star recoil "kicks," whereas progenitors with higher compactness that explode later and more aspherically leave neutron stars with larger kicks (Figure 8.4). In addition, the researchers found that these two classes correlate with the neutron star's gravitational mass, suggesting that the survival of binary neutron star systems may partly result from their lower kick speeds. This new 3D model suite provides a greatly expanded perspective and appears to explain some observed pulsar properties by default.

Summary

Core-collapse supernova explosions accompany the deaths of massive stars, leaving neutron stars, black holes, and solar mass ejections of heavy elements as artifacts. Although determining the mechanism of explosion has eluded science for half a century, the delayed neutrino-heating mechanism is emerging as a robust solution. Nevertheless, the models must lead not only to explosions: the asymptotic state of the blast must also be reached to determine many of the observables — all as a function of progenitor mass. The most comprehensive and well-developed theory for pulsar and black hole kicks at birth involves recoil due to the asymmetrical supernova explosion and accompanying aspherical neutrino emissions. Simultaneous accretion and explosion lead to interactions between ejecta and accreta, which influence the kick velocity. Once this process ceases seconds later, the core is in its final asymptotic kick state. For the first time, using a large and uniform collection of 3D supernova models ranging from 9- to 60-solar-mass stars, researchers obtained perspective on the overall dependence of the recoil kicks and induced spins on progenitor mass.

ALCF Contribution: The ALCF Catalyst provided support to transition the code to Polaris and is working with the INCITE team to port the Fornax GPU version to Aurora. Staff members from the ALCF Visualization and Data Analysis Team contributed to producing images for publications and presentations.

Contact

Adam Burrows
Professor of Astrophysical Sciences, Princeton University
burrows@astro.princeton.edu
ASCR Allocation PI: Adam Burrows (Princeton University)

Funding

Main sources of funding from: SciDAC4 program (DE- SC0018297, subaward 00009650), U.S. NSF AST-1714267 and PHY-1804048 grants (the latter via the Max-Planck/Princeton Center [MPPC] for Plasma Physics).

Publications

[1] A. Burrows, T. Wang, D. Vartanyan, and M. S. B. Coleman, “A Theory for Neutron Star and Black Hole Kicks and Induced Spins,” *The Astrophysical Journal*, 963, 63 (2024).

[2] A. Burrows, T. Wang, and D. Vartanyan, “Physical Correlations and Predictions Emerging from Modern Core-Collapse Supernova Theory,” *The Astrophysical Journal Letters*, 964, L16 (2024).

8.1.5 Three-Dimensional Shock Boundary Layer Interactions Over Flexible Walls

The Science

Supersonic and hypersonic (greater than five times the speed of sound) flows are often characterized by shocks or sharp discontinuities in the flow properties. When a sufficiently strong oblique shock impinges on a turbulent boundary layer, it creates a flow separation bubble that can oscillate in size and location and be detrimental to the structure on which the shock impacts. Developing a better understanding of shock/boundary-layer interactions (SBLIs) with coupled fluid-structure interaction (FSI) is important for the design and development of hypersonic aircraft (Figure 8.5).

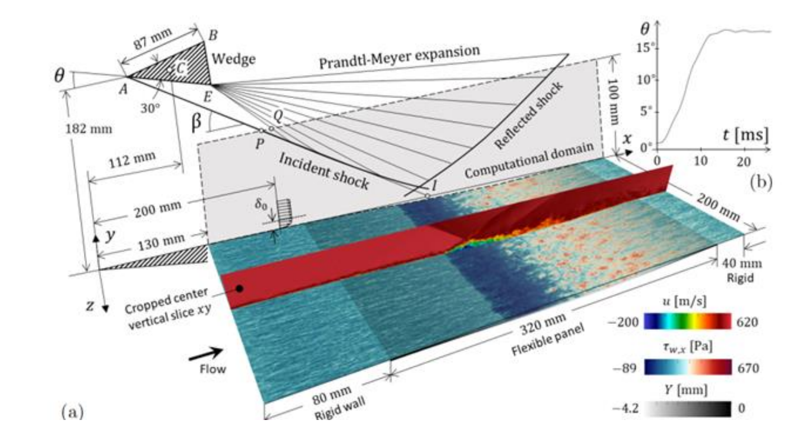
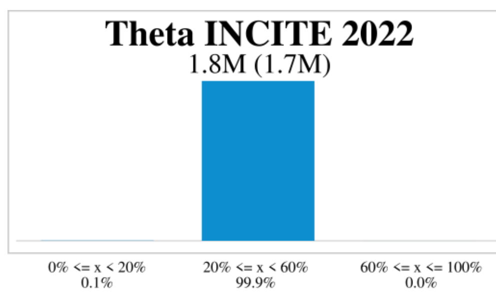


Figure 8.5 The computational setup of the full-span 3D SBLI simulations. By oscillating the wedge, the strength of the SBLI can be increased to study the effects of flow separation and that of an oscillating separation bubble coupled with the fluid-structure interaction. Simulation details compared well with measurements from flexible-panel experiments.

Image courtesy of Ivan Bermejo-Moreno (University of Southern California).

The Impact



The project team used Theta to perform full 3D simulations of the coupling between FSI and SBLI to replicate and complement wind tunnel experiments. This study is the first time in the open literature to incorporate wall-modeled large eddy simulations (WMLESs) and a Finite Element Method (FEM) solid mechanics solver to study the low-frequency motions that are characteristic of SBLI+FSI interactions on flexible panels. The use of WMLES in the flow solver greatly reduces the computational cost, relative to wall-resolved simulations, by modeling (rather than resolving) the effects of dominant turbulent motions in the very-near wall region. This optimization enables feasible computations for much longer times and thus for meaningful studies of the coupled interactions of flow and structure at the low frequencies that can lead to structural failure. This modeling strategy is also a steppingstone toward incorporating thermal coupling effects between the fluid and solid response, which often occurs at even longer time scales. Ultimately, these predictive simulation techniques will enable the development of more aggressively designed hypersonic vehicles by reducing the required safety factors due to simulation uncertainties.

Summary
Much of the allocation was used for completing a detailed study of the effects of FSI on SBLI over flexible walls. The WMLES simulations were conducted with the unstructured grid finite-volume/finite-element uPDE solver that was developed by the project co-PI. Validation of the high-fidelity simulation methodology was pursued by using WMLES to replicate experiments at different strengths of the incident (oblique) shock on the turbulent boundary layer (i.e., different flow Mach numbers, shock angles, flexible panel properties). The results were also used to assess the importance of 3D effects in the tested STBLIs by carrying out reduced-span simulations with imposed periodicity in the spanwise direction. The simulations replicated these coupled interactions observed experimentally with better accuracy than prior numerical studies, while also complementing the wind-tunnel experiments by providing additional key insights. The computations on Theta included a combination of large-scale simulations to characterize the 3D effects of the interaction over full-span panels matching the experimental configuration, and ensemble jobs of multiple span-wise periodic simulations on a reduced span width for parametric studies to evaluate the effects of, for example, the incoming boundary layer thickness, the compression wedge extent, and oscillation frequency.

ALCF Contribution: The ALCF Catalyst Team and Support staff provided timely responses to questions pertaining to the code's performance on Theta and the project team's jobs in the queue. The Catalyst also shared the HyPar_GPU ProxyApp for future code developments on Aurora.

Contact
Ivan Bermejo-Moreno
University of Southern California
bermejom@usc.edu
ASCR Allocation PI: Johan Larsson (University of Maryland)

Publications
[1] J. Hoy, I. Bermejo-Moreno (2022), "Fluid-Structural coupling of an impinging shock-turbulent boundary layer interaction at Mach 3 over a flexible panel," *Flow*, vol. 2, p. E35.
[2] J. Hoy, I. Bermejo-Moreno (2021), "Numerical study of STBLI on flexible panels with wall-modeled LES," AIAA SciTech 2021 Forum.

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 HPC community. 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 the ALCF, the JLSE centralizes Argonne's research activities aimed at evaluating future extreme-scale computing systems, technologies, and capabilities.

The list of testbeds supported at the JLSE and the projects that utilize them can be viewed on the newly designed website in 2024 at https://www.jlse.anl.gov.

In 2024, the JLSE added two Supermicro servers with two Intel 5th generation Xeon CPUs and eight Intel Data Center GPU Max 1550 GPUs each. These servers, consisting of PRQ parts and publicly available oneAPI software and drivers and requiring no NDA for access, are a steppingstone for projects intended to be run on Aurora. The JLSE also added NVIDIA's Grace Hopper (GH200) server with Grace Arm CPU and Hopper 100 GPU, and a Vast filesystem, replacing the old GPFS filesystem, for projects to use as scratch.

In 2024, the JLSE supported more than 400 users spanning more than 175 projects. The new projects that started in 2024 are listed below; a complete list of all active projects is available on the JLSE website:

- Scalable Accelerator System Memory Management – Explores performance and design of scalable memory management in accelerated computing.
- X-CELLENT – LLVM compiler and OpenMP runtime development for Intel, AMD, and NVIDIA GPUs.
- AI.Cure – Aims to build advanced AI models using the latest hardware for outcome predictions to compress the time needed from diagnosis to treatment for all cancers.
- LQCD – For testing, benchmarking, and evaluation of Lattice QCD codes on JLSE testbed architectures.
- AGHMC – A test using auto-grad methods for tuning HMC parameters in LQCD simulations.
- Scalable and Resilient Modeling for Federated-Learning-Based Complex Workflows – The outcome of this research will be a scalable and resilient FL simulation and modeling system that can support efficient simulation and modeling of complex workflow and related resilience issues during federated learning.
- UIC-HPC – A collaboration between Argonne and UIC graduate students and their research.
- SYCL Benchmarking – Evaluates the SYCL programming model for batched kernels with memory allocations to propose new implementations.

- NucStructPorting – INCITE project members porting to Intel GPUs and Aurora.
- CFM – Is developing a novel approach to continually pretrain a large language model. This project has two parts: first, we seek to develop a federated learning-driven approach to continually fine-tune a foundation model approach. Second, we seek to develop a foundation model for a molecular dynamics dataset.
- Porting the Fire Dynamics Simulation software suite to PVC – This project will advance the development of the open-source, Fire Dynamics Simulator (FDS) (https://github.com/firemodels/fds) and its companion visualization software, Smokeview, to run on the latest generation of HPC systems.
- ScaleStuds – The goal of this project is to understand the performance of various computational science kernels across a number of diverse architectures and project their performance on other architectures.
- Compact Binary Merger INCITE – An INCITE 2025 project that aims to port AthenaK code for simulations of binary black hole and binary neutron systems to PVC GPUs.
- Development and optimization of the NekRS CFD solver on accelerated platforms – Testing and optimization of NekRS across a wide range of hardware.

8.2.2 Research Activity - ALCF AI Testbed

With an eye toward the future of scientific computing, the ALCF has deployed an advanced AI platform testbed for the research community. The testbed enables the ALCF and its user community to help define the role of AI accelerators in next-generation scientific applications. It also helps shape vendor roadmaps and the 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. The ALCF actively worked with several AI accelerator systems and plans to include new systems in the testbed. The systems from Cerebras, Graphcore, Groq, and SambaNova are available via the Director's Discretionary allocation program to the ALCF user community. The AI Testbed systems were also made available to researchers through the National Artificial Intelligence Research Resource (NAIRR) Pilot program, and ALCF had multiple project allocations via the NAIRR program. In CY 2024, ALCF actively collaborated with new AI accelerator vendors, including D-Matrix, Esperanto, 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. The ALCF has conducted several user workshops on the AI testbed. The ALCF held a full-day tutorial at SC24 and two Birds of a Feather (BoF) sessions 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 (static random access 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: The 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, many opportunities have arisen for lessons learned regarding the 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. The ALCF has conducted a detailed evaluation to understand the performance of pre-training that considers key factors such as sequence lengths, scaling behavior, energy efficiency and sensitivity to gradient accumulation steps. ALCF has worked closely with various AI vendors for this study. (See: https://doi.org/10.1109/IPDPSW63119.2024.00016)
- As the ALCF increases support for AI for Science, AI inference is seen as a key application workload. Benchmarking the performance of inference of large language models (LLMs) across diverse hardware platforms is crucial to understanding their scalability and throughput characteristics. The ALCF developed LLM-Inference-Bench, a comprehensive benchmarking suite to evaluate the hardware inference performance of LLMs. The ALCF thoroughly analyzes diverse hardware platforms, including GPUs from NVIDIA and AMD and specialized AI accelerators Intel Habana and SambaNova. Our benchmarking results reveal the strengths and limitations of various models, hardware platforms, and inference frameworks. (See: https://doi.org/10.1109/SCW63240.2024.00178).
- 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 same model code that runs on a GPU. 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 the ALCF expects improved support going forward.
- Given the long compilation times for models, strategies wherein users can trade-off the compilation times to the performance of generated code have 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. The ALCF has worked closely with the AI vendors to incorporate these, and it has improved developers' productivity.
- AI Testbed architectures, including dataflow architectures, are very attractive for traditional HPC applications. The Monte Carlo XSBench Lookup has achieved over a 100x improvement in performance using a CS-2 compared to an A100. (See: https://doi.org/10.1016/j.cpc.2023.109072).

8.2.3 Vendor Engagements for Future Systems

The ALCF-4 project represents a strategic DOE Office of Science initiative to deploy the next-generation leadership computing capability at Argonne, scheduled to replace the Aurora system in 2030. This system will be essential to sustain U.S. leadership in advanced computing while addressing the evolving computational needs of the scientific community, particularly in areas of AI, modeling/simulation, data analytics, and complex scientific workflows.

As part of the project's planning and requirements-gathering phase, the ALCF-4 Leadership Team has conducted extensive engagements with leading technology vendors throughout 2024 and early 2025. These discussions have been instrumental in understanding the evolving technological landscape and evaluating potential pathways to meet the ALCF's mission requirements, which include:

- Ensuring continuity of leadership computing capabilities.
- Supporting both traditional HPC workloads and emerging AI-driven scientific discovery.
- Enabling seamless integration with DOE's Integrated Research Infrastructure (IRI).
- Advancing energy efficiency in high-performance computing.
- Facilitating high-impact, grand-challenge science and engineering.

The breadth of vendor engagement has been comprehensive, spanning major hardware manufacturers (AMD, Intel, NVIDIA), system integrators (HPE, Dell, Penguin, SuperMicro), cloud service providers (Microsoft, Amazon, Google), and specialized technology providers (Cerebras, SambaNova, ARM, WEKA, Lightmatter, Next Silicon). These discussions have revealed promising developments in several critical areas that align with our mission requirements.

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 the direction of 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 2024, staff participated in 161 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 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 a daily basis. ALCF staff serve on committees for NERSC and OLCF projects. The ALCF has worked closely with the APS as we engage how ALCF resources will support some of their needs with the APS Upgrade. The ALCF and the OLCF jointly run the flagship allocation program, INCITE, which is a year-round collaboration.

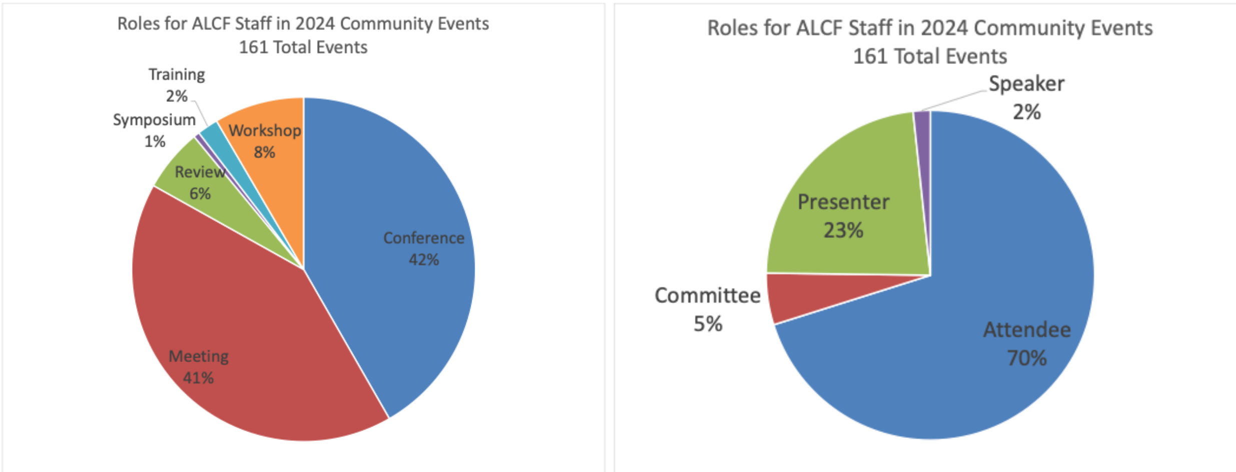


Figure 8.6 Breakdown of key activities by the ALCF in CY 2024. The first pie chart (left) breaks down the 161 events by type, primarily based on how the event identified itself. The second pie chart (right) breaks down the same events by the staff member's role.

8.3.1 Engagement Highlights

Supercomputing 2024

SC is a pivotal event in supercomputing, encompassing every aspect of the field. Participation serves as a major opportunity to document and share essential knowledge. The ALCF participates prominently in the event, as indicated in Table 8.3.

Table 8.3 Summary of ALCF Participation in SC24

Program	Total
Tutorials	2
Workshops	14
Posters	4
Birds of a Feather	9
Booth Talks	1
Papers	4

Summary of Engagements with the Exascale Computing Project

The ECP, which operated from 2016 to 2024, held its final CD-4 Independent Project Review (IPR) in April 2024. The ALCF continued to provide ECP access to Sunspot and Aurora from January to December 2024. However, access to Aurora was removed during the Top500 runs in the spring and several months in the fall due to acceptance testing.

A total of 43 ECP projects used allocations on ALCF resources in CY2024, and all three of ECP's broad research topics of Application Development, Software Technology, and Hardware and Integration were represented within that set of projects. Together, they accounted for 188.6K node-hours on Aurora, 8.2K node-hours on Sunspot, 2.3K node-hours on Polaris, and 15.0K node-hours on various testbed systems in Argonne's JLSE.

In the first half of CY 2024, four ALCF staff members allocated varying percentages of their efforts to ECP, mainly for project management closeout tasks. ECP Hardware and Integration Deputy Director Susan Coghlan managed ECP access to the Sunspot and Aurora resources throughout the year, prepared materials for and participated in the CD-4 IPR, and prepared materials for and participated in the ECP Science Expo on Capitol Hill. ECP Industry Council Co-Executive Director David Martin and ECP Hardware and Integration/Facility Resource Utilization subproject lead Haritha Som also prepared presentation materials and participated in the final CD-4 Review. David Martin also organized and ran the final ECP Industry and Agency Council, a much larger and involved meeting to celebrate the end of the program. He also organized industry participation and speakers for the ECP Science Expo on Capitol Hill.

Similarly, ECP Application Integration at the Facilities Lead Scott Parker concluded efforts that have contributed to the achievement of ECP Key Performance Parameters on Aurora and Frontier, along with contributions to the functionality and usability of both leadership-class systems. Parker prepared and presented slides on this work at the ECP CD-4 IPR, wrote the final ECP Project Execution Plan (PEP) milestone covering ECP Application Integration work, and communicated and coordinated ECP Application Integration efforts with ECP management.

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, MLCommons (HPC, Science, and Storage Working Groups), NITRD Middleware and Grid Infrastructure Team, Open Fabrics Alliance, Open Scalable File Systems (OpenSFS) Board, OpenMP Architecture Review Board, OpenMP Language Committee, OSTI ORCID Consortium Membership, Standard Performance Evaluation Corporation (SPEC) HPG (High-Performance Group), and Unified Communications Framework.

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

The International Performance, Portability & Productivity Workshop at SC24 was jointly organized by the 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.

8.4 Integrated Research Infrastructure Activities

The ALCF has numerous significant engagements and achievements related to the Integrated Research Infrastructure (IRI) effort, including working directly with pathfinder projects, participating in technical subcommittees, and representing the ALCF in the IRI Leadership Group. Details of each of these activities are given below.

Nexus

As IRI aims to deliver DOE-enterprise-wide infrastructure for computing, the ALCF has continued its work to link experimental facilities with ALCF computing resources. Work with the APS over recent years (described below) has been a primary driver for defining new functionality and services that the ALCF has deployed to satisfy experiment-time computing needs at APS beamlines. Service accounts enable APS users to leverage automated analysis of their data at the ALCF in a shared environment and in a streamlined fashion throughout their multi-day beamline campaigns, with jobs running immediately at experiment time in the on-demand queue on Polaris. Analysis results are available to scientists at the beamline via Globus Sharing enabled on the Eagle file system, at the time of experiment, and post-hoc. Building on these ALCF-deployed features, Globus Compute and Globus Flows manage application execution and data transfer in a frictionless manner for projects across the DOE-SC program offices.

Pathfinding Projects

The ALCF has partnered with several pathfinding projects to advance their work and also engaged with other projects with strategic IRI-relevant needs.

DIII-D

The DIII-D runs experiments on a twenty-minute cycle that requires time-sensitive analysis of experimental data from one experiment to inform and prepare for the next experiment. Working closely with the DIII-D team and combining efforts with a team at NERSC, the ALCF detailed the needs for an ion orbital simulation application (IonOrb) and a kinetic equilibrium application, CAKE. In weekly meetings throughout the year, the ALCF established service accounts for the DIII-D team, worked with them to leverage a discretionary IRI allocation, installed their software and required local proprietary databases, and enabled outbound connectivity to communicate with systems at the DIII-D facility to exchange input and output data. This work culminated in production-level analysis during experiments at DIII-D using Globus Flows to analyze data automatically between experiments, where CAKE was run on Perlmutter at NERSC and IonOrb was run on Polaris at the ALCF. The IonOrb application used up to 40 nodes in the demand queue on Polaris during a week-long experiment run; in fact, SC24 attendees were given a glimpse into this running experiment in the DOE conference booth during a talk by DIII-D's Sterling Smith.

APS

Following a year-long upgrade of the APS, which now delivers up to 500× greater brightness of its X-ray beams, the APS-ALCF collaboration is paying off with increased computational power at experiment-time — a requirement given the corresponding growth in data rates and, thus, in computing needs. In many ways, the APS use case is similar to the DIII-D case described above, but with an important distinction: whereas the DIII-D case represents a single experiment, the

APS powers more than 60 beamlines, each representing a unique experimental apparatus and research community. Of these, more than 20 beamline communities have identified significant computing needs and have engaged the full power of ALCF's Nexus services and functionality, using service accounts for transparent access to the ALCF and the demand queue for time-sensitive analysis of beamline data through integration with the APS Data Management system. With more beamlines coming online following the upgrade and having greater data analysis needs, the APS demand for computing at the ALCF will continue to grow and will also leverage the newly upgraded network connectivity between the two user facilities.

Advanced Light Source

Working with a team at the Advanced Light Source (ALS), ALCF staff helped to automate the analysis of data from a tomography beamline on Polaris. By using a service account to submit jobs to Polaris through Globus Compute and utilizing the demand queue for data analysis during experiment time, the team has progressed beyond the initial prototype and is now able to conduct analysis in a dedicated discretionary allocation. There are plans to utilize this production-ready capability in upcoming beamline experiments.

Technical Subcommittees

The ALCF participated in all of the IRI subcommittees, including Outreach and Engagement, Interfaces, and TRUSTID (dedicated to cross-facility authentication and authorization solutions). These groups are dedicated to designing and building functionality at the computing facilities to enable their use by experimental scientists.

Leadership Group

The ALCF has been involved in integrated research activities since 2018, well before the term IRI was used to describe this work. With staff in the Leadership Group, including Tom Uram serving as the co-chair of the group, ALCF staff participate in weekly meetings to direct overall efforts related to IRI, specific work by technical subcommittees, formation of new subcommittees, and work with the pathfinder projects. In 2024, ALCF staff served on the organizing committee for the IRI/High-Performance Data Facility (HPDF) kickoff meeting in Gaithersburg, Maryland, and produced follow-up materials describing outcomes from the meeting. ALCF staff also presented during the Leadership Group's participation in the DOE Advanced Scientific Computing Advisory Committee (ASCAC) meeting in May 2024.

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. As of March 18, 2025, researchers utilizing ALCF resources have published 254 papers in high-quality conferences and journals. ALCF projects have succeeded across various fields, employing many different computational approaches. These projects were able to reach their scientific goals and successfully use their allocations.

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*:**

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

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:

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

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*:**

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

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 Polaris.

Table A.1 Capability Definitions for Polaris

Polaris				
Capability	**High Capability**	**Range**	**Minimum Nodes**	**Maximum Nodes**
No	No	0% <= x < 20.0%	1	98
Yes	No	20.0% <= x < 60.0%	99	296
Yes	Yes	60.0% <= x	297	See: A.7, Polaris 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:

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

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

A.7 Polaris Nodes

The number of reportable nodes on Polaris is fewer than the total number of nodes (Table A.2).

Table A.2 Total and Reportable Nodes for Polaris

Polaris		
Data Range	Total Nodes	Reportable Nodes
August 9, 2022	560	504

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, 2024–December 31, 2024
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 Chicago	High-Resolution Ensemble 3D Structures of Genome across Tissues	Biological Sciences	8,397
ABFEBTK	Arjun Saha	University of Wisconsin-Milwaukee	Using Insilico Tools to Understand the Impact of Mutations	Chemistry	2,513
abir	Christine Marie Sweeney	University of New Mexico	AI-Based Image Reconstruction	Materials Science	2,602
adsdl	Noah Harris Paulson	Argonne National Laboratory	Autonomous Discovery – Self-driving Laboratories	Computer Science	1,000
AdvancingEnergyInt	Masood Shahverdi	Argonne National Laboratory	Advancing Energy Intelligence: Optimized Machine Learning Models for Onboard State of Charge Estimation Leveraging Stock Electric Vehicles' CAN Datasets	Energy Technologies	5,000
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	6,875
AIAstroEC	Eugenio Culurciello	Purdue University	AI and Astronomy: Training an Artificial Astronomer Assistant	Physics	371
AIHPC4Edu	Ravi Kiran Madduri	Argonne National Laboratory	AI and HPC for Science and Education	Biological Sciences	3,238
AIPower	Michel Schanen	Argonne National Laboratory	ECP and Beyond for AI in Power Systems	Energy Technologies	2,000
aisd-align	Justin Baker	University of California, Los Angeles	Explicit Frame Alignment of AISD Organic Molecule Dataset	Mathematics	1,026
AI_for_Ab_design	zizhang sheng	Columbia University	AI-based Design of Inescapable Antibodies against Coronavirus	Biological Sciences	7,606
ALCFAITP	Venkatram Vishwanath	Argonne National Laboratory	Argonne AI Training Program	Training	9,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
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	264
ALCF_for_DUNE	Aleena Rafique	Argonne National Laboratory	Enabling Precise Measurement of Neutrino Interactions Using Advanced Computing in DUNE	Physics	14,152
alcf_training	Yasaman Ghadar	Argonne National Laboratory	ALCF Training	Training	7,048
alcf_training_x	Wai Nim Alfred Tang	PhD Tutor Hub	ALCF Training Extension	Physics	14
ALEXIS	Jorge Rene Padial	Vanderbilt University Medical Center	Automatically Labeled EUV and X-Ray Incident Solar Flare Catalog	Physics	8,956
amr-direct	Damyn Maxwell Chipman	Boise State University	Scalable CPU/GPU Direct Solver for Adaptive Mesh Refinements	Mathematics	1,416
AMRMRBC	Mu-Hua Chien	New York University	AMRMRBC	Earth Science	3,000
ANLBSC	Vishal Kumar	Argonne National Laboratory	Wall-modeled LES and Hybrid Simulations for Wind Energy Applications	Engineering	5,000
APSDataAnalysis	Rafael Vescovi	Argonne National Laboratory	APS Beamline Data Processing and Analysis	Computer Science	8,546
APSDataProcessing	Nicholas Schwarz	Argonne National Laboratory	Advanced Photon Source (APS) Data Processing	Computer Science	2,996
APS_ILLUMINE	Nicholas Schwarz	Argonne National Laboratory	APS ILLUMINE Project Space	Physics	685
argonne_tpc	Venkatram Vishwanath	Argonne National Laboratory	Trillion Parameter Consortium	Computer Science	1,230
ArtISD	Tuhin Sahai	SRI International	Artificial Intelligence for Scientific Discovery	Computer Science	7,395
ASSBs_Microstructure	Qingsong Tu	Rochester Institute of Technology	Microstructural Optimization of All-Solid-State Batteries	Energy Technologies	1,000
AS_EN_TZ	Italo Moletto	AgroSpace	Water and Food Security for Tanzania Using Satellite Imagery + ML/AI	Earth Science	1,299
ATLAS_workflow_ALCF	Rui Wang	Argonne National Laboratory	Development and Preparation of the Working Environment for ATLAS Workflows	Physics	233

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
ATPESC2024	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing 2024	Training	21,000
ATPESC_Instructors	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing Instructors	Training	1,000
attpc	Daniel Pierre Bazin	Michigan State University	Active Target Time Projection Chamber	Physics	750
Auriga	Christine Mary Simpson	Argonne National Laboratory	The Auriga Project	Physics	686
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	1,510
BACTWA	Philippe Regis-Guy Piot	Argonne National Laboratory	Beam Acceleration and Control in Wakefield Accelerator	Physics	6,163
BenchmarkUNM	Amanda Bienz	University of New Mexico	Benchmarking and Optimizing Data Movement on Emerging Heterogeneous Architectures	Computer Science	1,450
beta_decay	Ante Ravlic	Michigan State University	Improving Theoretical Description of Nuclear Beta-decay Rates	Physics	2,991
BFTrainer	Rajkumar Kettimuthu	Argonne National Laboratory	Rescaling DNN Training Tasks to Fit Dynamically Changing Holes in Supercomputer Schedule	Computer Science	5,663
BioelectricalCircuit	Michael Levin	Tufts University	Exploration of Bioelectrical Circuits for Viable Interventions	Biological Sciences	4,810
BIP167	Philip Kurian	Howard University	Computing Superradiance and van der Waals Many-body Dispersion Effects from MD Simulations of Biomolecular Complexes	Biological Sciences	5,669
BPC	Christopher Michael Graziul	The University of Chicago (UChicago)	Optimization of Audio Processing Pipeline for Broadcast Police Communications	Computer Science	3,525
BRAIN	Getnet Dubale Betrie	National Renewable Energy Laboratory (NREL)	Scalable Brain Simulator for Extreme Computing	Biological Sciences	4,938

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
BrainImagingML	Thomas David Uram	Argonne National Laboratory	Large-scale Brain Imaging and Reconstruction	Biological Sciences	8,252
C-Star	Matthew Long	CWorthy Systems, Inc.	Computational Systems for Tracking Ocean Carbon	Earth Science	3,417
cac	Chang Liu	Princeton Plasma Physics Laboratory (PPPL)	Center for Advanced Computation: Disruption Avoidance	Fusion Energy	3,111
CaloDiffusion	Oz Mikael Amram	Fermi National Accelerator Laboratory (Fermilab)	Diffusion Models to Accelerate Detector Simulation	Physics	5,382
CampSwift	Marcos Vanella	National Institute of Standards and Technology (NIST)	Physics Based Modeling for Forced Fuel Management	Earth Science	866,702
Catalyst	Katherine M. Riley, Christopher James Knight, James Clifton Osborn, Timothy Joe Williams	Argonne National Laboratory	Catalyst	Internal	10,478
CD4DC	Laura Gagliardi	The University of Chicago (UChicago)	Rational High-throughput Design on Novel Iron–Sulfur-based Metal Organic Frameworks for Environmental Application	Chemistry	7,256
CEESD	Michael Todd Campbell	University of Illinois Urbana-Champaign	Performance Evaluation of Scramjet Internal Flow and Combustion Simulation Application	Engineering	4,499
cellAI	Tanzima Zerin Islam	Texas State University	Parallelize Cellular Automata on Cutting-Edge AI Architectures	Computer Science	1,485
CertRand	Minzhao Liu	JPMorgan Chase & Co.	(INCITE) Experimental Realization of Certified Randomness from Quantum Supremacy	Computer Science	848
CFS_UX_TEST	Haritha Siddabathuni Som	Argonne National Laboratory	Testing CFS	Support	15
CGModels-IonConduct	Elif Ertekin	University of Illinois Urbana-Champaign	Learning Coarse-Grained Models of Complex Atomistic Dynamics	Materials Science	4,513
CHAMPS_aero	Roberto Paoli	Polytechnique Montreal	Aerodynamics Simulations with Chapel Programming Language	Engineering	122

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
CharmRTS	Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia	University of Illinois Urbana-Champaign	Charm++ and Its Applications	Computer Science	843
Chi_Geo	Andrew Barbeau	The Accelerate Group, LLC	Sustainable Chicago Geothermal	Energy Technologies	5,000
climate_severe	Vittorio Angelo Gensini	Northern Illinois University (NIU)	Anticipating Severe Weather Events via Dynamical Downscaling	Earth Science	9,311
ClimaX	Paris Georgios Perdikaris	Microsoft Corporation	Large-scale Foundation Models for Weather and Climate Forecasting	Earth Science	5,000
cms_l1t_fm	Abhijith Gandrakota	Fermi National Accelerator Laboratory (Fermilab)	Advancing (HL-)LHC Physics: Foundational Models for HEP Experiments	Physics	5,051
CobaltDevel	Paul Michael Rich, William Edward Allcock	Argonne National Laboratory	Cobalt Development	Internal	59
community_ai	Emmanouil Koukoumidis	Learning Machines AI (Public Benefit Corporation)	Collectively, Efficiently, and Responsibly Advanced Multimodal Foundation Models	Computer Science	4,778
COMPASS-GLM	William James Pringle	Argonne National Laboratory	Coastal Observations, Mechanisms, and Predictions Across Systems and Scales – Great Lakes Modeling (COMPASS-GLM)	Earth Science	1,754
Compressible_Multi	Sanjiva K. Lele	Stanford University	Simulation and Analysis of Canonical Compressible Turbulent Flows with Complex Thermodynamics and Phase Changes	Engineering	977
CONUS-Carbon	Jinxun Liu	U.S. Geological Survey (USGS)	Terrestrial Ecosystem Carbon Cycle of the Conterminous U.S.	Earth Science	2,829
coreform-lattice	Greg John Vernon	Coreform LLC	HPC-enabled Geometry-compliant Lattice Structures for 3D Printing and Structural Simulation	Engineering	1,000
CosDiscover	Katrin Heitmann	Argonne National Laboratory	Enabling Cosmic Discoveries in the Exascale Era	Physics	521

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	36,574
CrO-Al	Peng Geng	University of California, Los Angeles	Interface Stability of Chromium Oxide and Aluminum with First-Principles Calculations	Materials Science	4,000
crocus	Dimitrios Fytanidis	Argonne National Laboratory	NekRS Scaling Urban Modeling – DOE BER CROCUS Proposal	Earth Science	3,279
CSTEELML	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Steelmaking Research at Lower Temperatures and Low CO_2 Processes	Chemistry	913
cug_scaling	Colleen Elizabeth Bertoni	Argonne National Laboratory	Scaling Comparison on Polaris and Aurora	Computer Science	2,560
CVD_CityCOVID	Jonathan Ozik	Argonne National Laboratory	Agent-based Model Called CityCOVID Capable of Tracking Detailed COVID-19 Transmission	Biological Sciences	1,967
cvg_sycl	Thomas Applencourt	Argonne National Laboratory	SYCL Bench-marking for CONVERGE CFD	Engineering	4,771
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	5,000
darkskyml_aesp	Salman Habib	Argonne National Laboratory	Dark Sky Mining	Physics	10,000
DARWIN	Tong Xie	The University of New South Wales Sydney	From Token to Discovery: Leverage Large Language Model for Material Design	Materials Science	3,733
datascience	Venkatram Vishwanath	Argonne National Laboratory	ALCF Data Science and Workflows Allocation	Internal	10,996
datascience_collab	Venkatram Vishwanath	Argonne National Laboratory	ALCF Data Science and Workflows Allocation for Collaborators	Computer Science	3,115
DataServicePrototype	Benoit Joseph Serge Cote	Argonne National Laboratory	Data Service Prototype	Computer Science	4,913
DAVSDK	Patrick O'Leary	Kitware Inc.	Data and Visualization SDK	Computer Science	3,273

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
DE-SC0020314	Jaime Marian	University of California, Los Angeles	Grain Boundary Microstates: Exploring the Metastability of Sink Efficiency	Materials Science	6,491
DeepOrb	Nathaniel Clinton Hudson	The University of Chicago (UChicago)	Deep Learning for Molecular Orbitals	Chemistry	5,848
deepspeed_collab	Venkatram Vishwanath	Argonne National Laboratory	Deepspeed Collaboration with Microsoft for Large Language Models Scaling	Computer Science	5,213
DERCybersecurity	Siby Jose Plathottam	Argonne National Laboratory	Adaptive Cybersecurity for Distributed Energy Resources: A Game-Theoretic and Machine Learning Approach for Real-Time Threat Detection and Mitigation	Energy Technologies	7,727
DesignPepPro2	Vikram Khipple Mulligan	Simons Foundation	Heteropolymer Design Harnessing New and Emerging Computing Technologies	Biological Sciences	6,000
DFTCalculations	Vishwas Hebbur Venkata Subba Rao	Argonne National Laboratory	DFTCalculations	Materials Science	2,222
Diaspora	Ryan Lloyd Chard	Argonne National Laboratory	Diaspora: Resilience-enabling Services for Science from HPC to Edge	Computer Science	1,076
DiNTGPU	Yeonjun Jeong	Argonne National Laboratory	Towards Exascale Chemical Dynamics via GPU-Accelerated Polynomial Expansions	Chemistry	1,621
DirectContacts2	Kevin Sheahan Drew	University of Illinois Chicago	Structural Modeling of High-Confident Direct Protein Interactions Using AlphaFold	Biological Sciences	2,770
dist_relational_alg	Sidharth Kumar	The University of Alabama at Birmingham	Distributed Relational Algebra at Scale	Computer Science	2,162
DLIO	Huihuo Zheng	Argonne National Laboratory	Deep Learning IO Performance Evaluation on HPC Storage and File System and AI Testbeds	Computer Science	5,000
DMPE	Shizhong Han	Lieber Institute for Brain Development	Deep Learning Models for Predicting Enhancers in Single Cells	Biological Sciences	4,166
dns-ftsi-bodony	Daniel J. Bodony	University of Illinois Urbana-Champaign	DNS of Fluid-Thermal-Structural Interactions in Hypersonic Flows	Engineering	5,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
DNS3D	Ramesh Balakrishnan	Argonne National Laboratory	Direct Numerical Simulation of Three-Dimensional Turbulence	Engineering	8,000
DriveDock	Vince G. Amoroso	University of Illinois Chicago	Simulation of Intrinsically Disordered Proteins	Chemistry	1,822
DSMC_NEQ	Jacqueline Chen	Sandia National Laboratories, California	Molecular-level Simulations of Reacting Flows under Thermal and Chemical Non-equilibrium	Chemistry	2,491
DTLLM	Bo Li	The University of Chicago (UChicago)	Trustworthiness Evaluation for Llama2	Computer Science	3,111
dynres	Sergio Iserte	Barcelona Supercomputing Center	Evaluation of Dynamic Resource Management Techniques	Computer Science	5,000
E2H	Michael E. Papka	Argonne National Laboratory	Edge to HPC	Computer Science	51
E3SM_RRM	Brandi Lee Gamelin	Argonne National Laboratory	Coupled High-resolution GPU-enabled E3SM Simulations for Mid-century Extreme Events	Earth Science	10,923
earnest	Robert Clyde Jackson	Argonne National Laboratory	An Equitable, Affordable, and Resilient Nationwide Energy System Transition (EARNEST)	Earth Science	2,700
EarthWorks	Richard Dana Loft	National Science Foundation	Preparing EarthWorks for GPU-based Climate Simulations at Global Storm-Resolving Scales	Earth Science	500
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	10,258
ECP_SDK	Sameer Suresh Shende	University of Oregon	Deploying the ECP SDK Software Stack at ALCF	Computer Science	571
EE-ECP	Xingfu Wu, Valerie Taylor	Argonne National Laboratory	Energy-efficient Trade-off among Execution Time and Power of ECP Applications	Computer Science	10,919
EfficientLLM	Mi Zhang	The Ohio State University	Efficient Large Language Model	Computer Science	1,654
efficios_thapi	Thomas Applencourt	Argonne National Laboratory	Improving Argonne Tracing Capabilities of Heterogeneous API	Computer Science	2,108

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
ElecDyComplexSys_DD	Yosuke Kanai	The University of North Carolina-Chapel Hill	Electron Dynamics of Complex Systems	Chemistry	657
electrolyte-chibueze	Chibueze Vincent Amanchukwu	The University of Chicago (UChicago)	Molecular Dynamics-driven Discovery of Novel Liquid Electrolytes	Energy Technologies	8,404
ELRDE	Praveen Kumar Ramaprabhu	The University of North Carolina-Chapel Hill	Euler-Lagrange Simulations of Liquid-fueled Rotating Detonation Engines	Engineering	3,111
EngineDNS	Christos Frouzakis	Eidgenössische Technische Hochschule Zürich (ETH Zurich)	Towards Reactive DNS in Complex Internal Combustion Engine Geometries	Engineering	9,761
ESGF2	Ian Foster	Argonne National Laboratory	Earth System Grid Federation 2 Project	Earth Science	1,191
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	2,614
EVITA	George K. Thiruvathukal, Nicholas John Eliopoulos, Nicholas Michael Synovic	Loyola University Chicago	Energy-efficient Visual Transformers: Investigations in Computer Vision and Empirical Software Engineering	Computer Science	8,912
ExaMol	Logan Timothy Ward	Argonne National Laboratory	Exascale Design of Molecules for Energy Storage	Materials Science	11,042
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	3,433
ExtraNuc	Krishnan Raghavan	Argonne National Laboratory	Extrapolation of Nuclear Observables	Physics	3,706
ExtremeConvection	Mahendra Kumar Verma	Indian Institute of Technology Kanpur	Simulation of Turbulent Convection at Extreme Parameters	Physics	10,103
fciqmc-julia	Mingrui Yang	Washington University–St. Louis	Full Configuration Interaction Quantum Monte Carlo in Julia	Physics	2,206
FD4GW	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Federated Learning for Gravitational Wave Astrophysics	Physics	6,607

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
FES_MLFF_MetaD	Orlando Andres Mendible	University of Notre Dame	Free Energy Surface Determination with Equivariant Neural Network	Materials Science	219
FE_Canopy	Timothy Wayne Juliano	The National Center for Atmospheric Research (NCAR)	Forest Canopy-Resolved Atmospheric Flow Simulations using FastEddy®	Engineering	3,660
filmcooling	Pinaki Pal	Argonne National Laboratory	Robust Film Cooling under Manufacturing Uncertainty for Improved Jet Engine Life Cycle Energy Efficiency	Energy Technologies	6,535
FinTut24	Junye Wang	Dassault Systemes	Finish ATPESC2024 Tutorials	Computer Science	1,000
FMScale	Ruixuan Gao	University of Illinois Chicago	Analysis of Large-scale Fluorescence Image Datasets	Biological Sciences	3,863
fm_electrolyte	Venkatasubramanian Viswanathan	Carnegie Mellon University	Foundation Models for Electrolyte Design	Materials Science	204
FOUND4CHEM	Ganesh Sivaraman	University at Buffalo	Porting and Finetuning GPU-based Transferable Models for Materials and Chemistry	Chemistry	5,043
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	3,192
FusAblator	Ivan Oleynik	University of South Florida (USF)	Predictive Simulations of Inertial Confinement Fusion Ablator Materials	Fusion Energy	948
fusiondl_aesp	William Ming-Wu Tang	Princeton University	Accelerated Deep Learning Discovery in Fusion Energy Science	Fusion Energy	4,000
generative_hydrology	John Kress Hutchison	Argonne National Laboratory	Generative AI Model for Hydrology and Near-Real-Time Flooding Forecasts for Enhanced Emergency Response	Earth Science	959
GeomicVar	Ravi Kiran Madduri	Argonne National Laboratory	Scaling Genomic Variant Callers to Leadership-class Systems: A Collaboration between VA-MVP and DOE	Biological Sciences	21,549
GPU-DG	Pinaki Pal	Argonne National Laboratory	GPU-enabled Discontinuous Galerkin Simulations of Complex Fluid Flows	Engineering	1,874

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
GPU-Trajectories	Maninder Singh Grover	University of Dayton	Molecular Trajectory Analysis on GPU Platform	Chemistry	4,000
GPUreconstructCMS	Dirk Hufnagel	Fermi National Accelerator Laboratory (Fermilab)	Using GPU to Reconstruct LHC Collisions Recorded with the CMS Detector	Physics	47,544
gpu_hack	Yasaman Ghadar	Argonne National Laboratory	GPU Hackathon	Training	9,000
GRACE	Sayan Ghosh	Pacific Northwest National Laboratory (PNNL)	Graph Analytics Codesign on GPUs	Computer Science	5,266
gtcarbon	Chao Xu	Argonne National Laboratory	Computational Modeling of Cost-effective Carbon Capture Technologies on Industrial Gas Turbines to Reduce CO_2 Emissions	Energy Technologies	5,005
GURU_AI_TESTBED	Allan Dougal Grosvenor	Microsurgeonbot Inc.	Advancing AI-Driven Scientific Software Navigation and 3D Geometry Synthesis	Computer Science	9,375
HACC_P3HPC	Esteban Miguel Rangel	Argonne National Laboratory	P3HPC Performance Results for CRK-HACC	Computer Science	1,755
Hdiffusion_MgNW	Miguel Molinos Pérez	Universidad de Sevilla	Atomistic Modeling of Long-term Hydrogen Diffusion in Magnesium Nanowires	Materials Science	900
HEA	Jaime Marian	University of California, Los Angeles	Integrated Thermomechanical Model of First Wall Components under Evolving Chemistry and Microstructure during Fusion Reactor Operation (ThermChem-FW)	Materials Science	5,000
HEDM_Viz	Chihpin Chuang	Argonne National Laboratory	Visualization for High-Energy Diffraction Microscopy	Materials Science	188
hep-cce	Charles George Leggett	Lawrence Berkeley National Laboratory (LBNL)	HEP-CCE Workflow Portability	Physics	2,586
hifisepturb	Ivan Bermejo Moreno	University of Southern California (USC)	High-fidelity Simulations of Compressible Separated Turbulent Flows	Engineering	1,919
HighReyTurb_PostProc	Robert D. Moser, Myoungkyu Lee	University of Houston	Data Analysis of Turbulent Channel Flow at High Reynolds Number	Engineering	8,803

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
HIV-DRUG	Ao Ma	University of Illinois Chicago	Understanding the Rigorous Molecular Mechanism of Drug Resistance to HIV Protease Inhibitors	Biological Sciences	9,669
hp-imagegen	Weijian Zheng	Argonne National Laboratory	Co-optimize AI Algorithm and Computer System for Accelerating High-resolution Image Generation	Computer Science	1,363
hp-ptycho	Tekin Bicer	Argonne National Laboratory	High Performance 3D Ptychographic Reconstruction and Image Enhancement	Materials Science	4,330
hpc-spectacle	Kevin Antoney Brown	Argonne National Laboratory	Evaluate and Optimize Data Movement Strategies in AI and Climate Science Workloads	Computer Science	2,550
hpcbdsm	Tanwi Mallick	Argonne National Laboratory	High-Performance Computing and Big Data Solutions for Mobility Design and Planning	Computer Science	4,279
HPCBot	Murat Keceli	Argonne National Laboratory	Large Language Models for High-Performance Computing	Computer Science	2,000
hpe_dragon_collab	Venkatram Vishwanath	Argonne National Laboratory	Collaboration with HPE on the Dragon Project Evaluation and Scaling	Computer Science	6,500
Hybrid-Sn-Diffusion	Madicken Munk	University of Illinois Urbana-Champaign	Development of a Hybrid SN-Diffusion Neutronics Method for Time-Dependent Multiphysics Nuclear Reactor Simulations	Nuclear Energy	13,407
hydrosm	Jeremy A. Feinstein	Argonne National Laboratory	Improving the Predictability of Hydrological Systems with AI	Earth Science	2,000
hypersonic01	Chonglin Zhang	University of North Dakota	Kinetic Modeling of Hypersonic Nonequilibrium Flows	Engineering	1,123
hyper_bound_layer	Carlo Scalo	Purdue University	Passive Control of Hypersonic Boundary Layers via Wall Treatments	Engineering	80
hyqu	John Christopher Steuben	United States Naval Research Laboratory	Solving Multi-Particle Quantum Wave Equations	Materials Science	600
IBM-GSS	Venkatram Vishwanath	Argonne National Laboratory	IBM GeoSpatial Software System	Earth Science	77

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
icarusHPC	Thomas Benjamin Wester	The University of Chicago (UChicago)	Neutrino Interaction Simulations for the ICARUS Experiment	Physics	3,895
IMPROVE	Nicholas Lee-Ping Chia	Argonne National Laboratory	IMPROVE Hackathon	Biological Sciences	5,000
IMPROVE_Aim1	Rick Lyndon Stevens	Argonne National Laboratory	IMPROVE: Innovative Methodologies and New Data for Predictive Oncology Model Evaluation	Computer Science	149,919
insitu	Silvio Humberto Rafael Rizzi	Argonne National Laboratory	In Situ Vis and Analysis at ALCF	Computer Science	3,743
Intel	Kalyan Kumaran, Scott Parker, Timothy Joe Williams, Venkatram Vishwanath	Argonne National Laboratory	Intel Employees in Support of Theta	Internal	1,659
introtohpc_faculty	Paige Carolyn Kinsley	Argonne National Laboratory	ALCF Faculty Curriculum Development Program	Computer Science	6
IQC	Murat Keceli	Argonne National Laboratory	Interactive Quantum Chemistry	Chemistry	4,316
IRI-ALS-832	David Abramov	Lawrence Berkeley National Laboratory (LBNL)	ALS BL832 Tomographic Reconstruction	Materials Science	21,479
IRIBeta	Thomas David Uram	Argonne National Laboratory	IRI Explorations	Computer Science	1,095
KaonCPV	Christopher Brendan Kelly	Brookhaven National Laboratory (BNL)	Continuum Limit Lattice Calculation of Direct CP-violation in Kaon Decays	Physics	18,971
Kim-2DM	Seong-Gon Kim	Mississippi State University	Quantum Mechanical First-Principles Simulation of Two-dimensional Materials for Novel Electronic Devices	Materials Science	8,559
kokkos_math	Luc Berger-Vergiat	Sandia National Laboratories, New Mexico	Kokkos-based Math Libraries Performance at ALCF	Computer Science	538
LasersBeamsPlasmas	Eduardo Paulo Jorge da Costa Alves	University of California-Los Angeles	Unraveling How Lasers and Beams with Arbitrary Spatial and Temporal Structure Interact with Plasmas	Physics	50,000
LASSCF_gpudev	Christopher James Knight	Argonne National Laboratory	GPU Development of LASSCF	Chemistry	5,937
LastJourney	Katrin Heitmann	Argonne National Laboratory	Extreme-scale Cosmology – The Last Journey	Physics	1,476

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
lbpm_sycl	James Edward McClure	Virginia Polytechnic Institute and State University (Virginia Tech)	LBPM Performance Optimization	Earth Science	1,210
LDON	Sharmila Karumuri	Johns Hopkins University	Physics- and Uncertainty-Informed Latent Operator Learning	Earth Science	1,000
LeadChalco	Burak Guzelturk	Argonne National Laboratory	Understanding of Local Structure in Lead Chalcogenides and Their Nanocrystals	Physics	11,005
libensemble	Stephen Tobias Paige Hudson	Argonne National Laboratory	libEnsemble: Project testing, Development, and Scaling study	Computer Science	364
LIGHTCONTROL	Sandra Gail Biedron, Trudy Beth Bolin	University of New Mexico	Light Sources and Their Control Using AI Techniques	Physics	4,788
lighthouse-internal	Paige Carolyn Kinsley	Argonne National Laboratory	Lighthouse Initiative – Internal Allocation	Internal	212
lighthouse-purdue	Arman Pazouki	Purdue University	Purdue University / ALCF Lighthouse Partnership Hub	Computer Science	13
lighthouse-uchicago	Hakizumwami Birali Runesha	Argonne National Laboratory	University of Chicago Research Computing / ALCF Partnership Hub	Computer Science	8,497
lighthouse-uic	Himanshu Sharma	University of Illinois Chicago	UIC ACER / ALCF Lighthouse Partnership Hub	Computer Science	271
lighthouse-uwyoming	Suresh S. Muknahallipatna	University of Wyoming	University of Wyoming / ALCF Lighthouse Partnership Hub	Computer Science	149
LitMinMatSci	Tanjin He	Argonne National Laboratory	Accelerating Battery Material Design via Automated Multimodal Electrochemical Data	Materials Science	191
LLM4CS	Franck Cappello	Argonne National Laboratory	LLMs Training with ACM Digital Library	Computer Science	714
LLM_multilang	Hanzhi Zhang	University of North Texas	Large Language Models' Hallucination Detection and Mitigation Ability across Multi Languages	Computer Science	2,963
LQCDdev	James Clifton Osborn	Argonne National Laboratory	Lattice QCD Development	Physics	1,584
LUCID	Thomas Scott Brettin	Argonne National Laboratory	Low-dose Understanding, Cellular Insights, and Molecular Discoveries	Biological Sciences	4,872

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
luminary-cloud-2024	Juan Alonso	Luminary Cloud, Inc.	Scalability and Performance of Exascale Aerodynamics Simulations for AAIA Workshop Cases and AutoCFDusing Luminary Cloud's GPU-based CFD Solver	Engineering	5,509
lunar_flow	Bodhinanda Chandra	University of California-Berkeley	Continuum Modeling of Lunar Regolith and Its Complex Multi-physics Thermo-fluid-Electromechanical Interactions	Engineering	2,400
MADNESS	George I-Pan Fann	Oak Ridge National Laboratory (ORNL)	Porting MADNESS to BlueGene P	Materials Science	1
MADNESS_Hurtado	Adrian Guillermo Hurtado	Stony Brook University	MADNESS Aurora Readiness	Chemistry	649
MAE	Seongha Park	Argonne National Laboratory	Multimodal Algorithms at the Edge	Computer Science	5,000
Maintenance	William Edward Allcock, John Francis O'Connell, John Patrick Reddy, Ryan Milner, Torrance Ivan Leggett	Argonne National Laboratory	LCF Operations System Maintenance	Internal	13,470
matml	Noa Marom	Carnegie Mellon University	INCITE Proposal for Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials	Materials Science	5,000
matml_aesp	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials	Materials Science	5,000
MatQIS2	Giulia Galli	The University of Chicago (UChicago)	Large-scale Simulations of Materials for Energy and Quantum Information Science	Materials Science	2,222
Mayo	Nicholas Lee-Ping Chia	Argonne National Laboratory	Colorectal Cancer Evolution Project – Mayo and Argonne	Biological Sciences	2,442
mayopath	Nicholas Lee-Ping Chia	Mayo Clinic-Minnesota	Mayo Pathology	Biological Sciences	8,143

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
MBXOpt	Christopher James Knight	Argonne National Laboratory	Enabling Chemical Accuracy through GPU-accelerated Large-Scale Many-body Molecular Dynamics Expansions	Chemistry	1,281
MechanoEar	Marcos Manuel Sotomayor	The University of Chicago (UChicago)	Large-Scale All-Atom Molecular Dynamics Simulations of Mechanotransduction Complexes	Biological Sciences	5,000
MEDDIAC	Ehud Strobach	Agricultural Research Organization	High-resolution Interactions of Mediterranean Cyclone with Ocean Eddies	Earth Science	5,000
metafreeform	Zin Lin	Virginia Polytechnic Institute and State University (Virginia Tech)	Freeform Inverse Design of Large Area Metalenses	Materials Science	2,272
metastable	Subramanian Sankaranarayanan	Argonne National Laboratory	Metastable Phase Diagram of Material	Materials Science	2,606
metsim	Siddharth Deshpande	University of Rochester	A Huckel Method Inspired Similarity Algorithm to Bridge the Materials Gap in Heterogeneous Metal Catalysis	Chemistry	15,310
MI2Dmaterials	Trevor David Rhone	Rensselaer Polytechnic Institute (RPI)	Materials Informatics Study of Two-dimensional Magnetic Materials and Their Heterostructures	Materials Science	4,953
Micrometer	Paris Perdikaris	University of Pennsylvania (UPenn)	Foundation Models for Computational Solid Mechanics	Engineering	5,042
micromix	Pinaki Pal	Argonne National Laboratory	High-fidelity CFD Modeling of Micromix Combustion Systems for Hydrogen-Fueled Gas Turbine Engines	Energy Technologies	2,391
ML-Coupling	Emil Mihai Constantinescu	Argonne National Laboratory	Data-driven Coupling Methods for Atmospheric-Ocean Interactions	Earth Science	1,331
MLP4THERMO	Cem Sevik	University of Antwerp	Machine Learning Potentials for Thermal Properties of Two-Dimensional Materials	Materials Science	2,278
mm-ChemFM	Dmitry Zubarev	IBM Corporation	Multi-modal Foundational Model Based on Data Modalities Offered by Chemical Simulations	Chemistry	7,305

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
MM_DFMs	Azton Ian Wells	Argonne National Laboratory	Multimodal Domain Specific Foundation Models across Scientific Disciplines	Physics	4,810
mm_protein	Andrew Lindsay Ferguson	The University of Chicago (UChicago)	Multimodal Contrastive Learning for Natural Language-prompted Data-driven Protein Design	Biological Sciences	9,472
MOAB_App	Vijay Subramaniam Mahadevan	Argonne National Laboratory	MOAB Algorithmic Performance Portability	Mathematics	5,869
ModelingCoronaVirus	Zhangli Peng	University of Illinois Chicago	Modeling Corona Virus	Biological Sciences	7,099
mosart_downscaling	Jemma Stachelek	Los Alamos National Laboratory (LANL)	Scoping River Model Downscaling for the Energy Exascale Earth System Model	Earth Science	384
mpiabc	Pariksheet Nanda	University of Michigan	Solving Bayesian Inverse Problems with ABC-SMC	Biological Sciences	587
MPICH_MCS	Kenneth James Raffenetti, Pavan Balaji	Argonne National Laboratory	MPICH – A High-performance and Widely Portable MPI Implementation	Computer Science	1,539
MSR_ML	Yiqi Yu	Argonne National Laboratory	AI-powered Toolkit for Advanced Nuclear Reactor Design	Nuclear Energy	829
MultiActiveAI	Dario Dematties	Northwestern Argonne Institute of Science and Engineering (NAISE)	Multimodal Intelligence for Federated Edge Computing Simulations	Computer Science	1,954
multiphysics_aesp	Amanda Randles	Duke University	Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations	Engineering	8,000
MVAPICH2	Dhabaleswar Kumar Panda	The Ohio State University	Optimizing and Tuning MVAPICH2-GDR Library and Studying Its Impact on HPC and AI Applications	Computer Science	7,007
N2M3O7	Iwnetim Iwnetu Abate	Massachusetts Institute of Technology (MIT)	Tunable Catalysis in Transition Metal Oxide: $NaxMn_3O_7$	Materials Science	16,818
nacatho1	Wenhua Zuo	Argonne National Laboratory	DeePMD Simulations of the Synthesis of Layered Oxide Cathodes for Advanced Batteries	Materials Science	2,363

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
NAIRR240032	Haritha Siddabathuni Som, Lijing Wang	University of Connecticut	Developing a Benchmark Hydrologic Dataset and a Fast AI Surrogate Model for Assessing Climate Impacts on Mountainous Hillslopes	Earth Science	2,606
NAISE-NIH-NIAMS	Marta Garcia Martinez	Argonne National Laboratory	Automatic MRI Segmentation for Upper Limb Muscles for Clinical Applications	Biological Sciences	4,153
NAISE_Spinal_Cord	Marta Garcia Martinez	Argonne National Laboratory	Large Volume Feline Spinal Cord Microtomography	Biological Sciences	2,111
nekrs-scaling	Pinaki Pal	Argonne National Laboratory	NekRS Scalability Studies for Gas Turbine Film Cooling High-fidelity Simulations	Energy Technologies	1,330
nekrsCombustion	Benjamin William Keeton	Argonne National Laboratory	NekRS Combustion – Gas Turbines	Engineering	3,505
NERSC_BLAS_S2um4	Adam Lavely	Lawrence Berkeley National Laboratory (LBNL)	Understanding the Performance Variations between BLAS Implementations	Computer Science	1,000
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	2,444
neutrinoGPU	Thomas Benjamin Wester	Argonne National Laboratory	GPU-Accelerated Neutrino Software	Physics	8,187
NeutronStarRemnants	Ore Gottlieb	Simons Foundation	Simulating Large-scale Long-lived Neutron Star Remnants from Binary Neutron Star Mergers	Physics	18,646
NN4PDEs	Boyang Chen	Imperial College London	AI-Powered Digital Twins and Exascale PDE-Based Simulations for Complex Environmental Solutions	Engineering	267
NNFF_mem_sys	Kasidet Trerayapiwat	Argonne National Laboratory	Exascale Framework for AI-enabled Coupled Electro-thermal Modeling of 3D Integrated Devices	Materials Science	800
novacosmics	Alexander I Himmel	Fermi National Accelerator Laboratory (Fermilab)	NOvA Cosmic Rejection	Physics	6,603

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
NSLS2DataProcessing	Stuart Ian Campbell	Brookhaven National Laboratory (BNL)	National Synchrotron Light Source II (NSLS-II) Data Processing	Computer Science	925
Nu3D	Yong Zhao	Argonne National Laboratory	3D Imaging of the Pion from Lattice QCD	Physics	38,950
NUCat_Micro-CT	Marta Garcia Martinez	Argonne National Laboratory	Large Volume Feline Spinal Cord Microtomography	Biological Sciences	695
NucStructReact_6	Gaute Hagen	Oak Ridge National Laboratory (ORNL)	Ab-initio Nuclear Structure and Nuclear Reactions	Physics	5,000
NuQMC	Alessandro Lovato	Argonne National Laboratory	Nuclear Quantum Monte Carlo	Physics	4,274
Nu_Novel	Zelimir Djurcic	Argonne National Laboratory	Novel LArTPC Simulation and Reconstruction	Physics	6,623
OBP	Christopher James Knight	Argonne National Laboratory	Operator Backpropagation Simulations	Computer Science	10,476
OmniverseEval	Joseph A. Insley	Argonne National Laboratory	NVIDIA Omniverse Evaluation	Computer Science	21,361
OnlineTurb	Kenneth Edward Jansen	University of Colorado-Boulder	Online Machine Learning for Large Scale Turbulent Simulations	Engineering	3,830
OpenCosmo	Katrin Heitmann	Argonne National Laboratory	HACC Simulation Data Portal	Physics	2,502
Operations	William Edward Allcock	Argonne National Laboratory	Systems Administration Tasks	Internal	15,051
OptADDN	Sandeep Madireddy	Argonne National Laboratory	Optimal Architecture Discovery for Deep Probabilistic Models and Neuromorphic Systems	Computer Science	1,443
Ortho-ParaConversion	Sohail Murad	Illinois Institute of Technology (IIT)	Understanding Ortho-para Conversion of H2 at a Magnetically Dense Solid Catalyst	Chemistry	8,000
ParaLLMs	Kibaek Kim	Argonne National Laboratory	Parallel Training and Fine-Tuning on Large Language Models	Computer Science	4,872
PARTURB3D	Ramesh Balakrishnan	Argonne National Laboratory	Simulating Turbulent Particulate Flows inside Enclosures	Engineering	8,000
PatentAnalysis	Yan Zhou	Argonne National Laboratory	Battery Technologies Patent Analysis	Energy Technologies	50
pCLAMD	Madhurima Vardhan	Argonne National Laboratory	pCLAMD – Parallelized Clinical Language Model to Aid Medical Decision-making	Biological Sciences	5,029

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
pde-hartley-operator	Sathya Ravi	University of Illinois Chicago	Fast Simulation of PDEs Using Neural Operators	Computer Science	1,875
PEDAL	Tom Peterka	Argonne National Laboratory	PEDAL: Parallel Extreme-scale Data Analysis	Computer Science	4,297
Performance	Scott Parker, Raymond M. Loy	Argonne National Laboratory	Performance	Internal	15,603
petsc-alcf	Junchao Zhang	Argonne National Laboratory	Supporting PETSc for ALCF Users	Mathematics	1,088
PHASTA_NCSU	Igor A. Bolotnov	North Carolina State University (NCSU)	Multiphase Simulations of Nuclear Reactor Thermal Hydraulics	Engineering	9,433
phoebus	Carl Fields	Los Alamos National Laboratory (LANL)	Accelerating CCSN Simulations with Phoebus	Physics	7,600
phoebus_ccsn	Jeremiah Wayne Murphy	Florida State University	The Phoebus Project	Physics	4,806
PHTMD	Cem Sevik	University of Antwerp	Phase Transition Properties of Transition Metal Dichalcogenides	Physics	9,185
PlasmaBlackHole	Dmitri Uzdensky	University of Colorado-Boulder	Electron Kinetic Plasma Physics of Black Hole Accretion Flows	Physics	585
Polaris	Torrance Ivan Leggett	Hewlett Packard Enterprise	Polaris Project for Installation and Related Work for the Vendors	Internal	139
PolyEFluc	Pamela Chenyang Cai	The University of Chicago (UChicago)	Theoretical Prediction of Dynamics from Density Fluctuations in Polyelectrolyte Complexes	Materials Science	487
PolyLLM	Juan Pablo	The University of Chicago (UChicago)	Fine Tuning LLM for Polymer Design	Materials Science	2,000
PorousMatCarbon	Mark S. Gordon	Ames Laboratory	Highly Scalable Ab Initio Simulations of N-Doped Porous Materials for Carbon Capture	Chemistry	2,039
Postcard	Michael E. Papka	Argonne National Laboratory	Postcard: Facility Data Mining	Computer Science	100,000
PPI	Arjun Saha	University of Wisconsin-Milwaukee	Simulation of Protein–Protein Interfaces for Skin Cancer	Biological Sciences	1,450
PPMstarGPU	Paul Ralph Woodward	University of Minnesota-Twin Cities	Moving PPMstar to Aurora	Physics	479

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
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	3,620
ProFESSA_rbfe	Darrin Michael York	Rutgers University	Adaptive, High-throughput Methods for Drug Discovery	Chemistry	5,000
prs-atlas	Ravi Kiran Madduri	Argonne National Laboratory	An Atlas of Multi-ancestry Polygenic Risk Scores for Heritable Human Disease	Biological Sciences	3,042
PTP_Evol	Shina Caroline Lynn Kamerlin	Georgia Institute of Technology (Georgia Tech)	Conformational Dynamics and Functional Evolution in Protein Tyrosine Phosphatases	Biological Sciences	4,375
q4bio	Katherine Inzani	The University of Nottingham	Quantum Circuit Emulation for Covalent-inhibitor Drug Discovery	Chemistry	5,000
QAOA-Simulation	Minzhao Liu	JPMorgan Chase & Co.	Fast Simulation of Quantum Approximate Optimization Algorithms	Computer Science	2,963
qbraid-q4bio	Kanav Setia	QBraid Co.	Quanta-bind: A Quantum Pipeline to Model Strong Correlation in Metalloproteins	Chemistry	5,848
QCS	Huan-Hsin Tseng	Brookhaven National Laboratory (BNL)	Quantum Centric-Supercomputing	Computer Science	2,000
QINC_Sims	Joeson Wong	The University of Chicago (UChicago)	Coherence Simulations of Spin Defects in Nanocrystal Hosts	Materials Science	5,000
QMC4Bio	Anouar Benali	Qubit Pharmaceuticals	Effects of High Accuracy Methods on the Torsion of Biomolecules	Chemistry	3,434
qmchamm	Lucas Kyle Wagner	University of Illinois at Urbana-Champaign	QMC-HAMM: High Accuracy Multiscale Models Using Quantum Monte Carlo	Materials Science	4,810
QSupremacy	Yuri Alexeev	Argonne National Laboratory	Quantum Supremacy	Computer Science	6,654
quantom_scidac	John Taylor Childers	Argonne National Laboratory	The QuantOm Event-Level Inference Framework	Physics	857
QuantumDS	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Quantum Mechanics and Data Science	Chemistry	1,457
radix-io	Philip Hutchinson Carns	Argonne National Laboratory	System Software to Enable Data-intensive Science	Computer Science	11,430
RAN	William Edward Allcock, Brian R. Toonen	Argonne National Laboratory	RAM Area Network	Computer Science	674

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
RAPINS	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Reproducible and Accelerated Physics-inspired Neural Networks	Physics	9,937
RaptorX	Jinbo Xu	Toyota Technological Institute at Chicago (TTIC)	Protein Folding through Deep Learning and Energy Minimization	Biological Sciences	190
RAPTOR_gpudev	Gregory Alan Voth	The University of Chicago (UChicago)	Enabling Large-scale Reactive Simulations of Complex Biophysical Systems	Chemistry	1,165
RBC_Conv_2	Janet Diane Scheel	Occidental College	Towards Ultimate Rayleigh-Bénard Convection, Part 2	Engineering	5,000
RECUP	Line Catherine Pouchard	Argonne National Laboratory	RECUP – Capture and Management of In-depth Provenance Data from HPC Workflows	Computer Science	7,552
Redfield	Dmitri Sergeevich Kilin	North Dakota State University (NDSU)	Machine Learning–Accelerated Nonadiatic Dynamics in Open Quantum Systems	Chemistry	7,513
ReForMerS	Sandeep Madireddy	Argonne National Laboratory	Robust and Reliable Foundation Models for Science	Computer Science	9,460
remote_offloading	Jose Manuel Monsalve Diaz	Argonne National Laboratory	Exploring Collective Operations with Remote Offloading	Computer Science	2,091
RL-fold	Arvind Ramanathan	Argonne National Laboratory	Targeting Intrinsically Disordered Proteins Using Artificial Intelligence-driven Molecular Simulations	Biological Sciences	1,967
RNAtoImage	Ravi Kiran Madduri	Argonne National Laboratory	RNA-to-Image Synthesis Project	Biological Sciences	4,266
RomanDESC	Katrin Heitmann	Argonne National Laboratory	Joint Roman-LSST DESC Time Domain Survey Simulations	Physics	175
RRAPLLM	Benjamin Alan Blakely	Argonne National Laboratory	Evaluating LLMs for SME Elicitation Support	Computer Science	2,000
S3D	Silvio Humberto Rafael Rizzi	Argonne National Laboratory	Porting S3D Code to Polaris	Chemistry	1,349
sachdeva	Rohan Sachdeva	University of California-Berkeley	Development of Plasmid Large Language Models	Biological Sciences	4,760
sbi-fair	Pete Beckman, Kamil Antoni Iskra	Argonne National Laboratory	FAIR Surrogate Benchmarks Supporting AI and Simulation Research	Computer Science	21,152

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
SBI_DE	Marco Gatti	University of Pennsylvania (UPenn)	Simulation-based Dark Energy Inference	Physics	4,778
ScalingATLAS	Walter Howard Hopkins	Argonne National Laboratory	Scaling ATLAS ML Workflows	Physics	2,897
Scaling_ALCC_2024	Liney Arnadottir	Pacific Northwest National Laboratory (PNNL)	Scaling and Benchmarking of ALCC Proposal: Structure and Reactivity of Metal/Metal-oxides under Synthesis and Operations Conditions	Chemistry	1,000
scat_lowd_super	Adrian Giuseppe Del Maestro	The University of Tennessee at Knoxville	Scattering in Low-Dimensional Superfluids	Physics	2,663
SCREAM_Calib	Jiali Wang	Argonne National Laboratory	Towards Neighborhood Scale Climate Simulations Using AI and Accelerated GPUs	Earth Science	81
SDR	Sheng Di	Argonne National Laboratory	Scalable Dynamic Scientific Data Reduction Framework (SDR)	Computer Science	4,474
SEEr-Polaris	Zhiling Lan	University of Illinois Chicago	AI-enabled Benchmarking on Polaris	Computer Science	12,527
shared-inference	Will Houser Engler	The University of Chicago (UChicago)	Proof of Concept Shared Foundation Model Inference APIs	Computer Science	300
shockdropbub	Eric Johnsen	University of Michigan	Shocks, Droplets, and Bubbles	Engineering	4,799
SimLiElectrolyte	Wei Jiang	Argonne National Laboratory	Microscopic Insight into Transport Properties of Li-battery Electrolytes	Chemistry	18,329
SLHMC	Wai Nim Alfred Tang	PhD Tutor Hub	Self Learning Hybrid Monte Carlo	Physics	285
SolarWindowsADSP	Jacqueline Manina Cole	University of Cambridge	Data-Driven Molecular Engineering of Solar-powered Windows	Materials Science	9,726
SolarWindowsASDP	Jacqueline Manina Cole	University of Cambridge	Language Models for Energy Science	Materials Science	8,045
solitons_simulations	Ashwyn Mathew Sam	Stanford University	Particle-in-Cell Simulations of Solitons in a Plasma	Physics	3,886
SpTensor	Jiajia Li	North Carolina State University (NCSU)	Accelerating Sparse Tensor Methods	Computer Science	4,474
SSLCos	Dario Dematties	Northwestern University	Self-supervised Learning for Cosmology	Physics	755

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
stabCNS	Ali Karakus	Middle East Technical University	GPU-accelerated High-order Compressible Flow Simulations	Engineering	2,053
StellTurbOpt	Walter Allen Guttenfelder	Type One Energy Group, Inc.	Optimizing Multi-channel Turbulent Transport in Stellarator Fusion Pilot Plants	Fusion Energy	9,966
SU2OTOC	Xiaojun Yao	University of Washington	OTOC Calculation for 2+1D SU(2) Lattice Gauge Theory	Physics	7,490
SuperBERT	Ian Foster	Argonne National Laboratory	Training of Language Models on Large Quantities of Scientific Text	Computer Science	20,457
superEddNS	Kyle Parfrey	Princeton Plasma Physics Laboratory (PPPL)	Radiation-GRMHD Simulations of Super-Eddington Accretion onto ULX Pulsars	Physics	6,994
Surrogate_LS_DON	Dibakar Roy Sarkar	Johns Hopkins University	Surrogate Modeling for Large-scale Simulation of Porous Media Flow	Earth Science	80
Sustainable_RecSys	Yong Zheng	Illinois Institute of Technology (IIT)	Sustainable Deployment of Recommendation Models	Computer Science	4,500
SYCLSupport	Kevin Harms	Argonne National Laboratory	SYCL Support on ALCF Systems	Computer Science	2,016
tachyon-drs	Kevin Antoney Brown	Argonne National Laboratory	Tachyon – Distributed Resilient System Modeling	Computer Science	1,951
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	5,250
ThroughFocal_DD	Jonathan Tyler Schwartz	University of Michigan	Aberration Corrected through Focal Electron Tomography ThroughFocal_DD	Materials Science	1,967
TMEM_DEL	Diomedes Elias Logothetis	Northeastern University	Molecular Dynamics Simulation on TMEM16A Chloride Channel	Biological Sciences	1,473
Tools	Scott Parker	Argonne National Laboratory	ALCF Performance Tools	Internal	1,852
topk-eigen	Ian Di Dio Lavore	Politecnico di Milano	Design of Multi-GPU Top-K Sparse Eigensolver	Computer Science	1,200
tpc	Venkatram Vishwanath	Argonne National Laboratory	Trillion Parameter Consortium	Computer Science	740

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
transformer_eval	Rick Lyndon Stevens	Argonne National Laboratory	Transformers on AI Accelerators	Computer Science	1,515
TRB	Parisa Mirbod	University of Illinois Chicago	Turbulent Rayleigh-Benard Convection in Suspensions of Bubbles	Engineering	942
TRIDENT	Marta Galbiati	Politecnico di Milano	invesTigation of laseR-drIven raDiation sourcEs with advaNced Targets	Physics	602
tstsallm	Weijie Su	University of Pennsylvania (UPenn)	Towards Safe and Trustworthy Scientific Applications of Large Language Models	Computer Science	829
TurbParticles	Luca Franci	Northumbria University	Simulations of Particle Dynamics in Turbulent Plasmas	Physics	9,732
TurbulentLiquidDrop	Arne Jacob Pearlstein	University of Illinois Urbana-Champaign	Computation of Transitional and Turbulent Drop Flows for Liquid Carbon Dioxide Drops Rising in Seawater	Engineering	6,000
uchicago-alcf-hub	Hakizumwami Birali Runesha	Argonne National Laboratory	University of Chicago Research Computing / ALCF Partnership Hub	Computer Science	121
UIC-CS455	Haritha Siddabathuni Som	University of Illinois Chicago	UIC-CS455-Sp2025	Computer Science	38
UIC-CS494-Sp2024	Michael E. Papka	University of Illinois Chicago	CS494 Introduction to High Performance Computing	Computer Science	6,000
UIC-HPC	Michael E. Papka, Zhiling Lan	University of Illinois Chicago	UIC High Performance Computing Research	Computer Science	3,221
UncertaintyDL	Maria Pantoja	Argonne National Laboratory	Quantifying Uncertainty in Deep Learning	Computer Science	945
UNNPredCloudHPC	Esma Yildirim	Brookhaven National Laboratory (BNL)	Using Neural Networks to Predict Runtime and Resource Utilization of Jobs on Integrated Cloud and HPC Systems	Computer Science	4,793
vacuumms	Franklin Ted Willmore	Boise State University	Extending VACUUMMS for Analysis of MD Trajectories	Materials Science	259
VeloC	Bogdan Florin Nicolae	Argonne National Laboratory	VeloC: Very Low Overhead Checkpointing System	Computer Science	10,900

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
Vendor_Support	William Edward Allcock, Andrew J. Cherry, Susan Marie Coghlan, Torrance Ivan Leggett, William R. Scullin	Argonne National Laboratory	Vendor Support	Internal	152
Veraaidrugdiscovery2	Julian Reyes	Vera Biotechnologies	HTS for Oncology and Neurology Targets	Chemistry	331
visualization	Joseph A. Insley, Michael E. Papka	Argonne National Laboratory	Visualization and Analysis Research and Development for ALCF	Internal	15,413
Viz_Support	Joseph A. Insley, William Edward Allcock	Argonne National Laboratory	Visualization Support	Computer Science	3,279
VLM-Plant-Disease	Xi Yu	Brookhaven National Laboratory (BNL)	Developing and Integrating Foundation Models and Causal Inference for Plant Disease Detection	Biological Sciences	961
von	Colin Reed Bundschu	Cornell University	Using Bayesian Learning and DFT to Discover Efficient M-N-C Electrocatalysts for Alkaline Hydrogen Fuel Cells and Electrolyzers	Chemistry	7,431
WallBoundedMHDTurb	Myoungkyu Lee	The University of Alabama	DNS of Wall-Bounded Magnetohydrodynamic Turbulence at High Reynolds Number	Engineering	6,977
wall_turb_dd	Ramesh Balakrishnan	Argonne National Laboratory	Wall-resolved Simulations of Canonical Wall-bounded Flows	Engineering	8,000
WALSforAll	Christine Mary Simpson	Argonne National Laboratory	Workflows across Labs	Computer Science	99
warpx_frc_stability	Roelof Erasmus Groenewald	TAE Technologies, Inc.	Global Stability Simulations of Advanced FRC Using ECP Tools	Fusion Energy	358
WRF_col	Jose Manuel Monsalve Diaz	Argonne National Laboratory	Exploring Weather Forecast Simulation in Tropical Regions	Earth Science	4,504
WRF_LES	Haochen Tan	Argonne National Laboratory	Impact of Sensible Heat Flux Due to Rainfall over Chicago Area	Earth Science	9,634
XMultiImage	Phay J. Ho	Argonne National Laboratory	Multimodal Imaging with Intense X-ray Pulses	Chemistry	1,436

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount
YouDiffuse	Nicholas Lee-Ping Chia	Argonne National Laboratory	YouDiffuse: Training Latent Diffusion Models by Leveraging Educational Medical Content from YouTube	Computer Science	5,131
ZrO2_TBC	Indranil Roy	General Electric Company (GE)	Exploring the Phase Stability of ZrO_2 and Rare Earth Sesquioxides for Advanced Thermal Barrier Coatings (TBCs)	Materials Science	7,360
				Total DD	**2855.0K**

This page intentionally left blank.

Appendix C – ALCF CY2024 Science Highlights

The following Science Highlights were submitted to ASCR for the 2024 OAR performance period.

Projected Long-Term Effects of Colorectal Cancer Screening Disruptions Following the COVID-19 Pandemic

The Science

Colorectal cancer (CRC) remains the second-leading cause of cancer deaths in the United States, and there is clear evidence that screening has a major impact on reducing the burden of CRC and is cost-effective. While CRC screening rates have increased in the past two decades, they are still below established population-level targets, and these low rates have been further exacerbated by the COVID-19 pandemic. Two independently developed microsimulation models of CRC were used to estimate the effects of pandemic-induced disruptions in colonoscopy screening for several population cohorts and variations on two commonly used screening strategies. The impact of disruptions was investigated for three scenarios: delays in screening, regimen switching, and screening discontinuation.

The Impact

Despite cancer screening reopening efforts, CRC screening has not yet returned to pre-pandemic levels. While short-term delays in screening of 3-18 months are predicted to result in minor decreases in life expectancy, discontinuing screening resulted in much more significant decreases. These findings demonstrate that unequal recovery of screening following the pandemic can further widen disparities and emphasize the importance of ensuring equitable recovery of screening following the pandemic. The worst-case scenario considered was that of 50-year-olds who postponed screening until the age of 65 when they became Medicare eligible, whereas other disruption scenarios for this group are predicted to have minor effects.

Screening benefits lost due to disruptions by cohort and scenario for two microsimulation models and for unscreened (U60), colonoscopy screening-adherent (C60), and fecal immunochemical test screening-adherent (F60)/semi-adherent (f60) 60-year-olds. (Image courtesy of Pedro Nascimento de Lima (RAND))

Contact PI: Pedro Nascimento de Lima (RAND)
ASCR Allocation PI: Jonathan Ozik (ANL)
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: January 16, 2024
Publication: P. Nascimento de Lima, et al., eLife, **12e85264** (2023).

RAND Corp., Erasmus MC, MSK Cancer Center, Argonne, and Fred Hutchinson Cancer Center

U.S. DEPARTMENT OF ENERGY | Office of Science

NCI, NIH/NCI

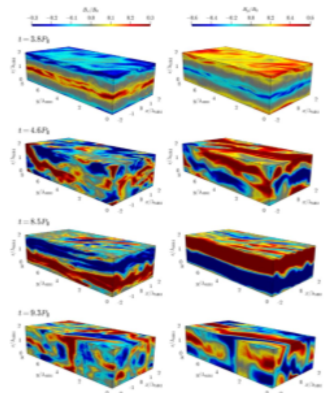
Fully Kinetic 3D Shearing-Box Simulations of Magnetorotational Turbulence in Pair Plasmas

The Science

Supermassive black holes (SMBH), with masses between several hundred thousand and several billion times the mass of our own Sun, are thought to inhabit the centers of nearly all large galaxies. While we cannot directly observe black holes because no light can escape them to reach us, the in-spiraling plasma (hot, ionized gas) in the "accretion disks" surrounding black holes produces photons that can escape the system and reach our telescopes. Within the past few years, a global network of radio telescopes called the Event Horizon Telescope (EHT) has used very-long-baseline interferometry (VLBI) to produce the first images of the accretion disks around SMBHs, including M87* at the center of the Messier 87 supergiant elliptical galaxy and Sagittarius A* (SgrA*) at the center of our own Milky Way galaxy. We need sophisticated models and supercomputers to make the most of these fantastic new observational capabilities to understand the connections between physical interactions in the accretion disk and observational signatures.

The Impact

The project team used their Zeltron code on Theta to generate the largest 3D simulation to date of plasma turbulence driven by magnetorotational instability (MRI). This simulation modeled the evolution of a section of a rotating accretion disk using the particle-in-cell (PIC) method, which reproduces the physics of turbulent plasmas from first principles. Previous studies in the literature employed 3D simulations using much less computationally expensive magnetohydrodynamic (MHD) methods to study the turbulence that develops on large scales in accretion disks due to MRI. However, MHD methods cannot model the full cascade of energy from macroscopic scales to the small-scale processes that produce the observable properties of the accretion disk. This project's large 3D PIC simulation agrees well with previous MHD results on larger scales, and the first-principles nature of the PIC methods allows them to model microscopic processes and start drawing quantitative connections between MRI-induced turbulence and observations of accretion disks.



Spatial distributions of the radial (B_x) and toroidal (B_y) magnetic fields during the development of sustained large-scale turbulence due to magnetorotational instability (MRI) in a simulation of pair-plasma in a large 3D box using Theta at ALCF. (Image courtesy of Fabio Bacchini)

Contact PI: Fabio Bacchini (Katholieke Universiteit Leuven)
ASCR Allocation PI: Dmitri A. Uzdensky (U. of Colorado Boulder)
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: February 2024
Publication: F. Bacchini et al., *ApJ*, **938**, 86 (2022).

U. of Colorado Boulder; IAS, Princeton; Flatiron Institute; JILA/NIST

U.S. DEPARTMENT OF ENERGY | Office of Science

NASA, NSF, FWO, Simons Foundation

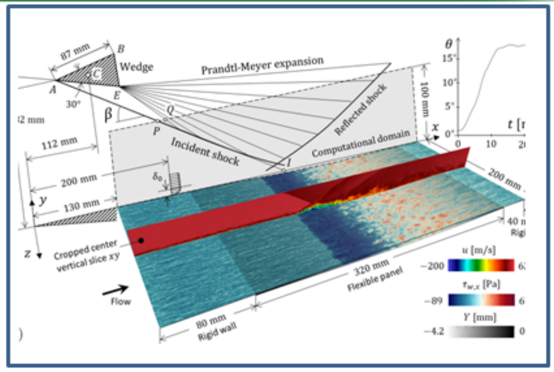
Three-Dimensional Shock Boundary Layer Interactions Over Flexible Walls

The Science

Supersonic and hypersonic (greater than five times the speed of sound) flows are often characterized by shocks or sharp discontinuities in the flow properties. When a sufficiently strong oblique shock impinges on a turbulent boundary layer, it creates a flow separation bubble that can oscillate in size and location and be detrimental to the structure on which the shock impacts. Developing a better understanding of Shock/Boundary-Layer Interactions (SBLI) with coupled Fluid-Structure Interaction (FSI) is important for the design and development of hypersonic aircraft.

The Impact

The project team used Theta to perform full 3D simulations of the coupling between FSI and SBLI to replicate and complement wind tunnel experiments. This study is the first time in the open literature to incorporate wall-modeled LES (WMLES) and an FEM solid mechanics solver to study the low-frequency motions that are characteristic of SBLI+FSI interactions on flexible panels. The use of WMLES in the flow solver greatly reduces the computational cost, relative to wall-resolved simulations, by modeling (rather than resolving) the effects of dominant turbulent motions in the very-near wall region. This cost reduction enables feasible computations for much longer times, for meaningful studies of the coupled interactions of flow and structure at the low frequencies that can lead to structural failure. This modeling strategy is also a stepping stone towards incorporating thermal coupling effects between the fluid and solid response, which often occurs at even longer time scales. Ultimately, these predictive simulation techniques will enable the development of more aggressively designed hypersonic vehicles by reducing the required safety factors due to simulation uncertainties.



The computational setup of the full-span 3D SBLI simulations. By oscillating the wedge, the strength of the SBLI can be increased to study the effects of flow separation and that of an oscillating separation bubble coupled with the fluid-structure interaction. Simulation details compared well with measurements from flexible-panel experiments. (Image courtesy of Ivan Bermejo-Moreno)

Contact PI: Ivan Bermejo-Moreno
ASCR Allocation PI: Johan Larsson
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: March 2024
Publication: J. Hoy, I. Bermejo-Moreno, *Flow* (2022), **2** E35

University of Maryland, College Park; University of Southern California, Los Angeles

U.S. DEPARTMENT OF ENERGY | Office of Science

NSF

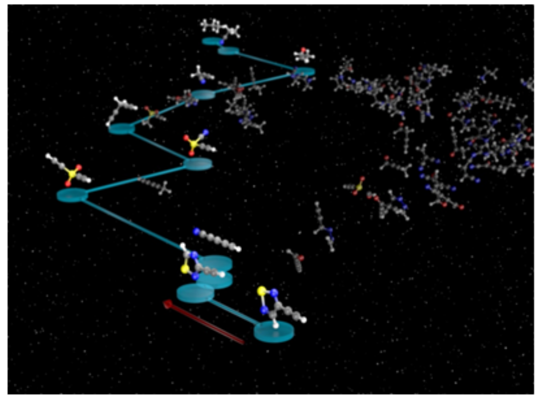
Understanding the Relationships between Quantum Mechanical Properties Reveals the Pathways of Molecular Design

The Science

Chemical and materials properties are dictated by the structural arrangements of atoms. The rational design of molecules with targeted properties for applications requires an advanced understanding of structure-property and property-property relationships that exist across a vast chemical compound space. Accessing large data sets of molecules and their quantum mechanical calculations can assist in unlocking such rational molecular designs aiming to maximize desired properties, such as excitation energies and polarizability. Researchers characterized regions of molecular property space, revealing that one has a substantial degree of flexibility or "freedom of design" when searching for a single molecule with a desired pair of properties or a set of molecules sharing several properties.

The Impact

The chemical compound space is known as the universe of molecules that could exist and is practically infinite. Effective screening of this space is key to designing molecules with tailored properties for applications spanning high-performance electronic materials, better pharmaceuticals, and high-yield catalysts. The "freedom of design" discovered by the researchers has important implications in advancing the molecular design process. It is estimated that 7-10 properties are needed to uniquely identify small (primarily organic) molecules. This work also emphasizes the critical importance of high-quality structure-property data in the training and development of next-generation machine-learning approaches.



Conceptual representation of the chemical compound space navigated with multi-property optimization. (Image courtesy of Victor Mateevitsi (ANL))

Contact PI: Alexander Tkatchenko (U. Luxembourg)
ASCR Allocation PI: Alexander Tkatchenko (U. Luxembourg)
ASCR Program/Facility: DD/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: April 16, 2024
Publication: L. Medrano Sandonas et al., *Chem. Sci.*, **14** 10702 (2023).

University of Luxembourg, University of Graz, Cornell University, Argonne

U.S. DEPARTMENT OF ENERGY | Office of Science

ERC-CoG, NSF, Sloan Research Fellowship

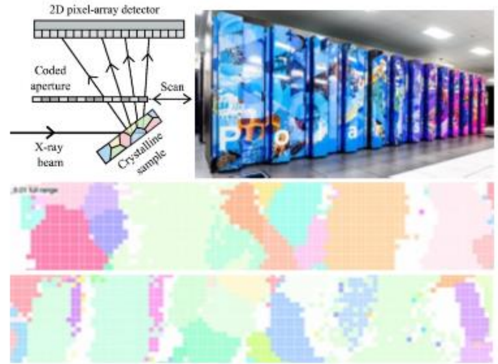
Demonstrating Cross-Facility Data Processing at Scale with Laue Microdiffraction

The Science

Laue microdiffraction is a technique employed at synchrotron facilities, such as the Advanced Photon Source (APS), to analyze the local structure of materials with crystalline structures. This method requires a significant amount of computation, where a single 12-hour scan being processed on a CPU workstation computer is estimated to require over 250 days to complete the analysis. The APS upgrade project will offer users X-ray beams 500 times more powerful than the previous facility, with commensurate increases in experimental data generated for processing. A team from APS, ALCF, and Globus developed high-performance software tools and data infrastructure to integrate APS data processing and analysis in near-real-time with ALCF computing resources.

The Impact

The team demonstrated the on-demand use of ALCF resources to reconstruct data obtained from the APS beamline, returning reconstructed scans to the APS within 15 minutes of being sent to the ALCF. Using large-scale computing resources, the APS will be able to reconstruct coded aperture Laue datasets automatically in near-real-time, allowing users to collect data about an order of magnitude faster than is possible today, accelerating time to science. Building a common infrastructure such as this enables the APS, and potentially other facilities, to use ALCF supercomputing resources.



Data from the new coded aperture at an APS beamline (top left) is automatically transferred to the Polaris supercomputer (top right), where it is reconstructed on demand in real time (bottom). (Image courtesy of Michael Prince et al. (ANL))

Contact PI: Nicholas Schwarz (ANL)
ASCR Allocation PI: Nicholas Schwarz (ANL)
ASCR Program/Facility: DD/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: May 6, 2024
Publication: M. Prince et al., Proc. SC-W 2023 (pp. 2133-2139), Denver, CO, USA, November 12–17, 2023.

Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE BES

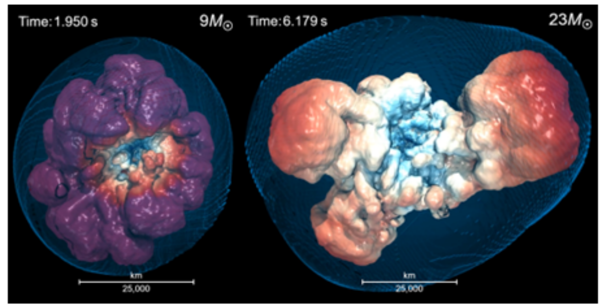
A Theory for Neutron Star and Black Hole Kicks and Induced Spins

The Science

Supernovae shape our universe and life as we know it, but how exactly a massive star explodes as a supernova remains a mystery. Images from NASA's James Webb Space Telescope provide an unprecedented view of supernovae remnants and other mysterious cosmological phenomena, but scientists need to peer deep inside massive stars to understand the internal mechanisms behind the distant cosmic explosions. The project team has been generating one of the largest collections of 3D supernova simulations with the aim of illuminating how and why a massive star goes supernova.

The Impact

Using 20 long-term three-dimensional (3D) core-collapse supernova simulations, the team found that lower compactness progenitors that explode quasi-spherically due to the short delay to explosion, experience smaller neutron star recoil kicks, while higher compactness progenitors that explode later and more aspherically, leave neutron stars with larger kicks. In addition, the researchers found that these two classes correlate with the neutron star's gravitational mass. This correlation suggests that the survival of binary neutron star systems may partly be due to their lower kick speeds. This new 3D model suite provides a greatly expanded perspective and appears to explain some observed pulsar properties by default.



An isosurface (10%) of the ^{56}Ni abundance in the explosion of a 9 $M_\odot$ (left) and 23 $M_\odot$ (right) progenitor model at ~1.95 and ~6.2 seconds, respectively, after the bounce, painted by Mach number. Red/purple is positive and white/blue is negative, indicating infall. (Image courtesy of Joseph Insley (ANL))

Contact PI: Adam Burrows
ASCR Allocation PI: Adam Burrows
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: June 2024
Publication: A. Burrows *et al.*, ApJ **963**(1):63 (2024)

Princeton University, Carnegie Observatories

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE, NSF, Princeton IAS

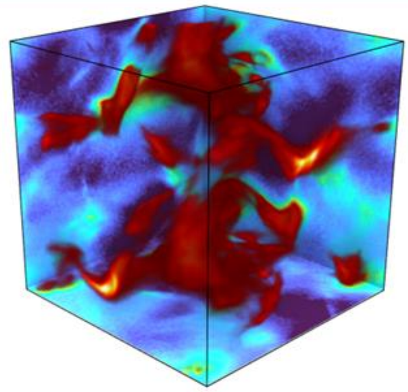
Simulations of Turbulence in Black Hole Coronae

The Science

Matter accreting onto black holes makes up some of the brightest X-ray sources in the Universe. Yet, since the early days of X-ray astronomy in the 1960s, researchers have struggled to explain the origin of the observed emission. A research group is now using powerful supercomputers to study the self-consistent interplay of turbulence and radiation in the most magnetized regions around luminous black hole systems to assess the origin of the observed hard X-ray emission.

The Impact

A new theoretical study by the project team proposes that the X-rays are produced in a strongly turbulent medium threaded by intense magnetic fields near the black hole. Their model is supported by first-of-a-kind computer simulations of electromagnetic turbulence performed on Theta, which track the interactions between billions of charged particles and photons. The researchers compared their simulation results to the black hole Cygnus X-1 in the Milky Way, one of the brightest persistent sources of X-rays in the sky. They demonstrated good agreement with the observed emission. In the future, their novel method could be used to study X-ray emission from various astronomical sources, from stellar-mass black holes to the supermassive black holes in the centers of galaxies.



Spatial structure of the photon energy density in a simulation of a turbulent plasma with moderate optical depth. (Image courtesy of Daniel Grošelj)

Contact PI: Daniel Grošelj (KU Leuven & Columbia U.)
ASCR Allocation PI: Luca Comisso (Columbia U.)
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: August 2024
Publication: D. Grošelj et al., *Phys. Rev. Lett.*, **132**, 085202 (2024)

KU Leuven, Columbia University, PPPL, Max Planck Institute for Astrophysics, University of Maryland

U.S. DEPARTMENT OF ENERGY | Office of Science

FWO, DOE, NASA, NSF, Simons Foundation

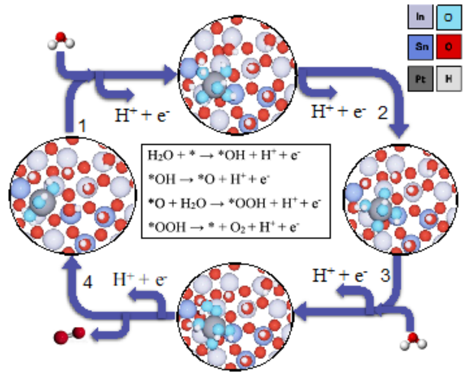
Elucidation of Active Site for the Oxygen Evolution Reaction on a Single Pt Atom

The Science

Single-atom catalysts based on noble metals represent an ultimate dispersion of the metal with all atoms exposed and, if activity remains high, optimal utilization and cost. They have attracted attention for their high catalytic activity and selectivity, but the nature of their active sites under realistic reaction conditions is not well understood. These catalysts respond to reaction conditions by adapting their set of ligands and their mode of interaction with the support, making the determination of the active site challenging. The team employed a rigorous computational methodology to investigate the active site of the oxygen evolution reaction, including the influence of the electrochemical potential.

The Impact

It was shown that the nature of the ligands on the platinum atom changed in the presence of electrochemical potential. The operando computational approach was crucial to identifying the composition and geometry of the active site under reaction conditions, which may greatly change because of the numerous adsorbates from the original initial structure. It was only with the active site determined under electrochemical conditions that excellent agreement with experiment was found. The first-principles determination of the nature of the active site in electrocatalytic conditions, coupled with experimental characterization, opens major perspectives for our fundamental understanding of electrocatalysis and the design of efficient catalysts.



The four stages of the OER reaction mechanism (blue, Tin; light gray, Indium; gray, Platinum; red, Oxygen; white, Hydrogen; oxygens of adsorbed OH groups and O species binding the Pt atom are shown in light blue). (Image courtesy of Philippe Sautet).

Contact PI: Philippe Sautet (UCLA)
ASCR Allocation PI: Anastasia Alexandrova (UCLA)
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: September 9, 2024
Publication(s) for this work: S. Kumari and P. Sautet, *J. Phys. Chem. Lett.* **14**, 2635 (2023).

University of California, Los Angeles

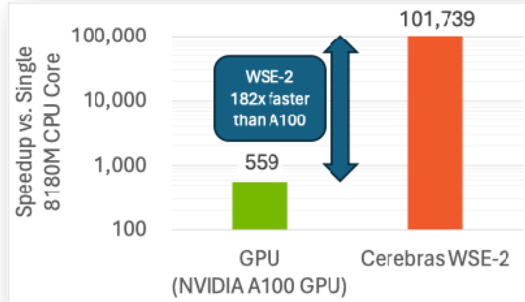
U.S. DEPARTMENT OF ENERGY | Office of Science

DOE BES

Efficient Algorithms for Monte Carlo Particle Transport on AI Accelerator Hardware

The Science

Artificial Intelligence (AI) accelerators, such as the Cerebras Wafer-Scale Engine 2 (WSE-2), are custom-made to efficiently perform AI training tasks at high speed. While not intended for traditional modeling and simulation workloads, aspects of these accelerators make them attractive for some simulation algorithms, nonetheless. A team has developed new algorithms and performance optimization strategies to enable a key Monte Carlo (MC) particle transport simulation kernel to effectively use the device. Speedups of 182x over a single GPU were observed for the algorithm using a production-quality workload typical of a nuclear reactor simulation.



Speedup vs. serial CPU execution for macroscopic cross section lookup kernel (adapted from XSBench). (John Tramm ANL).

The Impact

Significant speed and power advantages compared to highly optimized CPU and GPU implementations were found, suggesting that acceleration of a full MC particle transport code on WSE-2 would be possible. New algorithms for minimizing communication costs and for handling load balancing were developed and evaluated, which could aid in the development of optimized algorithms for related simulation methods. AI accelerators, such as the WSE-2, could offer significant advantages to traditional simulation workloads and the development of higher-level programming models to more readily enable software development and exploration could have a tremendous impact for HPC simulations.

Contact PI: John Tramm (ANL)
ASCR Allocation PI: John Tramm (ANL)
ASCR Program/Facility: DD/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: October 7, 2024
Publications: J. Tramm, et. al., *Comput. Physics Commun.* **298**, 109072 (2024); J. Tramm, et. al., *PHYSOR 2024*. (2024).

ANL, U. Chicago, Cerebras

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE ANL

From Quantum Mechanics To Hypersonic Aerodynamics

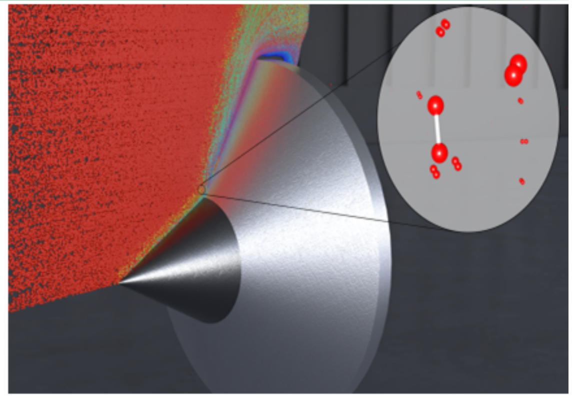
The Science

In recent years, there has been interest in developing hypersonic flight capabilities globally as these capabilities hold immense importance for national security and developing a healthy space industry. At the core of these developments is the pursuit of accuracy in predicting the environment surrounding the hyper-velocity vehicle. At high Mach numbers, the shock encapsulating the vehicle is strong enough to cause chemical reactions and excitation of internal energy modes in the shock-heated gas. An accurate description of these processes is essential to many aspects of flight, especially the heat loads experienced by the vehicle. In this work, for the first time, researchers replicated a hypervelocity ground-test experiment using the novel direct molecular simulation (DMS) method. The DMS method uses molecular dynamics guided by interaction potentials grounded in quantum mechanics to construct this complex flow field, providing a unified solution to this multi-physics environment by seamlessly linking the field of quantum chemistry to nonequilibrium fluid mechanics.

The Impact

Using Theta, for the first time, a full-scale reactive hypersonic flow studied experimentally was simulated using a method grounded in quantum mechanics. Due to the fundamental nature of the simulation technique, this work provides a molecular-level description of internal energy excitation and reaction mechanisms throughout the system. A simulation of such fidelity and detail can be used as a benchmark solution to verify computational fluid mechanics (CFD) codes and assess fundamental gaps and opportunities of improvement for the physics simulated by lower fidelity models. This, in turn, enables robust and accurate CFD solutions to mission-critical scenarios where the flow field may be in thermal and chemical nonequilibrium.

Distribution Statement A: Approved for Public Release; Distribution is Unlimited. AFRL-2024-5606



Direct Molecular Simulation (DMS) of hypervelocity flow over a double cone. The molecular dynamics are guided by interaction potentials grounded in quantum mechanics to construct this complex flow field. The solution to this multi-physics environment seamlessly links the field of quantum chemistry to nonequilibrium fluid mechanics.

ASCR Allocation PI: Maninder S. Grover
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: November 2024
Publication: Grover, M. S., et al., J. Fluid Mech (2024).

Air Force Research Laboratory, University of Dayton Research Institute

U.S. DEPARTMENT OF ENERGY | Office of Science

Air Force Office of Scientific Research

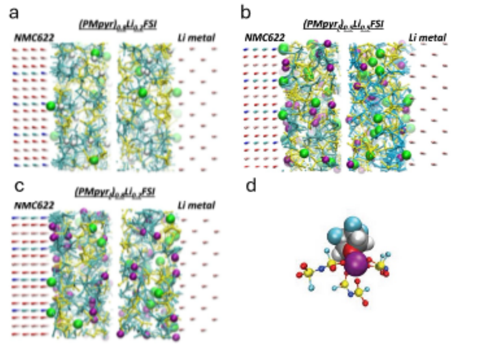
Computer Simulation-Aided Design of High-Performance Electrolytes for Lithium Batteries

The Science

For scientists working on future lithium-battery chemistries of high energy densities, designing a robust protective “interphase” formed by sacrificial electrolytes becomes the central mission of developing better electrolytes. It has been experimentally recognized that electrolyte fluorination is a promising solution to create such an interphase. However, the challenge is the ability to rationally design a fluorination strategy without relying on demanding lab implementations. Researchers have unlocked the fundamental mechanism of fluorination effects on battery stability, lifetime, and capacity through extensive molecular dynamics simulations under realistic experimental conditions. This study opens a broad design space for next-generation electrolytes for these coveted battery chemistries.

The Impact

Fundamental knowledge of the prevailing fluorination process, particularly at the atomic scale, speeds up the discovery of new electrolytes without extensive wet lab synthesis and validations, and minimizes related hazards. Developing a high-fidelity, all-atom simulation and modeling methodology that can predict fluorination effects on interphase formation and charge conduction, as well as physicochemical mechanisms within a realistic electrolyte-electrode matrix, is a key step towards the rational design of fluorinated electrolytes. The novel all-atom models for cathode and anode and the ensemble-enhanced sampling algorithm open the door to virtually characterize fluorination effects on interphase formation and lithium-ion transport properties.



(a-c) Simulation snapshots of electrolyte distribution on NMC62 and Li electrodes: cyan-IL cation; white-H on the PMpyr+ backbone; purple-F on the PMpyrf+; yellow-FSI−; green-Li+. (d) A representative snapshot of Li+ ion solvation structure with TfTHF that results in better mobility. (Image courtesy of Wei Jiang (ANL))

Contact PI: Zhengcheng Zhang (ANL)
ASCR Allocation PI: Wei Jiang (ANL)
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Saswata Hier-Majumder
Date submitted to ASCR: December 7, 2024
Publications: Q. Liu et al., *Adv. Sci.* 2409662 (2024); Q. Liu, et al., *Nat. Commun.*, **14**, 3678 (2023).

Argonne, PNNL, U.S. Army Research Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

EERE - VTO

Appendix D – Acronyms and Abbreviations

2D, 3D	two (three)-dimensional
AED	automated external defibrillator
AFD	Argonne Fire Department
AI	artificial intelligence
ALCC	ASCR Leadership Computing Challenge
ALCF	Argonne Leadership Computing Facility
ALS	Advanced Light Source
API	application programming interface
APS	Advanced Photon Source (Argonne)
Argonne, ANL	Argonne National Laboratory
ASCR	Advanced Scientific Computing Research
ASO	Argonne Site Officer
ATPESC	Argonne Training Program on Extreme-Scale Computing
ATO	Authority To Operate
BoF	Birds of a Feather
CELS	Computing, Environment, and Life Sciences (Argonne)
CFD	computational fluid dynamics
CPS	Computational Science Division (Argonne)
CPU	central processing unit
CRC	colorectal cancer
CSPO	Cyber Security Program Office (Argonne)
CY	calendar year
DAOS	Distributed Asynchronous Object Storage
DART	Days Away Restricted or Transferred
DD	Director's Discretionary
DFT	density-functional theory
DL	deep learning
DOE	U.S. Department of Energy
DOI	Digital Object Identifier
DSAV	Division Site Assist Visit
ECP	Exascale Computing Project
ESH, ES&H	Environmental, Safety and Health (Argonne)
ESP	Early Science Program, Electrical Safety Program
FedRAMP	Federal Risk and Authorization Management Program
FEM	Finite Element Method
FL	federated learning
FNAP	Foreign National Access Program

FSI	fluid-structure interaction
FY	fiscal year
GPU	graphics processing unit
HACC	Hardware/Hybrid Accelerated Cosmology Code
HDF5	Hierarchical Data Format version 5
HEP	high energy physics
HEP-CCE	High Energy Physics Center for Computational Excellence
HPC	high-performance computing
HPE	Hewlett Packard Enterprise
HPCM	HPE Performance Cluster Manager
HPSS	high-performance storage system
HTML	HyperText Markup Language
IEEE	Institute of Electrical and Electronics Engineers
I/O	input/output
INCITE	Innovative and Novel Computational Impact on Theory and Experiment
IPU	intelligence processing unit
IRI	Integrated Research Infrastructure
ISC	International Supercomputing Conference
ISM	Integrated Safety Management (Argonne)
ISSF	Interim Supercomputing Support Facility (formerly of Argonne)
JHQ	job hazard questionnaire
JLSE	Joint Laboratory for System Evaluation
JSA	Job Safety Analysis
LCF	Leadership Computing Facility
LES	large eddy simulation
LL	lesson learned
LLM	large-language model
MC	Monte Carlo
MD	molecular dynamics
ML	machine learning
MPI	Message Passing Interface
MTBI	Mean Time Between Interrupt
MTTF	Mean Time to Failure
MTTI	Mean Time to Interrupt
NAIRR	National Artificial Intelligence Research Resource
NAS	Neural Architecture Search
NDA	nondisclosure agreement
NERSC	National Energy Research Scientific Computing Center
NIH	National Institutes of Health

NIST	National Institute of Standards and Technology
NP	Nuclear Physics (Argonne)
NRE	Non-recurring Engineering
NSF	National Science Foundation
Oak Ridge, ORNL	Oak Ridge National Laboratory
OAR	Operational Assessment Report
OJT	on-the-job training
OLCF	Oak Ridge Leadership Computing Facility
OS	Operating System
P3HPC	International Workshop on Performance, Portability and Productivity in HPC
PB	Petabyte
PF	Petaflop
PDSA	Plan-Do-Study-Act
PDU	power distribution unit
PI	principal investigator
PM	preventative maintenance
PMBS	Performance Modeling, Benchmarking and Simulation of High-Performance Computer Systems
PPE	personal protective equipment
PPI	Protein-Protein Interaction
PPPL	Princeton Plasma Physics Laboratory
PT	Physical Therapy (Argonne)
PY	previous year
QCD	Quantum Chromodynamics
QMC	Quantum Monte Carlo
QoS	quality-of-service
RAM	Random Access Memory
RDMA	Remote Direct Memory Access
RDU	Reconfigurable Dataflow Unit
RMP	risk management plan
RTR	Response to Recommendation
SBLI	shock/boundary-layer interaction
SC	Office of Science (DOE)
SC23	2023 International Conference for High Performance Computing, Networking, Storage and Analysis (annual supercomputing conference)
SC24	2024 International Conference for High Performance Computing, Networking, Storage and Analysis
SciDAC	Scientific Discovery through Advanced Computing
SME	subject matter expert
SRAM	static random access memory

SSL	Secure Sockets Layer
STEM	science, technology, engineering and math
TB	Terabyte
TCS	Theory and Computing Sciences (Argonne)
THAPI	Tracing Heterogeneous API
TMS	Training Management System (Argonne)
TRCs	Total Recordable Cases
TSP	Tensor Streaming Processor
UAC	User Advisory Council (ALCF)
UB3	Userbase 3
UChicago	University of Chicago
UI	user interface
UIC	University of Illinois at Chicago
UX	User Experience
VLM	Vision-Language Model
VM	virtual machine
WCD	Work Control Documents
WMLES	wall-modeled large eddy simulation
WPC	work planning and control
WRF	Weather Research and Forecasting
WSE2	Wafer-Scale Engine 2

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

DOCUMENT AVAILABILITY

Online Access: U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free at OSTI.GOV (http://www.osti.gov/), a service of the US Dept. of Energy's Office of Scientific and Technical Information.

Reports not in digital format may be purchased by the public from the National Technical Information Service (NTIS):
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312
www.ntis.gov
Phone: (800) 553-NTIS (6847) or (703) 605-6000
Fax: (703) 605-6900
Email: **orders@ntis.gov**

Reports not in digital format are available to DOE and DOE contractors from the Office of Scientific and Technical Information (OSTI):
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
www.osti.gov
Phone: (865) 576-8401
Fax: (865) 576-5728
Email: **reports@osti.gov**

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

www.anl.gov

U.S. DEPARTMENT of ENERGY

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