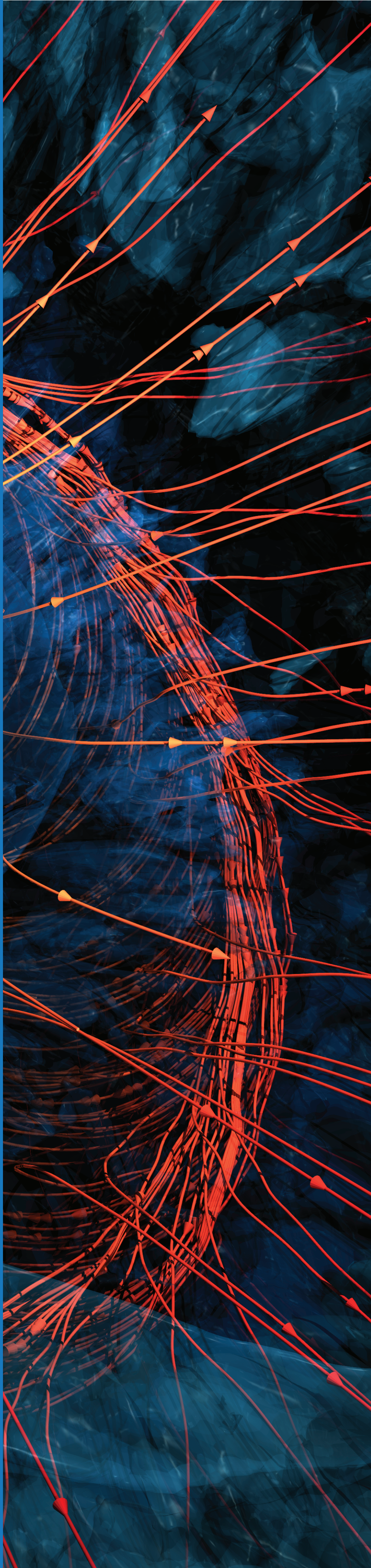
Argonne Leadership Computing Facility



2021 Operational Assessment Report

On the cover: *A team led by researchers from the Kavli Institute for Theoretical Physics is using Argonne's Theta supercomputer to perform radiation hydrodynamic simulations of massive stars with rotation. This image shows the structures of the outer envelope of a rotating 80 solar mass star. The color represents the gas density while the streamlines show the flow velocity. The envelope is convective around the yellow region driven by the strong radiation field from the core and the iron opacity peak. Convection transports most of the angular momentum outward and makes the photosphere rotate rapidly. Rotation also enhances the mass loss rate from the envelope.*

Image credit: *ALCF Visualization and Data Analysis Team; Yan-Fei Jiang, Center for Computational Astrophysics (CCA), Flatiron Institute; Lars Bildsten, Kavli Institute for Theoretical Physics (KITP) and University of California, Santa Barbara.*

Contents

Contents (Cont.)

Contents (Cont.)

Contents (Cont.)

Contents (Cont.)

Figures

Figures (Cont.)

Tables

Tables (Cont.)

Background

In 2004, the U.S. Department of Energy's (DOE's) Advanced Scientific Computing Research (ASCR) program founded 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 capability 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 petascale machines that are among the most powerful systems in the world today. Strategically, the LCF ranks among the top U.S. scientific facilities delivering impactful science. The work performed at these centers informs policy decisions and advances innovations in far-reaching topics such as energy assurance, ecological sustainability, and global security.

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

This page intentionally left blank.

Executive Summary

This Operational Assessment Report describes how the Argonne Leadership Computing Facility (ALCF) met or exceeded every one of its goals for calendar year (CY) 2021 as an advanced scientific computing center.

In CY 2021, the ALCF operated its production resource, Theta, an Intel-based Cray XC40 system (11.7-petaflops) augmented with 24 NVIDIA DGX A100-based nodes (3.9-petaflops) that supports diverse workloads, integrating data analytics with artificial intelligence (AI) training and learning in a single platform. In 2021, we began deploying Polaris, our newest 40-petaflops system, and augmented this powerful testbed system with an additional 28 nodes to support the integration of real-time experiments and HPC resources. We also deployed our two largest storage systems yet, named Grand and Eagle, that will bring new services to our users and will power data-driven research for years to come.

Last year, Theta delivered a total of 20.8 million node-hours to 16 Innovative and Novel Computational Impact on Theory and Experiment (INCITE) projects and 7.2 million node-hours to ASCR Leadership Computing Challenge (ALCC) projects (32 awarded during the 2020–2021 ALCC year and 17 awarded during the 2021–2022 ALCC year), as well as substantial support to Director's Discretionary (DD) projects (5.5 million node-hours). As Table ES.1 shows, Theta performed exceptionally well in terms of overall availability (95.1 percent), scheduled availability (99.4 percent), and utilization (98.1 percent; Table 2.1).

As of the submission date of this document, ALCF's user community has published 249 papers in high-quality, peer-reviewed journals and technical proceedings. At the 2021 International Conference for High Performance Computing, Networking, Storage and Analysis (SC'21), Argonne researchers won two HPCwire Readers' Choice Awards and were part of a Gordon Bell Prize finalist team recognized for developing an AI-enabled, multi-resolution simulation framework for studying complex biomolecular machines. Their framework was used to observe the SARS-CoV-2 replication-transcription machinery in action, by directly integrating experimental data. ALCF also provided a comprehensive program of high-performance computing (HPC) support services to help our community make productive use of the facility's diverse and growing collection of resources.

We are now entering the exascale era, with exascale machines being planned for national laboratories across the country, including Aurora at Argonne National Laboratory (Argonne) in 2023. ALCF researchers have been leading and guiding numerous strategic activities that will push the boundaries of what's possible in computational science and engineering and allow us to deliver science on day one.

Table ES.1 Summary of the Target and Actual Data for the Previous Year (PY) (CY 2021) Metrics

Area	Metric	2021 Target	2021 Actual
User Results	User Survey – **Overall Satisfaction**	3.5/5.0	4.5/5.0
	User Survey – **User Support**	3.5/5.0	4.6/5.0
	User Survey – **Problem Resolution**	3.5/5.0	4.6/5.0
	User Survey – **Response Rate**	25.0%	42.3%
	% of User Problems Addressed within Three Working Days	80.0%	95.0%
Business Results	Theta with expansion Overall Availability	80.0%	95.1%
	Theta with expansion Scheduled Availability	90.0%	99.4%
	% of INCITE node hours from jobs run on 20.0% or more of Theta (800–4008 nodes)	20.0%	84.8%
	% of INCITE node hours from jobs run on 60.0% or more of Theta (2400–4008 nodes)	N/A	22.7%
	Lustre File System Overall Availability	90.0%	96.1%
	Grand File System Overall Availability	N/A	96.5%
	Eagle File System Availability	N/A	96.5%
	HPSS [a] Overall Availability	90.0%	96.5%

[a] HPSS = high-performance storage system

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 Argonne Leadership Computing Facility (ALCF) has processes in place to effectively support its customers, to resolve user problems, and to conduct outreach. The 2021 annual user survey measured overall satisfaction, user support, and problem resolution, and thereby served 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 that were made to those processes during calendar year (CY) 2021.

Table 1.1 All 2021 User Support Metrics and Results [a]

		2020 Target	2020 Actual	2021 Target	2021 Actual
Number Surveyed		N/A [c]	1,174	N/A [c]	1,141
Number of Respondents (Response Rate)		25.0%	500 (42.6%)	25.0%	483 (42.3%)
Overall Satisfaction	Mean	3.5	4.6	3.5	4.5
	Variance	N/A	0.5	N/A	0.5
	Standard Deviation	N/A	0.7	N/A	0.7
Problem Resolution	Mean	3.5	4.6	3.5	4.6
	Variance	N/A	0.4	N/A	0.5
	Standard Deviation	N/A	0.7	N/A	0.7
User Support	Mean	3.5	4.6	3.5	4.6
	Variance	N/A	0.4	N/A	0.5
	Standard Deviation	N/A	0.7	N/A	0.7
		2020 Target	**2020 Actual**	**2021 Target**	**2021 Actual**
% of Problems Addressed Within Three Working Days [b]		80.0%	97.7%	80.0%	95.0%

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

[c] N/A = not applicable.

Survey Approach

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

The 2021 user survey closely resembled the 2020 survey, with a few minor modifications to the infrastructure, performance, and debugging tools/services listed in various questions. Two new questions were added: one that captured the use or planned use of workflow tools on ALCF systems, and another that checked which accelerator programming model was currently used in users' codes; and two questions were removed: one that captured users' enthusiasm for an ALCF Slack channel, and another that gauged their interest in participating in an annual meeting.

The 2021 survey was administered through a new contract with Statistical Consulting Services at the Department of Statistics and Actuarial Science at Northern Illinois University and consisted of six required questions and 21 optional questions. The survey and associated e-mail campaign ran from November 16, 2021, through December 31, 2021. Each reminder e-mail included a user-specific link to the online survey. Most respondents were able to complete the survey in 10 minutes or less.

Likert Scale and Numeric Mapping

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

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

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

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

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

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

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

Comments

1.1 User Support Metrics

Everyone who met the definition of a facility user during the fiscal year (FY) 2021—that is, users who would be counted under the Facility's annual user statistics submission to the Office of Science—was asked to complete a user survey, which came to 1,168 individuals in total. Of those individuals, 27 did not receive the email due to undeliverable messages and opt-outs. Of the 1,141 remaining users, 483 responded, for a 42.3 percent response rate. The ALCF surpassed all targets for the survey metrics.

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

Table 1.2 User Survey Results in 2021 by Allocation Program

Metrics by Program		INCITE	ALCC	INCITE + ALCC	DD	All
Number Surveyed		194	124	318	823	1,141
Number of Respondents		75	68	143	340	483
Response Rate		38.7%	54.8%	45.0%	41.3%	42.3%
Overall Satisfaction	Mean	4.7	4.6	4.7	4.5	4.5
	Variance	0.3	0.7	0.5	0.5	0.5
	Standard Deviation	0.5	0.8	0.7	0.7	0.7
Problem Resolution	Mean	4.7	4.6	4.6	4.5	4.6
	Variance	0.3	0.6	0.4	0.5	0.5
	Standard Deviation	0.5	0.8	0.7	0.7	0.7
User Support	Mean	4.6	4.6	4.6	4.5	4.6
	Variance	0.4	0.6	0.5	0.5	0.5
	Standard Deviation	0.6	0.7	0.7	0.7	0.7
All Questions	Mean	4.7	4.6	4.6	4.5	4.6
	Variance	0.3	0.6	0.5	0.5	0.5
	Standard Deviation	0.6	0.8	0.7	0.7	0.7

As Table 1.3 shows, in 2021, the ALCF again exceeded the targets for overall satisfaction and user support.

Table 1.3 User Support Metrics

Survey Area	2020 Target	2020 Actual	2021 Target	2021 Actual
Overall Satisfaction Rating	3.5/5.0	4.6/5.0	3.5/5.0	4.5/5.0
Avg. of User Support Ratings	3.5/5.0	4.6/5.0	3.5/5.0	4.6/5.0

1.2 Problem Resolution Metrics

Table 1.4 shows the target set for the percentage of user problem tickets or "trouble tickets" addressed in three days or less, which ALCF exceeded. A trouble ticket, which encompasses incidents, problems, and service requests, is considered "addressed" once (1) the ticket is accepted by a staff member; (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

	2020 Target	2020 Actual	2021 Target	2021 Actual
% of Problems Addressed Within Three Working Days [a]	80.0%	97.7%	80.0%	95.0%
Avg. of Problem Resolution Ratings	3.5/5.0	4.6/5.0	3.5/5.0	4.6/5.0

[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 and E-mail Support

In 2021, ALCF addressed and categorized 5,034 user support tickets, a 10 percent increase from the previous year. The biggest increases were in the Accounts and Allocations categories (Table 1.5). There are two possible explanations for this change: First, ALCF introduced a new training program in October 2021, a popular AI for Science series aimed at undergraduate and graduate students that generated 250 new accounts. Before that, ALCF hosted a GPU Hackathon that generated 75 new accounts. Second, there were more account requests due to the increased number of project allocations. Last January, as ALCF began rolling out its new file systems, Grand and Eagle, internal processes were used to track storage allocations as awards, which resulted in an increase in allocations. Every new project awarded compute time was also given an allocation on Grand or Eagle.

Table 1.5 Ticket Categorization for 2020 and 2021

Category	2020	2021
Access	806 (18%)	808 (16%)
Accounts	1,830 (40%)	2,084 (40%)
Allocations	629 (14%)	856 (16%)
Applications Software	195 (4%)	172 (3%)
Automated E-mail Responses	205 (4%)	205 (4%)
Compilers	61 (1%)	37 (1%)
Data Transfer	45 (1%)	54 (1%)
Debugging and Debuggers	9 (0%)	9 (0%)
File System	162 (4%)	162 (4%)
HPSS [a] and Quota Management	102 (2%)	102 (2%)
Libraries	24 (1%)	36 (1%)
Miscellaneous	148 (3%)	148 (3%)
Network	7 (0%)	9 (0%)
Performance and Performance Tools	15 (0%)	15 (0%)
Reports	181 (4%)	181 (4%)
Scheduling	154 (3%)	154 (3%)
Visualization	5 (0%)	2 (0%)
TOTAL TICKETS	4,578 (100%)	5,034 (100%)

[a] HPSS = high-performance storage system.

1.3.1.2 Community Data Sharing

In early 2021, ALCF deployed Eagle, a 100PB Lustre file system to support data sharing across high-performance computing (HPC) facilities using Globus. Now principal investigators (PIs) and their collaborators without ALCF accounts can both read and store data in a shared space. ALCF's User Experience (UX) team validated the Globus services, added anonymous https read-only controls (a requirement many projects have), produced user-facing documentation, and compiled an internal workflow process document. The team presented a talk, "Sharing data using Eagle," at the 2021 ALCF Computational Performance Workshop in May and was later awarded a 2021 Impact Argonne award in recognition of their efforts.

1.3.1.3 Argonne Collaborator Accounts for Joint Laboratory for System Evaluation (JLSE) Users

ALCF helped ease the transition of JLSE users from the JLSE UserBase 2 (UB2) environment to the CELS (Argonne's Computing, Environment and Life Sciences directorate) Account Management System by proactively loading their identity data. Users were able to login and, with the click of a button, queue up their data for migration into the new system. This effort is an example of a collaborative project involving several teams across several Argonne divisions successfully working with stakeholders.

1.3.1.4 Project Award On-boarding Guides

ALCF created step-by-step, on-boarding guides for the PIs of all major project awards. Links to these online guides will be included in ramp-up emails sent to PIs going forward.

1.3.1.5 User Base Improvement

ALCF streamlined the project management process in UserBase3 (UB3) to reduce the turnaround time needed for users to start using their allocations on ALCF resources. Now, a user's project membership is determined automatically based on project allocation type and the user's primary affiliation. The series of what used to be manual checks for master user agreements (MUAs) with user institutions has now been automated and membership is set to active, and the user can start using the projects allocation immediately without any further administrative processing.

1.3.2 Application Support for Individual Projects

Improving LAMMPS code for an ALCC project

This physics DD project aims to improve the performance of the Monte Carlo/molecular dynamics (MC/MD) method on Theta to accelerate simulations of chemical systems radiated by intense X-ray free-electron laser pulses, such as those provided by the Linac Coherent Light Source-II (LCLS-II). ALCF staff worked with the project PI to port the LAMMPS code to Theta. ALCF staff then refactored the pair calculation and achieved an 80 percent scaling efficiency with 64 OpenMP threads on Theta compute nodes with production input decks. The benchmark data from these code improvements were included in the team's submission that resulted in the 2021–2022 ALCC award.

Modeling polarization of water in quantum accuracy

Based on training data from a highly accurate quantum-mechanical method, DFT-SCAN, the team's neural network (NN) MD simulations coupled with an NN model to compute polarization yields an excellent description of liquid water: the computed dielectric constants are in good agreement with experimental data over a range of temperatures. The physics-based integrated approach can be further improved by systematically increasing the quality of the training data for the NNs. Computational speedups of ~5,000x were observed for small systems of water in this work, and speed increases with system size. ALCF staff and a postdoc worked closely with this project team on an INCITE project and an Aurora Early Science Program (ESP) project to develop a scalable implementation of the RXMD-NN code for the NNQMD calculation. Staff also assisted with the scheduling of large jobs, resulting in 97% of the allocation being used in jobs on >60% of Theta nodes.

Enabling WRF code to run on Theta

This ALCC project aims to better understand the interactions of mesoscale convective system (MCS) processes on a multi-spatiotemporal scale with the focus of improving their representation in state-of-the-art models and to develop novel cloud parameterizations. The code used, the Weather Research and Forecasting (WRF) Model, had previously run error-free on the National Center for Atmospheric Research's (NCAR's) Cheyenne system; however, early simulation job runs on Theta had produced a message passing interface (MPI) hang issue. ALCF staff worked with the science team to track down and resolve the problem and begin production runs. ALCF staff also assisted with debugging a post-processing script on Cooley.

1.3.3 User Engagement

1.3.3.1 General Outreach

User Advisory Council

ALCF's User Advisory Council (UAC) provides guidance on proposed policy changes and services and gives feedback on the experiences and needs of the user community in general. Members are appointed by the ALCF Director and share an expert knowledge of the tasks and requirements of specific applications or domain areas.

Two new members, Sibendu Som and Aiichiro Nakano, were appointed to the 2021–2022 UAC roster, and two others, James King and Sean Dettrick, rotated off. 2020–2021 chair Tom LeCompte agreed to serve another term, which officially commenced June 1, 2021. The first quarterly meeting was held on June 14, 2021, where the council heard updates on transition plans from Theta through Aurora (and post-Aurora). Council members asked questions and discussed a range of topics that included common workflows and the future role for abstractions like those offered in Python and Julia.

Connection to Science and Technology Partnerships and Outreach directorate

ALCF interacts with Argonne's Commercialization and Capture (C&C) group within the Science and Technology Partnerships and Outreach (S&TPO) directorate and regularly supports S&TPO-organized meetings with potential industry partners. S&TPO, in turn, provides potential partners, even those without immediate HPC needs, with a complete picture of Argonne computing capabilities and partnership opportunities, including those at ALCF.

Diversity, Equity, and Inclusion

Through participation in annual Argonne-sponsored outreach events, such as Introduce a Girl to Engineering Day and Science Careers in Search of Women, ALCF staff members connect with young women and introduce them to potential career paths in science, technology, engineering, and math (STEM). ALCF also promotes STEM careers to women through contributions to Argonne's Women in Science and Technology program, AnitaB.org's Top Companies for Women Technologists program, and the Grace Hopper Celebration of Women in Computing. ALCF staff also attend the Richard Tapia Celebration of Diversity in Computing Conference to recruit from a diverse set of backgrounds and ethnicities.

1.3.3.2 Training Activities

ALCF offers workshops and webinars on various tools, systems, and frameworks. These hands-on training programs are designed to help PIs and their project members take advantage of leadership-class computers available at ALCF and enhance the performance and productivity of their research programs. ALCF also collaborates with peer institutions and vendor partners to offer training that strengthens community competencies and promotes best practices. Below is a list of ALCF 2021 training activities:

Computational Performance Workshop (virtual)

This annual workshop helps science teams achieve computational readiness for INCITE and other ALCF allocation programs. Participants work with ALCF and vendor staff to debug/benchmark codes on ALCF resources and expand their data science skills. The following is a breakout of 2021 attendees by institution: ALCF (48), other Argonne (18), industry (14), DOE labs (9), and universities (28). Of the non-Argonne attendees who completed the feedback survey, 36 percent said they plan to apply for an INCITE award. The 2021 workshop registration deadline was extended this year to inform attendees of upcoming INCITE proposal writing webinars. (Dates: May 4–6, 2021)

2022 INCITE Proposal Writing Webinars (virtual)

The INCITE program office, ALCF, and the Oak Ridge Leadership Computing Facility (OLCF) hosted two webinars, in May and in June, covering how to write an effective INCITE proposal. For 2022, up to 60% of the INCITE allocations will go to Summit, Theta, and Polaris. Additionally, the coming exascale systems Aurora (ALCF) and Frontier (OLCF) will be available for out-year requests. (Dates: May 4 and June 4, 2021)

Argonne Training Program on Extreme-Scale Computing (ATPESC) (virtual)

ATPESC, now part of the Exascale Computing Project, is a two-week intensive training program on the key skills and tools needed to use supercomputers for science. The program features talks given by leading computer scientists and HPC experts and hands-on training using DOE leadership-class systems at Argonne, Oak Ridge National Laboratory (Oak Ridge or ORNL), and the National Energy Research Scientific Computing Center (NERSC). In 2021, ATPESC attracted 79 attendees from 58 different institutions worldwide. Video playlists from past ATPESC programs are available on Argonne's YouTube training channel. (Dates: August 1–13, 2021)

Simulation, Data, and Learning (SDL) Workshop (virtual)

This annual interactive workshop is aimed at researchers who are planning to apply for a major allocation award in the near term. Participants learn to scale data-centric science on ALCF systems, set up workflows and use containers, and test and debug codes. The following is a breakout of 2021 SDL Workshop attendees by institution: ALCF (23), other Argonne (23), DOE labs (8), universities (21), and industry/other (10). (Dates: October 5–7, 2021)

2021 ALCF Aurora Workshop (virtual)

This invitation-only workshop focused on helping ESP and Exascale Computing Project (ECP) researchers prepare applications and software technologies for Aurora. The workshop was geared toward developers and emphasized using the Intel software development kit to get applications running on testbed hardware. Teams were also given the opportunity to consult with ALCF staff and provide feedback. Part one of the workshop was held October 26–27 and focused on presentations and status updates on Aurora's hardware and software. Part two was held December 7–9 and provided significant hands-on time with Aurora's testbed technologies. ALCF staff also held dedicated office hours on a range of topics from programming models to profiling tools. Attendance breakdowns for part one and part two are as follows: October 26–27: ESP 43 (21 from Argonne) and ECP 63 (20 from Argonne). December 7–9: ESP 43 (21 from Argonne) and ECP 62 (20 from Argonne).

Aurora Early Science Program (ESP) Dungeons and Hackathons (virtual) [funded by ALCF-3 and ECP]

Throughout 2021, the Intel Center of Excellence (COE), in collaboration with ALCF Early Science Program, held multiday events where select project teams worked on developing, porting, and profiling their codes with help from Intel and Argonne experts. The 2021 program included the following activities:

- Hackathon 11 (Atlas) (February 8, 2021)
- Hackathon 12 (Kokkos) (Teams: LAMMPS, Kokkos Kernel, XGC, HYPRE/UINTAH, VTK-m) (March 5, September 16, and October 14, 2021)
- Hackathon 13 (Phasta) (May 20–21, 2021)
- Dungeon 2 (projects LQCD, XGC, RxMD/QxMD, NAMD) (June 15–17, 2021)
- Hackathon 14 (Connectomics) (September 23, 2021)
- Hackathon 15 (CANDLE) (December 13–14, 2021)
- Hackathon 16 – Day 1 (Fusion Energy) (December 16, 2021)

SC21 Student Cluster Competition (virtual)

ALCF staff participate in the annual International Conference for High Performance Computing, Networking, Storage, and Analysis (SC) student cluster competition as team mentors. In 2021, ALCF was part of the competition committee and took the lead in providing support for three applications from ALCF and Lawrence Livermore National Laboratory (LLNL): Quantum Espresso, Cosmic Tagger, and Cardioid. Ten teams participated in the 2021 competition. (Dates: November 15–17, 2021)

ALCF Webinars (virtual)

The 2021 ALCF webinar program consisted of two tracks: *ALCF Developer Sessions* and *Aurora Early Adopter Series*. The Aurora Early Adopter Series was focused on public discussions related to Aurora. All talks are posted to ALCF's YouTube channel, and the associated training materials can be found on the ALCF Events website. ALCF also participates in useful community events such as the INCITE proposal writing webinars, the IDEAS productivity project webinar series, and Intel webinars. The 2021 webinar program was as follows:

- Running on ThetaGPU with NVIDIA HPC SDK (January 27, 2021)
- High-Performance Data Science with RAPIDS (February 25, 2021)

- SYCL 2020 & DPC++: Improvements to the SYCL Programming Model (March 24, 2021)
- Transitioning from Cobalt to PBS at the ALCF (April 30, 2021)
- NVIDIA Profiling Tools Ecosystem (May 26, 2021)
- Performance, Portability, and Productivity (June 30, 2021)
- Inside the NVIDIA Ampere A100 GPU in ThetaGPU and Perlmutter (July 28, 2021)
- Introduction to Kernel Performance Analysis with NVIDIA Nsight Compute (August 26, 2021)
- Preparing Applications for Aurora: OpenACC to OpenMP Migration Tool (September 29, 2021)
- SYCL Webinar (December 7, 2021)
- Reduce Cross-platform Programming Effort and Write Performant Parallel Code with oneDPL (December 17, 2021)

ALCF GPU Hackathon (virtual)

ALCF and NVIDIA hosted a GPU hackathon on April 20 and April 27–29, 2021. Thirteen small teams of developers worked with mentors to accelerate their codes on ThetaGPU using a portable programming model, such as OpenMP, or an artificial intelligence (AI) framework of their choice.

AI/Machine Learning (M/L) Workshops/Training

ALCF invited researchers from across the laboratory to introduce researchers to the resources available in ALCF's AI Testbed and how they will provide next-generation capabilities for science.

- Cerebras – 37 attendees (ALCF [4], other Argonne [33]) (March 21–23, 2021)
- SambaNova – 45 attendees (ALCF [8], other Argonne [32], industry [2], universities [3]) (June 15–16, 2021)
- Groq – 48 attendees (ALCF [11], other Argonne [34], universities [3]) (July 6–7, 2021)

ALCF AI for Science Training Series (virtual)

Building on ALCF's robust training programs in HPC and AI, ALCF launched a new webinar-based training series in October aimed at undergraduates and graduates enrolled in U.S. universities. The ALCF AI for Science Training Series teaches the fundamentals for how to use supercomputers to develop and apply AI solutions to highly challenging problems. Topics covered in 2021 so far (to continue into 2022) include the following, each attracting around 100 attendees from 22 universities:

- Introduction to AI training on ThetaGPU Supercomputer (October 28, 2021)
- Machine Learning (November 4, 2021)
- Introduction to Deep Learning (November 11, 2021)
- Data Pipelines for Deep Learning (December 2, 2021)

MolSSI Workshop on HPC in Computational Chemistry and Materials Science (virtual)

ALCF and Argonne's Computing, Environment and Life Sciences (CELS) directorate helped organize a 3-day conference sponsored by the Molecular Sciences Software Institute (MolSSI). The event attracted about 50 people and 25 invited speakers from universities, DOE, and industry. (Dates: December 13–15, 2021)

1.3.3.3 Community Outreach

ALCF supports a variety of outreach activities from giving tours to industry groups and DOE leadership to participating in STEM efforts directed at K-12 students. From running several coding camps each summer to participating in annual computer science education events such as the Hour of Code, ALCF staff contribute to a wide range of activities aimed at sparking students' interest in scientific computing and promoting career possibilities in STEM fields. Additionally, the ALCF's annual summer student program gives college students the opportunity to work side-by-side with staff members on real-world research projects and utilize some of the world's most powerful supercomputers, collaborating in areas like computational science, system administration, and data science.

STEM Activities

Summer 2021 Research Internships (virtual)

ALCF hosted 24 student participants of DOE's Science Undergraduate Laboratory Internships (SULI) program and Argonne's Research Aide program to work on mentored research projects in the field of scientific computing. These junior researchers used online collaboration tools to meet and conduct hands-on activities throughout the summer and presented their findings to the ALCF community in a seminar series at the end. The 2021 summer student projects included using Jupyter Notebooks for simulation science; developing infrastructure for data and AI models; experimenting with floating point operations per second (FLOPs)-aware deep learning strategies; and exploring the application of reinforcement learning techniques to fluid dynamics problems for specific controls and designs.

2021 Illinois Regional Middle School Science Bowl (virtual)

Argonne hosted the 2021 Illinois Regional Science Bowl, held virtually for the first time, which allowed more students to compete than ever. Ten middle school teams from seven schools competed in answering science-related trivia questions. The winner of the Illinois regional science bowl advanced to the National Science Bowl Competition in Washington, D.C., on May 9, 2021. (Date: January 30, 2021)

2021 Introduce a Girl to Engineering Day (virtual)

Argonne hosted 78 eighth-grade girls for its 19th annual Introduce a Girl to Engineering Day (IGED) on February 18, 2021. Twenty-three Argonne mentors, including ALCF staff members, volunteered for an engaging day of presentations and hands-on activities focused on STEM careers. (Date: February 18, 2021)

2021 Science Careers in Search of Women (virtual)

Science Careers in Search of Women (SCSW), a long-standing lab-sponsored event held in April, returned in a virtual format in 2021 after being canceled in 2020 due to the pandemic. The conference provides female students from area high school an opportunity to explore their desired STEM profession or area of interest through various interactions with Argonne's women scientists and engineers. Teachers and counselors who accompany the students have a separate program. ALCF staff members annually volunteer for various SCSW roles. (Date: April 15, 2021)

2021 CodeGirls @ Argonne Camp (virtual)

Argonne held its fifth annual CodeGirls @ Argonne summer camp, a weeklong STEM course for 6th- and 7th-grade girls taught by Argonne Learning Center staff and ALCF research staff. The campers arrive with little or no experience in programming and, over the course of a week, learn Python coding fundamentals, experiment with robotics, and meet women scientists who use code to solve problems. The group also virtually toured the ALCF machine room and learned about the future Aurora supercomputer. (Dates: June 28–July 2, 2021)

Argonne-NIU AI Camp (virtual)

Argonne and Northern Illinois University (NIU) hosted its first AI-focused summer camp, *Science and Inquiry: Exploring Artificial Intelligence*, taught by learning science experts from both institutions and ALCF computer scientists and attended by regional high schoolers in NIU's precollegiate program Upward Bound. The four-week camp provided a foundational introduction to AI and machine learning, and several hands-on activities explored real datasets using real analysis tools. (Date: July 2021)

2021 Coding for Science Camp (virtual)

Coding for Science Camp is a five-day enrichment experience for high school freshmen and sophomores who are new to coding. The camp curriculum promotes problem-solving and teamwork skills through hands-on coding activities, such as coding with Python and programming a robot, and interactions with Argonne staff members working in HPC and visualization. The camp is a joint initiative of Argonne's Educational Programs Office and ALCF. (Dates: July 19–23, 2021)

2021 Big Data Camp (virtual)

Argonne's Educational Programs Office and ALCF partnered to host the fourth annual Big Data Camp for high school juniors and seniors to discover professional uses for their STEM and data skills. Camp participants come with prior coding experience and learn techniques for probing and analyzing massive scientific datasets, including data visualization methods. Campers accessed data from the Array of Things (AoT) urban sensor project at Argonne and learned about data science first-hand from ALCF staff. (July 26–30, 2021)

Hour of Code (virtual)

During Computer Science Education Week (CSEdWeek) in December each year, Argonne computer scientists visit Chicagoland schools and assist teachers in celebrating the Worldwide Hour of Code initiative by giving short presentations and interacting with students one-on-one for related coding activities. In 2020, Argonne converted its K-12 Hour of Code activities into a virtual format. (Dates: December 6–12, 2021)

Facility Tours

Visitors to Argonne can request a tour of the ALCF. ALCF visitors include student groups, Congressional representatives and other government officials, industry representatives, summer research students, visiting researchers, and journalists. Tours are guided by various staff members and include the machine room, data center, and visualization lab.

ALCF has suspended all in-person tours as of March 17, 2020, due to the pandemic. Staff have transitioned to virtual tours for participants of CodeGirls @ Argonne, Big Data camp, and ATPESC 2021. On August 10, 2021, members of ALCF's leadership welcomed White House National Climate Advisor Gina McCarthy, U.S. Representative Sean Casten, and U.S. Representative Marie Newman for a tour; and on September 17, 2021, welcomed DOE Chief Commercialization Officer and Director of the Office of Technology Transitions Vanessa Chan and Senior Advisor to the Director of the Office of Technology Transitions James Fritz for a tour.

1.3.4 Communications

1.3.4.1 Website Support Center Continuous Improvement

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

ALCF uses Google Analytics to collect data on how users interact with ALCF's website and have tracked a steady growth in page views across user resources and documentation since launching the redesigned website in December 2019. In 2021, the Support Center's landing page was viewed 90,243 times, an increase of 17,372 page views over CY 2020, and the training and events webpage had 21,401 page views, an increase of 4,508 page views.

1.3.4.2 Consistent Cadence of ALCF Impact on Exascale and AI Efforts

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

In 2021, ALCF collaborated with ECP and Intel on several Aurora Early Science Program articles, "Let's Talk Exascale Code Development" podcasts, and Twitter Live events; and with Groq and Cerebras to jointly promote AI research results on machines in ALCF's AI Testbed, resulting in extra media exposure for ALCF overall.

1.3.4.3 Communicating Scientific Impact

ALCF produces science stories and articles and promotes HPC training opportunities year-round. Furthermore, ALCF plans marketing campaigns around major annual HPC conferences and events such as the ECP Annual Meeting, ISC, Exascale Day, and SC.

In 2021, ALCF placed 72 original science stories in various news outlets in coordination with Argonne's Communications & Public Affairs (CPA) Division and other ALCF direct relationships and tracked its media hits through its media monitoring service, Meltwater. In 2021, Meltwater reported 950 unique ALCF media hits, 238 of which were chronicled on the ALCF website, and an audience reach of 810.66 million (an increase of 383.96 million). Note: Meltwater defines "reach" as estimating the potential viewership of any article based on the number of visitors to the specific source on both desktop and mobile devices.

ALCF also produces various high-quality 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.

Table 1.6 Publications Designed for Print

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

1.3.4.4 Messaging for Users & Community

The facility maintains several communication channels, including direct e-mail campaigns, scriptable e-mail messages, social media postings (Facebook, Twitter, and LinkedIn), and website postings (Table 1.7; target audiences are identified in Table 1.8). Users can opt out of the system notify and newsletter mailing lists.

ALCF's monthly e-newsletter, Newsbytes, highlights ALCF-supported research or advancements, promotes training events and allocation program announcements, and links to relevant news stories. Special announcements about certain training opportunities and fellowships are sent throughout the year, as needed.

Table 1.7 Primary Communication Channels

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

Table 1.8 Target Audiences

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

Conclusion

Maximizing the use of our resources is at the forefront of all improvements to ALCF's customer service. In 2021, ALCF continued to support project teams in adapting their codes to new architectures and prepare others to apply for major project allocations. We continued to partner with other national laboratories and the ECP and present our work in premier scientific journals and at top professional meetings and conferences. ALCF users gained access to two new Lustre file systems, Grand and Eagle, and one community filesharing service. We launched a new training series (ALCF AI for Science Training) aimed at undergraduate and graduate students. There were multiple improvements made to the account and project management software which helped our users and administrative staff.

Section 2. Operational Performance

Did the facility's operational performance meet established targets?

2.1 ALCF Response

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

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

Table 2.1 Summary of Operational Performance of HPC Computational Systems

Theta with expansion Theta (Cray XC40): 4008-node, 251K-core, 64 TB of MCDRAM, 770 TB of DDR4 Theta expansion (NVIDIA DGX): 24-node, 26 TB of DDR4 RAM, 8.32 TB of GPU memory				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
Scheduled Availability	90.0%	98.3%	90.0%	99.4%
Overall Availability	90.0%	93.8%	90.0%	95.1%
Theta: System MTTI	N/A	11.45 days	N/A	12.07 days
Theta: System MTTF	N/A	32.74 days	N/A	45.51 days
Theta expansion: System MTTI	N/A	N/A	N/A	13.46 days
Theta expansion: System MTTF	N/A	N/A	N/A	121.54 days
INCITE Usage	17.8M	20.9M	17.8M	20.8M
Total Usage	N/A	33.1M	N/A	33.5M
System Utilization	N/A	97.7%	N/A	98.1%
INCITE Overall Capability [a]	20.0%	84.8%	20.0%	84.8%
INCITE High Capability [b]	N/A	24.1%	N/A	22.7%

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

[b] Theta with expansion High Capability = Jobs using ≥ 60.0 percent (2,400 nodes) of Theta.

Table 2.2 Summary of Operational Performance of HPC Storage Systems

Lustre File System Cray Sonexion 2000 with 9.2 PB of usable storage				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
Scheduled Availability	N/A	98.8%	90.0%	100.0%
Overall Availability	N/A	94.8%	90.0%	96.1%
System MTTI	N/A	18.26 days	N/A	15.25 days
System MTTF	N/A	188.95 days	N/A	182.47 days
Grand File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
Scheduled Availability	N/A	N/A	N/A	100.0%
Overall Availability	N/A	N/A	N/A	96.5%
System MTTI	N/A	N/A	N/A	15.31 days
System MTTF	N/A	N/A	N/A	182.45 days
Eagle File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
Scheduled Availability	N/A	N/A	N/A	100.0%
Overall Availability	N/A	N/A	N/A	96.5%
System MTTI	N/A	N/A	N/A	16.01 days
System MTTF	N/A	N/A	N/A	365.00 days
HPSS Archive LTO8 tape drives and tape with 350 PB of storage capacity				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
Scheduled Availability	N/A	98.8%	95.0%	100.0%
Overall Availability	N/A	94.2%	90.0%	96.5%
System MTTI	N/A	19.51 days	N/A	16.01 days
System MTTF	N/A	366.00 days	N/A	365.00 days

2.2 ALCF Production Resources Overview

During CY 2021, the ALCF operated several production resources: Theta with expansion, Sonexion Lustre file system, and HPSS.

- Theta is a 4,392-node, ~281K-core, 11.69-PF Cray XC40 with 891 TB of RAM. Theta supported both ALCC and INCITE campaigns all year.
- ThetaGPU is a 24-node expansion to Theta consisting of NVIDIA DGX A100-based system. The DGX A100 comprises eight NVIDIA A100 GPUs with AMD EPYC 7742 CPUs for a total of 26 TB of DDR4 RAM and 8.32 TB of GPU memory at a speed of 3.9PF.
- A Sonexion 2000 Lustre file system is mounted by Theta and the data and analysis resource with 9.2 PB of usable space.
- Grand and Eagle Lustre file systems provide a global file system and a community file system, with 200 PB of storage.
- The facility-wide HPSS (high-performance storage system) tape archive is comprised of three 10,000-slot libraries with LTO8 drives and tapes, with some legacy LTO6 drives and tapes. Currently, the tape libraries have a maximum storage capacity of 305 PB.

(For more information on performance metrics calculations, see Appendix A.7.)

2.3 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 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 that 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 window is categorized as an unscheduled outage. A significant event that delays a return to scheduled production will be counted as an adjacent unscheduled outage if the return to service is four or more hours later than the scheduled end time. For storage resources, the system is online and available if any user can read and write any portion of the disk space. The availability metric provides measures that are indicative of the stability of the systems and the quality of the maintenance procedures.
- *MTTI, mean time to interrupt,* is the time, on average, to any outage on the system, whether unscheduled or scheduled. Also known as MTBI (Mean Time Between Interrupts).
- *MTTF, mean time to failure,* is the time, on average, to an unscheduled outage on 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 kept when it was available).
- *Total System Utilization* is the percent of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user group, including staff and vendors.

2.4 Theta

2.4.1 Scheduled and Overall Availability

Theta entered full production on July 1, 2017. The GPU expansion to Theta entered production on January 1, 2021. In consultation with the DOE Program Manager, ALCF agreed to a target of 90 percent overall availability and a target of 90 percent scheduled availability. (ASCR requested that all user facilities use a target of 90 percent for scheduled availability for the lifetime of a production resource). Table 2.3 summarizes the availability results for Theta (with expansion in CY 2021).

Table 2.3 Availability Results

Theta with expansion Theta (Cray XC40): 4008-node, 251K-core, 64 TB of MCDRAM, 770 TB of DDR4 Theta expansion (NVIDIA DGX): 24-node, 26 TB of DDR4 RAM, 8.32 TB of GPU memory				
	CY 2020		CY 2021	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	90.0	98.3	90.0	99.4
Overall Availability	90.0	93.8	90.0	95.1

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

2.4.1.1 Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability for the period January 1, 2021, through December 31, 2021, as shown in Figure 2.1.

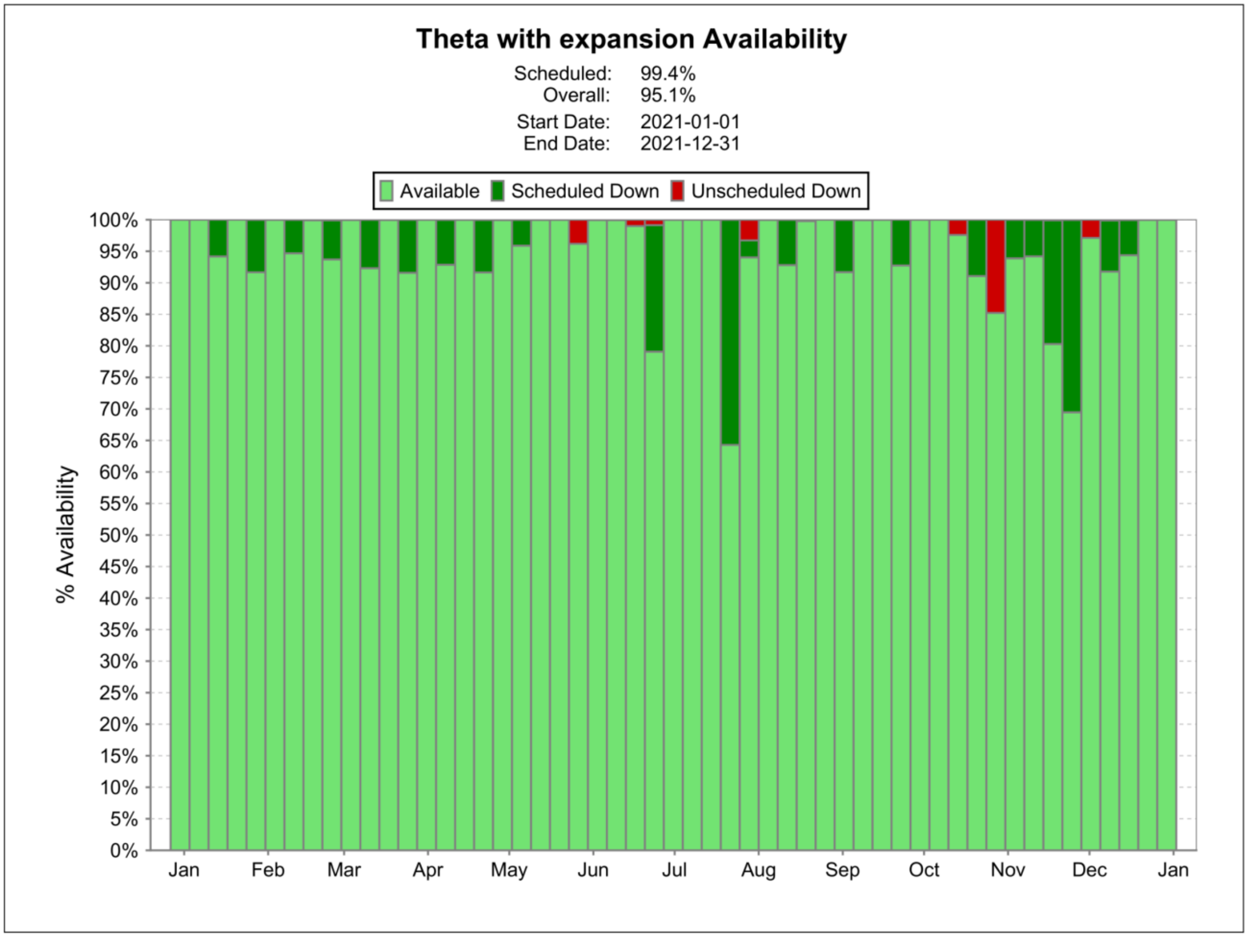


Figure 2.1 Theta with expansion Weekly Availability for CY 2021

Graph Description: Each bar in Figure 2.1 represents the percentage of the machine available for seven days. Each bar accounts for all of 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 red represents unscheduled downtime. Significant loss events are described in detail below.

May 26, 2021: Unscheduled outage – cooling failure

There was a water line break that affected both chilled water plants (CWPs) on campus, impacting water-cooled systems in the Theory and Computing Sciences (TCS) datacenter. At 9:00 a.m., ALCF staff shut down compute resources to protect hardware from rising temperatures and to head off automated protections, which are more abrupt than human intervention. By the time shutdown was complete, water temperatures had risen from the normal 17–19ºC to 29–30ºC. At noon, the all-clear was given for resumption of services, and systems were returned to normal operation.

June 16, 2021: Unscheduled outage – Theta blower failure

During a routine replacement of a blower in Theta, a breaker tripped, causing the ungraceful shutdown of multiple racks of Theta compute nodes. The compute environment was shut down, and power was safely restored and then brought back online. The blower was successfully replaced. Several compute nodes had to be replaced because of damage caused by the power fluctuation.

June 23–24, 2021: Scheduled extended maintenance – Upgrade Theta Lustre to NEO 3.5
This multiday scheduled maintenance was for a vital upgrade of the Theta Lustre file system's control software to NEO 3.5. This upgrade had been deferred to allow for sufficient testing. During this window, other opportunistic maintenance activities were performed.

July 18–20, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for the power needs of Aurora, TCS building management shut down power to perform maintenance and upgrades on power feeds coming into the building. This work was carried out over the weekend, leading to ALCF performing a controlled shutdown on Friday and beginning startup procedures on the following Monday at 8:00 a.m.

July 28, 2021: Unscheduled outage – Cabinet C11 down
Theta compute cabinet C11 experienced a power failure, resulting in an interruption of all jobs running on the system at the time. The interruption lasted several hours, and Theta was returned to service that evening. 32 of 36 power rectifiers were replaced during the repairs to cabinet C11.

October 10, 2021: Unscheduled outage – Cabinet C3 down
Hewlett Packard Engineering (HPE) staff reported detection of failed power rectifiers in cabinet C3. Theta remained online, although in a degraded state. Cabinet C3 was returned to service on October 19, after receiving replacement parts and release testing. This outage did not affect the availability of the machine and is not shown in Figure 2.1 above as the cabinet was one of the two extras in Theta used for redundancy.

October 26, 2021: Unscheduled outage – Cabinet C4 down
Theta cabinet C4 suffered a loss of 29 of its 36 power rectifiers, causing all nodes in the cabinet to fail. This cabinet unfortunately contained the node that hosts the scheduler resource manager, Cobalt. Loss of Cobalt caused operations on Theta to halt, interrupting all running jobs.

November 19–22, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for the power needs of Aurora, TCS building management shut down power to perform maintenance and upgrades on power feeds coming into the building. This work was carried out over the weekend, leading to ALCF performing a controlled shutdown on Friday and beginning startup procedures on the following Monday at 8:00 a.m.

December 1, 2021: Unscheduled outage – Grand file system self-check
The Grand file system's RAID array commenced its monthly automated self-checks of the storage system. This event coincided with many Aries network link errors and triggered errors in Lustre's LNET router daemons, which subsequently caused instability in Theta. The eventual root-cause analysis by HPE determined that on Aries networks, the LNET processes on certain nodes become busy processing errors and show reduced capacity to handle already established connections. This instability required the system to be rebooted.

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

2.4.2.1 ALCF MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.4 summarizes the current MTTI and MTTF values for Theta and Theta expansion, respectively.

Table 2.4 MTTI and MTTF Results

Theta (Cray XC40): 4008-node, 251K-core, 64 TB of MCDRAM, 770 TB of DDR4				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
System MTTI	N/A	11.45 days	N/A	12.07 days
System MTTF	N/A	32.74 days	N/A	45.51 days

Theta expansion (NVIDIA DGX): 24-node, 26 TB of DDR4 RAM, 8.32 TB of GPU memory				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
System MTTI	N/A	N/A	N/A	13.46 days
System MTTF	N/A	N/A	N/A	121.54 days

Theta currently has a biweekly maintenance schedule to perform Cray driver upgrades, hardware replacements, operating system (OS) upgrades, etc. ALCF uses these preventative maintenance (PM) opportunities to schedule other potentially disruptive maintenance, such as facility power and cooling work and storage system upgrades and patching. ALCF canceled or postponed several PMs this year in favor of preparatory work for the installations of Polaris and Aurora. Although Theta's biweekly maintenance schedule does not directly affect MTTF, it generally tends to cap MTTI at 14 days.

2.4.3 Resource Utilization

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

2.4.3.1 System Utilization

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

Table 2.5 System Utilization Results

Theta with expansion Theta (Cray XC40): 4008-node, 251K-core, 64 TB of MCDRAM, 770 TB of DDR4 Theta expansion (NVIDIA DGX): 24-node, 26 TB of DDR4 RAM, 8.32 TB of GPU memory				
	CY 2020		CY 2021	
	Target	**Actual**	**Target**	**Actual**
System Utilization	N/A	97.7%	N/A	98.1%

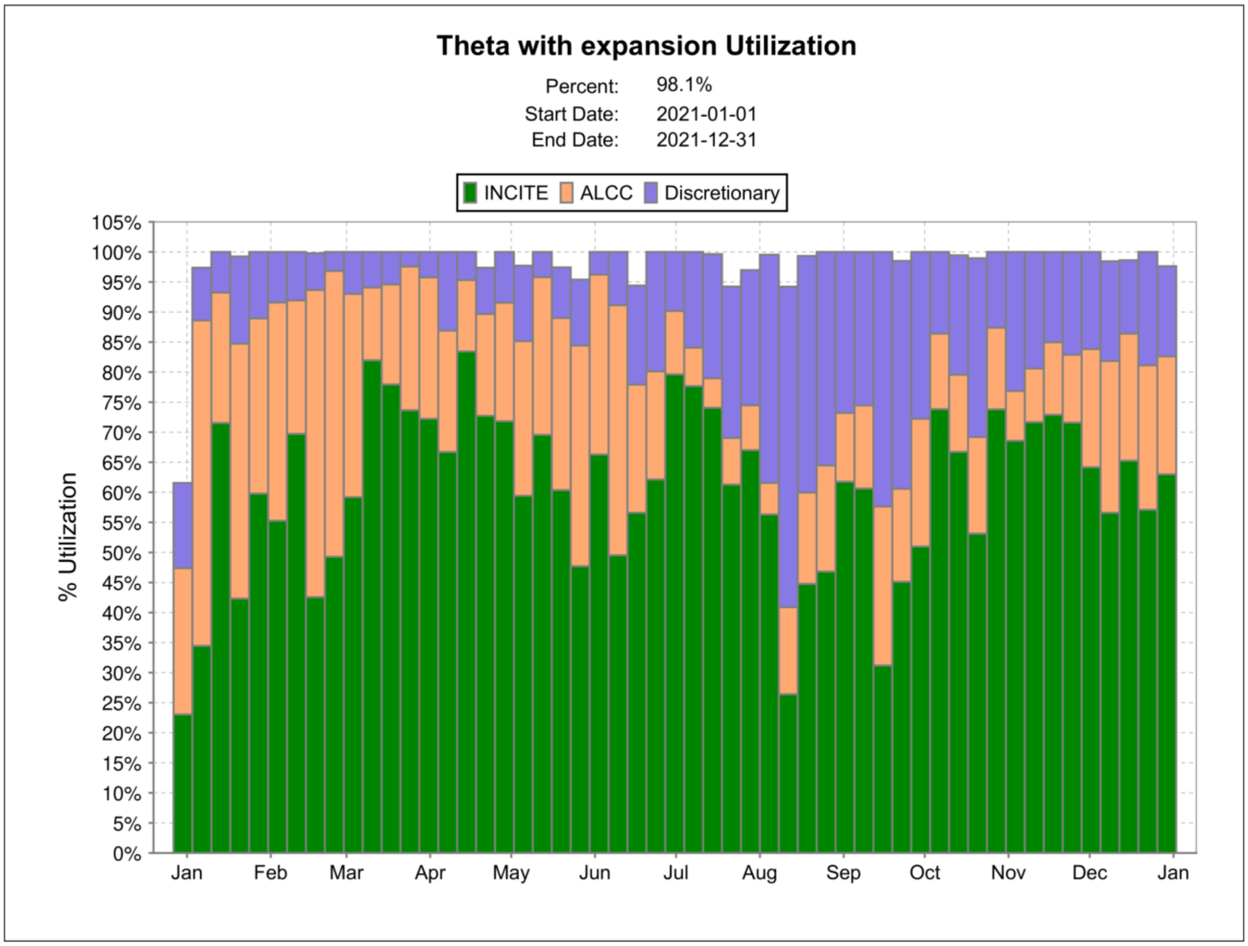


Figure 2.2 System Utilization over Time by Program

The system utilization for Theta was 98.1 percent for its 2021 production period of January 1, 2021, through December 31, 2021.

Table 2.6 shows how Theta's system hours were allocated and used by allocation type. Multiplying the theoretical hours by availability and system utilization values that were agreed upon with ALCF's DOE Program Manager determines the hours available. Of the hours available, 60 percent were allocated to the INCITE program, up to 20 percent to ALCC program allocations, and 20 percent to DD allocations. The allocated values for the DD allocations appear higher than expected because they represent a rolling allocation. Most DD projects are exploratory investigations, so the time allocated is often not used in full. DD allocations are discussed in detail in Section 3.2. In CY 2021, Theta delivered a total of 33.5 million node-hours.

Table 2.6 Node-Hours Allocated and Used by Program

Theta with expansion Theta (Cray XC40): 4008-node, 251K-core, 64 TB of MCDRAM, 770 TB of DDR4 Theta expansion (NVIDIA DGX): 24-node, 26 TB of DDR4 RAM, 8.32 TB of GPU memory						
	CY 2020			CY 2021		
	Allocated	Used		Allocated	Used	
	Node-hours	Node-hours	%	Node-hours	Node-hours	%
INCITE	17.8M	20.9M	63.2%	17.8M	20.8M	62.1%
ALCC	6.7M	7.4M	22.4%	7.3M	7.2M	21.6%
DD	9.1M	4.7M	14.4%	7.8M	5.5M	16.3%
Total	33.6M	33.1M	100.0%	32.9M	33.5M	100.0%

Summary: For CY 2021, the system usage and system utilization values were in line with general expectations. The calculations for system utilization are described in Appendix A.

2.4.3.2 Capability Utilization

On Theta, capability is defined as using greater than 20 percent of the machine, or 800 nodes, and high capability is defined as using greater than 60 percent of the machine, or 2,400 nodes. See Table A.1 in Appendix A for more detail on the capability calculation. Table 2.7 and Figure 2.3 show that ALCF has substantially exceeded these targets set for INCITE. Although no targets are set, data are also provided in the table for ALCC and DD projects as reference. 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

Theta with expansion Theta (Cray XC40): 4008-node, 251K-core, 64 TB of MCDRAM, 770 TB of DDR4 Theta expansion (NVIDIA DGX): 24-node, 26 TB of DDR4 RAM, 8.32 TB of GPU memory						
	CY 2020			CY 2021		
Capability Usage	**Total Hours**	**Capability Hours**	**Percent Capability**	**Total Hours**	**Capability Hours**	**Percent Capability**
INCITE Overall [a]	20.9M	17.7M	84.8%	20.8M	17.6M	84.8%
INCITE High [b]	20.9M	5.0M	24.1%	20.8M	4.7M	22.7%
ALCC Overall	7.4M	4.7M	63.8%	7.2M	3.0M	41.1%
ALCC High	7.4M	0.2M	3.3%	7.2M	0.6M	7.7%
Director's Discretionary Overall	4.7M	1.7M	35.2%	5.5M	1.1M	20.9%
Director's Discretionary High	4.7M	0.8M	16.7%	5.5M	0.2M	4.2%
TOTAL Overall	33.1M	24.1M	72.9%	33.5M	21.8M	65.0%
TOTAL High	33.1M	6.1M	18.4%	33.5M	5.5M	16.4%

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

[b] Theta with expansion High Capability = Jobs using ≥ 60.0 percent (2,400 nodes) of Theta.

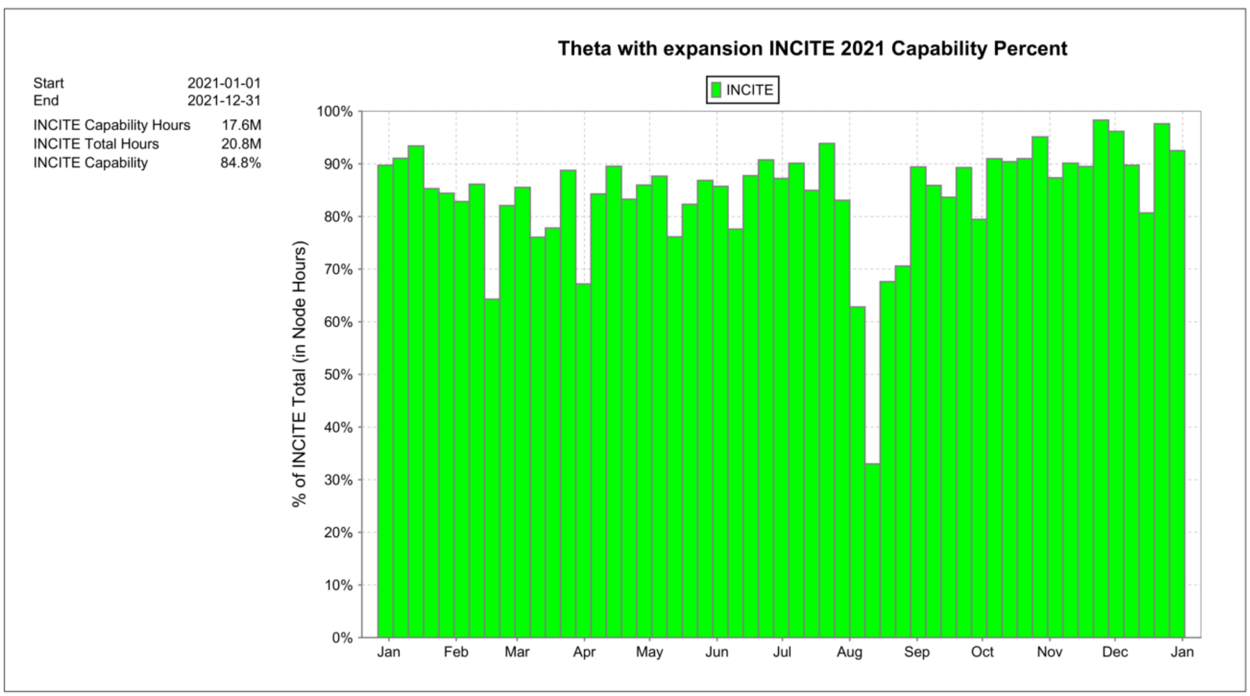


Figure 2.3 Theta INCITE Overall Capability

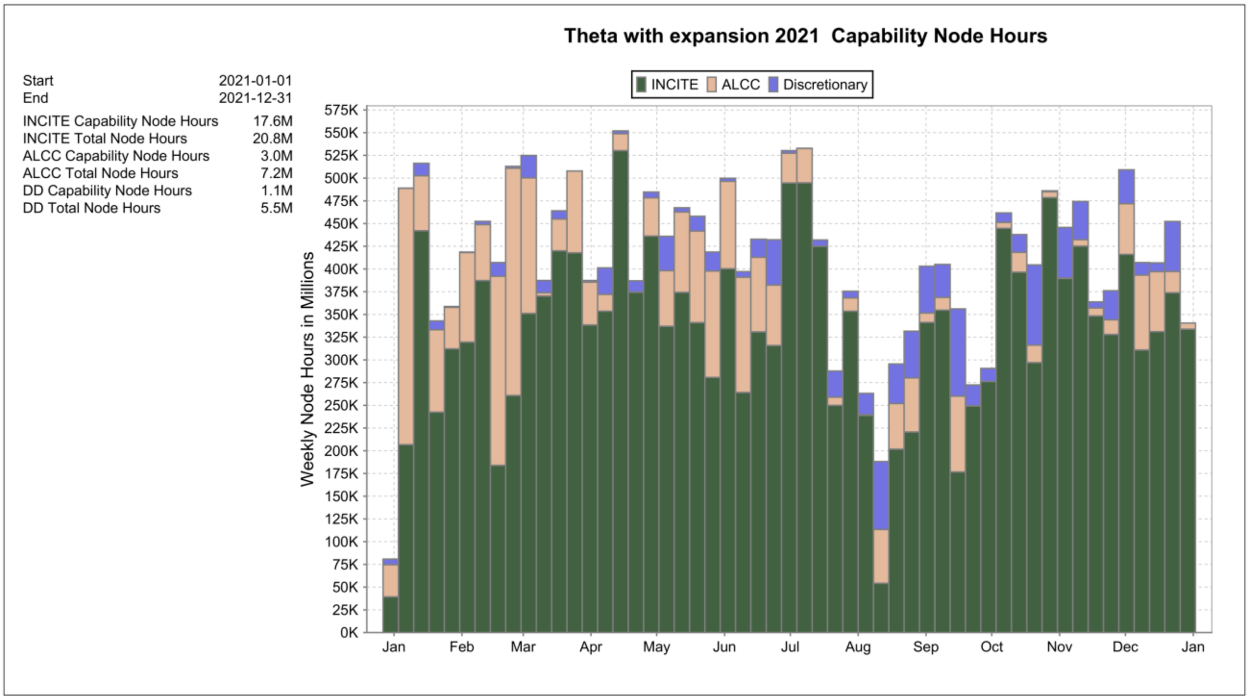


Figure 2.4 Theta Capability Node-Hours by Program

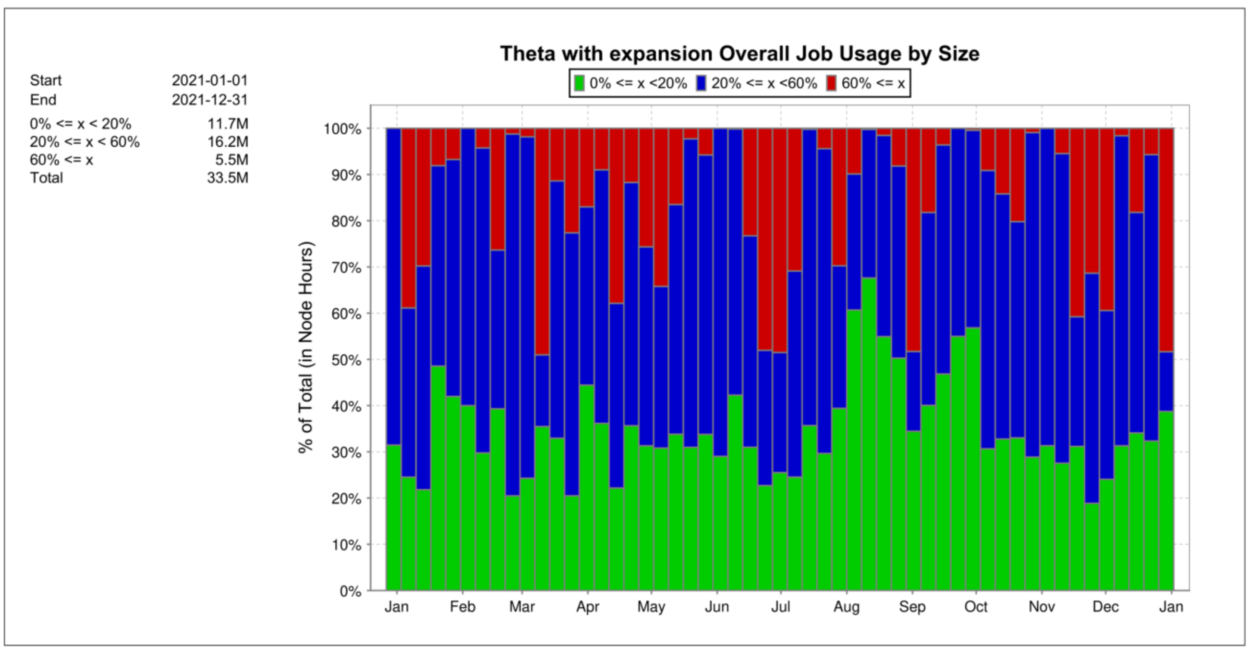


Figure 2.5 Theta Job Usage by Size

2.5 Storage

This section covers availability and MTTI/F metrics for the production storage resources. This is the first year that the global filesystem and second year that the storage resources are being covered separately.

2.5.1 Theta Lustre File System

Theta Lustre, also referred to as "theta-fs0," is a Cray Sonexion 2000 Lustre file system that is mounted by Theta and ThetaGPU, and also by the data and analysis resource with 9.2 PB of usable space. Theta and Theta Lustre were installed together, and both entered full production on July 1, 2017. Theta Lustre and Theta were not previously treated as separate entities.

2.5.1.1 Scheduled and Overall Availability

ALCF uses the target metrics of 90 percent overall availability and 90 percent scheduled availability as proposed in the 2020 OAR. Theta Lustre is tightly tied to Theta, and this follows ASCR's request that all user facilities use a target of 90 percent for scheduled availability for the lifetime of a production resource. Table 2.8 summarizes the availability results.

Table 2.8 Availability Results

Lustre File System Cray Sonexion 2000 with 9.2 PB of usable storage				
	CY 2020		CY 2021	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A	98.8	90.0	100.0
Overall Availability	N/A	94.8	90.0	96.1

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

Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability of the Sonexion for the period of January 1, 2021, through December 31, 2021, as annotated in Figure 2.6.

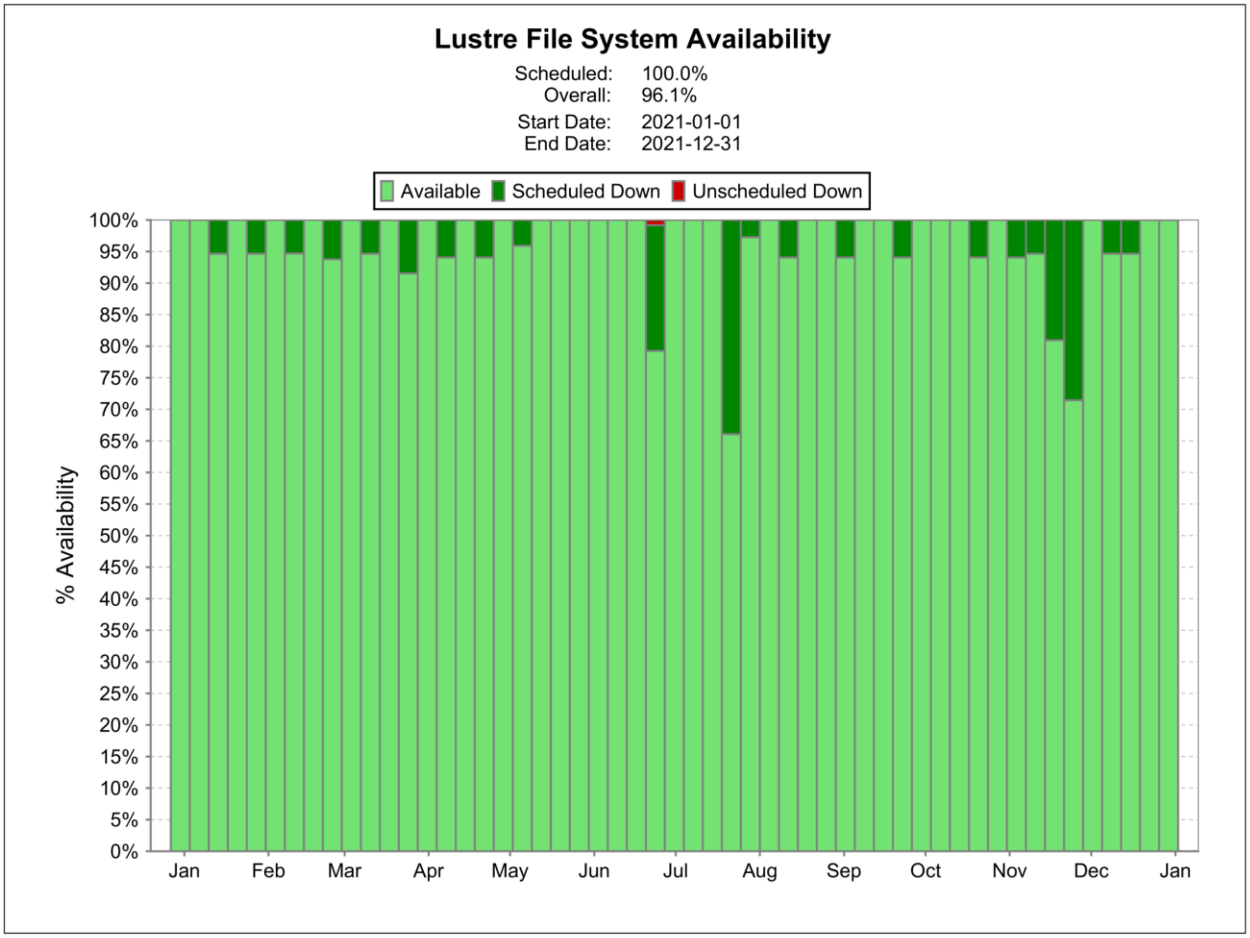


Figure 2.6 Theta Lustre Weekly Availability for CY 2021

Graph Description: Each bar in Figure 2.6 represents the percentage of the machine's availability for seven days. Each bar accounts for all of 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 red represents unscheduled downtime. Each of the significant loss events is described in detail below. (These also appeared in the Theta section).

June 23–24, 2021: Scheduled extended maintenance – Upgrade Theta Lustre to NEO 3.5
This multiday scheduled maintenance was for a vital upgrade of Theta Lustre's control software to NEO 3.5. This upgrade had been deferred to allow for sufficient testing. Other opportunistic maintenance activities were also performed during this window.

July 18–20, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

November 19–22, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

2.5.1.2 MTTI and MTTF

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

Lustre File System Cray Sonexion 2000 with 9.2 PB of usable storage				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
System MTTI	N/A	18.26 days	N/A	15.25 days
System MTTF	N/A	188.95 days	N/A	182.47 days

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

2.5.2 Grand and Eagle Lustre File Systems

The ALCF installed a new set of Lustre file systems in 2020 running a Cray E1000 storage solution. Grand and Eagle each provide 100 PB of storage at 650 GB/s and provide availability protection if one fails. Additionally, the file systems have the capability of sharing data via Globus, a move toward providing a community file system. The file systems went into production starting on January 1, 2021.

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

2.5.2.1 Grand Scheduled and Overall Availability

Table 2.10 summarizes the availability results for the Grand file system.

Table 2.10 Availability Results – Grand

Grand File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2020		CY 2021	CY 2021
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A	N/A	N/A	100.0
Overall Availability	N/A	N/A	N/A	96.5

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

This section briefly describes the causes of major losses of availability of Grand for the period of January 1, 2021, through December 31, 2021, as annotated in Figure 2.7.

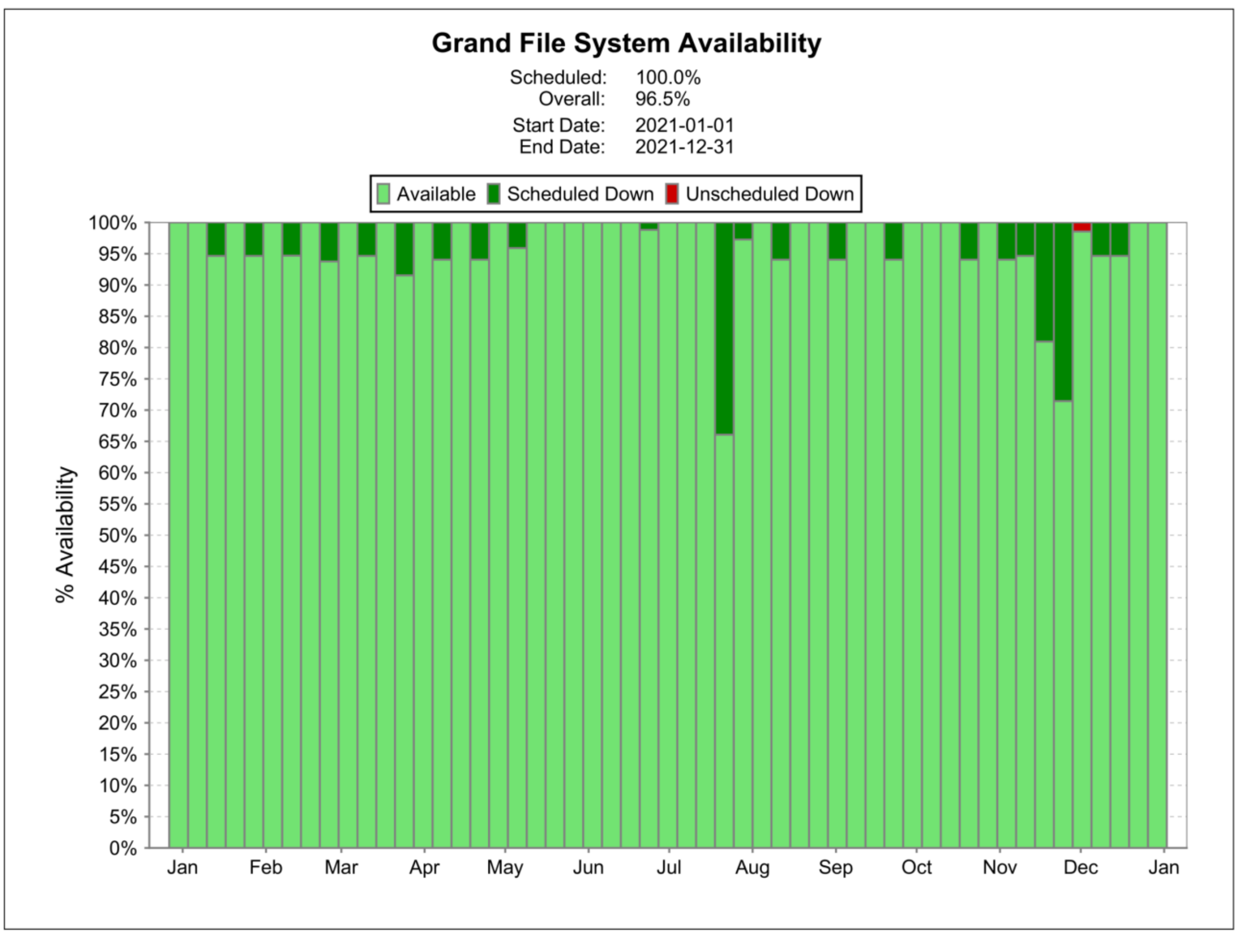


Figure 2.7 Grand Weekly Availability for CY 2021

Graph Description: Each bar in Figure 2.7 represents the percentage of the machine's availability for seven days. Each bar accounts for time in one of three categories. The pale-green portion represents available node-hours; the darker green represents scheduled downtime for that week; and red represents unscheduled downtime. Each of the significant loss events is described in detail below (these also appeared in the Theta section).

July 18–20, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

November 19–22, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

Grand – 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 - Grand

Grand File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
System MTTI	N/A	N/A	N/A	15.31 days
System MTTF	N/A	N/A	N/A	182.45 days

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

2.5.2.2 Eagle – Scheduled and Overall Availability

Table 2.12 summarizes the availability results for the Eagle file system.

Table 2.12 Availability Results – Eagle

Eagle File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2020		CY 2021	CY 2021
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A	N/A	N/A	100.0
Overall Availability	N/A	N/A	N/A	96.5

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.

Eagle – Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability of Eagle for the period of January 1, 2021, through December 31, 2021, as annotated in Figure 2.8.

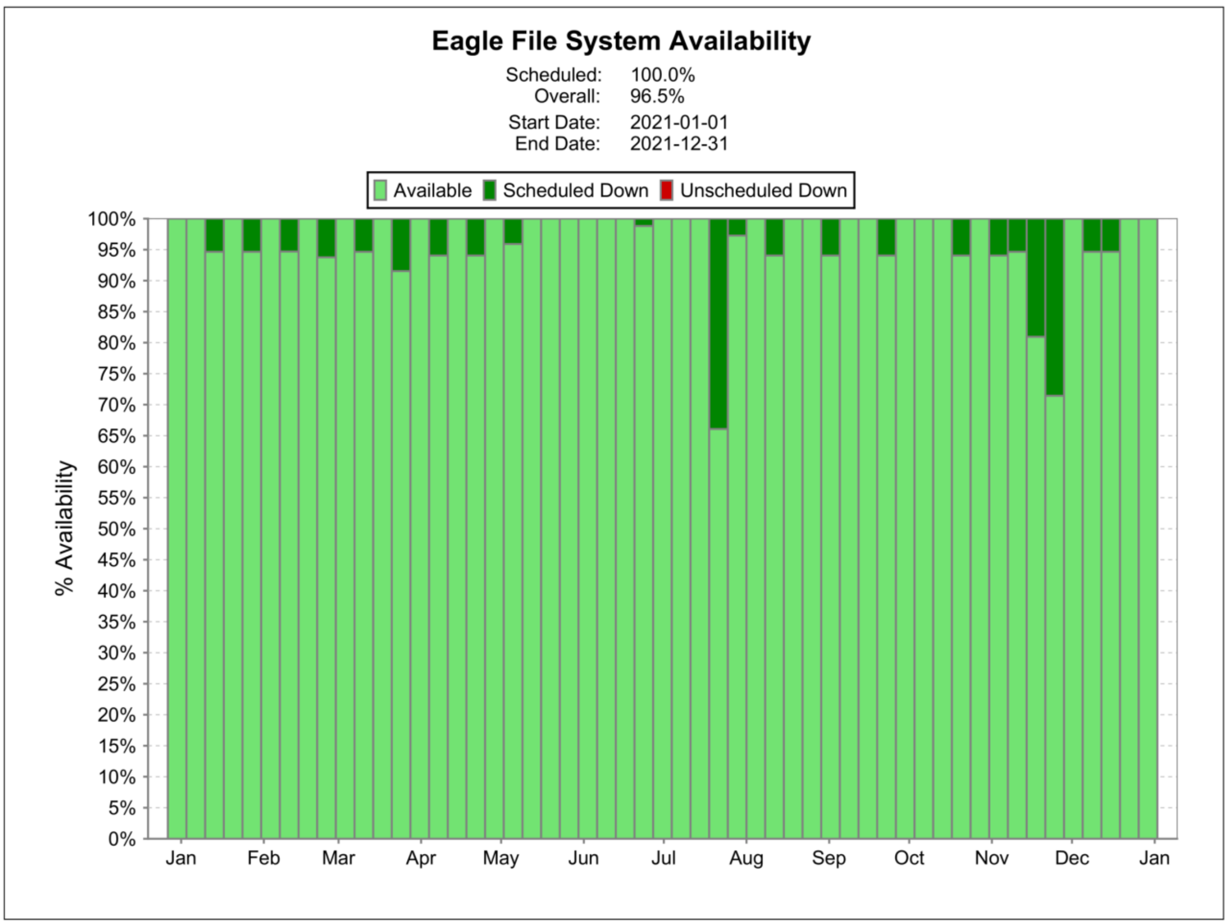


Figure 2.8 Eagle Weekly Availability for CY 2021

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 node-hours; the darker green represents scheduled downtime for that week; and red represents unscheduled downtime. Each of the significant loss events is described in detail below (these also appeared in the Theta section).

July 18–20, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

November 19–22, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

2.5.2.3 MTTI and MTTF

Eagle – 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 – Eagle

Eagle File System Cray E1000 with 100 PB of storage at 650 GB/s				
	CY 2020		CY 2021	
	Target	Actual	Target	Actual
System MTTI	N/A	N/A	N/A	16.01 days
System MTTF	N/A	N/A	N/A	365.00 days

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

2.5.3 Tape Storage

The facility-wide HPSS (high-performance storage system) tape archive was available to all ALCF users from all compute resources in 2021, 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 Interim Scientific Supercomputing Facility (ISSF) datacenter, and the second followed in 2010 in the TCS datacenter, while the third library went into production in 2016. In 2019, all of the tape libraries were moved to another building to provide separation of the archive data from the data center while also permanently vacating the ISSF datacenter. The HPSS disk cache and data movers are in the TCS datacenter. With the LTO8 drives and tape technology, the tape libraries have a maximum storage capacity of 305 PB.

2.5.3.1 Scheduled and Overall Availability

ALCF uses the target metrics of 90 percent overall availability and 95 percent scheduled availability as proposed in the 2020 OAR. Table 2.14 summarizes the availability results.

Table 2.14 Availability Results

HPSS Archive LTO8 tape drives and tape with 350 PB of storage capacity				
	CY 2020		CY 2021	
	Target (%)	Actual (%)	Target (%)	Actual (%)
Scheduled Availability	N/A	98.8	95.0	100.0
Overall Availability	N/A	94.2	90.0	96.5

Note that HPSS is considered unavailable when users cannot retrieve or access files via logins or data transfer nodes even though the HPSS libraries were unaffected during the scheduled maintenance periods,[1] and still could do system functions like data migration. Therefore, HPSS overall availability will reflect that users could not access it during scheduled maintenance.

Explanation of Significant Availability Losses

This section briefly describes the causes of major losses of availability of HPSS for the period of January 1, 2021, through December 31, 2021, as annotated in Figure 2.9.

[1] SpectraLogic quarterly maintenance occurs during the maintenance windows where one library is offline for a short time.

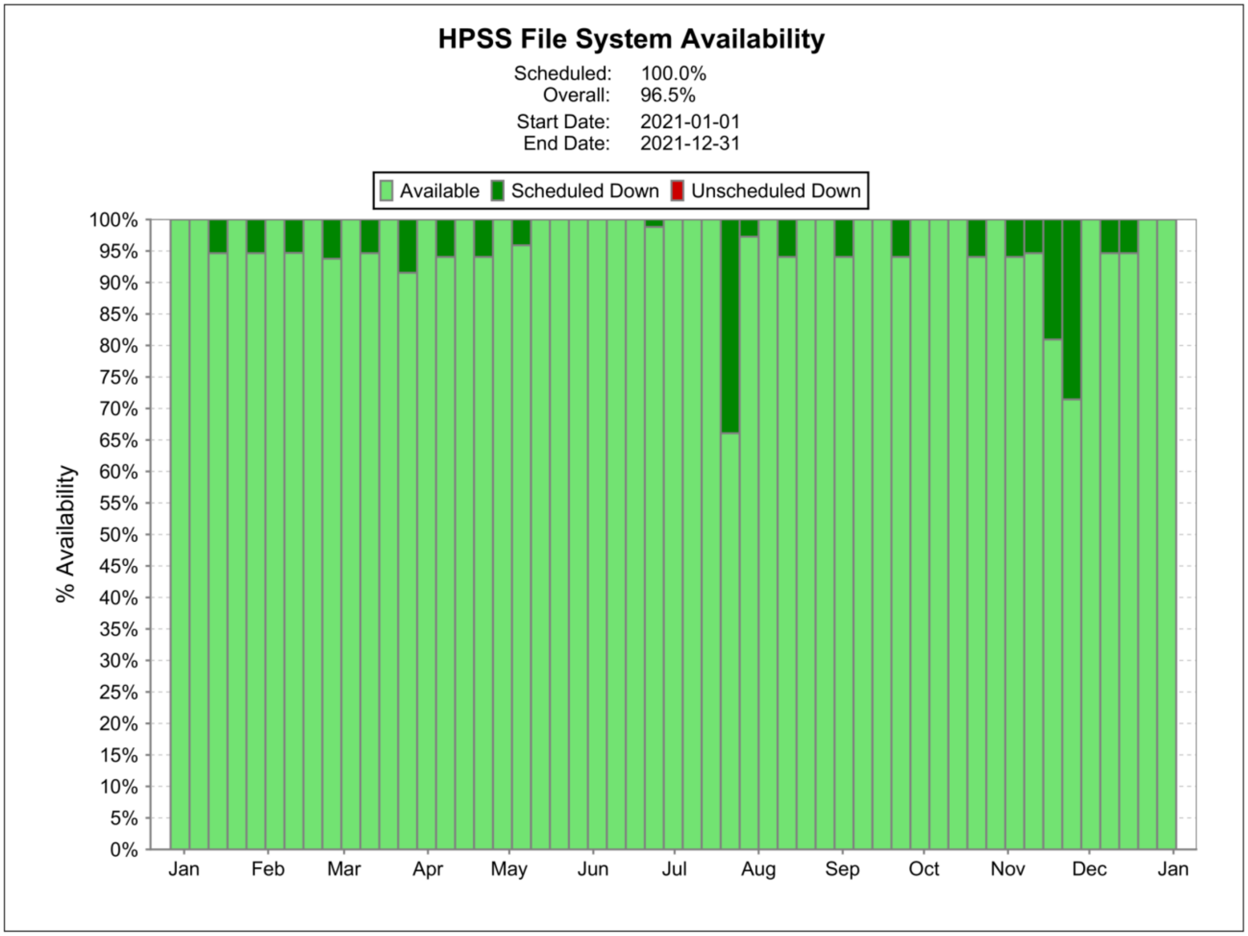


Figure 2.9 HPSS Weekly Availability for CY 2021

July 18–20, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

November 19–22, 2021: Scheduled extended maintenance – Power work for Aurora
In preparation for Aurora's power needs, TCS building management shut down power to the building over the weekend to perform power feed maintenance and upgrades. ALCF initiated a controlled shutdown on Friday and began startup procedures on the following Monday at 8:00 a.m.

2.5.3.2 MTTI and MTTF

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

HPSS Archive LTO8 tape drives and tape with 350 PB of storage capacity				
	CY 2020		CY 2021	
	Target	**Actual**	**Target**	**Actual**
System MTTI	N/A	19.51 days	N/A	16.01 days
System MTTF	N/A	366.00 days	N/A	365.00 days

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

2.6 Center-Wide Operational Highlights

2.6.1 Management Storage Refresh

For more than a decade, ALCF has leveraged NetApp storage arrays attached via storage area network (SAN) to provide high-speed and fault-tolerant data stores. In April of 2021, ALCF finished the migration to a new set of solid-state drive (SSD)-based storage systems to better support databases and other management systems.

2.6.2 Ubiquitous Secure Sockets Layer (SSL) Certificates

For years ALCF has used Entrust SSL certificates on systems that required encrypted connections. These were costly and time-intensive to acquire. In 2021, ALCF participated in the Argonne pilot program for use of InCommon SSL certificates. This self-service program allows ALCF to deploy SSL certificates trivially and at no additional cost.

2.6.3 Community File System Live

In 2021, Argonne deployed its community file system, Eagle, whose primary purpose is to enable researchers to share their data with the research community using the Globus file transfer and data sharing service. On top of the Lustre configuration previously detailed, Eagle uses four 100-GB data transfer nodes running Globus Connect Server v5. Globus provides flexible data sharing options to our researchers, ranging from small collaborations with specific colleagues to broad sharing with the public at large. Eagle is also directly accessible from all ALCF HPC resources, which makes it possible for researchers to share their data in-place as soon as it is produced.

Conclusion

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

ALCF closely tracks hardware and software failures and their impact on user jobs and metrics. These data are used as a factor in the selection of troubleshooting efforts and improvement projects. In CY 2021, this regular failure analysis has continued to drive code changes to software infrastructure at ALCF (like Cobalt) and has provided details to support debugging of storage system problems.

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 published allocation policies. The ratio of allocation policies available to INCITE, ALCC, DD, and ECP is 60 percent, 20 percent, 10 percent, and 10 percent, respectively. The expectation of 30 percent of the facility's resources available to ASCR is provided through the 20 percent to ALCC and the 10 percent to ECP.

The INCITE program fully allocates the 60 percent of the time available to it. The Discretionary time tends to be overallocated, but substantially underused due to the exploratory nature of the projects in the Discretionary program.

As the results show in Section 8, these are reasonable allocations of resources. Below are a few data points ALCF looks at when analyzing usage statistics for the various allocation programs.

3.1 Usage of the INCITE and ALCC Hours

The INCITE 2021 program allocated 17.8M node-hours on Theta to 16 projects. The allocation usage by project for Theta is shown in Figure 3.1. A total of 20.8M node-hours was delivered to INCITE on Theta (Table 3.1). Of the 16 INCITE projects, only 3 projects used less than 75 percent of their allocation. The other 13 projects used more than 90 percent of their allocations, including 10 projects that used their entire allocations or more. These projects used the extra node-hours to achieve additional milestones. INCITE and ALCC projects were permitted to exceed their node-hour allocations up to 125 percent on Theta, enabled by the ALCF overburn policy that permitted projects to continue running capability-sized jobs after their allocations are completely exhausted (as explained in section 3.3).

For the 2020–2021 ALCC year, 32 projects had allocations on Theta for a total of 6.9M node-hours. The allocation usage is shown in Figure 3.2. Nine of these projects used under 50 percent of their allocations, 9 projects used between 50 and 90 percent of their allocations, and 14 projects used their entire allocations or more.

The 2021–2022 ALCC year is approximately halfway through its allocation cycle. So far, the 17 projects have received allocations of 6.9M node-hours on Theta. The 2021–2022 ALCC projects have used a total of 2.3M node-hours from July 1, 2021, through December 31, 2021. The allocation usage is shown in Figure 3.3.

Table 3.2 summarizes the ALCC node-hours allocated and used in CY 2021 and includes COVID-19 research projects run on Theta per ASCR which typically had off-cycle start and end dates. The 7.3M ALCC node-hours allocated are calculated by adjusting the 2020–2021 and 2021–2022 Theta ALCC year allocations by the percentage of their award cycle occurring in CY 2021, then summing these two values. The total ALCC node-hours used, 7.2M, is the sum of all node-hours charged against any Theta ALCC project in CY 2021.

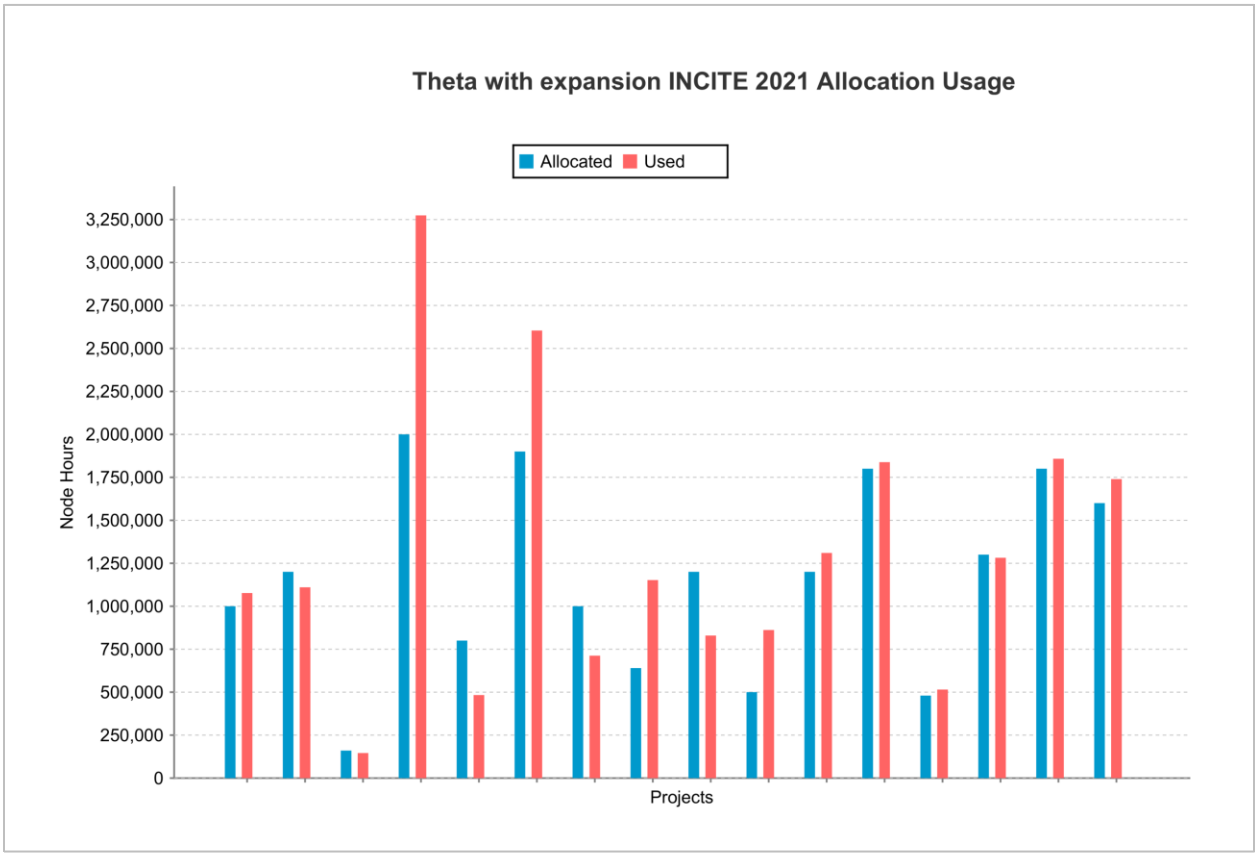


Figure 3.1 Theta INCITE 2021 Allocation Usage (Note: Projects are randomly ordered.)

Table 3.1 INCITE 2021 Time Allocated and Used on Theta with expansion, 2021

Projects	Theta with expansion
Allocated Node-Hours	17.8M
Used Node-Hours	20.8M

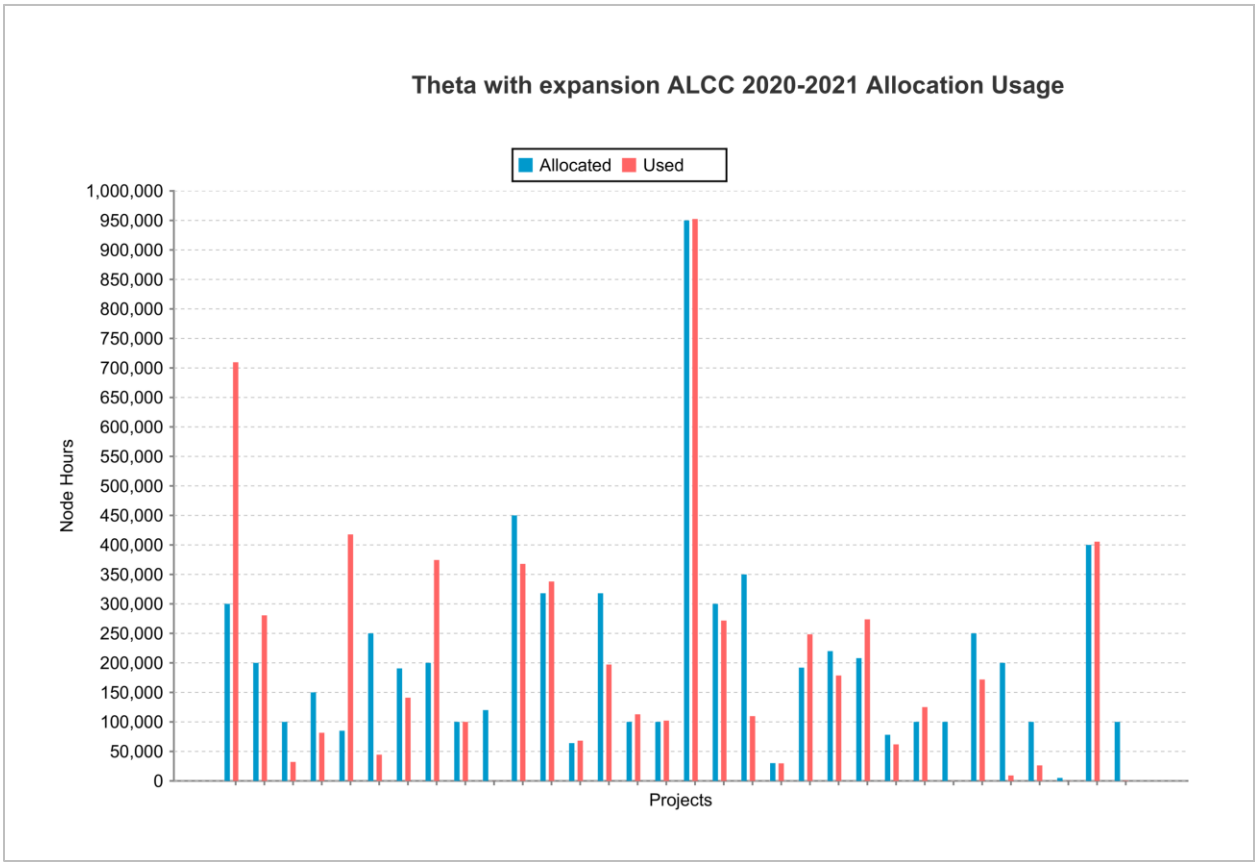


Figure 3.2 Theta ALCC 2020–2021 Allocation Usage (Note: Projects are randomly ordered.)

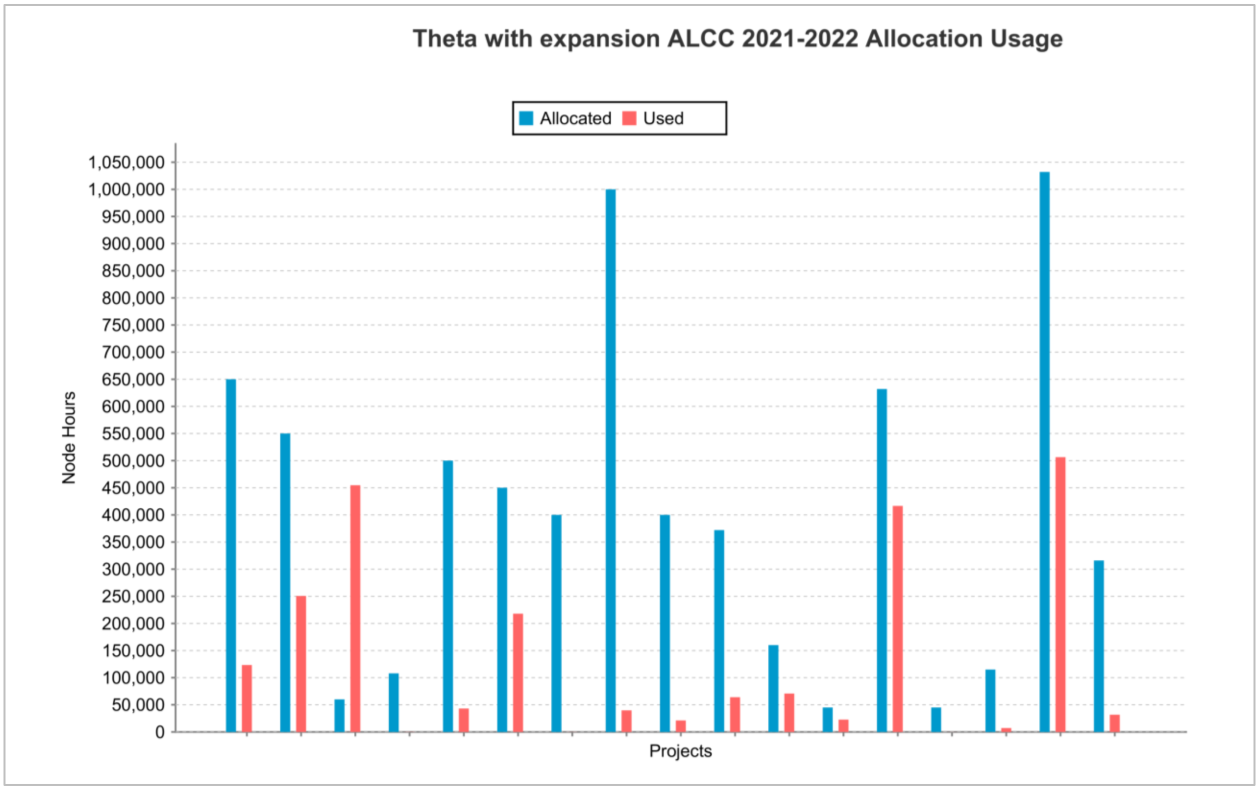


Figure 3.3 Theta ALCC 2021–2022 Allocation Usage as of Dec. 31, 2021 (Note: Projects are randomly ordered.)

Table 3.2 ALCC Time Allocated and Used on Theta with expansion in CY 2021

Projects	Theta with expansion
Allocated Node-Hours	7.3M [a]
Used Node-Hours	7.2M [b]

[a] Allocation total is obtained by adjusting each ALCC cycle allocation (2020–2021, 2021–2022) to prorate it for the amount of time allocated in CY 2021, then summing.

[b] Usage total is the number of node-hours charged for jobs run against any ALCC allocation in CY 2021.

3.2 Facility Director's Discretionary Reserve Time

The Director's Reserve, or Director's Discretionary (DD) program, serves members of the HPC community who are interested in testing science and applications on leadership-class resources. Projects are allocated in five categories:

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

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

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

ALCF also allocates time to projects that might still be some time away from submitting an INCITE proposal, or that offer a “strategic science” problem worth pursuing. Examples include supporting projects from DOE’s Scientific Discovery through Advanced Computing (SciDAC) program, industry research efforts, and emerging use cases, such as coupling experimental and computing facilities. The ALCF Data Science Program (ADSP) is allocated through the DD pool and is focused on developing the technical capabilities of data-driven projects that need leadership-class resources.

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

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

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

Table 3.3 shows the number of projects and total time allocated in the DD program on Theta during 2021. By its very nature, the DD program is amenable to over-allocation because often time is left unused; however, it should be noted that these totals do not represent open 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 allocated time. DD projects run at a lower priority than INCITE or ALCC projects, which reduces the amount of time available for their use. Exceptions are made for some internal projects that support acceptance of new hardware or support of users, which is in line with the ALCF core mission.

Table 3.3 DD Time Allocated and Used on Theta with expansion, 2021

Projects	Theta with expansion
Allocated Node-Hours	7.8M
Used Node-Hours	5.5M

A list of the CY 2021 DD projects on Theta is provided in Appendix B.

Figure 3.4 provides a breakdown of the CY 2021 DD allocations by domain for Theta.

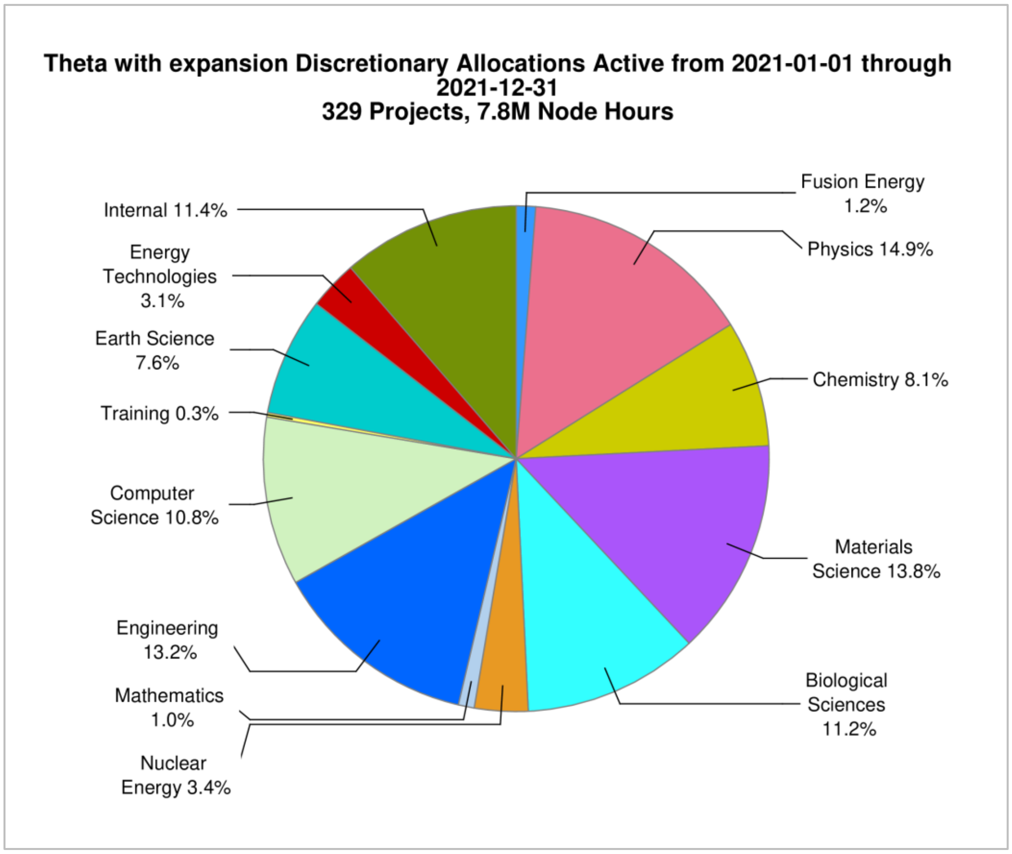


Figure 3.4 Theta CY 2021 DD Allocations by Domain

3.3 User Allocation Utilization

Inevitably, some projects will use less time than allocated and other projects will want to use more time than allocated. 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 allocations, 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 running. At this point, 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 facility with a short description of what they plan to do with the additional hours, highlighting specific goals or milestones and the time needed to accomplish them. These requests are reviewed by the scheduling committee, allocations committee, and ALCF management. Non-capability jobs from projects that have exhausted their allocation will run in the backfill queue.

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

The ALCF has multiple methods of managing under- and over-utilization of user allocations. The overall goal of all of the policies is to ensure that user projects have the greatest chance to accomplish their science goals.

Under-utilization earlier in the allocation year is primarily managed through personal contact with the projects to understand and help resolve. Examples of these challenges can include data movement, scheduling challenges, porting problems and bugs. Nearing the end of allocation years, projects with lower utilization are given higher scheduling priority through both the regular batch queuing system and through reservations.

3.4 Other Large-Scale Managed Resources

Last January, ALCF rolled out two new filesystems, Grand and Eagle, as part of the ALCF Community Data Co-Op. Every new project awarded compute time was also given an allocation on Grand or Eagle. Supplementary information on the technical specifications and availability for Grand and Eagle can be found in section 2.5.2. As part of the larger mission to support collaborative and data-driven scientific discoveries, ALCF is providing researchers with services to securely share and reliably transfer data using Globus. Since this capability was announced and made available, 19 data-only projects were created using these services. Additional information on the ALCF Community Data Co-Op can be found in section 4.1.2.

Conclusion

The ALCF delivered the following node-hours to the allocation programs in CY 2021: 20.8 million to INCITE, 7.2 million to ALCC, and 5.5 million to DD. The DD Reserve has been used not only to develop INCITE and ALCC proposals but also to conduct real science of strategic importance and to drive development and scaling of key INCITE and ALCC science applications. Excellent ALCF support and solid, high-performing ALCF resources have enabled INCITE and ALCC projects to run simulations efficiently and to achieve science goals that could not otherwise have been reached.

This page intentionally left blank.

Section 4. Operational Innovation

(a) Have technical innovations been implemented that have improved the facility's operations?

(b) Have management/workforce innovations been implemented that have improved the facility's operations?

(c) Is the facility effectively utilizing their postdoctoral fellows?

ALCF Response

Listed below are the innovations and best practices carried out at ALCF during CY 2021. ALCF innovations and best practices have helped to prepare for future systems, have enabled more efficient operations, and have strengthened collaboration and engagement, both across ASCR facilities and beyond.

4.1 Operational Innovation – Technical

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

4.1.1 Drastically Improved Memory Collection in PBS

Challenge: Because of the scale of ALCF's operations, ALCF pushes PBS harder than most facilities do. The existence of a memory leak in PBS has been known to the community for several years—one that even Altair has been unable to track down. ALCF staff have been using a periodic interpreter restart as a workaround; however, our testing, and in particular, our extreme level of operational monitoring, was causing 64-GB systems under heavy workloads to crash in a matter of hours.

Approach: After two weeks of deep effort, ALCF staff tracked down and corrected the elusive memory leak by writing debug "hooks" to track reference counts, garbage collector activity, and circular references. The discovery of a circular reference between the class and class objects which are cached revealed that any object that had an attribute reference could not be garbage-collected, thereby resulting in continuous memory growth. Replacing the cached dictionaries with weakref.WeakKeyDictionary objects completely eliminated the issue; memory consumption now follows the typical "up and down" pattern as memory is used and then garbage collected.

Impact/Status: Memory consumption by PBS has drastically changed, increasing the operational efficiency. This bug fix has been submitted as a pull request (PR) to the OpenPBS repository, and Altair has designated incorporation of this PR as a blocker for the V22 PBS Pro release.

4.1.2 ALCF Community Data Co-Op (ACDC)

Challenge: Eagle, ALCF's recently deployed 100PB file system, is configured to allow PIs to broadly share their data using Globus. However, simplified interfaces for sharing, discovering, and accessing these data sets were still needed.

Approach: The ALCF Community Data Co-Op (ACDC) portal is a deployment of the Django Globus Portal Framework customized for a variety of different projects. The portal enables many possibilities for data control and distribution, including straightforward management of permissions and access to data. The portals also allow users to query and filter data and to use faceted search for data discovery, and they provide a framework for other interfaces including data processing capabilities, all secured with authentication and configured authorization policy.

Figure 4.1 Snapshot of the ALCF Community Data Co-Op (ACDC) Portal page.

Impact/Status: The ACDC powers data-driven research by providing a platform for data access and sharing, and value-added services for data discovery and analysis, enhancing user experience. ACDC's fully supported production environment includes several project-specific data portals that enable search and discovery of data hosted on Eagle.

4.1.3 Accelerating Data Pipelines of Scientific Deep Learning Applications

Challenge: Data Input/Output (I/O) often is a bottleneck for scientific deep learning (DL) applications due to the repeated loading of large data sets from the file system at each iteration of the training process. Identifying and mitigating these I/O bottlenecks can be a challenge, particularly as the size and complexity of the resources being used increase.

Approach: ALCF developed a new benchmark, DLIO, to identify I/O bottlenecks and offer optimization strategies to improve the overall time-to-solution for DL applications (Figure 4.2). DLIO uses high- and low-level I/O profiling tools (profilers from DL frameworks, like TensorFlow, and I/O profilers like Darshan) to provide a holistic view of how data is accessed in applications. DLIO incorporates the observed I/O behavior of these applications and can emulate an application's I/O performance. It also provides numerous tunable mechanisms to evaluate the efficacy of multiple I/O access optimization strategies and can project the behaviors of DL applications at scale.

Impact/Status: DLIO was deployed on ALCF systems and evaluated with DL applications from the ECP, ADSP, and ESP programs, and helped improve application performance at scale. The work resulted in a Best Paper Award at the 2021 Institute of Electrical and Electronics Engineers Association for Computing Machinery (IEEE ACM) International Symposium on Cluster, Cloud, and Internet Computing (CCGrid). ALCF plans to investigate operational performance soon and eventually share the benchmark with other facilities.

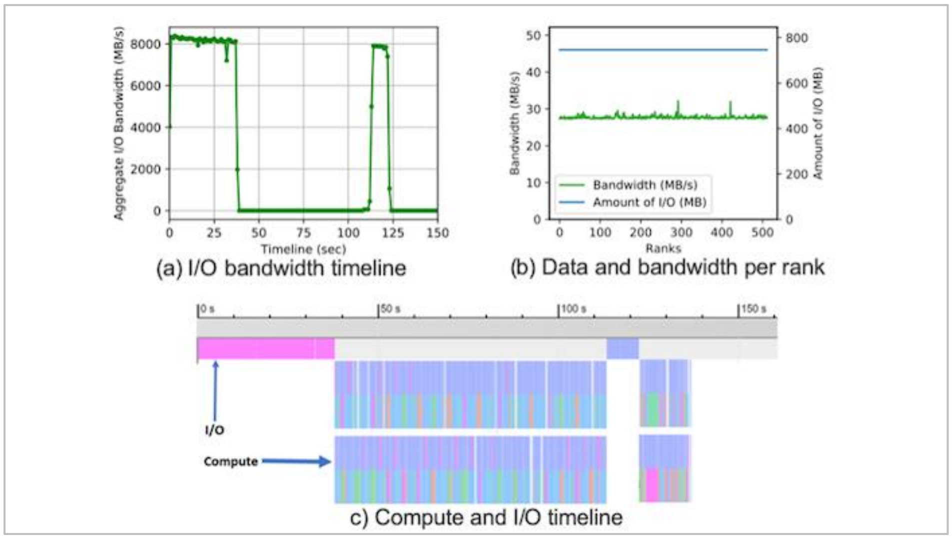


Figure 4.2 I/O Behavior of CANDLE NT3 DL application: Graph (a) shows the aggregate bandwidth of 8 GB/s achieved for the applications. Graph (b) depicts the distribution of I/O (i.e., 700 MB) and the achieved bandwidth (28 MB/s) across ranks. Graph (c) shows the merged timeline of I/O and compute, which shows that I/O and compute do not overlap in the application. Image courtesy of H. Devarajan, H. Zheng, A. Kougkas, X.-H. Sun, and V. Vishwanath, ALCF.

4.1.4 Splitting and Streamlining Theta's GPUs to Support More Users

Challenge: ThetaGPU consists of 24 NVIDIA DGX A100-based nodes capable of supporting diverse workloads. While some workloads may require all eight A100 GPUs on a node, others may need significantly fewer. Dedicating a full A100 GPU to a job that needs only a fraction of its cores limits the number of users who can access the system, while leaving resources unutilized. The NVIDIA Multi-Instance GPU (MIG) enables a single A100 GPU to be partitioned into as many as seven independent instances. In addition to dividing up the GPU, other resources within a node, such as memory, local filesystem, and CPU cores, must also be divided and managed among multiple users.

Approach: To facilitate the use of the MIG mode on Theta, a number of scripts were developed to help streamline this process. These scripts leverage pluggable authentication modules (PAMs) for allocation and isolation of CPU cores, memory, and GPUs via MIG. Mount namespaces are used to isolate portions of the local filesystem, giving exclusive access to a specific job running on the resource. Use of these scripts has made it less cumbersome and disruptive to divide Theta's GPU nodes to support multiple users.

Impact/Status: The development of these MIG mode scripts took several iterations as NVIDIA's software stack evolved. This process has been completed and the scripts are used in production for easily reconfiguring a subset of the ThetaGPU nodes into, and out of, MIG mode to accommodate user needs. For example, to support hands-on activities at recent AI for Science workshops hosted at ALCF, a small number of nodes were put into MIG mode, where each node was capable of supporting 56 users (eight A100s, each divided into seven GPU instances). This configuration enabled ALCF to support a large number of users engaged in these workshops, while keeping the majority of the nodes in their standard configuration, eliminating disruption to other facility users, helping both with user experience and increased operational performance.

4.1.5 Continuous Integration Standardization and Usage Increase

Challenge: Continuous integration (CI) testing on HPC resources is critical to ensuring that users' codes are ready for exascale systems when they arrive at DOE facilities. Standardizing and simplifying this process improves ALCF's ability to support a growing number of projects.

Approach: ALCF works with the ECP-CI project and collaborates with other national laboratories on the development of Jacamar (gitlab.com/ecp-ci/jacamar-ci), an executor for supporting Gitlab CI on HPC resource schedulers, such as Cobalt. Gitlab runners are remote servers that execute CI jobs based on configurations in the source repository. Use of the Jacamar executor allows ALCF to use Gitlab's official releases for the runners, rather than a forked version of Gitlab runners, which was out-of-date and not supported. Jacamar also includes quite a few security improvements tailored to work in HPC environments.

Results/Status: ALCF maintains two Gitlab instances for supporting CI: one for running on Theta, and the other for JLSE to facilitate the use of CI testing on early access hardware. Both instances have been updated to use the official Gitlab release and support the Jacamar executor. Both have also experienced increased usage, with 20 new projects on the Theta instance, and 50 new projects on JLSE in 2021.

4.2 Operational Innovation – Management/Workforce

ALCF works to prepare for next-generation systems through collaboration with vendors and other DOE facilities. The Facility reports on participation in research projects in Section 1.3.2.1 and professional community activities in Section 8.3.1.

4.2.1 New Hands-on AI Training for New Target Audience

The convergence of AI and HPC is transforming science and engineering, empowering researchers with new tools and capabilities to accelerate discoveries and innovation. To realize the full potential of AI for science, ALCF is committed to training and mentoring a diverse, new generation of AI practitioners and innovators. This training is a tool for future recruitment in the field of scalable AI.

Building on the ALCF's robust training programs in both these areas, ALCF hosted a series of hands-on AI for Science training sessions specifically aimed at undergraduate and graduate students enrolled at U.S. universities (Figure 4.3). These courses taught attendees to use leading-edge supercomputers to develop and apply AI solutions for the world's most challenging problems. Attendees gained knowledge and hands-on experience in using AI tools and frameworks on the ThetaGPU system for a variety of science cases. ALCF developed training materials focused on scientific AI topics that include data parallel training, hyperparameter and network architecture search, profiling and optimization of DL applications, and others.

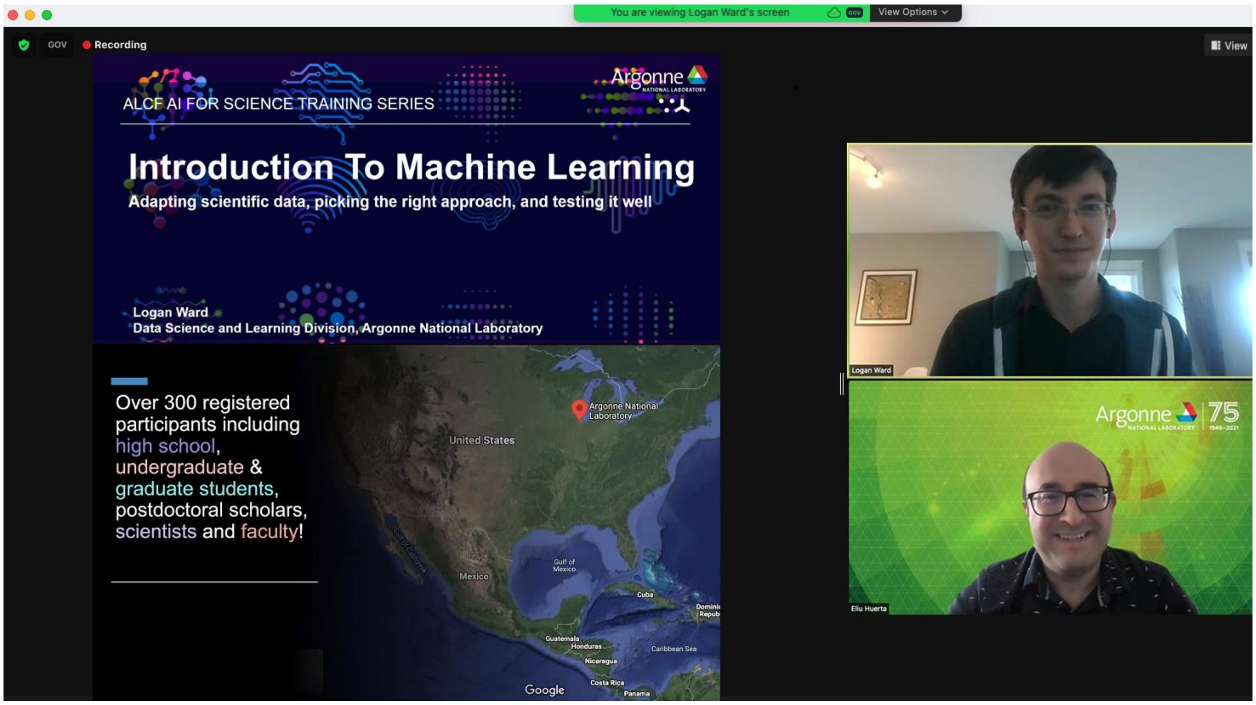


Figure 4.3 Screengrab from an AI for Science training workshop presentation.

4.2.2 ServiceNow Migration

Challenge: ALCF continually strives to enhance its user services. The service desk software used to track and respond to user requests is an integral part of that service. Request Tracker (RT), the software used for this purpose at ALCF, was an older version that was no longer supported and in need of an update.

Approach: After an evaluation of RT and other similar platforms, ServiceNow was found to offer additional features over RT. In addition, not only was ServiceNow already in use in other Argonne divisions, but it is also the laboratory's official service desk software.

Results/Status: The testing phase of ServiceNow in the ALCF environment started in CY 2020. That testing has been completed, and ALCF's service desk software has now been fully migrated to ServiceNow, going into production in November 2021. With this migration, ALCF moved away from an older, unsupported version of RT to a platform that is up-to-date and in use laboratory-wide, bringing ALCF in line with the rest of the laboratory.

4.2.3 Use of Virtual Platforms for Collaboration and User Support

The pivot to virtual platforms like Teams and Slack was very instrumental in strengthening and advancing collaborations with project PIs. The CPSFM_2 INCITE project, for example, is a renewed project, now in its fifth year at ALCF. The collaborators in the project are spread across four national laboratories (Argonne, Oak Ridge, Sandia, and Livermore), multiple universities, and four countries (the United States, South Korea, Japan, and France). With Covid-19 meeting and travel restrictions in place, new schemes were needed to keep communication channels open and maintain the high productivity of the project. In addition to moving their bi-annual face-to-face project meetings to Teams, they leveraged Teams and Slack for training newcomers to the project. They also implemented a bi-weekly “office hour” where the ALCF Catalyst is guaranteed to be available on Slack with an option to use Teams if needed.

Virtual platforms have also been instrumental in maintaining and improving user support activities. Prior to Covid, most user issues were handled via e-mail or phone. With the shift to virtual platforms with video conferencing capabilities, many user issues are more adeptly addressed, and ALCF has continued to use more of those services in 2021. Now a user can share their screen with support staff in Teams to quickly diagnose and resolve problems. User support staff have also leveraged Slack during ALCF-sponsored workshops and training events to provide real-time support.

4.2.4 Use of Virtual Platforms for Workforce Engagement

As Argonne continued in Minimum Safe Operations during FY 2021, the majority of ALCF staff continued to work remotely. In addition, the ALCF team continued to grow, adding five new staff members, as well as six new postdoctoral fellows over this time. Starting a new position in a new environment under normal circumstances can be stressful enough. But adding to that the fact that everyone is remote can make it difficult to feel engaged and become integrated with the rest of your new team. Virtual platforms like Teams and Zoom have helped to ease the onboarding process through video conferencing capabilities. However, ALCF is more than just the technical expertise of its team. While valuing diversity and inclusion, it also fosters a sense of community, which can be lost while in isolation. Virtual platforms have been used not only for the exchange of technical information, but also to promote social interaction. Outside of working hours, ALCF hosts a bi-weekly virtual social hour via Zoom. Members from across the division come together to discuss a range of topics, from cars to movies to sports, and enjoy each other’s company. It has greatly improved the morale of team members, both old and new.

4.3 Postdoctoral Fellows

ALCF supports a steady-state postdoctoral fellowship program. Within this program, ALCF supports one named postdoctoral fellow, the ALCF’s Margaret Butler for Computational Science Fellow. Postdocs are awarded 1-year appointments with an option to be renewed for an additional year (this is typically the case), with a similar option to renew for a third year. The major goal of the program is to either convert the postdocs to a regular staff appointment, place them at another DOE laboratory, or support their efforts to find an academic or industry appointment. The objective, in all cases, is for these postdocs to continue as lifelong users of DOE compute resources.

The general process is that applications for postdoctoral positions are handled by Argonne's Postdoctoral Program Office. In CY 2021, ALCF hired six new postdoctoral researchers (fellows) and supported a total of 18, representing a range of scientific domain backgrounds. Of those, three postdocs were terminated, four transferred to another division, and two were promoted to ALCF staff. Of the nine active postdocs, 4 are ESP, 2 are ECP, and 3 are steady state. The ALCF postdocs work on various research topics, including computational chemistry; computational fluid dynamics; ML/AI for particle physics; ATLAS, simulation, and deep learning; exascale interconnect performance evaluation; fusion and deep learning; Lattice QCD; electronic structure; climate-scale simulation; materials science; cosmological n-body simulations; dark matter halo modeling and clustering analysis and cosmological parameter estimation; and mixed precision optimization for next-generation AI accelerator systems.

Once hired, each postdoc is assigned both a direct research supervisor and an Argonne staff mentor. The mentor, initially selected by the division or the supervisor, can be changed by the postdoc. The supervisor meets with the postdoc on a weekly basis and engages in the postdoc's research effort. The supervisor then evaluates the progress and completes a standardized review process that is submitted to ALCF management for review, including the Division Director, who reviews and authorizes all appointment renewals. The mentor is responsible for meeting with the postdoc to discuss career development milestones and personal goals; this interaction happens as needed, but no less than once a quarter. The guidance for these discussions includes key skills the postdoc should focus on over the next year; opportunities for development; and, if entering the third year, what skills or experience will be most beneficial to enabling a smooth career transition. The Division Director also meets monthly with the postdocs as a group to hear progress updates, address any issues specific to the postdoc community, and solicit feedback in general.

ALCF supported the following 18 postdoctoral researchers in CY 2021:

Abhishek Bagusetty (Ph.D., Chemical Engineering, University of Pittsburgh). **Hired:** January 2020. **Promoted to Argonne staff:** June 2021. **Current role:** ALCF performance engineer. **Research area:** computational chemistry. **Postdoctoral research project:** Aurora ESP project *NWChemEx*. **Accomplishments included:** (1) contributing to several pieces of the NWChemEx suite and other software in the ECP ecosystem; (2) mastering porting codes with CUDA to DP++SYCL and tuning them to perform in different architectures in Aurora testbeds; and (3) publishing articles in *Parallel Computing* and in *Nature Communications;* and (4) co-authoring a poster presented at IWOCL'21: International Workshop on OpenCL.

Riccardo Balin (Ph.D., Computational Fluid Dynamics and Turbulence Modeling, University of Colorado Boulder). **Hired:** January 2021. **Research area:** coupling large-scale, high-fidelity simulations of turbulent flows with machine learning to perform in-situ training of, as well as inference with, data-driven turbulence closure models. **Current projects:** (1) *CFDML_aesp*: Data Analytics and Machine Learning for Exascale Computational Fluid Dynamics (CFD) (primary project), and (2) *PHASTA_aesp*: Development of the PHASTA CFD Code for Exascale Simulations on Aurora. **Scientific goal:** to develop a framework for applying in-situ machine learning to subgrid modeling for hybrid Reynolds-averaged Navier-Stokes (RANS)/large eddy simulation (LES) of high Reynolds number separated flows. These coupled simulations will be among the first to run on Aurora. **Accomplishments:** developed an in-situ machine learning framework for large-scale turbulence simulations on Aurora, including adopting SmartSim (Cray/HPE Aurora NRE development) and connecting to CFD simulation software (PHASTA).

Denis Boyda (Ph.D., Theoretical Physics, National Research Center Kurchatov Institute, Moscow, Russia). **Hired:** September 2020. **Research area:** machine learning for particle physics using HPC. **Project:** Aurora ESP project "Machine Learning for Lattice Quantum Chromodynamics" project (PI**:** William Detmold). Currently working to prepare it for Aurora (porting code, writing tests for early-access hardware and software, trying alternative mathematical approaches to speed up the code, and more.) **Scientific goal:** to simulate the interactions of dark matter candidate particles and nuclei, leading to insights into dark matter and fundamental particle physics. **Accomplishments include:** (1) improving the normalizing of flows for sampling lattice field theories with fermions such as occur in the Yukawa model; (2) developing normalizing flows for sampling lattice field theories in three and four dimensions (U(1) model in four dimensions was developed); (3) developing an approach for simulating Hamiltonian Monte Carlo (HMC) dynamics with linearized formalism of Koopman operator; and (4) developing memory optimizations and coding for special separated convolutions and convolutions in 4D. Published three articles in *Physical Review D*, and one in the proceedings of the *XXXIII International Workshop on High Energy Physics*.

Tyler Burch (Ph.D., Physics, Northern Illinois University). **Hired:** May 2020. **Terminated:** January 2021. **Current job:** Baseball analyst for the Boston Red Sox. **Research area:** ATLAS, simulation, deep learning. **Past projects:** Aurora ESP: ATLAS (PI: Walter Hopkins). **Accomplishments included:** (1) porting the MadGraph event generation framework processes from CUDA to SYCL; and (2) developing a neural network that takes a Bayesian posterior as input to perform approximate uncertainty quantification in ATLAS physics analyses.

Kevin A. Brown (Ph.D., Mathematical and Computer Science, Tokyo Institute of Technology). **Hired:** October 2019. Won Argonne's very first Walter Massey Fellowship in 2021. Transferred to the Mathematics and Computer Science (MCS) Division in October 2021. **Current role:** Argonne Fellow and project PI of a 3-year Laboratory Directed Research and Development (LDRD) Program project titled "Improving Data Movement Performance for Emerging AI and Climate Science Workloads on Future Supercomputers." **Research area at ALCF:** optimizing network designs and communication strategies for exascale systems. This work will guide the configuration of Aurora and other Slingshot machines at DOE facilities. **Accomplishments:** network QoS simulation work and other network model development and support work, including: (1) making contributions to the CODES simulations toolkit; (2) developing and simulating routing strategies that can improve packet latencies on large supercomputers; and (3) supporting the Threadwork LDRD project by simulating human-assisted computer control (HACC) network traffic using CODES. Talks and papers included one publication (paper) in *2021 ISC High Performance Conference*, one presentation at the ECP Annual Meeting, and one workshop presentation at the Illinois Louis Stokes Alliance for Minority Participation (ILSAMP) Annual Spring Symposium & Student Research Conference.

Kyle Felker (Ph.D., Physics, Princeton University). **Hired:** September 2019. **Promoted:** to Assistant Computational Science in Argonne's Computational Science Division in July 2021. **Current role:** member of ALCF's Catalyst team. **Research area:** fusion, deep learning. **Project:** Aurora ESP project "Fusion RNN" (PI: Bill Tang). **Scientific goal:** scaling-up studies of fusion Tokamak reactors using deep learning to predict operational failures. **Accomplishments include:** (1) multiple publications on plasma physics applications with deep learning; (2) collaborations with AI vendors including Groq for deterministic machine learning for latency critical control systems; and (3) multiple publications on the applications of AI in plasma physics control systems.

Sam Foreman (Ph.D., Physics, University of Iowa). **Hired:** August 2019. **Research areas:** machine learning, data science, high energy physics, Lattice QCD. **Scientific goal:** to use neural networks to overcome bottlenecks in Lattice QCD workflows. **Projects:** (1) continued development on l2hmc-qcd, a Python library for building and training improved sampling techniques for simulations in Lattice QCD; and (2) primary developer of nftqcd/fthmc, a larger collaborative effort geared toward studying trainable field transformations for generating gauge configurations in lattice gauge models. **Accomplishments include:** (1) presenting four invited talks: at Brookhaven National Laboratory-High Energy Theory (BNL-HET) & RIKEN BNL Research Center (RBRC) Joint Workshop "DWQ@25"; at Machine Learning for High Energy Physics, on and off the Lattice at ECT* -Trento; at the Massachusetts Institute of Technology (MIT) Lattice Group Seminar; and at the Machine Learning Techniques in Lattice QCD Workshop; (2) contributing to ThetaGPU MLPerf HPC submission for ALCF; (3) writing and presenting material at ALCF's 2021 SDL Workshop, 2021 Computational Performance Workshop, and AI for Science Training Series; and (4) presenting a poster at the Deep Learning for Simulation (SimDL) workshop at International Conference on Learning Representations (ICLR) 2021.

Kevin Gasperich (Ph.D., Chemistry, University of Pittsburgh). **Hired:** November 2019. **Research area:** computational materials science. **Current projects:** collaborates in porting and improving QMCPACK to exascale through OpenMP offload. **Accomplishments included:** development of QMCPACK mini app for OneAPI testing for Aurora.

Geng Liu (Ph.D., Mechanical Engineering, City College of New York). **Hired:** October 2021. **Research area:** computational fluid dynamics, HPC. **Current projects:** Aurora ESP project: *multiphysics_aesp* (PI: Amanda Randles) ─ the goal of this project is to advance the use of data science to drive in situ visualization of extreme-scale fluid-structure-interaction simulations. **Accomplishments included:** successfully porting lattice Boltzmann kernels from CUDA to Kokkos, with SYCL backend, and testing on Aurora early access hardware; working on performance evaluation study comparing implementations in CUDA, Kokkos (CUDA and SYCL backends), and Raja (CUDA backend).

Romit Maulik (Ph.D., Mechanical & Aerospace Engineering, Oklahoma State University). **Fellowship Award:** Margaret Butler Postdoctoral Fellow (2019–2021). **Promoted:** to Assistant Computational Scientist in the MCS Division in June 2021. Also holds a joint appointment as a Research Assistant Professor in the Department of Applied Mathematics at Illinois Institute of Technology. **Research area:** scientific machine learning algorithm development for various computational problems in aerospace, geophysical, and fusion applications. **Accomplishments included:** (1) being awarded an Impact Argonne Award for "tackling several climate model challenges and advancing the field of downscaled climate modeling and impact analysis"; (2) having numerous publications, including in *Physics of Fluids* and *Physica D*; and (3) being awarded DOE and National Science Foundation (NSF) funding to study mathematics of surrogate modeling for nonlinear dynamical systems.

Nathan Nichols (Ph.D., Material Sciences, University of Vermont). **Hired:** September 2021. **Research area:** Quantum Chemistry and Computer Science. **Project:** Postdoc for the ATLAS Aurora ESP program. **Scientific goal:** to adapt research codes to a variety of architectures, including Aurora. **Accomplishments included:** running SYCL software for ATLAS Aurora Early Science project on experimental Intel hardware for testing and measurement; (2) working on development of SYCL port of MadGraph5 software package; and (3) engaging with an AI research project with the local ATLAS group members in the High Energy Physics (HEP) division.

Andrea Orton (Ph.D., Geophysics/Atmospheric Sciences, Purdue University). **Hired:** October 2020. **Terminated:** November 2021. **Current role:** appointed as a Visiting Professor/Continuing Lecturer in Purdue University's Department of Earth, Atmospheric, and Planetary Sciences. **Research area:** climate scale simulation. **Project:** DD: Earth System Modeling project. **Accomplishments included:** assisting with pre-processing phase (initial model setup and scaling activity) on ThetaGPU and Cori for convective resolved climate simulations over the North American continent.

Pankaj Rajak (Ph.D., Computational Materials Science, University of Southern California). **Hired:** October 2018. **Terminated:** May 2021. **Current role:** Applied Scientist at Amazon. **Research areas:** materials science, HPC, DL, reinforcement learning (RL). **Projects:** (1) offline RL to build an autonomous AI agent for material synthesis and design of atomic structure with desired physical properties; and (2) Graph Neural Network Force Field development for molecular simulation and RL for accelerated discovery of dielectric polymer. **Accomplishments:** numerous publications, including two in *npj Comput. Mater.;* one in *J. Phys. Chem. Lett.;* one in *Phys. Rev. Lett.*; one in *J. Chem Inf Model.;* one in *Applied Phys. Lett.,* and two in *Bulletin of the American Physical Society.*

Esteban Rangel (Ph.D., Computer Science, Northwestern University). **Hired:** July 2018. **Transferred:** to the Computational Science (CPS) Division in July 2021. **Research area:** computational science, cosmological n-body simulations, HPC. **Projects:** ESP/HACC, ECP Application Integration, ECP/ExaSky, and Threadwork. **Accomplishments include:** (1) being the main developer of HACC's halo merger tree construction and analysis framework using a new algorithm that employs the recently developed "core-tracking" methodology; (2) publishing in *ApJ* and in *ApJS*; (3) leading work on a new parallel tessellation-based density estimation code that is now being used for work on weak and strong gravitational lensing; and (4) working at present on HACC's implementation on exascale platforms.

Siddhisanket (Sid) Raskar (Ph.D., Computer Science, University of Delaware). **Hired:** June 2021. **Research goals:** to evaluate the efficacy of AI architectures for scientific machine learning and the design of next-generation AI architectures for science. **Current projects:** working with dataflow-based AI architectures to explore their efficacies for scientific applications as a member of ALCF's Data Science group.

Umesh Unnikrishnan (Ph.D., aerospace engineering, Georgia Institute of Technology). **Hired:** November 2021. **Research area:** HPC, performance engineering, software portability, CFD, turbulence modeling. **Current projects:** (1) working with ALCF performance engineers to optimize the performance portability layer, OCCA, for Aurora. This work will enable codes like libParanumal and NekRS to run efficiently on Aurora, benefitting multiple science teams; and (2) working with the developers of libParanumal to leverage the existing library features and develop new capabilities for applications that are of interest to science teams at Argonne (e.g., Energy Systems [ES] Division) and beyond. **Scientific goal:** to develop a scientific application targeting CFD applications of high-speed compressible flows that is performant and portable to Aurora and other upcoming exascale architectures.

Antonia Sierra Villarreal (Ph.D., Astrophysics, University of Pittsburgh). **Hired:** August 2018. **Transferred:** to HEP Division in August 2021. **Research area:** dark matter halo modeling and clustering analysis and cosmological parameter estimation for the Large Synoptic Survey Telescope (LSST)-Dark Energy Science Collaboration (DESC); HPC. **Current projects:** works with the ALCF and HEP divisions to realize key deliverables for DESC in the form of end-to-end simulation results. **Accomplishments include:** (1) working with the Vera C. Rubin Observatory to generate state-of-the-art synthetic observations; (2) developing a cross-facility workflow and demonstrated execution on two DOE supercomputers; (3) having one manuscript in review and one in progress; and (4) winning a 2021 NERSC HPC Achievement Award for "Innovative Use of High-Performance Computing" for the image simulation campaign.

Zhen Xie (Ph.D., Computer Science and Technology, The Institute of Computing Technology of the Chinese Academy of Sciences, Beijing, China). **Hired:** August 2021. **Research area:** mixed precision optimization for next-generation AI accelerator systems. **Current projects:** (1) performance optimization on HPC and AI/DL applications; (2) parallel computing on various architectures; and (3) heterogeneous computing and memory systems. **Accomplishments include:** publication submitted on throughput-oriented and accuracy-aware deep neural network (DNN) training with bfloat16 on GPU; and evaluation of critical computer vision kernels on GPU and AI hardware.

This page intentionally left blank.

Section 5. Risk Management

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

ALCF Response

The overview of the risk management process that ALCF follows (and laid out in Section 5.1) clearly demonstrates that ALCF successfully managed both its project risks and operational risks in CY 2021. As part of the ALCF's Risk Management Plan (RMP), all risks (proposed, open, and retired) are tracked, along with their triggers and mitigations (proposed, in progress, and completed), in a risk register managed by risk managers. All risk ratings in this report are post-mitigation ratings. ALCF currently has 37 open risks, with two high operational risks: (1) Funding/Budget Shortfalls, which is managed by careful planning with the DOE program office and the implementation of austerity measures as necessary; and (2) Staff Recruitment Challenges, which is managed by ongoing recruiting and re-tasking of current staff as needed. The major risks tracked for the past year are listed in Section 5.2, along with the details of these risks in Table 5.1. Events stemming from those risks and the mitigations for them are described in greater detail in Section 5.3. Section 5.6 and Table 5.1 provide details on the major risks that will be tracked in CY 2022.

Of primary interest here is a description of the most significant operational risks and the Risk Management Plan's effect on the Facility's day-to-day operations.

The Facility should provide:

- ***A brief overview of the risk management process employed by the Facility, including the cycle for identifying, mitigating, and retiring risks;***
- ***A brief summary of the key risks and their mitigations, including:***
  - ***The 3–5 most important operational risks for the review year;***
  - ***Any significant new operational risks since the previous review;***
  - ***The operational risks that were retired during the review year;***
  - ***The major risks that will be tracked in the next year; and***
  - ***For the risk events that occurred, how the Risk Management Plan was implemented and an assessment of the mitigations' effectiveness.***

5.1 Risk Management Process Overview

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

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

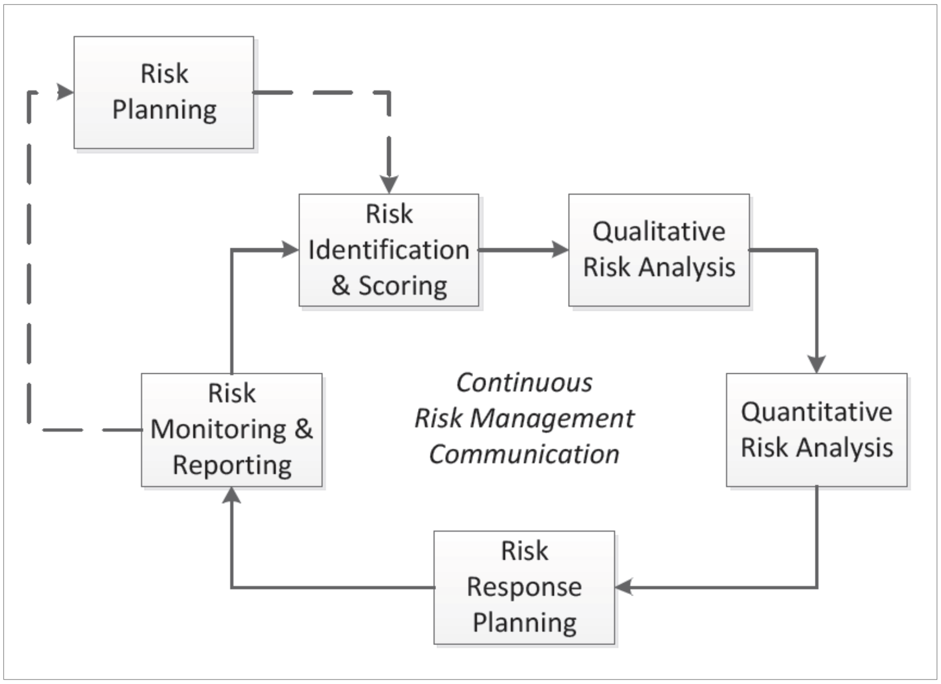


Figure 5.1 Illustration of the Iterative Risk Management Process from ALCF's Risk Management Plan

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

1. Plan, implement, and revise the RMP.
2. Identify threats and opportunities to the 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 the risks.
4. Monitor risks, mitigation plans, and management reserve and contingency until the risks are retired or a project is closed.

The objective 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 and shared cloud-based storage system, and risk forms and the risk register are formatted using Excel. Risk owners continuously monitor the risks they own and submit monthly reports on all risks through the ALCF online risk reporting form.

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 that serve 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 adding a proposed risk 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 and/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, including because of changes in the technology landscape.
  - Works with the risk owner to modify the risk statement should any risk information indicate a need for changes to risk mitigation strategies, risk triggers, or risk scope.
- 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 identifies any that apply to ALCF.

5.1.2 Risk Management in Day-to-Day Operations

ALCF currently has 37 open risks in the facility operations risk register and uses the post-mitigated risk scoring to rank the risks. These 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), and, as such, all risks are distributed throughout the division and owned by group leads or managers. On the operations side, subject matter experts estimate risk mitigation costs and use them to inform management.

In addition to formal and individual risk meetings and the Risk Review Board meetings, wherein risk owners discuss whether a risk continues to exist and whether ownership of a risk can change, the ALCF holds many informal risk discussions. Risks are identified and evaluated, and mitigation actions are developed, for all changes that occur at the Facility—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, and also can be triggered by an upcoming decommissioning of a resource. If the risks identified are short term or minor, they are not added to the registry. New significant risks identified during the activity planning are added to the registry and reviewed at the next risk meeting. Any changes made to an existing risk—whether recharacterizing it, changing the factors affecting it, or retiring it entirely, requires a meeting which can be proposed by any of the risk owners.

Other tools beyond the risk register are used for managing risks in day-to-day operations. An example is the use of Work Planning and Controls (WPCs) and Job Hazard Questionnaires (JHQs) to manage risks for activities where safety is a potential concern. WPCs are developed in consultation with safety and subject matter experts. JHQs are used for all staff and all contractors and covers all work. During planning meetings for any activities, staff members review the planned actions and evaluate 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 then documented in the WPC. The WPC is then used during the activity to direct the work.

Beyond the operations of the machines, risk management is used in such diverse ways as evaluating and managing INCITE and ALCC proposal risks (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 (including the opportunity risk that electricity costs could be lower than budgeted), etc.

5.1.3 Continuation of the ALCF-3 Project

The project to procure and deploy ALCF's next supercomputer, known as ALCF-3, continued in CY 2021. Risk Register managers continue to maintain a project risk register and track a set of detailed risks. Risk mitigation costs on the project side are developed using a bottom-up cost analysis, then are input to the commercial project risk analysis tool Oracle Primavera Risk Analysis (OPRA) to set 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 for the Review Year

The ALCF operated both Mira and Theta during CY 2021 and planned the growth of both the staff and the budget to bring the facility to full strength. As such, ALCF continues to monitor a large number of major risks for the facility. No risks were retired in CY 2021.

Four major operations risks were tracked for CY 2021, two with a risk rating of High, one with a risk rating of Moderate, and one with a risk rating of Low. None of these were encountered and all of them were managed. The four major operational risks are described in Table 5.1. All risk ratings shown are post-mitigation ratings. The risks are color-coded as follows:

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

ID	Title	Encountered	Rating	Notes
1059	Funding/Budget Shortfalls	No	High	ALCF regularly worked with the program office to plan a budget for handling the impact of a Continuing Resolution in FY 2022, new hires, and changes in the laboratory indirect expense rate. This risk remains a major concern as the facility moves forward with Theta and ALCF-3 in CY 2022.
25	Staff Recruitment Challenges	No	High	The ALCF added five new staff members in CY 2021: two external new hires and three reclassified employees. ALCF continues to have staff available who can be re-tasked as needed. With difficulty competing with industry for new hires, staff hiring remains a concern.
1049	Staff Retention	No	Moderate	ALCF lost three staff members during CY 2021: two terminated and one transferred to another division. 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	

5.3 Risks Encountered in the Review Year and Their Mitigations

ALCF encountered four risks in CY 2021. The risk owners are identified below, along with an assessment of the risk's probability and impacts, a brief description of what transpired, and how the risk was ultimately managed. One risk has a rating of Low (shown first), while three have a rating of Very Low (shown by code number).

5.3.1 Facility Power Interruptions

0031: Facility Power Interruptions	
Risk Owner	Mark Fahey
Probability	Low
Impact	Cost: Low; Technical Scope: Low
Risk Rating	Low
Primary Management Strategies	ALCF's director of operations participates in the Data Center management group. ALCF pays part of the cost of an Argonne Data Center liaison. Improve power system bus transfer by investigating the practicality of modifications to the power system to weather bus transfers without interruption.
Triggers	Electrical failure; multiple events related to power quality; scheduled power outages.

Description

On February 18, 2021, the first of several ComEd power line sags (decreases in voltage) occurred at 9:46 a.m., causing Theta PDU1 to trip and all ThetaGPU nodes to lose power.

Evaluation

ALCF staff waited until the sags ended before bringing up ThetaGPU, which was around 2:00 p.m.

Management

ALCF staff recognize that the compute nodes continue to be susceptible to power distribution unit (PDU) tripping and added a second uninterruptible power supply (UPS) feed to the networking equipment in ThetaGPU in December 2021.

5.3.2 Interruptions to Facility that Provides Cooling

0030: Interruptions to Facility that Provides Cooling	
Risk Owner	Mark Fahey
Probability	Very Low
Impact	Cost: Very Low; Technical Scope: Low
Risk Rating	Low
Primary Management Strategies	Increased redundancy. Site cooling piping is now interconnected to enable chiller plants to provide backup to each other. Implement emergency shutdown scripts to limit impact of increased heat on hardware and to reduce heat generation. Initiate root cause analysis on any failures or degradation and remediate accordingly.
Triggers	Rising temperatures on equipment in the Data Center; planned maintenance; monitoring notification of outages or rising temperatures.

Description

At 7:00 a.m. on May 26, 2021, a broken pipe at Argonne resulted in the loss of chilled water to the entire campus. Argonne plant operations notified all stakeholders via e-mail.

Evaluation

The operations team shut down Theta and all equipment in Building 221. Running jobs were killed and an e-mail notification was sent to users.

Management

The pipe was repaired around 8:30 a.m. and the chilled water plant that supplies ALCF was refilled by shortly after noon. Once cooling returned and stabilized, the team brought Theta back up and released it to users around 6:00 p.m. This event exposed an initial communication breakdown about the outage, which is being addressed through reinforcement of established processes.

5.3.3 System Stability Issues

1078: System Stability Issues	
Risk Owner	Ryan Milner
Probability	Very Low
Impact	Cost: Very Low; Technical Scope: Low
Risk Rating	Low
Primary Management Strategies	Conduct regression testing to monitor system health. Monitor and analyze system failure data. Activate swat team including vendor and ALCF Operations staff to do root cause analysis and develop workaround/resolve problem.
Triggers	User complaints; system failure rate increasing.

Event 1:

Description
On February 18, 2021, an upstream circuit breaker tripped, cutting power to ThetaGPU. Even after system power was restored, several of the nodes were unresponsive due to a firmware issue triggered by the power cycling.

Evaluation
ALCF staff waited until the sags ended before attempting to bring up ThetaGPU, which was around 2:00 p.m.

Management
NVIDIA support was engaged regarding the faults, which identified issues with the GPU firmware. The chassis firmware was upgraded shortly after this power event.

Event 2:

Description
On July 28, 2021, around 1:00 p.m., Theta entered a severely degraded state due to a cascade of compute rack power rectifier failures. These failed rectifiers are an older model with a known issue.

Evaluation
ALCF staff located enough spares on site to replace the failures.

Management
The hardware replacement was performed quickly. While jobs running at the time of failure were impacted by the outage, the machine was powered on and returned to users the same day.

5.3.4 Rack Coolant Monitor Fails to Function

1068: Rack coolant monitor fails to function	
Risk Owner	Mark Fahey
Probability	Very Low
Impact	Cost: Very Low; Technical Scope: Very Low
Risk Rating	Very Low
Primary Management Strategies	Consider increasing coolant flow. Contact other users for lessons learned. Root cause analysis. Activate swat team including vendor TCS management, FMS, and ALCF Operations to debug problem, develop workaround, and/or resolve problem.
Triggers	Temperatures rising in compute nodes; monitoring notifications indicate problems with water flow.

Description
On October 26, one cabinet of Theta shut off following a solenoid failure. Jobs running on the machine were lost and scheduling was temporarily halted.

Evaluation
The cabinet remained switched off for more than a week awaiting repairs. The rest of Theta was brought back online and continued to operate in the interim.

Management
Hewlett Packard Enterprise (HPE) engineering staff replaced the failed solenoid and brought the cabinet back online. Because this was the first such failure on the XC40, after analysis, no additional action was necessary.

5.4 Retired Risks

No risks were retired in CY 2021.

5.5 New and Recharacterized Risks since the Last OAR

There are no new risks and no recharacterized risks to report since the last OAR.

5.6 Top Operating Risks Monitored Closely for the Next Year

Table 5.1 lists the current top operating risks being closely monitored for CY 2022, along with the current risk rating and management strategies for each risk. These are the risks that experience has shown are most likely to be encountered in any fiscal year.

Table 5.1 Top Operating Risks Monitored for CY 2022

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.
25	Staff Recruitment Challenges	High	Evaluate possible additional recruiting avenues. Prioritize staffing needs. Adjust work planning. Retrain staff to meet ALCF needs. Re-task staff as needed.
1049	Staff Retention	Moderate	Make salaries as competitive as feasible. Identify promotion opportunities. Develop flexible work schedules. Implement flexibility in work assignments.
1091	Injury to Workers/Overall Safety of the Division	Low	Promote safety culture at all levels of the division. Follow Argonne Integrated Safety Management Plan. Monitor work areas for potential safety concerns. Enforce use of personal protective equipment.

Conclusion

ALCF uses a proven risk management strategy that is documented in its RMP. This plan is regularly 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 a part of the ALCF's culture and applies equally to all staff, from senior management to summer students. A formal risk assessment is performed for every major activity within ALCF, with informal assessments used for smaller activities. Risks are monitored and tracked using a secure and shared cloud-based storage system, along with risk forms and a risk register that are both formatted using Excel. Beyond these activities, many tools are used to manage risks at ALCF, particularly in the area of safety. ALCF's effective risk management plan has contributed to the successful management of all significant risks encountered in 2021.

This page intentionally left blank.

Section 6. Environment, Safety, and Health

(a) Does the Facility exhibit a culture of continual improvement in Environment, Safety, and Health (ES&H) practices to benefit staff, users, the public, and the environment?

(b) Has the Facility implemented appropriate Environment, Safety, and Health measures?

ALCF Response

ALCF is a leader in health and safety at Argonne. The facility itself has never experienced a lost-time incident. In 2021, ALCF experienced zero recordable injuries, zero first aid incidents, and a single contractor near-miss event occurring August 10, 2021, during the delivery and installation of Polaris.

All major ALCF installations follow the work planning and control (WPC) manual, which help ensure that work is performed according to Integrated Safety Management (ISM) guiding principles and various approaches from DOE-HDBK-1211-2014: *Activity-Level Work Planning and Control Implementation*, which contain the Site Office safety metrics. ALCF uses the Argonne AWARE software application for work planning and control. In addition, ALCF uses laboratory-employed subject matter experts (SMEs) to analyze hazards when needed, and also uses control measures to help ensure that work is performed while mitigating risks to the safety, environment, mission, security, and health of the users, public, and workers. In 2021, all ALCF work control documents (WCDs) were reviewed by the work planners and updated, as was the WPC manual, per the 36-month review cycle. ALCF uses a Job Safety Analysis (JSA) to plan for work performed by contractors, as required by ALCF's WPC manual and detailed in the laboratory-wide procedure *Contractor Safety*. All JSAs, including those used in 2021, are retained in the Laboratory's official cloud resource, Box.

The contractor near-miss incident occurred in ALCF's data center on August 10. Immediately following the incident, ALCF paused work and held a safety meeting. An account of the near-miss incident is as follows: One of two subcontractors using a 4 ft. x 8 ft. aluminum sheet as floor tile protection during rack installations inadvertently stepped in a 7 inch x 11 inch rear-door heat exchanger (RDHX) cutout opening in the floor. These RDHX cutouts had been identified as hazards in the pre-job briefing, pointed out in the walkthrough, and were tagged by safety marker plugs. The subcontractors were using floor tile pullers to move the aluminum sheeting horizontally between locations. Due to the orientation of the sheeting, a view of the floor was momentarily obstructed for one of the workers, who then stepped into a cutout. The worker was uninjured. The safety pause was enforced, and a corrective action plan was developed that included implementing safety marker plugs for future cut tiles when the tiles cannot be replaced with whole tiles.

As part of ALCF's commitment to safety and continuous improvement, a lesson-learned document was developed with the assistance of the laboratory's incident investigations SMEs. The lesson-learned document included a causal analysis that helped develop the action plan and follow-up, all of which drives the ISM process and will be used to improve future projects. A quad chart was developed and shared with the lab to share best practices and prevent accidents through a process of continuous organizational learning.

ALCF performs formal management assessments, such as the 2021 Working Alone Assessment, to seek continuous improvement within the division. Based on the findings from the Working Alone Assessment, changes to the division WPC manual were made to include a documented process for notifications of working alone/off hours and an escalation procedure for a non-response by a worker or supervisor. Additionally, bi-annual health and safety inspections conducted in 2021 with COVID controls by the ALCF's Division Director and Environment, Safety, and Health (ESH) Coordinator continue to function as a great tool to observe work, seek worker feedback, and identify areas for improvement.

Worker feedback is also captured at all ALCF operations (Ops) meetings. The bi-weekly Ops meeting begins with a safety discussion and safety share. All feedback is used to improve operations, for example, ways to improve communication in the data center. ALCF Ops implemented the use of hearing protection with integrated hands-free two-way communication based on worker feedback. The Data Center Facility Manager also attends these meetings to help plan and communicate any upcoming work or projects.

ALCF supports a strong safety culture at Argonne and implements appropriate ESH measures by encouraging and using pause work protocols to help ensure that all non-normal events, no matter the impact, are reviewed and integrated into the ISM process. ALCF shares best practices through lessons-learned, seeks feedback, and utilizes assistance of lab SMEs to benefit staff, users, the public, and the environment.

Section 7. Security

(a) Does the Facility exhibit a culture of continual improvement in cyber security practices?

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

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

ALCF Response

ALCF works to continually improve its cyber security practices by developing and maintaining relationships between facility personnel and ALCF's internal security personnel, and between ALCF and the Argonne Cyber Security Program Office (CSPO). ALCF's efforts are also evident in the commitment to improving relationships with sister laboratories, auditing new software deployed to ALCF's environment, staying up to date with emerging threats, and maintaining foresight of future facility needs as they relate to security controls. This posture has led to the ALCF having zero security incidents this year and allows ALCF to keep making forward progress in reducing risk in its environment. The ALCF has a valid Authority to Operate (ATO) and an effective process for compliance with export controls and foreign visitor access.

7.1 Continual Improvement in Cyber Security Practices

ALCF continuously works to maintain a strong cyber security posture and is vigilant in its efforts to stay informed about active and evolving threats, vulnerabilities, and potential exploitation. ALCF's security team partners with the OLCF, NERSC, and other facilities to quickly assess the risks of security incidents unique to HPC facilities. Working together, the facilities can quickly identify, coordinate, and verify mitigations for new vulnerabilities on HPC systems.

ALCF is also vigilant in making certain that the software used in its environment is audited and tested for security. ALCF has worked with the Exascale Computing Project's Continuous Integration team to help perform security audits and analysis on their Gitlab Runner Custom Executor that is being deployed at ALCF. By having the security team involved early on in assessing the design and implementation, ALCF can help ensure that the development team is aware of the security requirements at the site and that there are no surprises when it is time to transition to a production service.

Additionally, ALCF's security staff looks toward the future in determining the needs of the facility. The security team works to break down potential procedural barriers to enable ground-breaking science and improved workflows at the facility. In 2021, it was identified that the facility might want to leverage Globus's authentication framework for workflows. ALCF security staff worked with Argonne's CSPO to review the framework and agreed that the backend Globus infrastructure met requirements and that any future implementation using Globus's authentication framework would be reviewed and approved by CSPO and ALCF security. Having this established relationship and common understanding will allow significantly faster approvals of workflows as they are determined to be production ready, and when ALCF wants to explore making them available to users.

In CY 2021, there were **zero** cyber security incidents on ALCF-managed systems. ALCF's cyber security personnel take a proactive approach to problem management. Examples of proactive measures taken by the security personnel include efforts to:

1. Conduct privileged access reviews across the environment to help ensure that everyone has the appropriate level of access.
2. Conduct reviews of how and where data are stored to help ensure that data are accessible only to those with proper authorization.
3. Educate users and staff about how to prevent password exposure.
4. Educate developers on secure coding best practices via internal discussions/reviews and external training courses.
5. Integrate security auditing into developer workflows to identify security issues early in the development life cycle.
6. Work to keep the National Institute of Standards and Technology (NIST) Certification package up to date, including NIST 800-53, 800-34, 800-30, and 800-18 compliance documents.
7. Archive and delete obsolete data.
8. Set password rotation policies for ALCF systems and verify compliance.
9. Monitor new vulnerabilities to ALCF systems.
10. Conduct penetration testing of both internal- and external-facing web applications and recommend security improvements.

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

1. Cross-laboratory data sharing that CSPO integrates into automatic blocking scripts to keep systems secure 24/7/365.
2. Log ingestion capabilities that allow CSPO to block IP addresses or bad actors in near real time. This provides ALCF additional protection from attacks on Argonne networks as they are being detected.
3. CSPO provides Cloudflare, a powerful web application firewall (WAF). This WAF protects our publicly available sites from known attacks and ensures encryption requirements are met.
4. CSPO provides a Tenable security center (Tenable.sc) instance that we can tie our Nessus scanners into and allows us to leverage CSPO's public and internal scanners. This provides a holistic view of network vulnerabilities in our environment.
5. A laboratory-managed network border firewall protects ALCF applications from the public Internet and other networks within Argonne.

6. CSPO helps to manage publicly visible vulnerabilities by automatically alerting ALCF when new vulnerabilities are detected and provides guidance for patching the system or removing public conduits.
7. CSPO helps with cyber security policy by managing the Cyber Security Program Plan and allows ALCF to make documented alternative security implementations to suit ALCF's needs.

Some of ALCF proactive measures revealed security vulnerabilities that were promptly addressed and fixed, usually within days of their discovery. Immediately upon detection, ALCF staff also investigated all relevant logs to determine whether the security vulnerability had been exploited. In CY 2021, none of the issues that were investigated were found to have been exploited. Examples of the security issues that were detected, and their ensuing mitigations, are as follows:

1. **Issue:** The CSPO identified new security issues with public-facing ALCF services.
   Mitigation: ALCF staff reviewed the CSPO information and worked to address the issues in a timely manner.

2. **Issue:** Passwords in some applications were found to be stored insecurely.
   Mitigation: Evaluation of the application data integrity showed no unauthorized access. Passwords were rotated as a precaution against potential exploit.

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

1. Retiring obsolete services and data.

2. Verifying that strong encryption is used everywhere in the environment and that plain text protocols are not used for production systems.

3. Improving real-time log analysis techniques.

CSPO conducts an annual internal audit called a Division Site Assist Visit (DSAV) to assess divisional compliance with NIST-800-53 controls. Each DSAV covers roughly one-third of the controls. In 2021, CSPO's DSAV with ALCF resulted in ALCF passing 38 of the 38 assessed NIST controls. CSPO also identified opportunities for improvement in 11 areas that ALCF is working to assess. These opportunities for improvement include the following:

1. Integrating ALCF accounts with Argonne Domain accounts to better allow Argonne to enforce domain account policies around account deactivation.

2. Making improvements in network segmentation on ALCF internal networks.

3. Leveraging logging for operational and security analysis instead of designating them as primarily for archiving purposes.

4. Improving user approval workflows with security in mind.

CSPO has committed to continuing the DSAV process in 2022. ALCF will continue to work with CSPO to verify that all Argonne security standards and practices are met.

7.2 Cyber Security Plan

Argonne's Authority To Operate (ATO) includes the ALCF as a major application and was granted on January 22, 2018. It is valid as long as Argonne maintains robust, continuous monitoring of its Cyber Security Program as detailed in the ATO letter, which is included at the end of this section.

7.3 Foreign Visitor Access and Export Controls

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

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

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

The ANL-593 form records the type of work the user will be performing, including the sensitivity of the data used and generated. The ANL-593 must be approved by Argonne Cyber Security, FV&A, the Argonne Office of Counterintelligence, and the Argonne Export Control Office. Argonne's foreign visitor and assignments process integrates with the DOE Foreign Access Central Tracking System (FACTS), which documents and tracks access control records of international visits, assignments, and employment at DOE facilities and contractor sites. Once the ANL-593 form for the user is approved, UB3 is automatically updated with the user's ANL-593 start and end dates. The ALCF Accounts team then creates the user account and notifies the user. Any changes to the ANL-593 dates are automatically updated in UB3. Accounts are suspended if the user does not have an active ANL-593.

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

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

JAN 2 2 2018

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

Dear Dr. Kearns:

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

Reference: Letter, J. Livengood to P. Littlewood, dated November 21, 2016, Subject: Authority to Operate for the Argonne National Laboratory Information Technology Infrastructure

Over the past year, the Laboratory has conducted regular continuous monitoring briefings and has kept me informed of changes in cyber security risk in accordance with the Risk Management Framework. The Laboratory has revised system security documentation to incorporate NIST SP800-53 Revision 4 security controls and has been testing at least 60 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 (ATO) for the General Computing – Low enclave and the General Computing – Moderate enclave. The IT Infrastructure continues to contain the following sub-component major applications, which have components in both enclaves:

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

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

A component of the Office of Science

If I can be of any assistance, please contact me or have your staff contact Francis Healy at (630) 252-2827 or e-mail frank.healy@science.doe.gov.

Sincerely,

Joanna M. Livengood

Joanna M. Livengood
Manager

cc: S. Hannay, ANL-BIS
M. Skwarek, ANL-BIS
M. Kwiatkowski, ANL-BIS
B. Helland, SC-21
N. Masincupp, SC-OR
F. Healy, SC-CH

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

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

(b) Is the Facility collaborating with technology vendors and /or advancing research that will impact next generation [high performance computing platforms]/[high performance networking]?

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

ALCF Response

The science accomplishments of INCITE, ALCC, and DD projects clearly demonstrate ALCF's impact in supporting scientific breakthroughs. ALCF staff members have worked effectively with individual project teams to adapt their simulation codes to run efficiently in an HPC environment and have enabled scientific achievements that would not have been possible otherwise.

In this section, ALCF reports:

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

8.1 Science Highlights and Accomplishments

ALCF employs various methods and processes for monitoring its science accomplishments. Monthly scientific highlights mostly originate from the catalyst team and are based on an outcome documented in a quarterly report. The determination and coordination of scientific highlights is managed by ALCF's applications team, made up of members of both the catalyst team and the performance engineering team, and in consultation with ALCF's director of science. Other sources of scientific highlights include technical communications between ALCF staff members and a project PI or co-PI; significant findings reported in a high-impact publication or conference presentation; and a catalyst's own involvement in a publication.

ALCF annually tracks and reports the number of peer-reviewed publications resulting (in whole or in part) from use of the facility's resources. For the LCFs, tracking takes place for a period of five years following the project's use of the facility. This number may include publications in press or accepted but does not include papers submitted or in preparation. This is a reported number, not a metric. In addition, the facility may report other publications where appropriate. Methods used for gathering publication data include asking users to verify or update ALCF's online publications database and conducting Google Scholar and Crossref searches. ALCF also collects approximately one-third of its users' ORCID iDs in any given year and has been investigating ways to use this method to identify more user publication data.

Table 8.1 shows the breakdown of refereed publications based, in whole or in part, on the use of ALCF resources, and highlights those appearing in major journals and proceedings. These include one publication in *Nature*, one in *Nature Astronomy*, four in *Nature Physics*, six in *Nature Communications*, one in *Nature Machine Intelligence*, one in *Scientific Reports*, and one in *Scientific Data* (combined in the *Nature* journals category in the table below); two in *Science Advances* and one in *Science Direct* (listed under the *Science* journals category in the table below); three in the *Proceedings of the National Academy of Sciences (PNAS)*; seven in *Physical Review Letters*; and one in the proceedings of the *2020 International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*. Table 8.2 shows updated publication counts from prior years and are based on new information received after the prior year's OAR deadline.

In addition to publications, ALCF projects earned multiple awards in 2021, including HPCwire awards. The HPCwire Readers' Choice Award for Best Use of HPC in Physical Sciences was given to ALCF staff member Eliu Huerta and other Argonne researchers for developing the first class of AI models that outperform humans in classifying thousands of compact star clusters observed by the Hubble Space Telescope. The HPCwire Readers' Choice Award for Best Use of HPC in Industry was given to Argonne in recognition of its collaboration with Aramco Americas and Convergent Science for helping evaluate cold operations for propulsion systems. Researchers used ALCF's Theta to model and resolve uncertainties at cold conditions to better understand where pollutants originate.

Table 8.1 Summary of Users' Peer-Reviewed Publications in CY 2021

Nature Journals	Science	PNAS	Physical Review Letters	SC	Total 2021 Publications
15	3	3	7	1	249

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

OAR Year	CY 2017	CY 2018	CY 2019	CY 2020	CY 2021
Total Publications	225	276	288	257	249

Science Highlights

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

Each project highlight includes a figure and a bar graph showing time allocated and time used: the first number in the graph title is the allocation total and the second (in parentheses) shows what the project actually 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 High-Fidelity Gyrokinetic Simulation of Tokamak and ITER Edge Physics

The Science

ITER is the world’s largest magnetic fusion experiment, currently under construction in France. To prove the feasibility of fusion as a virtually inexhaustible source of carbon-free energy for generating electricity, successful operation must avoid damage to the material surface of the divertor plates from deposition of exhaust plasma heat loads that are excessively narrow in width. A simple extrapolation from present experimental results to ITER predicts a dangerously narrow heat-load width when ITER is producing 10 times more energy than was input. Because ITER is much larger in size and far stronger in magnetic field strength, its heat-load width may not be subject to the same physics as in present tokamaks, yielding large uncertainty in the extrapolation results. The team pursued a first-principles-based prediction for the heat-load width to reduce this dangerous uncertainty.

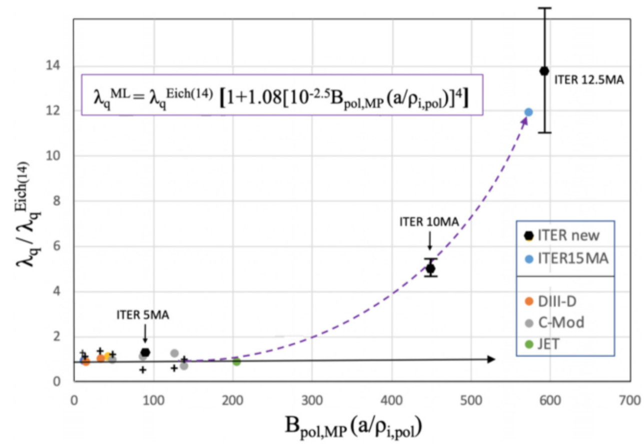
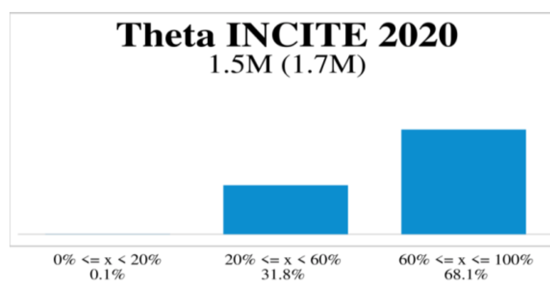


Figure 8.1 A supervised machine-learning program finds the hidden kinetic parameter $a/r_{i,pol}$ and a simple predictive formula for l_q/l_q^{Eich} where l_q is the divertor heat-load width normalized using the Eich regression formula. Three more ITER simulations (black symbols) have been performed to verify the validity of the new ML-found formula (black dots). The 10MA case was obtained on Theta utilizing 4,352 nodes.

Image courtesy of C.-S. Chang.

The Impact

If the simple extrapolation from present tokamak experiments to full-power ITER were correct, no known material could withstand the extreme heat-load density, an order-of-magnitude higher than what occurs on the surface of vehicles returning from space. Extraordinary methods would be needed to reduce the heat-load density before the exhaust plasma could reach the divertor plates, limiting ITER science and increasing operation cost. Thus, establishing an accurate predictive formula for the exhaust heat-load width of future tokamak fusion reactors can enable operating ITER in a more comfortable and cost-effective way toward the goal of producing 0.5 GW of fusion power from 50 MW of input power. A more accurate formula can also enable more reliable reactor designs, which are limited by the exhaust heat-load width on the divertor plates.



Summary

The team used the extreme-scale modeling code XGC to solve kinetic equations for the tokamak edge by modeling plasma with a large number of particles. Predictions from XGC for the divertor heat-load width on present tokamaks agreed with experimental data within the experimental error bar, yielding narrower divertor heat-loads than what are hoped for. A simple extrapolation to full-power ITER would thus give a pessimistically narrow divertor heat-load width. However, applying the same XGC code to the full-power ITER produced a divertor heat-load width more than six times wider than the formulas developed from the simple extrapolation. These results suggested the existence of hidden parameters not realized in previous exercises using only observable fluid quantities as input. To find them, the team used a supervised machine learning (ML) program to anchor or direct the extrapolation from the existing tokamak data to future ITER data obtained from the high-fidelity kinetic XGC simulation. The ML program identified a kinetic quantity as the hidden parameter and produced an analytic formula as a simple modification of the existing formula. The new predictive formula has been successfully verified by simulating three more ITER plasmas on Theta: 5MA, 10MA, and 12.5MA cases. The new 10MA data point enabled the team to verify the predictive formula line's correctness, raising the confidence level for it significantly and possibly affecting the ITER operation scenario in an important way.

ALCF Contribution: ALCF staff assisted in planning job submissions and improved job throughput on Theta. Staff also helped in setting up a reservation for a crucial run on the full Theta machine.

Contact

Choong-Seock Chang
Princeton Plasma Physics Laboratory
cschang@pppl.gov

Publication

C. S. Chang, S. Ku, R. Hager, R. M. Churchill, J. Hughes, F. Köchl, A. Loarte, V. Parail, and R. A. Pitts, "Constructing a new predictive scaling formula for ITER's divertor heat-load width informed by a simulation-anchored machine learning," *Phys. Plasma.* (2021). doi: 10.1063/5.0027637

Highlight Categories

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: June 2021

8.1.2 Optimized Structure and Electronic Band Gap of Monolayer GeSe from Quantum Monte Carlo Methods

The Science

Two-dimensional (2D) semiconductor nanomaterials are good candidates for use in electronic and optical sensors given their unique properties and tunability via altering their structures, such as by pulling, pushing, or twisting their geometries. Quantum Monte Carlo (QMC) methods were used to find the structure and electronic band gap of 2D germanium selenide (GeSe), determining that the gap and its nature are indeed highly tunable by strain. The geometry of the material was undetermined experimentally and beyond the reach of standard/common theoretical methods. Using a newly developed computational technique, the detailed atomic structure was determined, from which accurate optical properties of the 2D GeSe nanomaterial were obtained.

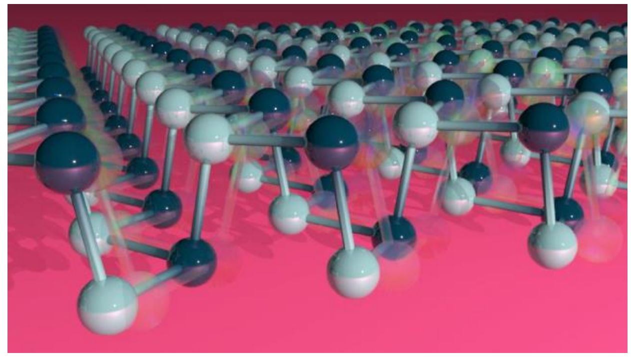
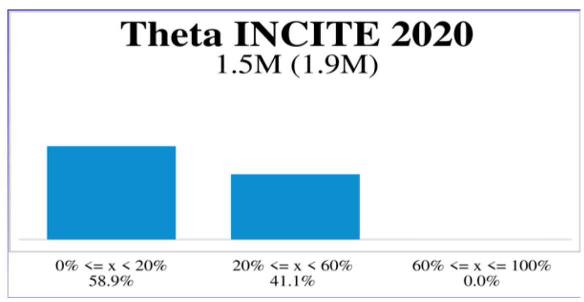


Figure 8.2 Optimized geometry of GeSe monolayer using the newly developed structural optimization algorithm within Quantum Monte Carlo (colored structure) compared to the initial Density Functional Theory-optimized structure (clear structure).

Image courtesy of Janet Knowles, Joseph A. Insley, Silvio Rizzi, and Victor Mateevitsi, Argonne National Laboratory.



The Impact

GeSe is considered a promising material for such devices as solar cells and photodetectors that are used to detect the presence of light. Highly accurate QMC methods were used to obtain the full geometry of a complex 2D nanomaterial for the first time. The newly developed QMC-based algorithm can accurately determine the structure of a material without calculating the atomic forces. The high tunability with strain of the band gaps indicates potential optical applications in this class of materials. This work also clearly demonstrated the need for highly accurate structural and electronic structure methods to reliably assess the properties of these materials for future applications.

Summary

The team used a newly developed computational technique to predict the detailed atomic structure and optical properties of the 2D GeSe nanomaterial. In the bulk, this material forms a 3D structure with layers bonded by van der Waals interactions. Different density functional theory (DFT) results have yielded significantly different geometries and band gaps. The team used highly accurate many-body diffusion Monte Carlo (DMC) methods, after first verifying that DMC yielded the correct experimental structure and electronic properties for bulk GeSe, to optimize the GeSe structure and calculate the charged quasiparticle and neutral excitonic gaps. The team then found the GeSe monolayer's structure using a newly developed optimization method within DMC (Figure 8.2), which offers a much lower computational cost and application to heavier elements.

All of the atomic positions and lattice constants are fully relaxed, marking the first full structural relaxation of a periodic nanostructure with QMC techniques. Further, the DMC electronic properties vary strongly with strain: not only does the magnitude of the band gap change with strain, but a transition from direct to indirect gap can be induced with strain. This result shows that the optical absorption properties of monolayer GeSe are highly tunable. Theta was used primarily for the equation of state for bulk GeSe, bandgap calculations, and geometry optimization of the monolayer (~100 calculations).

ALCF Contribution: ALCF staff and postdocs have made significant contributions to QMCPACK code since the Mira Early Science Project and continue to do so with an eye toward exascale resources. The ALCF catalyst worked closely with the project team and scheduling committee to coordinate job submissions throughout the year and made good use of the overburn policy. An ALCF-funded postdoc also contributed to the work with Configuration Interaction using a Perturbative Selection done Iteratively (CIPSI) calculations.

Contact

Olle Heinonen
Argonne National Laboratory
heinonen@anl.gov

ASCR Allocation PI: Paul R. C. Kent, Oak Ridge National Laboratory

Publication

H. Shin, J. T. Krogel, K. Gasperich, P. R. C. Kent, A. Benali, and O. Heinonen, "Optimized structure and electronic band gap of monolayer GeSe from quantum Monte Carlo Methods," *Phys. Rev. Materials* **5**, 024002 (2021). doi: https://doi.org/10.1103/PhysRevMaterials.5.024002

Highlight Categories

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: June 1, 2021

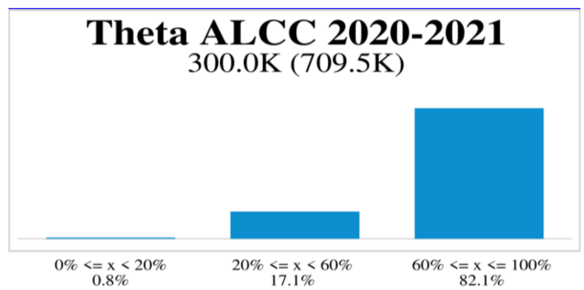
8.1.3 Expanding the Scale of Urban Building Energy Modeling

The Science

The energy consumption of buildings can be modeled and simulated to gain insights into possible energy efficiencies that would help to reduce carbon emissions. However, the scale of the problem is vast. The simulation of a typical urban area involves 10,000s of buildings, and each building model involves approximately 3,000 parameters. Researchers have harnessed supercomputing power using their software suite, Automatic Building Energy Modeling (AutoBEM), to rapidly create building energy models and study them in unprecedented detail, providing utilities and industry with the information they need to improve energy efficiency.

Figure 8.3 A digital twin of Chattanooga, Tennessee, processed with the AutoBEM software. All 178,337 buildings and simulation results can be searched, selected, and visualized using flexible regular expressions at bit.ly/virtual_epb.

Image courtesy of Joshua New.



The Impact

In the United States, residential and commercial buildings account for 40 percent of energy consumption, 73 percent of electrical consumption, and 39 percent of emissions. In 2019 alone, the energy bill for the 125 million buildings in the United States was $412 billion. As the threat of climate change looms, the U.S. Department of Energy has set aggressive goals for energy-efficiency—a 30 percent reduction in energy use intensity (in units of kilowatt-hours per square foot) of all U.S. buildings by 2030 compared to 2010. In their studies, researchers found that 99 percent of buildings in Chattanooga, Tennessee, would benefit from the efficiency technologies evaluated with an average savings of $28,500 per building per year. If implemented, such efficiencies could reduce greenhouse gas emissions and reduce the need for new power plants.

Summary

ORNL researchers developed AutoBEM and have modeled 122.9 million buildings across the U.S. Building models include data on the materials; floors; roof type; number of windows; heating, ventilation, and air conditioning (HVAC) systems; the hourly lighting schedule of each room; and so on. Some building data are publicly available—such as satellite images and street views—while others are obtained from utility companies or tax assessors.

Theta has been key to accelerating the research, with AutoBEM runs typically using more than 80 percent of Theta. The current study simulated 178,337 buildings in Chattanooga (Figure 8.3). Researchers found that each efficiency technology—such as improved HVAC efficiency, space sealing, insulation, and lighting—has the potential to offset 500 to 3,000 lbs. of carbon dioxide (CO_2) per building. In addition, the models can be used to help understand the impact of reducing energy consumption during a utility’s times of peak demand for each month at critically loaded substations. Significant relative demand savings were available in seasonal transition months (March/November) with smart thermostats, while switching from electric to gas water heaters resulted in the greatest savings in April and October.

ALCF Contribution: ALCF staff helped with increased priority to complete science runs as well as to provide support for I/O-bound data handling on ALCF filesystems when processing results for all buildings.

Contact

Joshua New
Oak Ridge National Laboratory
newjr@ornl.gov

Publications

B. Bass, J. R. New, and W. Copeland, “Potential energy, demand, emissions, and cost savings distributions for buildings in a utility’s service area,” *Energies* **14**, 132 (2021). DOI: https://doi.org/10.3390/en14010132

Highlight Categories

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: September 13, 2021

8.1.4 Finding Druggable Sites in SARS-Cov-2 Proteins Using Molecular Dynamics and Machine Learning

The Science

COVID-19 is caused by the novel Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and was declared a pandemic on March 11, 2020 by the World Health Organization. For new drug development, the challenge is accurately and efficiently determining the locations of drug-binding sites on target proteins. Using ML, a molecular biophysics group at Johns Hopkins School of Medicine developed a joint computational and experimental approach that locates known small-molecule binding sites and predicts locations of sites not previously observed in SARS-Cov-2 experimental structures and other proteins.

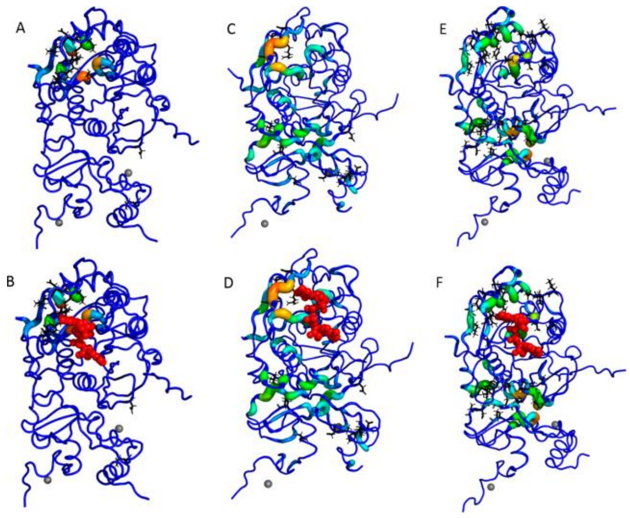
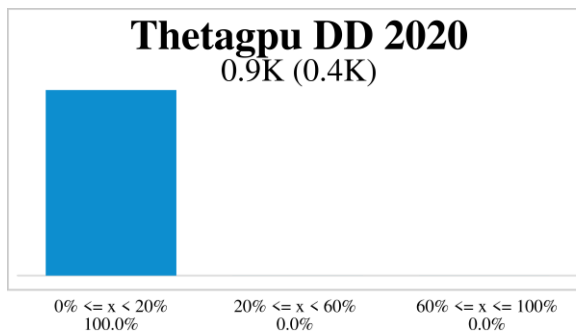


Figure 8.4 Druggable sites of SARS-Cov-2 MTase protein predicted by TACTICS workflow: open pockets (top) and ligand-bound pockets (bottom); residues of drug binding sites and ligands emphasized in bold color.

Image reproduced from Figure 3 of *J. Chem. Inf. Model* **61**, 2897 (2021).

The Impact

General oral medications are expected to present significant usage flexibility and reduced manufacturing, transportation, and storage costs over the same metrics for vaccines for COVID-19 control. A key step toward rational drug design is development of a high-fidelity, high-resolution, all-atom simulation and modeling methodology that can predict all drug binding sites as well as their local conformations. The team's TACTICS workflow is capable of detecting "cryptic" binding sites that are difficult to detect without a binding ligand and opens the door for identifying potential druggable sites. TACTICS does not require specialized simulations, can analyze both single structures and pre-computed trajectories, is self-contained within a single software package, and is freely available to the wider research community for finding druggable sites in other proteins.



Summary

The REST2 (Replica Exchange Solute Tempering – 2nd generation) is a powerful sampling enhancement algorithm that accelerates infrequent conformational transitions of macromolecules by augmenting interaction energy fluctuations of a simulated system. REST2 simulations were critical in this project to search all important conformational transitions that require timescales beyond general simulation methodologies. The MD trajectory generated with non-augmented

interactions was used to analyze conformations that possess significant weights by an ML method. Individual calculations typically ran on 144, 288, or 432 nodes of Theta for 12 hours.

The team then applied TACTICS to explore the infrequent transitions among important conformations. The ML model of TACTICS, trained with a large database of drug-binding crystal structures, successfully identified druggable sites observed by previous experiments and refining the local residues and conformations (Figure 8.4). TACTICS also predicted several additional druggable sites and characterized their drug bindability for later drug screening. The drugging of these binding sites may prevent engineering replication of the SARS-CoV-2 virus.

ALCF Contribution: ALCF staff helped build the NAMD code for Theta, determining optimal usage of the resources by coordinating production job runs, setting optimal parameters for the REST2 simulation algorithm, and managing the large amount of generated simulation data.

Contact

Albert Lau
Johns Hopkins University School of Medicine
alau@jhmi.edu

Publication

D. J. Evans, R. A. Yovanno, S. Rahman, D. W. Cao, M. Q. Beckett, M. H. Patel, A. F. Bandak, and A. Y. Lau, "Finding druggable sites in proteins using TACTICS," *J. Chem. Inf. Model* **61**, 2897 (2021). DOI: https://doi.org/10.1021/acs.jcim.1c00204

Highlight Categories

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: August 16, 2021

8.1.5 Hadronic Light-by-Light Scattering and Vacuum Polarization Contributions to the Muon Anomalous Magnetic Moment from Lattice QCD with Chiral Fermions

The Science

One of the current top priorities in high energy physics research is to understand the discrepancy between the theoretical prediction and experimental results for the magnetic moment of the muon (a heavier relative of the electron). A recent experiment at Fermilab has confirmed a 20-year-old result obtained at Brookhaven National Laboratory (BNL), increasing confidence that a discrepancy exists. As the measurement is refined, errors from the theoretical calculation must be understood and reduced to determine whether or not the difference is due to new physics. The largest uncertainty in the theoretical calculation comes from particles that interact through the strong force, quantum chromodynamics (QCD), known as hadronic contributions. This project aimed to quantify and reduce the largest uncertainties associated with this value to obtain the most precise calculation of the anomaly.

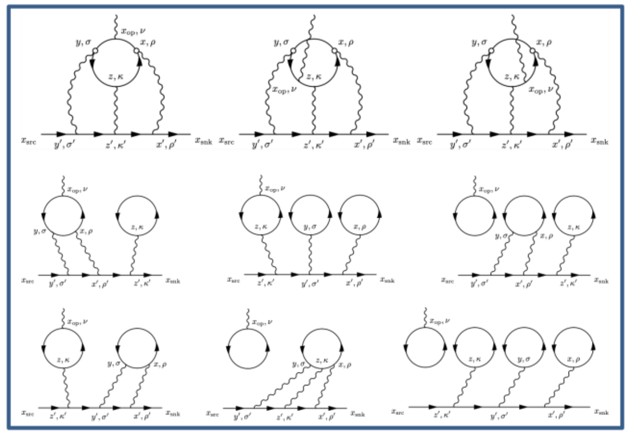
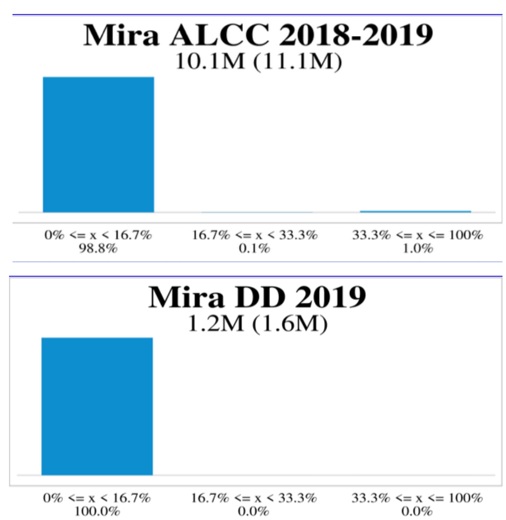


Figure 8.5 Hadronic light-by-light scattering contributions to the muon anomaly. In all cases, the muon (horizontal line at bottom) interacts with hadrons (quark loops shown as circles) through emission of virtual photons (wavy lines from muon). The diagrams represent different interaction possibilities with varying numbers of quark loops. The quark loops interact with each other through the exchange of gluons (not shown).

Image courtesy of Tom Blum, University of Connecticut.

The Impact

Making extensive use of Mira until its decommissioning in early 2020, the team produced the first theoretical result for the hadronic light-by-light scattering contribution to the muon anomalous magnetic moment, with all errors systematically controlled. The result is consistent with previous model results and essentially rules out the uncertainty in the hadronic light-by-light scattering as being a source of the discrepancy. This result shifts the theoretical focus to the hadronic vacuum polarization contribution as the last remaining source of theoretical uncertainty that could account for the discrepancy without requiring new physics.



Summary

The muon's magnetic moment describes how this fundamental particle interacts with a magnetic field; it depends on all particles that can couple to the muon—including as-yet-undiscovered particles. The muon moment has been measured in experiments and calculated theoretically, but those two values do not quite match up. In April 2021, the Fermilab g-2 experiment released its initial result, agreeing with BNL's previous experimental result and reinforcing the disagreement between theory and experiment. The disagreement is at 4.2 sigma (standard deviations)—just below the 5-sigma threshold typically used to signal the discovery of new physics.

The largest uncertainty in the theoretical calculation comes from hadronic light-by-light scattering and hadronic vacuum polarization. A major goal of this effort was to produce first-principles calculations of these hadronic contributions that do not rely on any model or experimental input. In its *Physical Review Letters* paper, the team presented the first theoretical result for the hadronic light-by-light scattering contribution; this result has helped rule out its contribution as a significant source of the discrepancy, shifting the focus to the hadronic vacuum polarization contribution for which the leading result still relies on experimental input.

ALCF Contribution: ALCF staff assisted with optimizing and debugging code on Blue Gene/Q on many occasions.

Contact

Tom Blum
University of Connecticut
thomas.blum@uconn.edu

Publication

T. Blum, N. Christ, M. Hayakawa, T. Izubuchi, L. Jin, C. Jung, and C. Lehner, "Hadronic light-by-light scattering contribution to the muon anomalous magnetic moment from lattice QCD," *Phys. Rev. Lett.* **124**, no.13, 132002 (2020).

Highlight Categories

Performer/Facility: ASCR-ALCF

Date Submitted to ASCR: December 2021

8.2 Research Activities / Vendor Engagements for Future Operations

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

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

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

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

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

- **Arcticus:** Intel Development Chassis with XeHP GPU and Future Intel Xeon CPUs (Intel development GPU card [code name XeHP]). A similar cluster (called DevEP) at Intel was also made available to JLSE users.
- **Yarrow:** Supermicro SYS-1029GQ-TNRT with Intel development GPUs (code name DG1) and Xeon Gold 6226R CPUs (Intel's first discrete GPU). Yarrow was decommissioned at the end of 2021.
- **Iris:** SuperMicro X11SSH-GF-1585 Server Motherboard with Intel Xeon E3-1585 v5 CPU and Iris Pro Graphics P580 GPU (Intel integrated Gen9 GPUs).
- **Intel Pre-production Development Platform** includes 2x Next Gen Intel Xeon Scalable processor.
- **Presque:** Intel DAOS nodes (DCPMM and NVMe storage) with Intel DAOS file system.

Other JLSE testbeds include:

- Intel Xeon Phi Knights Landing (KNL) Cluster
- NVIDIA GPUs:
    - DGX–1 (V100 GPUs)
    - Gigabyte NVIDIA A100 and A40 cluster
    - Supermicro NVIDIA V100 and P100 cluster
- AMD GPUs:
    - AMD GPU MI100 and MI50 cluster
- Intel Xeon Clusters: Skylake, Cascade Lake, and Cooper Lake
- ARM resources:
    - HPE Apollo 70 – Comanche Prototype ARM64 Cluster
    - HPE Apollo 80 – 8 node Fujitsu A64X CPU
    - NVIDIA ARM Dev Kit – Ampere Altra Q80-30 ARM CPU, NVIDI A100 GPU
- IBM Power System AC922 (Power9 CPU, V100 GPU)
- Supermicro NVIDIA V100 and P100 cluster
- Atos Quantum Learning Machine

In 2021, the JLSE supported more than 400 users spanning more than 80 projects. These projects ranged from application portability to software development to tools and compiler development for an ESP project. Teams from within the ECP's Application Development and Software Technology groups have been using the JLSE Aurora testbeds and the Aurora SDK to develop applications and software for Aurora. The following summaries represent a sampling of current JLSE projects:

- **ALCF Data Science Program:** Application teams from the ALCF Data Science Program use JLSE resources to explore and improve data science techniques, such as data mining, graph analytics, machine learning, and complex and interactive workflows.

- **ALCF Early Science:** ESP Application teams use JLSE resources to prepare and optimize applications for the next-generation supercomputers in advance of the systems becoming available. For example, researchers from the Aurora ESP projects access the Xeon Skylake Iris nodes with Intel's integrated GPUs and the early versions of OneAPI software to develop and test their applications for Aurora.

- **Argo:** Argo is an exascale operating system and runtime system designed to support extreme-scale scientific computation. Researchers from the Argo project used JLSE resources to prototype the GlobalOS distributed resource management and to evaluate the performance of NodeOS. They also used JLSE resources to develop and optimize a lightweight, low-level threading and task framework for OpenMP and other programming models (e.g., Cilk, Quark, Charm++).

- **Big Data:** Researchers are using JLSE testbeds to study the layering of HPC programming models beneath big data programming models. Specifically, they are researching the development of a software environment with a Spark user interface (Java and Scala) that can run on a supercomputer, cluster, or a cloud with a back end for executing data-intensive communication patterns.

- **CANDLE:** Using the NVIDIA DGX-1 system and other JLSE computing resources, researchers are developing the CANcer Distributed Learning Environment (CANDLE), a computational framework designed to facilitate breakthroughs in the fight against cancer.

- **Deep Learning:** Multiple projects are using JLSE systems to investigate the potential of deep learning. One research team is focused on understanding how deep learning can be used to improve lossy compression of scientific data from simulations and instruments. Another effort is exploring the performance of different machine learning frameworks that have implemented deep learning and neural networks on KNL systems.

- **LLVM:** Researchers used the JLSE's IBM power systems to advance LLVM compiler development. The instruction set architecture for these systems is the same as for the IBM Blue Gene/Q system, with the only difference being in vectorization. LLVM and Clang builds were carried out on the Intel Xeon Phi systems for quality assurance (QA) purposes. Researchers can complete these builds in 10 minutes using JLSE resources (compared to hours on a laptop).

- **MPI:** Several MPI Chameleon (MPICH) improvements were tested on JLSE systems, including the memory scalability of MPI communicators by exploiting regular patterns in rack-address mapping, enhanced threading support through locking optimizations, and communication-aware thread scheduling.

- **Quantum Computing:** A research team is using the JLSE's Atos Quantum Learning Machine and other resources to develop an artificial neural network for spectral analysis called Spectranne. This tool will automate the analysis of vast amounts of data being produced by state-of-the art, chirped-pulse spectroscopy experiments.

Portable Programming Models

ALCF continues to push portable programming models forward across various ecosystems to allow greater application portability across all the leading DOE systems. ALCF is leading an ECP-funded project, HIPLZ, to provide a "Level 0" backend which allows codes written with the HIP programming model to be used on Intel GPUs. In collaboration with NERSC and OLCF, ALCF is working on two projects to enable DPC++ (Intel's open source SYCL implementation) to run on NVIDIA A100 GPUs and AMD's MI-100 GPUs. Additionally, ALCF staff are contributing backend implementations to Kokkos, RAJA, and OCCA to support Intel GPUs.

A critical element to the success of portable programming models will be the ability for different models to interoperate. ALCF has spearheaded an effort with Intel to ensure the ability of open standard SYCL and OpenMP to interoperate with each other so that a library implementor who uses SYCL can confidently provide it to an application developer who chose to use OpenMP and know that the two will work together.

Early Hardware and Testing Framework

In collaboration with Intel, ALCF staff have deployed multiple testbeds with Intel pre-production hardware to allow ECP and ESP teams to develop and test their codes for the Intel hardware and software that will be used on Aurora. The ALCF has developed a method to quickly deploy updated releases from Intel in a streamlined manner that allows users to easily pick up the latest releases. In addition, the ALCF has developed a test set and automated testing framework to help improve and ensure the quality of the new Intel software stack for Aurora. The test set has continually expanded, with ALCF staff adding tests to it based on issues found when using the Intel software to prepare their code for Aurora. The testing framework is run nightly and allows the ALCF to monitor and provide rapid feedback to Intel. The test set currently contains 540 tests, for which the pass, fail, and timeout rates are tracked.

8.2.2 Research Activity - ALCF AI Testbed

With an eye toward the future of scientific computing, ALCF is building a testbed of advanced AI platforms for the research community. This testbed will enable the facility and its user community to help define the role of AI accelerators in next-generation scientific machine learning. The testbed's innovative AI platforms will complement Argonne's next-generation 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.

Currently open to Argonne researchers, the ALCF expects to open the doors to the broader research community in early 2022. Active users today include application teams from nuclear physics, materials science, biosciences, astrophysics, cosmology, imaging sciences, and precision medicine. Two highlights in the past year include scaling a deep learning (DL) application to understand protein-protein interactions for COVID-19 by coupling the ThetaGPU system to the Cerebras system (PASC 2021), and evaluation of the SambaNova system for applications, including neutrino physics (IEEE CS&E).

The ALCF AI Testbed includes the following systems:

- The **Cerebras** CS-2 is a wafer-scale deep learning accelerator comprising 850,000 processing cores, each providing 48KB of dedicated SRAM memory for an on-chip total of 40 GB and interconnected to optimize bandwidth and latency. 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 SambaNova is a half-rack system consisting of two 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 standard machine learning frameworks, including TensorFlow and PyTorch.

- The **Graphcore** Colossus, designed to provide state-of-the-art performance for training and inference workloads, consists of 1,216 IPU tiles, each of which has an independent core and tightly coupled memory. The Dell DSS8440, the first Graphcore IPU server, features 8 dual-IPU C2 PCIe cards, all connected with IPU-Link technology in an industry standard 4U server for AI training and inference workloads. The server has two sockets, each with 20 cores and 768GB of memory
- 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 able to achieve 250 TFlops in FP16 and 1 PetaOp/s in INT8 performance. The Groq accelerators are PCIe gen4-based, and multiple accelerators on a single node can be interconnected via a proprietary chip-to-chip interconnect to enable larger models and data parallelism.

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 of the vendors, and conducts in-person training and hackathon events.

Common Software Environment: The AI Testbed software environment is based on an Ubuntu Linux distribution. Support for DL frameworks (such as TensorFlow and PyTorch), compilers, and the Python ecosystem, among others, is provided. Each system typically supports a custom DL software environment, such as Poplar for Graphcore and SambaFlow for SambaNova, and ALCF installs and supports these. The environment also provides support for Singularity-based containers and for executing Jupyter notebooks.

8.2.3 Vendor Engagement – Next Silicon Ltd. (NextSilicon)

Next Silicon is a company that is in stealth mode developing a new processor. ALCF, in collaboration with OLCF and PNNL, is working with the company to benchmark the science workloads of interest to DOE on their simulators and is providing design feedback.

8.2.4 Vendor Engagement – Codeplay Software Ltd. (Codeplay)

ALCF collaborated closely with Codeplay Software Ltd., OLCF, and NERSC in supporting the SYCL 2020 programming model on AMD and NVIDIA GPUs. Codeplay creates software based on open programming standards so that application developers can program complex processors using familiar standards and tools. It is the first company to deliver a product that is SYCL conformant, giving developers the most mature, reliable standards-based development platform on the market.

8.2.5 Vendor Engagement – Altair and OpenPBS

ALCF collaborates closely with Altair and the OpenPBS community on scheduler developments as part of ALCF's transition to using PBS as a production resource scheduler. The goal of this collaboration is to ensure that OpenPBS has sufficient support to meet ALCF's needs and has capabilities for scheduling increasingly complex workflows.

8.3 DOE Program Engagements / Requirements Gathering

To help ensure that the ALCF delivers on its mission of delivering breakthrough science, staff need to closely engage with domain science and keep a close eye on directions of supercomputing technologies. 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 in the future.

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 community and domain activities, including conferences, workshops, reviews, and meetings. In 2021, staff participated in more than 105 events. Figure 8.6 breaks down these events by both type and community. Staff members support DOE mission needs by serving on review committees and advisory boards and by participating and organizing DOE and broader community workshops. ALCF staff are regular participants in DOE and National Science Foundation (NSF) workshops and reviews. Staff are engaged in standards committees and boards for both future and current software and hardware technologies.

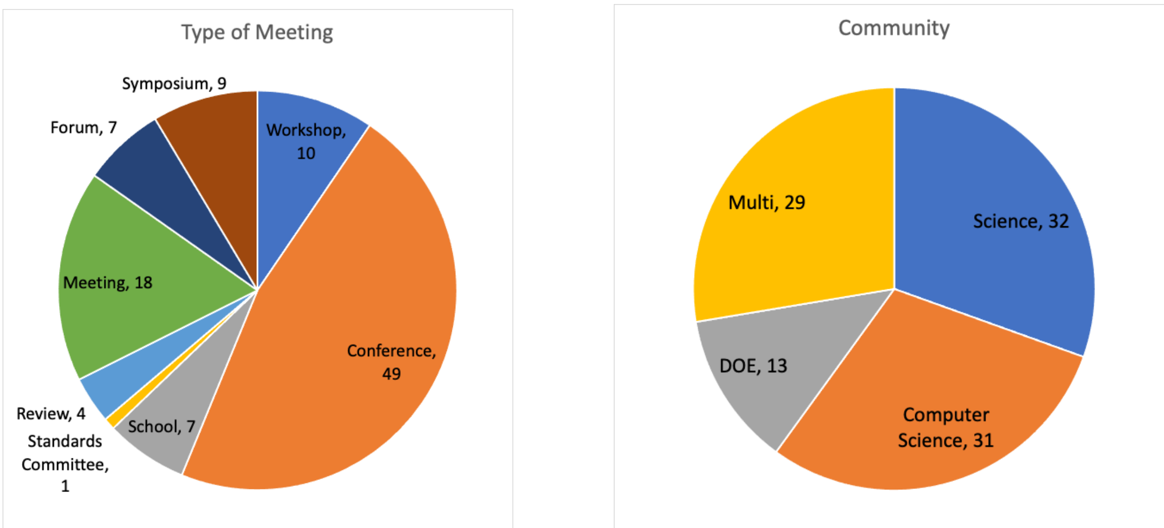


Figure 8.6 Breakdown of some key activities by ALCF in CY 2021. The first pie chart (left) breaks down the 105 events by type, primarily derived by how the event identified itself. The second pie chart (right) breaks down the same 105 events by the community being engaged. "Multi" refers to events such as Supercomputing (SC), where multiple communities are engaged.

Not only do these activities maintain expertise of the staff, but they show the respect that ALCF staff have in the community.

8.3.1 Engagement Highlights

Supercomputing 2021

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

Table 8.3 Summary of ALCF Participation in SC21

Program	Total
Tutorials	2
Workshops	14
Papers	1
Posters	1
Birds of a Feather	5
Booth Talks	1

Engagement in Standards and Community Groups

The ALCF participates in several HPC standards and community groups to promote ALCF interests, educate the community about ALCF resources, and increase collaboration with ALCF staff. These include the following: HPC User Forum; Cray User Group; Intel Extreme Performance Users Group (IXPUG); C++ Standards Committee; OpenMP Architectural Review Board; Khronos OpenCL and SYCL working groups; OpenSFS; and MPI Forum.

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

P3HPC events are more than User Training. Originally focused on sharing definitions, best practices, and ideas for performance portability across DOE, these events have grown into broader events open to the public. The impact of these events has grown as the landscape of HPC platforms has made performance portability an important target for supercomputing applications. In 2021, ALCF co-organized the SC21 P3HPC workshop as a key opportunity for the HPC community to discuss progress in performance portability.

Summer Student Research Programs

As mentioned in Community Outreach (Section 1.3.3.3), every summer, ALCF staff members mentor undergraduate and graduate students on real-world research projects through DOE's Science Undergraduate Laboratory Internship (SULI) program and Argonne's Research Aide Appointments. In 2021, 24 students worked on projects ranging from system administration and data analytics to computational science and performance engineering. ALCF engagements with these students, other student activities and STEM outreach, and ALCF postdocs are a valuable part of finding the next generation of scientists.

8.3.2 Summary of Engagements with the Exascale Computing Project

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

In 2021, more than 40 ALCF staff attended the Virtual ECP Annual Meeting held April 12–15, 2021, to participate in technical conversations, project discussions, and facility-specific breakouts. In addition, ALCF participated in several planning meetings with ECP and the other computing facilities (NERSC, OLCF) to augment and execute the ECP/Facilities engagement plan and worked with ECP's Training Lead to promote ECP training activities to ALCF users.

ECP-Funded Positions in ALCF

The ALCF's ECP/HI Applications Integration effort made great strides in 2021, continuing the team's work in porting and testing ECP applications across many GPUs, including Intel GPUs. There are 26 staff members funded at various levels to work with ECP Application Development and Software Technology projects. One staff member is in place to focus on planning training and Intel's Center of Excellence (COE) for Aurora is staffed with seven people. Additionally, five staff members have been funded to develop and deploy ECP continuous integration (CI) capabilities, support software technologies, and work with others within ECP on containers. Two additional contractors were hired to develop specific enhancements to the Gitlab-CI platform used for continuous integration. As new ECP project teams were onboarded at JLSE, additional staff members and a contractor were funded to support these project teams. Finally, five staff members were funded to explore the HPE/Cray Shasta software stack.

Continuous Integration (CI) Pipeline

ALCF has enabled the deployment and growth of the CI Pipeline to its user community through the ECP-CI project using Gitlab-CI. This resource provides a key tool for projects to have regular, automated testing on ALCF resources. Through Gitlab-CI, users can enable their CI pipelines on Theta as well as early hardware available through JLSE. More details on this resource can also be found in section 4.1.5.

Communication between the ALCF and the ECP Resource Allocation Council

In 2018, the ECP ALCC allocation completed and the compute facilities and ECP switched to the Resource Allocation Council (RAC) to support ECP computing needs. This group, composed of representatives of the facilities and the ECP, meets monthly to review project progress and to assess new project needs.

To help automate how the RAC consumes this data, the ALCF sends allocation and usage data in a CSV (comma separated values) file to the ECP each day. The file is uploaded to a Box folder accessible by ECP from where it is downloaded, processed, and merged into the data pipeline that feeds into the ECP User Program dashboard.

Conclusion

The ALCF continues to enable scientific achievements, consistent with DOE's strategic goals for scientific breakthroughs and foundations of science, through projects carried out on ALCF resources. In CY 2021, researchers participating in projects using ALCF resources published 249 papers in high-quality conferences and journals. ALCF projects have had success in a variety of fields, using many different computational approaches. They have been able to reach their scientific goals and successfully use their allocations. Several of the projects and PIs have subsequently received awards or have been recognized as achieving significant accomplishments in their fields.

This page intentionally left blank.

Appendix A – Calculations

A.1 Scheduled Availability

Scheduled availability is the percentage of time a designated level of resource is available to users, excluding **scheduled outage** time for maintenance and upgrades. To be considered a scheduled outage, the user community must be notified of the need for a maintenance event window no less than 24 hours in advance of the outage (emergency fixes). Users will be notified of regularly scheduled maintenance in advance, on a schedule that provides sufficient notification, and no less than 72 hours prior to the event—and preferably as much as seven calendar days prior. If the regularly scheduled maintenance is not needed, users will be informed of the cancellation of the maintenance event in a timely manner. Any interruption of service that does not meet the minimum notification window is categorized as an **unscheduled outage**.

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

Formula:

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

Where

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

A.2 Overall Availability

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

***Formula*:**

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

A.3 System Mean Time to Interrupt (MTTI)

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

Formula:

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

A.4 System Mean Time to Failure (MTTF)

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

Formula:

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

A.5 Total System Utilization

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

Formula:

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

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 Theta.

Table A.1 Capability Definitions for Theta

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

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

***Formula*:**

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

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

A.7 Theta Nodes

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

Table A.2 Total and Reportable Nodes for Theta

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

The reportable node count is used in the following calculations:

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

This page intentionally left blank.

Appendix B – ALCF Director's Discretionary Projects

January 1, 2021 - December 31, 2021
Director's Discretionary (DD) Projects on Theta

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
3DChromatin	Jie Liang	University of Illinois at Chicago	Large Ensemble Model of Single-Cell 3D Genome Structures	Biological Sciences	46,157
3DMPCK	Harold Schock	Michigan State University	Three-Dimensional Multiphase Methodology to Study Piston Cylinder-kit Tribology	Energy Technologies	14,577
ACO2RDS	John J Low	Argonne National Laboratory	Adsorptive CO2 Removal from Dilute Sources	Materials Science	2,486
AIElectrolytes	Logan Timothy Ward	Argonne National Laboratory	Reinforcement Learning Driven Automated Optimization of Battery Electrolytes	Materials Science	20,128
ALCFAITP	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Argonne AI Training Program	Training	0
Allinea	Raymond M. Loy, Kalyan Kumaran	Argonne National Laboratory	Improved debugging memory usage for BG/Q	Internal	818
APSDataAnalysis	Rafael Vescovi	Argonne National Laboratory	APS Beamline Data Processing and Analysis	Computer Science	12,082
arfc-moltres	Kathryn Dorsey Huff	University of Illinois at Urbana-Champaign	Advancement of modeling coupled physics in fluid-fueled molten salt reactors, fluoride salt cooled reactors, and high-temperature-gas-reactors.	Nuclear Energy	6,382
ARPA-E-NPM-2019	Mahmoud I Hussein	University of Colorado-Boulder	Thermal Conductivity of Doped Nanophononic Metamaterials Using Massively Parallel Molecular Dynamics Simulations	Materials Science	54,913
athena	Sylvester Johannes Joosten	Argonne National Laboratory	ATHENA EIC Detector Simulation and Optimization	Physics	16,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
atlas_aesp	Walter Howard Hopkins	Argonne National Laboratory	Simulating and Learning in the ATLAS detector at the Exascale	Physics	30,999
ATPESC2021	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing 2021	Training	5,001
ATPESC_Instructors	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing for ALL Instructors	Training	301
AutoBEM	Joshua Ryan New	Oak Ridge National Laboratory (ORNL)	Automatic Building Energy Modeling and analysis	Energy Technologies	118,000
aviansolardata	Yuki Hamada	Argonne National Laboratory	Storage Allocation for Avian Solar Monitoring Project	Earth Science	1,000
BIP167	Philip Kurian	Howard University	Computing superradiance and van der Waals many-body dispersion effects for biomacromolecules	Physics	34,718
BLawB	Lucian Ivan	Canadian Nuclear Laboratories	Application of Maximum-Entropy Moment Methods to Turbulent and Multiphase Flow Prediction: Software Package Preparation	Engineering	62,375
bloodflow_dd	Jifu Tan	Northern Illinois University (NIU)	Multiphysics modeling of biological flow with cell suspensions	Engineering	33,558
BRAIN	Getnet Dubale Betrie	Argonne National Laboratory	Scalable Brain Simulator for Extreme Computing	Biological Sciences	24,108
BS-SOLCTRA	Esteban Meneses	Costa Rica National High Technology Center	Plasma Physics Simulations for SCR-1 Stellarator	Physics	22,665
Bubblecollapse	Eric Johnsen	University of Michigan	Dynamics and heat transfer in bubble collapse near solid surfaces	Engineering	22,743

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
BubbleOcean	Parviz Moin	Stanford University	Multiscale bubble breakup and gas transfer in turbulent oceanic environments	Engineering	21,333
Bubble_Collapse	Charlotte Noemie Barbier	Oak Ridge National Laboratory (ORNL)	High Definition Simulation of cavitating bubble near a wall with a shear flow	Engineering	5,333
CAIDS	Julio Cesar Mendez Carvajal	North Carolina State University (NCSU)	Consistent Averaging Procedure for solving the fundamental equations of fluid dynamics	Engineering	5,774
candle_aesp	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	Virtual Drug Response Prediction	Biological Sciences	7,999
Carbon_composites	Hendrik Heinz	University of Colorado-Boulder	Designing Functional Nanostructures and Carbon-Based Composite Materials	Materials Science	54,000
CASPER	John Paul Nelson Walters	University of Southern California (USC)	CASPER: Award: CASPER: Compiler Abstractions Supporting High Performance on Extreme-scale Resources	Computer Science	12,885
catalysis_aesp	David Hamilton Bross	Argonne National Laboratory	Exascale Computational Catalysis	Chemistry	99,999
Catalyst	Katherine M Riley, Christopher James Knight, James Clifton Osborn, Timothy Joe Williams	Argonne National Laboratory	Catalyst	Internal	78,017
cfdml_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Data Analytics and Machine Learning for Exascale CFD	Engineering	77,999
CFS_UX_TEST	Haritha Siddabathuni Som	Argonne National Laboratory	TESTING CFS	Support	0
CharmRTS	Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia	University of Illinois at Urbana-Champaign	Charm++ and its applications	Computer Science	7,994

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
climate_severe	Vittorio Angelo Gensini	Northern Illinois University (NIU)	Anticipating Severe Weather Events via Dynamical Downscaling	Earth Science	1,562
Clouds	Ian Foster	Argonne National Laboratory	Unsupervised analysis of satellite cloud imagery	Earth Science	104,767
CobaltDevel	Paul Michael Rich, William Edward Allcock	Argonne National Laboratory	Cobalt Development	Internal	7,316
Comp_Perf_Workshop	Raymond M. Loy, Yasaman Ghadar	Argonne National Laboratory	ALCF Computational Performance Workshop	Training	8,000
Conductivity_Polymer	Hanning Chen	American University	Simulated Conductivity of Polymer Thin Films for Solar Cells	Chemistry	31,625
connectomics_aesp	Nicola Joy Ferrier, Thomas David Uram	Argonne National Laboratory	Enabling Connectomics at Exascale to Facilitate Discoveries in Neuroscience	Biological Sciences	155,999
CONUS-Carbon	Jinxun Liu	U.S. Geological Survey (USGS)	Terrestrial ecosystem carbon cycle of the conterminous U.S.	Earth Science	28,305
covid-ct	Ravi Kiran Madduri	Argonne National Laboratory	Medical Imaging Domain-Expertise Machine Learning for Interrogation of COVID	Computer Science	1,500
Cray	Torrance Ivan Leggett, Mark R Fahey, Susan Marie Coghlan, Timothy Joe Williams	Cray Inc.	Cray Installation	Internal	78,017
cray-hpo	Michael Adnan Salim	Argonne National Laboratory	Scaling Studies of CrayAI Hyperparameter Optimization	Computer Science	23,866
CRRCS	Veerabhadra Rao Kotamarthi	Argonne National Laboratory	Convection Resolved Regional Scale Climate Simulations	Earth Science	28,800
CSC249ADCD01	Ian Foster	Argonne National Laboratory	2.2.6.03 ADCD01-CODAR	Computer Science	30,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
CSC249ADCD02	Susan Marie Mniszewski, Timothy C. Germann	Los Alamos National Laboratory (LANL)	2.2.6.04 ADCD02-COPA: Co-Design Center for Particle Applications	Physics	1,300
CSC249ADCD04	Tzanio Valentinov Kolev, Misun Min, Paul Fischer	Lawrence Livermore National Laboratory (LLNL)	2.2.6.06 CEED: Center for Efficient Exascale Discretizations	Computer Science	13,276
CSC249ADCD05	Mahantesh Halappanavar	Pacific Northwest National Laboratory (PNNL)	2.2.6.07 ADCD05-ExaGraph	Computer Science	26,000
CSC249ADCD08	Francis Joseph Alexander	Brookhaven National Laboratory (BNL)	2.2.6.08 ADCD08-ExaLearn	Physics	99,000
CSC249ADCD502	Kenneth John Roche	Pacific Northwest National Laboratory (PNNL)	2.2.6.02 ADCD502 Application Assessment	Computer Science	4,000
CSC249ADCD504	Jeanine Cook, Shirley Victoria Moore	Lawrence Livermore National Laboratory (LLNL)	2.2.6.01 ADCD504-Proxy Applications	Computer Science	1,300
CSC249ADOA01	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	2.2.4.03 ADOA01 CANDLE: Exascale Deep Learning Enabled Precision Medicine for Cancer	Biological Sciences	4,500
CSC249ADSE03	Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie	Fermi National Accelerator Laboratory (Fermilab)	2.2.1.01 ADSE03-LatticeQCD: Exascale Lattice Gauge Theory Opportunities/Reqmts for Nuclear & High Energy Physics	Physics	146,000
CSC249ADSE04	Danny Perez	Los Alamos National Laboratory (LANL)	2.2.1.04 ADSE04-EXAALT - Molecular dynamics at the exascale	Nuclear Energy	1,300

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
CSC249ADSE05	David Trebotich	Lawrence Berkeley National Laboratory (LBNL)	2.2.3.04 ADSE05-Subsurface	Earth Science	1,300
CSC249ADSE08	Steven Hamilton, Paul Kollath Romano	Oak Ridge National Laboratory (ORNL)	2.2.2.03 ADSE08 ExaSMR	Nuclear Energy	2,200
CSC249ADSE09	Paul Richard Charles Kent, Anouar Benali	Oak Ridge National Laboratory (ORNL)	2.2.1.06 QMCPACK: Predictive and Improvable Quantum-mechanics Based Simulations	Materials Science	4,100
CSC249ADSE11	Theresa Windus	University of Washington	2.2.1.02 ADSE11-NWChemEx: Tackling Chemical, Materials, & Biomolecular Challenges in Exascale	Chemistry	1,700
CSC249ADSE12	Amitava Bhattacharjee	Princeton Plasma Physics Laboratory (PPPL)	2.2.2.05 ADSE12 WDMAPP	Computer Science	95,000
CSC249ADSE14	Jacqueline Chen	Sandia National Laboratories, California	2.2.2.02 ADSE14-Combustion-Pele: Transforming Combustion Science & Technology with Exascale Simulations	Engineering	2,200
CSC249ADSE16	Mark S Gordon	Ames Laboratory	2.2.1.03 ADSE16-GAMESS	Chemistry	111,300
CSC249ADSE17	Charlie Catlett, Melissa Ree Allen, Rajeev Jain, Scott A Ehling	Argonne National Laboratory	(OBSOLETE) 2.2.4.01 Urban: Multiscale Coupled Urban Systems	Computer Science	1,200
CSC249ADSE18	Daniel Kasen	Lawrence Berkeley National Laboratory (LBNL)	2.2.3.01 ADSE18 Exastar	Physics	1,500
CSC249ADSE22	Slaven Peles, Andres Marquez, Zhenyu Huang	Pacific Northwest National Laboratory (PNNL)	2.2.4.02 ADSE22-ExaSGD	Energy Technologies	1,300

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
CSC249ADTR01	Daniel Edward Laney	Lawrence Livermore National Laboratory (LLNL)	2.3.5.10 ADTR01-ExaWorks	Computer Science	1,300
CSC249ADTR02	Ashley Barker	Oak Ridge National Laboratory (ORNL)	2.4.6.02 ADTR02-Productivity	Computer Science	1,300
CSC250STDA05	Kenneth Dean Moreland	Oak Ridge National Laboratory (ORNL)	2.3.4.13 STDA05-ECP/VTK-m	Computer Science	1,300
CSC250STDM10	Surendra Byna, Venkatram Vishwanath	Lawrence Berkeley National Laboratory (LBNL)	2.3.4.15 ExaIO - Delivering Efficient Parallel I/O on Exascale Computing Systems with HDF5 and Unify	Computer Science	18,000
CSC250STDM11	Scott Klasky, Norbert Podhorszki	Oak Ridge National Laboratory (ORNL)	2.3.4.09 STDM11-ADIOS Framework for Scientific Data on Exascale Systems	Computer Science	4,000
CSC250STDM12	Robert B. Ross, Robert J Latham	Argonne National Laboratory	2.3.4.10 STDM12-DataLib: Data Libraries and Services Enabling Exascale Science	Computer Science	4,000
CSC250STDM14	Franck Cappello	Argonne National Laboratory	2.3.4.14 STDM14 - VeloC-SZ: Very Low Overhead Transparent Multilevel Checkpoint/Restart/SZ: Fast, Effective, Parallel Error-bounded Exascale Loss....	Computer Science	3,100
CSC250STDM16	James Paul Ahrens, Terece Louise Turton	Lawrence Livermore National Laboratory (LLNL)	2.3.4.16 STDM16-ALPINE/ZFP	Computer Science	1,700
CSC250STDT10	Jeffrey S Vetter	Oak Ridge National Laboratory (ORNL)	2.3.2.10 STDT10 PPROTEAS-TUNE	Computer Science	1,300

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
CSC250STDT11	Sunita Chandrasekaran, Doss ay Oryspayev	Stony Brook University	2.3.2.11 SOLLVE: Scaling OpenMP with LLVm for Exascale	Computer Science	1,300
CSC250STDV01	Charles Vernon Atkins	Kitware Inc.	2.3.4.01 STDV01-Data and Visualization Software Development Kit	Computer Science	2,200
CSC250STMS05	Ulrike Meier Yang, Satish Balay	Argonne National Laboratory	2.3.3.01 STMS05-Extreme-scale Scientific xSDK for ECP	Mathematics	1,300
CSC250STMS07	Todd S. Munson, Hong Zhang, Richard Tran Mills, Satish Balay	Argonne National Laboratory	2.3.3.06 STMS07-PETSc/TAO for Exascale	Mathematics	5,000
CSC250STNS01	Michael Lang, Terece Louise Turton	Los Alamos National Laboratory (LANL)	2.3.6.01 - STNS01 -LANL ATDM ST Projects	Computer Science	1,300
CSC250STPM09	Yanfei Guo	Argonne National Laboratory	2.3.1.07 STPM09-Exascale MPI	Computer Science	1,300
CSC250STPM11	George Bosilca, Earl Luther Carr, Jack Dongarra, Thomas Herault	The University of Tennessee at Knoxville	2.3.1.09 STPM11 ParSEC: Distributed Tasking	Computer Science	1,300
CSC250STPM17	Paul Hamilton Hargrove, Erich Strohmaier	Lawrence Berkeley National Laboratory (LBNL)	2.3.1.14 STPM17-UPC++ & GASNet	Computer Science	4,000
CSC250STPR19	Peter Hugh Beckman	Argonne National Laboratory	2.3.1.19 STPR19 Argo: Argo/Power Steering	Computer Science	1,300
CSC250STPR27	David Edward Bernholdt	Oak Ridge National Laboratory (ORNL)	2.3.1.17 STPR27-OMPI-X: Open MPI for Exascale	Materials Science	4,000
CSC250STTO09	Jack Dongarra, Anthony Danalis, Earl Luther Carr, Heike Jagode	The University of Tennessee at Knoxville	2.3.2.06 STTO09 EXAPAPI	Computer Science	4,000
CSC250STTO11	John Michael Mellor-Crummey	Rice University	2.3.2.08 STTO11 HPCToolkit	Computer Science	2,200

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
CSC251HIHE05	Scott Dov Pakin, Simon David Hammond	Los Alamos National Laboratory (LANL)	2.4.2.01 HIHE05-Analytical Modeling - Hardware Evaluation Working Groups	Computer Science	1,300
CSC251HISD01	Ryan Adamson	Los Alamos National Laboratory (LANL)	2.4.4.01 HISD01-Software Integration	Computer Science	2,300
CSCSTDT12345	Patrick McCormick	Los Alamos National Laboratory (LANL)	2.3.2.12 Flang: open-source Fortran front end for the LLVM infrastructure	Computer Science	2,200
CTOP	Devesh Tiwari	Northeastern University	Cost-Efficient and Throughput-Oriented Power Capping on Production HPC Systems	Computer Science	8,530
D6_XNET	Robert Fisher	University of Massachusetts-Dartmouth	Explorations of the D6 Scenario of Type Ia Supernovae on Theta	Physics	10,000
darkskyml_aesp	Salman Habib	Argonne National Laboratory	Dark Sky Mining	Physics	30,999
datascience	Venkatram Vishwanath	Argonne National Laboratory	ALCF Data Science and Workflows Allocation	Internal	78,017
DAWN	Xin-Zhong Liang	University of Maryland	USDA Funds UMD Project to Sustain Agricultural Production in the Corn Belt with $10 Million	Earth Science	32,000
DDES_JET	Roberto Paoli	University of Illinois at Chicago	DDES of compressible jet flows with OpenFoam	Engineering	49,237
DDICF-Dev	Duc Minh Cao	University of Rochester	Direct-Drive Inertial Confinement Fusion Code Porting and Proposal Preparation	Fusion Energy	6,285
DFT-FE	Phani Sudheer Motamarri	University of Michigan	Large-scale real-space electronic structure calculations for understanding energetics of complex defects in materials	Materials Science	15,538

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
Diagnostics-CNN	Lander Ibarra	Argonne National Laboratory	Inverse Problem CNN Computation for Diagnostics	Energy Technologies	810
dist_relational_alg	Sidharth Kumar	The University of Alabama at Birmingham	Distributed relational algebra at scale	Computer Science	12,107
DL4MFC	Petro Junior Milan	Georgia Institute of Technology (Georgia Tech)	Deep-Learning Enhancement of Multiphysics Flow Computations for Propulsion Applications	Engineering	1
dl_am	Satish Karra	Los Alamos National Laboratory (LANL)	Scalable Deep Learning Workflow for Additive Manufacturing	Engineering	14,626
DL_MODEX	MARUTI KUMAR MUDUNURU	Pacific Northwest National Laboratory (PNNL)	Towards a robust and scalable deep learning workflow for fast, accurate, and reliable calibration of watershed models	Earth Science	10,503
DNS3D	Ramesh Balakrishnan	Argonne National Laboratory	Direct Numerical Simulation of Three Dimensional Turbulence	Engineering	63,478
DNSforLIS	Antonino Ferrante	University of Washington	DNS investigation for the law of incipient separation	Engineering	23,747
DNS_SV_Turb_2WC	Josin Tom	Duke University	DNS study of particle settling velocities in turbulence in the presence of two-way coupling	Engineering	5,774
Drug_FEP_Data	Wei Jiang	Argonne National Laboratory	Machine elarning of drug binding and toxicity based on high throughput free energy computations	Biological Sciences	50,002
DynCap	Zhiling Lan	Illinois Institute of Technology (IIT)	Dynamic Power Capping for Scientific Applications	Computer Science	23,817
DynStall	Anupam Sharma	Iowa State University (ISU)	Analysis and Mitigation of Dynamic Stall in Energy Machines	Engineering	1,545

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
Eagle_Testing	Avanthi Madduri	Argonne National Laboratory	Eagle Testing purposes - updated	Internal	0
ECP_SDK	Sameer Suresh Shende	University of Oregon	Deploying the ECP SDK software stack at ALCF	Computer Science	871
EE-ECP	Xingfu Wu, Valerie Taylor	Argonne National Laboratory	Energy efficient tradeoff among execution time, power, and resilience of two ECP applications	Computer Science	19,157
ELSI_dev	Alvaro Vazquez Mayagoitia	Duke University	ELSI Developers	Computer Science	830
EngineDNS	Christos Frouzakis	Eidgenössische Technische Hochschule Zürich (ETH Zurich)	Towards reactive DNS in complex internal combustion engine geometries	Engineering	44,430
EstopSim_DD	Yosuke Kanai	The University of North Carolina-Chapel Hill	Massively Parallel Electronic Stopping Simulations of High Energy Particles in Solvated DNA	Chemistry	1,515
extrap-noise-ai	Marcus Ritter	Technische Universität Darmstadt	Noise Resilient Performance Modeling with Deep Learning	Computer Science	1,559
FASTMath2	Richard Kim Archibald	Oak Ridge National Laboratory (ORNL)	FASTMath Software Platform	Mathematics	50,000
FDTD_Cancer_2a	Allen Taflove	Northwestern University	Computational Physical Genomics: Exploring Potential Novel Cancer Therapies	Biological Sciences	158,554
FDTD_GPU	Wei Jiang	Argonne National Laboratory	Migration of FDTD modeling to GPU	Computer Science	10,000
fhi_etc1p	Thomas Alexander Reichmanis Purcell	Fritz-Haber-Institut der Max-Planck-Gesellscha ft (FHI)	Benchmarking a High-Throughput Framework for the Thermal Conductivity of Perovskites	Physics	6,108
FTI_Tribo_AS_DD	Matthias Baldofski	Aalto University	Benchmarking atomistic simulations of tribological systems	Materials Science	34,868

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
fusiondl_aesp	William Tang	Princeton University	Accelerated Deep Learning Discovery in Fusion Energy Science	Fusion Energy	3,999
GA-bfg3d	Mark Kostuk	General Atomics	General Atomics Next Generation Skew-Symmetric Fluid Solver	Physics	8,241
gas-turbine	Pinaki Pal	Argonne National Laboratory	Deep Learning-Augmented Flow Solver to Improve the Design of Gas-Turbine Engines	Engineering	16,035
GNPMem	Tarak K Patra	Indian Institute of Technology Madras	Computational design of polymer grafted nanoparticle membrane	Materials Science	8,301
GrainBoundaries	Wissam A Saidi	University of Pittsburgh	Structure and Properties of Grain Boundaries in Materials for Energy Applications	Materials Science	39,713
HACC_aesp	Katrin Heitmann	Argonne National Laboratory	Extreme-Scale Cosmological Hydrodynamics	Physics	15,999
HARDXBIOCANCEL	Carles Serrat	Universitat Politècnica de Catalunya-Barcelona Tech	Coherent hard X-ray core nonlinear selective cancellation of the effect of biological target molecules in pathogens	Chemistry	11,756
HED_Flows	Hussein Aluie	University of Rochester	Scale-Aware Modeling of Instabilities and Mixing in HED Flows	Fusion Energy	6,121
HEPCloud-FNAL	Burt Holzman	Fermi National Accelerator Laboratory (Fermilab)	High Energy Physics Computing for Fermilab experiments via HEPCloud	Physics	3,324
Hephaestus	Andrey Beresnyak	Naval Research Laboratory	Hephaestus heterogeneous mode	Physics	7,087
HEP_on_HPC	Jim B Kowalkowski, Marc Francis Paterno	Fermi National Accelerator Laboratory (Fermilab)	HEP analysis workflows on HPC	Physics	18,774
HighMachTurbulence	Sanjiva K Lele	Stanford University	Simulations of Turbulence at High Compressibility	Engineering	51,935

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
HighReyTurb_PostProc	Robert D. Moser, Myoungkyu Lee	The University of Texas at Austin	Data analysis of turbulent Channel Flow at High Reynolds number	Engineering	13,191
HiMB_Beamline	Eremey Vladimirovich Valetov	Michigan State University	Development and Optimization of a Novel High-Intensity Muon Beamline	Physics	14,644
HIV-PR	Ao Ma	University of Illinois at Chicago	Understanding the mechanism of ligand-induced conformational dynamics of HIV-1 protease and the effects of mutations	Biological Sciences	15,455
HNPballistics	Sinan Keten	Northwestern University	Engineering Nanocellulose based Hairy Nanoparticle Assemblies for High Ballistic Impact Performance	Engineering	23,383
hpcbdsm	Tanwi Mallick	Argonne National Laboratory	High-Performance Computing and Big Data Solutions for Mobility Design and Planning	Computer Science	15,625
HTAlloyDesign	Dongwon Shin	Oak Ridge National Laboratory (ORNL)	Supercomputing for automotive high-temperature alloy design	Materials Science	60,000
HumanVAN	Grant Addison Hartung	Athinoula A. Martinos Center for Biomedical Imaging	Biophysical modeling of the functional MRI signal through parametric variations in neuronal activation and blood vessel anatomy using realistic synthetic microvascular networks	Biological Sciences	5,418
ICE_InSitu	Muhsin Mohammed Ameen,Saumil Sudhir Patel	Argonne National Laboratory	Towards Exascale Internal Combustion Engine Simulations with In-Situ Analysis	Chemistry	39,564
ImageNanoX_DD	Phay J Ho	Argonne National Laboratory	X-ray Imaging of Transient Structure and Dynamics of Nanoparticles	Physics	7,501

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
Intel	Kalyan Kumaran, Scott Parker, Timothy Joe Williams, Venkatram Vishwanath	Argonne National Laboratory	Intel employees in support of Theta	Internal	78,017
IntelVis	Joseph A Insley	Argonne National Laboratory	Intel Visualization Development	Computer Science	5,672
inversedesign_oerorr	Rafael Gomez-Bombarelli	Massachusetts Institute of Technology (MIT)	Inverse design of oxide catalysts for OER/ORR	Materials Science	874
JCESR	Larry Curtiss, Anubhav Jain	Argonne National Laboratory	Development of High Throughput Methods	Materials Science	78,000
Job_Interference	Zhiling Lan	Illinois Institute of Technology (IIT)	Workload Interference Analysis on Theta	Computer Science	19,427
lalitha_test	Haritha Super User	3Q-Lab GmbH	testing	Internal	1
LatticeQCD_aesp	Paul Mackenzie, Norman Howard Christ	Fermi National Accelerator Laboratory (Fermilab)	Lattice Quantum Chromodynamics Calculations for Particle and Nuclear Physics	Physics	390,999
LESDNSHTECESHE2021	Lane Benjamin Carasik	Virginia Commonwealth University (VCU)	LES and DNS of Heat Transfer Enhancements in Clean Energy System Heat Exchangers	Engineering	4,705
LES_CommTransAir	Parviz Moin	Stanford University	Large-Eddy Simulation of a Commercial Transport Aircraft Model	Engineering	30,324
LIGHTCONTROL	Sandra Gail Biedron	University of New Mexico	Light sources and their control using AI techniques	Physics	6,857
lipid-sampling	Yun Lyna Luo	Western University of Health Sciences	Development of enhanced sampling approach for heterogenous membrane	Biological Sciences	17,812
LMPMBXOpt	Francesco Paesani	University of California-San Diego	Enabling Chemical Accuracy Through Large-Scale Many-body Molecular Dynamics	Chemistry	15,625

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
LQCDdev	James Clifton Osborn	Argonne National Laboratory	Lattice QCD development	Physics	1,640
lqcdml_aesp	William Detmold	Massachusetts Institute of Technology (MIT)	Machine Learning for Lattice Quantum Chromodynamics	Physics	54,999
LTC_Aramco_theta	Roberto Torelli	Argonne National Laboratory	Investigation of Gasoline-Range Fuels for a Heavy-Duty Diesel Engine in a Low-Temperature Combustion Regime	Engineering	16,017
LTP-Opt	Vyaas Gururajan	Argonne National Laboratory	Low Temperature Plasma Optimization	Chemistry	19,975
magnetotail	Samuel Richard Totorica	Princeton University	Kinetic simulations of the dynamic magnetotail	Physics	54,717
Maintenance	Mark R Fahey	Argonne National Laboratory	LCF Operations System Maintenance	Internal	78,017
marine-twin	Flavio Dal Forno Chuahy	Oak Ridge National Laboratory (ORNL)	Marine Digital-Twin Full-Scale Simulation	Energy Technologies	41,184
matml_aesp	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials	Materials Science	219,999
metastable	Subramanian Sankaranarayanan	Argonne National Laboratory	Metastable phase diagram of material	Materials Science	52,996
MI2Dmaterials	Trevor David Rhone	Rensselaer Polytechnic Institute (RPI)	Materials informatics study of two-dimensional magnetic materials and their heterostructures	Materials Science	142,942
microwave_catalysis	fanglin che	University of Massachusetts-Lowell	Characterizing Non-Thermal Effects of Microwave Accelerated Heterogeneous Catalysis	Chemistry	1,711
MITboiling	Jinyong Feng	Massachusetts Institute of Technology (MIT)	Modeling of Boiling Heat Transfer in Nuclear Reactors	Nuclear Energy	31,999

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
MKM_catal	Wilfred T Tysoe	University of Wisconsin-Milwaukee	Enantioselectivity in Heterogeneous Catalysts via the Addition of Chiral Modifiers	Chemistry	100,000
ML-Coupling	shinhoo kang	Argonne National Laboratory	Data-driven Coupling Methods for Atmospheric-Ocean Interactions	Earth Science	526
ML-target	Lianshan Lin	Oak Ridge National Laboratory (ORNL)	Application of ML to the Liquid Mercury Target	Physics	67,933
ML4MPF	Gina Maureen Sforzo	Argonne National Laboratory	Enabling predictive simulations of reacting multiphase flows via data-driven emulation	Engineering	12,527
MLP4THERMO	Cem Sevik	Eskisehir Technical University	Machine Learning Potentials for Thermal Properties of Two-Dimensional Materials	Materials Science	9,587
mmaADSP	Eliu Antonio Huerta Escudero	University of Illinois at Urbana-Champaign	Deep Learning at Scale for Multimessenger Astrophysics through the NCSA-Argonne Collaboration	Physics	25,660
MOAB_App	Vijay Subramaniam Mahadevan	Argonne National Laboratory	MOAB Algorithmic Performance Portability	Mathematics	16,503
ModelingCoronaVirus	Zhangli Peng	University of Illinois at Chicago	Modeling Corona Virus	Biological Sciences	160,755
MoltenSalts	Nicholas Everett Jackson	University of Illinois at Urbana-Champaign	Automated Active Learning on ALCF for Machine Learning Forcefield Automation	Nuclear Energy	22,530
MPICH_MCS	Kenneth James Raffenetti, Pavan Balaji	Argonne National Laboratory	MPICH - A high performance and widely portable MPI implementation	Computer Science	36,741
mriPic	Gregory R Werner	University of Colorado-Boulder	Particle-in-Cell simulations of the magneto-rotational instability	Physics	8,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
MSEMNeuro	Thomas David Uram	Argonne National Laboratory	Reconstruction of neuronal connectivity from MSEM imaged tissue	Biological Sciences	6,461
Mu2e_HEPCloud	Iuri Artur Oksuzian	Argonne National Laboratory	Mu2e simulations through the HEPCloud Project	Physics	25,628
multimode_comb	Pinaki Pal	Argonne National Laboratory	High-Fidelity CFD Simulations of Multi-Mode Combustion	Energy Technologies	35,455
MultiphaseNuc	Igor A Bolotnov	North Carolina State University (NCSU)	Multiphase Flow Simulations of Nuclear Reactor Flows	Nuclear Energy	30
multiphysics_aesp	Amanda Randles	Duke University	Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations	Engineering	38,999
MyVirtualCancer	Leili Shahriyari	University of Massachusetts-Amherst	My Virtual Cancer	Biological Sciences	20,000
NAMD_aesp	Benoit Roux, James C. Phillips	The University of Chicago (UChicago)	Free Energy Landscapes of Membrane Transport Proteins	Biological Sciences	77,999
NAQMC_RMD_aesp	Aiichiro Nakano	University of Southern California (USC)	Metascalable Layered Materials Genome	Materials Science	38,999
nek52rs	Aleksandr V. Obabko	Argonne National Laboratory	Nek5000/NekRS for NRC and COVID LES	Engineering	7,529
Nek5RS_covid	Aleksandr V. Obabko	Argonne National Laboratory	Nek5000/NekRS for COVID particle transport and NRC	Engineering	10,000
Nek_Boost	Pinaki Pal	Argonne National Laboratory	Development of High-Fidelity and Efficient Modeling Capabilities for Enabling Co-Optimization of Fuels and Multi-Mode Engines	Energy Technologies	34,502

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
networkbench	Devi Sudheer Kumar Chunduri, Elise Jennings, Kevin Harms, Misbah Mubarak	Argonne National Laboratory	network benchmarking and modeling	Computer Science	17,303
neutrinoADSP	Andrzej Michał Szelc, Corey J Adams, Diego Garcia-Gamez	The University of Edinburgh	Developing High Performance Computing Applications for Liquid Argon Neutrino Detectors	Physics	26,785
neutrino_osc_ADSP	Marco Del Tutto	Fermi National Accelerator Laboratory (Fermilab)	Machine Learning for Data Reconstruction to Accelerate Physics Discoveries in Accelerator-Based Neutrino Oscillation Experiments	Physics	48,328
NextGenReac	Yiqi Yu	Argonne National Laboratory	Toward the Future: High fidelity simulation for next Generation Nuclear Reactors	Nuclear Energy	16,841
NGM_EHT	Charles Forbes Gammie	University of Illinois at Urbana-Champaign	Next-generation Models for the Event Horizon Telescope	Physics	20,000
niubmrk	Sergey A Uzunyan	Northern Illinois University (NIU)	Benchmarks of applications running at NIU compute clusters using modern hardware	Computer Science	5,000
novacosmics	Alexander I Himmel	Fermi National Accelerator Laboratory (Fermilab)	NOvA Cosmic Rejection	Physics	450
NucContainmentMix	Christopher Fred Boyd	U.S. Nuclear Regulatory Commission (NRC)	LES Simulations of Severe Accident Conditions in Nuclear Containment	Nuclear Energy	30,358
NWChemEx_aesp	Theresa Windus, Alvaro Vazquez Mayagoitia	Pacific Northwest National Laboratory (PNNL)	NWChemEx: Tackling Chemical, Materials & Biochemical Challenges in the Exascale Era	Chemistry	7,999
Operations	Mark R Fahey	Argonne National Laboratory	Systems administration tasks	Internal	78,017

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
OptADDN	Sandeep Madireddy	Argonne National Laboratory	Optimal Architecture discovery for deep probabilistic models and neuromorphic systems	Computer Science	60,000
oqs_int	David Beratan	Duke University	Efficient Simulation of Open Quantum Systems with Matrix Product States in the Interaction Picture	Chemistry	32,000
PARTURB3D	Ramesh Balakrishnan	Argonne National Laboratory	Simulating turbulent particulate flows inside enclosures	Engineering	5,418
Performance	Scott Parker, Raymond M. Loy	Argonne National Laboratory	Performance	Internal	78,017
perf_research	Devi Sudheer Kumar Chunduri	Argonne National Laboratory	Performance group external facing research	Computer Science	0
PerovskiteMachine	Volker Wolfgang Blum	Duke University	High-Precision Dynamical Properties of Complex Perovskites by Interpolative Machine Learning	Materials Science	4,608
PHASTA_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Extreme Scale Unstructured Adaptive CFD: From Multiphase Flow to Aerodynamic Flow Control	Engineering	77,999
PHASTA_NCSU	Igor A Bolotnov	North Carolina State University (NCSU)	Multiphase Simulations of Nuclear Reactor Thermal Hydraulics	Engineering	32,000
pnc_ml	Boran Ma	Duke University	Machine Learning-assisted Polymer Nanocomposite Microstructure Design	Materials Science	64,000
poly-ion-dd	Juan Pablo	The University of Chicago (UChicago)	Ion Conduction Through Polymeric Interfaces	Chemistry	5,144
PPE-CV-NTM	Sandra Jean Bittner	Individual	Computational Study and Visualization Models for Non-Traditional Materials for Personal Protective Equipment	Computer Science	16,900

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
psr001	Ronald Otis Grover	General Motors Company	Electric Motor Thermal Management Analysis	Engineering	52,948
PTLearnPhoto	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Materials for Organic Photovoltaics	Materials Science	73,966
PUR-IRL	Nicholas Lee-Ping Chia	Mayo Clinic-Minnesota	Inferring the Reward Function of Cancer	Biological Sciences	476
Q-Pix	Jonathan Asaadi	The University of Texas at Arlington	QPix: Achieving kiloton scale pixelated readout for Liquid Argon Time Projection Chambers	Physics	937
QCProxyApps	Graham Donald Fletcher	Argonne National Laboratory	Quantum Chemistry Proxy Applications	Chemistry	25,087
QMCPACK_aesp	Anouar Benali	Argonne National Laboratory	Extending Moore's Law computing with Quantum Monte Carlo	Materials Science	77,999
qsars_qm_vae	Brad Reisfeld	Colorado State University	Discovering quantitative structure activity relationships using quantum chemical descriptors and variational autoencoding	Biological Sciences	742
QSim	Yuri Alexeev	Argonne National Laboratory	Quantum Simulations	Computer Science	48,468
QuantumDS	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Quantum mechanics and Data Science	Chemistry	17,850
radix-io	Philip Hutchinson Carns	Argonne National Laboratory	System software to enable data-intensive science	Computer Science	6,579
RaptorX	Jinbo Xu	Toyota Technological Institute at Chicago (TTIC)	Protein Folding through Deep Learning and Energy Minimization	Biological Sciences	36,445
RCM_4km	Jiali Wang	Argonne National Laboratory	Generation of a next level dataset for regional scale climate modeling: convective resolving spatial scales	Earth Science	269,862

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
Redox_ADSP	Logan Timothy Ward	Argonne National Laboratory	Autonomous Molecular Design for Redox Flow Batteries	Materials Science	17,116
REI_Flares	Marc Cremer	Reaction Engineering International	Leveraging the UCF for Simulation of Industrial Flares	Chemistry	52,144
ReservoirQuality	James Edward Guilkey	The University of Utah	Numerical prediction of sandstone reservoir quality	Earth Science	9,765
RL-fold	Arvind Ramanathan	Argonne National Laboratory	Targeting intrinsically disordered proteins using artificial intelligence driven molecular simulations	Biological Sciences	29,080
rnn-robustness	Liam Benjamin Johnston	University of Wisconsin-Madison	Large-scale Factorial Experiment on RNN Robustness	Computer Science	1,790
Rxn_Diff	Neeraj Rai	Mississippi State University	Modeling reaction-diffusion processes for liquid-phase heterogeneous catalysis	Chemistry	50,000
sbi-fair	Pete Beckman, Kamil Antoni Iskra	Argonne National Laboratory	FAIR Surrogate Benchmarks Supporting AI and Simulation Research	Computer Science	3,802
scalablemoose	Fande Kong	Idaho National Laboratory (INL)	MOOSE scaling study	Nuclear Energy	3,168
SCPlasma	RANGANATHAN GOPALAKRISHNAN	University of Memphis	Thermodynamics and Transport Models of Strongly Coupled Dusty Plasmas	Physics	16,994
SDL2021x	Wai Nim Alfred Tang	Phd Tutor Hub	SDL2021 Extension	Computer Science	47
SDL_Workshop	Raymond M. Loy	Argonne National Laboratory	ALCF Simulation, Data, and Learning Workshop	Training	3,000
SENSEI	Silvio Humberto Rafael Rizzi, Joseph A Insley, Nicola Joy Ferrier, Venkatram Vishwanath	Argonne National Laboratory	Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery	Computer Science	61,761

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
SimMCSProce	Andreas Franz Prein	The National Center for Atmospheric Research (NCAR)	Using ARM Observations to Evaluate Process-Interactions in MCS Simulations Across Scales	Earth Science	33,043
SI_IBM	Saurabh Chawdhary	Argonne National Laboratory	Scalable, efficient sharp-interface immersed boundary method for fluid-structure interaction problems	Engineering	9,410
SolarWindowsADSP	Jacqueline Manina Cole	University of Cambridge	Data-Driven Molecular Engineering of Solar-powered Windows	Materials Science	31,553
spentFuel	Angela Di Fulvio	University of Illinois at Urbana-Champaign	Cask Mis-loads Evaluation Techniques	Nuclear Energy	163,900
SU2_PadeOps_aesp	Sanjiva K Lele	Stanford University	Benchmark Simulations of Shock-Variable Density Turbulence and Shock-Boundary Layer Interactions with Applications to Engineering Modeling	Engineering	99
SuperBERT	Ian Foster	Argonne National Laboratory	Training of language models on large quantities of scientific text	Computer Science	4,767
THGSupport	Kevin Harms	Argonne National Laboratory	The HDF Group Support	Computer Science	245
TNContract	James Clifton Osborn	Argonne National Laboratory	Tensor Network Contractions for QIS	Physics	21,313
TomoEncoders	Aniket Tekawade	Argonne National Laboratory	TomoEncoders: Computer Vision Framework for 4D X-ray Tomography	Engineering	2,887
Tools	Scott Parker	Argonne National Laboratory	ALCF Performance Tools	Internal	10,000
TotalView	Peter Michael Thompson, Raymond M. Loy	Rogue Wave Software, Inc.	TotalView Debugger on Blue Gene P	Internal	1,301

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
TurbNet	Romit Maulik	Argonne National Laboratory	TurbNet: Scaleable physics-informed deep learning for turbulence model development	Engineering	8,893
UINTAH_aesp	Martin Berzins, John Andrew Schmidt	The University of Utah	Design and evaluation of high-efficiency boilers for energy production using a hierarchical V/UQ approach	Chemistry	62,999
Uintah_GPU_benchmark	Marta Garcia Martinez	Argonne National Laboratory	Uintah benchmarks on GPU	Engineering	32,000
Ultrafast_X-ray	Jin Wang	Argonne National Laboratory	Ultrafast_X-ray	Engineering	20,425
User_Services	Haritha Siddabathuni Som, Sreeranjani Ramprakash	Argonne National Laboratory	User Services	Internal	0
VeloC	Bogdan Florin Nicolae	Argonne National Laboratory	VeloC: Very Low Overhead Checkpointing System	Computer Science	16,895
visualization	Joseph A Insley, Michael E. Papka	Argonne National Laboratory	Visualization and Analysis Research and Development for ALCF	Internal	15,603
Viz_Support	Joseph A Insley, William Edward Allcock	Argonne National Laboratory	Visualization Support	Computer Science	3,900
wall_turb_dd	Ramesh Balakrishnan	Argonne National Laboratory	Wall Resolved Simulations of Canonical Wall Bounded Flows	Engineering	16,584
WaterHammer	Hong Zhang, Hong Zhang	Argonne National Laboratory	Water Hammer Simulation	Mathematics	20,000
WGSanalysis	Elizabeth McNally	Northwestern University	Large scale alignment and analysis of whole human and mouse genomes, with focus on realigning to HG 38 and harmonized workflows	Biological Sciences	2,571

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
WindScenarios	Ignas Vasilijus Satkauskas	National Renewable Energy Laboratory (NREL)	wind scenario generation for two-stage stochastic optimization	Energy Technologies	2,585
WRLCMF	Krishnan Mahesh	University of Minnesota-Twin Cities	Wall-Resolved LES of complex maneuvering flows	Engineering	49,888
XGC_aesp	Choongseok Chang	Princeton Plasma Physics Laboratory (PPPL)	High fidelity simulation of fusion reactor boundary plasmas	Fusion Energy	77,999
Xin-ZhongLiang	Xin-Zhong Liang, Chao Sun	University of Maryland	Xin-Zhong Liang	Earth Science	31,076
yingtao	Yingtao Wang	AAAS Fellowship Programs, Inc.	On the Role of Crystal Defects on the Lattice Thermal Conductivity of Monolayer WSe2 (P63/mmc) Thermoelectric Materials by DFT Calculation	Chemistry	476
				Total DD	7,389,111

January 1, 2021 - December 31, 2021
Theta Expansion DD

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
ACO2RDS	John J Low	Argonne National Laboratory	Adsorptive CO2 Removal from Dilute Sources	Materials Science	1,176
AGI-for-Science	Rick Lyndon Stevens	Argonne National Laboratory	Large scale multi modal language models for science comprehension	Computer Science	601
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	114
AIASMAAR	Rui Hu	Argonne National Laboratory	Artificial Intelligence Assisted Safety Modeling and Analysis of Advanced Reactors	Nuclear Energy	2,536
AI_Acceleration	Ian Foster	Argonne National Laboratory	Exploration of AI Accelerators for Neural Networks	Computer Science	741
ALCFAITP	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Argonne AI Training Program	Training	1,566
AP4GPU	Kevin Harms	Argonne National Laboratory	AutoPerf for Nvidia GPU	Computer Science	749
APSDataAnalysis	Rafael Vescovi	Argonne National Laboratory	APS Beamline Data Processing and Analysis	Computer Science	1,008
APSPolarisI2E	Nicholas Schwarz	Argonne National Laboratory	APS Beamline Processing and Workflows using ALCF Polaris I2E	Computer Science	3,504
ArgonneGPUAccess	Craig Stacey	Argonne National Laboratory	GPU access for Argonne researchers	Internal	181
athena	Sylvester Johannes Joosten	Argonne National Laboratory	ATHENA EIC Detector Simulation and Optimization	Physics	500

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
ATPESC2021	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing 2021	Training	2,001
ATPESC_Instructors	Raymond M. Loy	Argonne National Laboratory	Argonne Training Program on Extreme-Scale Computing for ALL Instructors	Training	201
AuroraPIC	Frank Shih Yu Tsung	University of California-Los Angeles	Code preparation for Aurora	Physics	1,000
AutoPhase	Yudong Yao	Argonne National Laboratory	Real-time X-ray coherent imaging with a self-trained neural network	Computer Science	485
BES-AXMAS	Charlotte Lisa Haley	Argonne National Laboratory	AI for quality control of single crystal X-ray scattering experiments	Materials Science	473
biolearning	Chongle Pan	University of Oklahoma	Development of large-scale biomedical machine learning models	Biological Sciences	1,021
BioMed	Yuri Alexeev	Argonne National Laboratory	Research hypotheses from existing biomedical papers	Biological Sciences	153
BioML	Yuri Alexeev	Argonne National Laboratory	Biomedical machine learning	Biological Sciences	2,000
BIP167	Philip Kurian	Howard University	Computing superradiance and van der Waals many-body dispersion effects for biomacromolecules	Physics	3,428
BirdAudio	Nicola Joy Ferrier	Argonne National Laboratory	Machine Learning for Classification of Birdsong	Computer Science	683
BLawB	Lucian Ivan	Canadian Nuclear Laboratories	Application of Maximum-Entropy Moment Methods to Turbulent and Multiphase Flow Prediction: Software Package Preparation	Engineering	600
BlazingSQLforHPC	Benjamin Hernandez Arreguin	Oak Ridge National Laboratory (ORNL)	Optimizing BlazingSQL for DOE's Leadership Computing Facilities	Computer Science	751

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
bloodflow_dd	Jifu Tan	Northern Illinois University (NIU)	Multiphysics modeling of biological flow with cell suspensions	Engineering	1,016
BNN-Scale	Murali Krishna Emani	Argonne National Laboratory	Optimizing Bayesian Neural Networks for Scientific Machine Learning Applications	Computer Science	1,840
BRAIN	Getnet Dubale Betrie	Argonne National Laboratory	Scalable Brain Simulator for Extreme Computing	Biological Sciences	3,013
BS-SOLCTRA	Esteban Meneses	Costa Rica National High Technology Center	Plasma Physics Simulations for SCR-1 Stellarator	Physics	3,052
bubble-ai	Ben J Blaiszik	Argonne National Laboratory	Discovery of Novel Fuel Cell Catalyst Materials via Development of High-Throughput AI-Guided Characterization Methods	Materials Science	83
Bubblecollapse	Eric Johnsen	University of Michigan	Dynamics and heat transfer in bubble collapse near solid surfaces	Engineering	947
CAIDS	Julio Cesar Mendez Carvajal	North Carolina State University (NCSU)	Consistent Averaging Procedure for solving the fundamental equations of fluid dynamics	Engineering	577
caloml	Douglas Paul Benjamin	Argonne National Laboratory	Fast simulation of the ATLAS calorimeter system with Machine Learning	Physics	865
candle_aesp	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	Virtual Drug Response Prediction	Biological Sciences	25,000
Carbon_composites	Hendrik Heinz	University of Colorado-Boulder	Designing Functional Nanostructures and Carbon-Based Composite Materials	Materials Science	100
Catalyst	Katherine M Riley,Christopher James Knight,James Clifton Osborn,Timothy Joe Williams	Argonne National Laboratory	Catalyst	Internal	15,303

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
ceed-app	Misun Min	Argonne National Laboratory	Aerosol Transport Modeling Towards Exascale	Engineering	6,459
cfdml_aesp	Kenneth Edward Jansen	University of Colorado-Boulder	Data Analytics and Machine Learning for Exascale CFD	Engineering	50
CharmRTS	Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia	University of Illinois at Urbana-Champaign	Charm++ and its applications	Computer Science	5,022
Clouds	Ian Foster	Argonne National Laboratory	Unsupervised analysis of satellite cloud imagery	Earth Science	54,767
compsensingADSP	Robert Hovden	University of Michigan	Dynamic Compressed Sensing for Real-time Tomographic Reconstruction	Materials Science	2,684
Comp_Perf_Workshop	Raymond M. Loy, Yasaman Ghadar	Argonne National Laboratory	ALCF Computational Performance Workshop	Training	2,000
connectomics_aesp	Nicola Joy Ferrier, Thomas David Uram	Argonne National Laboratory	Enabling Connectomics at Exascale to Facilitate Discoveries in Neuroscience	Biological Sciences	177
coreform-lattice	Greg Vernon	Coreform LLC	HPC-enabled geometry-compliant lattice structures for 3d printing and structural simulation	Engineering	1,113
covid-ct	Ravi Kiran Madduri	Argonne National Laboratory	Medical Imaging Domain-Expertise Machine Learning for Interrogation of COVID	Computer Science	5,384
CSC249ADCD502	Kenneth John Roche	Pacific Northwest National Laboratory (PNNL)	2.2.6.02 ADCD502 Application Assessment	Computer Science	2,000
CSC249ADOA01	Rick Lyndon Stevens, Thomas Scott Brettin	Argonne National Laboratory	2.2.4.03 ADOA01 CANDLE: Exascale Deep Learning Enabled Precision Medicine for Cancer	Biological Sciences	6,771

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
CSC250STDM14	Franck Cappello	Argonne National Laboratory	2.3.4.14 STDM14 - VeloC-SZ: Very Low Overhead Transparent Multilevel Checkpoint/Restart/SZ: Fast, Effective, Parallel Error-bounded Exascale Loss....	Computer Science	1,000
CSC250STDT11	Sunita Chandrasekaran, Dossay Oryspayev	Stony Brook University	2.3.2.11 SOLLVE: Scaling OpenMP with LLVm for Exascale	Computer Science	6,000
D6_XNET	Robert Fisher	University of Massachusetts-Dartmo uth	Explorations of the D6 Scenario of Type Ia Supernovae on Theta	Physics	2,000
datascience	Venkatram Vishwanath	Argonne National Laboratory	ALCF Data Science and Workflows Allocation	Internal	1,549
Deep_WF	Zhi Qiao	Argonne National Laboratory	AI-enabled real-time super-resolution X-ray wavefront sensing and advanced beamline control	Physics	1,079
dendritesegmentation	Ben J Blaiszik	Argonne National Laboratory	Machine Learning for Automated Dendrite Segmentation to Accelerate Experiments at the Advanced Photon Source	Materials Science	88
Diagnostics-CNN	Lander Ibarra	Argonne National Laboratory	Inverse Problem CNN Computation for Diagnostics	Energy Technologies	810
DL4VIS	Hanqi Guo	Argonne National Laboratory	Deep Learning for In Situ Analysis and Visualization	Computer Science	3,162
DLHMC	Sam Alfred Foreman	Argonne National Laboratory	Deep Learning HMC	Physics	6,099
dl_am	Satish Karra	Los Alamos National Laboratory (LANL)	Scalable Deep Learning Workflow for Additive Manufacturing	Engineering	914

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
DL_MODEX	MARUTI KUMAR MUDUNURU	Pacific Northwest National Laboratory (PNNL)	Towards a robust and scalable deep learning workflow for fast, accurate, and reliable calibration of watershed models	Earth Science	656
DNS_SV_Turb_2WC	Josin Tom	Duke University	DNS study of particle settling velocities in turbulence in the presence of two-way coupling	Engineering	577
DPCPPA100	Kevin Harms	Argonne National Laboratory	DPC++ on A100	Computer Science	2,508
Drug_FEP_Data	Wei Jiang	Argonne National Laboratory	Machine elarning of drug binding and toxicity based on high throughput free energy computations	Biological Sciences	1,062
ecp_omp2021	Colleen Elizabeth Bertoni	Argonne National Laboratory	ECP OpenMP Virtual Hackathon 2021	Computer Science	24
ECP_SDK	Sameer Suresh Shende	University of Oregon	Deploying the ECP SDK software stack at ALCF	Computer Science	1,980
EE-ECP	Xingfu Wu, Valerie Taylor	Argonne National Laboratory	Energy efficient tradeoff among execution time, power, and resilience of two ECP applications	Computer Science	1,526
ELSI_dev	Alvaro Vazquez Mayagoitia	Duke University	ELSI Developers	Computer Science	415
EngineDNS	Christos Frouzakis	Eidgenössische Technische Hochschule Zürich (ETH Zurich)	Towards reactive DNS in complex internal combustion engine geometries	Engineering	3,094
ES_AI	Himanshu Sharma	Pacific Northwest National Laboratory (PNNL)	Deep Neural Network Model for Modeling Aqueous Aerosol Chemistry for Climate Science	Earth Science	329
ExaLearn	Ian Foster	Argonne National Laboratory	ExaLearn applications and libraries	Computer Science	2,586

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
ExaPF	Michel Schanen	Argonne National Laboratory	Optimal Power Flow on GPUs	Mathematics	214
FASTMath2	Richard Kim Archibald	Oak Ridge National Laboratory (ORNL)	FASTMath Software Platform	Mathematics	2,000
FDTD_GPU	Wei Jiang	Argonne National Laboratory	Migration of FDTD modeling to GPU	Computer Science	2,000
fibregpu	Davide Di Giusto	University of Udine	DNS of fibre-laden turbulent channel flow	Engineering	577
Fornax_GPU	Adam Seth Burrows	Princeton University	Porting the Fornax Supernova Code to GPUs	Physics	1,052
gccy3	Marc Francis Paterno	Fermi National Accelerator Laboratory (Fermilab)	Cosmological Parameter Inference from Galaxy Clusters	Physics	2,129
gnn_uq	Shengli Jiang	University of Wisconsin-Madison	Molecular Property Uncertainty Quantification via Automated Graph Neural Networks	Computer Science	444
GNPMem	Tarak K Patra	Indian Institute of Technology Madras	Computational design of polymer grafted nanoparticle membrane	Materials Science	1,303
GPU-DG	Pinaki Pal	Argonne National Laboratory	GPU-enabled Discontinuous Galerkin simulations of complex fluid flows	Engineering	2,005
gpu_hack	Yasaman Ghadar, Raymond M. Loy	Argonne National Laboratory	GPU Hackathon	Training	2,000
GRACE	Sayan Ghosh	Pacific Northwest National Laboratory (PNNL)	Graph Analytics Codesign on GPUs	Computer Science	1,154
GrainBoundaries	Wissam A Saidi	University of Pittsburgh	Structure and Properties of Grain Boundaries in Materials for Energy Applications	Materials Science	4,368

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
GraphNeuralNetwork	Qi Yu	University of California-San Diego	Accelerating Traffic Simulation via Graph Neural Networks	Computer Science	1,967
hp-ptycho	Tekin Bicer	Argonne National Laboratory	High Performance 3D Ptychographic Reconstruction and Image Enhancement	Materials Science	2,622
IMEXLBM	Saumil Sudhir Patel	Argonne National Laboratory	ECP ProxyApp Development for the Lattice Boltzmann Method	Computer Science	2,213
img_ai_exp	Venkatram Vishwanath	Argonne National Laboratory	Benchmarking Xray and EM based Deep Learning AI expedition applications	Physics	2,000
inversedesign_oerorr	Rafael Gomez-Bombarelli	Massachusetts Institute of Technology (MIT)	Inverse design of oxide catalysts for OER/ORR	Materials Science	1,714
IonTransES-ML	Boris Kozinsky	Harvard University	IonTransES - Machine learning of quantum energies	Energy Technologies	3,800
IVEM-AIDefects	Logan Timothy Ward	Argonne National Laboratory	Realtime Analysis of Intermediate Voltage Electron Microscopy Images	Nuclear Energy	200
JCESR	Larry Curtiss, Anubhav Jain	Argonne National Laboratory	Development of High Throughput Methods	Materials Science	4,900
Job_Interference	Zhiling Lan	Illinois Institute of Technology (IIT)	Workload Interference Analysis on Theta	Computer Science	471
les120	George Karniadakis	Brown University	Learning the sub-grid model in Large Eddy Simulations using domain-decomposition based parallel physics-informed neural networks (PINNs)	Mathematics	1,013
LES_CommTransAir	Parviz Moin	Stanford University	Large-Eddy Simulation of a Commercial Transport Aircraft Model	Engineering	1,895
LIGHTCONTROL	Sandra Gail Biedron	University of New Mexico	Light sources and their control using AI techniques	Physics	488

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
LQCD-ML	Wai Nim Alfred Tang	Phd Tutor Hub	LQCD parametric regression by neural networks	Physics	1,108
LQCDdev	James Clifton Osborn	Argonne National Laboratory	Lattice QCD development	Physics	991
lqcdml_aesp	William Detmold	Massachusetts Institute of Technology (MIT)	Machine Learning for Lattice Quantum Chromodynamics	Physics	2,861
M4DA	Zichao Di	Argonne National Laboratory	Enabling Large-scale Multimodal Data Analysis for the APS	Materials Science	1,534
Maintenance	Mark R Fahey	Argonne National Laboratory	LCF Operations System Maintenance	Internal	1,998
matml_aesp	Noa Marom	Carnegie Mellon University	Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials	Materials Science	3,050
MBPTA100	Marco Govoni	Argonne National Laboratory	Efficient implementation of MBPT on heterogeneous architectures	Materials Science	559
MI2Dmaterials	Trevor David Rhone	Rensselaer Polytechnic Institute (RPI)	Materials informatics study of two-dimensional magnetic materials and their heterostructures	Materials Science	1,663
MILC_GPU	Steven Arthur Gottlieb	Indiana University (IU)	Exploring the muon anomalous magnetic moment using ThetaGPU nodes	Physics	1,384
ML-Coupling	shinhoo kang	Argonne National Laboratory	Data-driven Coupling Methods for Atmospheric-Ocean Interactions	Earth Science	1,447
MLP4THERMO	Cem Sevik	Eskisehir Technical University	Machine Learning Potentials for Thermal Properties of Two-Dimensional Materials	Materials Science	135
MLPerf-Science	Murali Krishna Emani	Argonne National Laboratory	Performance Evaluation of MLPerf Scientific Machine Learning Benchmarks	Computer Science	2,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
MOAB_App	Vijay Subramaniam Mahadevan	Argonne National Laboratory	MOAB Algorithmic Performance Portability	Mathematics	2,000
MPICH_MCS	Kenneth James Raffenetti, Pavan Balaji	Argonne National Laboratory	MPICH - A high performance and widely portable MPI implementation	Computer Science	531
MSEMNeuro	Thomas David Uram	Argonne National Laboratory	Reconstruction of neuronal connectivity from MSEM imaged tissue	Biological Sciences	1,720
MultiActiveAI	Dario Dematties	Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina	Multimodal Intelligence from Active Foveated Vision	Computer Science	2,624
MVAPICH2	Dhabaleswar Kumar Panda	The Ohio State University	Optimizing and Tuning MVAPICH2-GDR Library and Study Its Impact on HPC and AI Applications	Computer Science	2,588
nanoct	Viktor Valerievich Nikitin	Argonne National Laboratory	Multi-GPU nanoscale 3D reconstructions of dynamically evolving materials	Materials Science	2,682
nek52rs	Aleksandr V. Obabko	Argonne National Laboratory	Nek5000/NekRS for NRC and COVID LES	Engineering	470
Nek5RS_covid	Aleksandr V. Obabko	Argonne National Laboratory	Nek5000/NekRS for COVID particle transport and NRC	Engineering	1,500
networkbench	Devi Sudheer Kumar Chunduri, Elise Jennings, Kevin Harms, Misbah Mubarak	Argonne National Laboratory	network benchmarking and modeling	Computer Science	471
neutrino_osc_ADSP	Marco Del Tutto	Fermi National Accelerator Laboratory (Fermilab)	Machine Learning for Data Reconstruction to Accelerate Physics Discoveries in Accelerator-Based Neutrino Oscillation Experiments	Physics	1,342

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
NGM_EHT	Charles Forbes Gammie	University of Illinois at Urbana-Champaign	Next-generation Models for the Event Horizon Telescope	Physics	2,000
niubmrk	Sergey A Uzunyan	Northern Illinois University (NIU)	Benchmarks of applications running at NIU compute clusters using modern hardware	Computer Science	50
NNM	Yongchao Yang	Michigan Technological University	Deep Learning for Strongly Nonlinear Dynamical Systems	Engineering	2,503
NODE-PETSc	Hong Zhang	Argonne National Laboratory	Training neural ordinary equations using PETSc	Mathematics	2,000
novacosmics	Alexander I Himmel	Fermi National Accelerator Laboratory (Fermilab)	NOvA Cosmic Rejection	Physics	2,232
NSCS	Tanwi Mallick	Argonne National Laboratory	Towards Neighborhood Scale Climate Simulations using AI and Accelerated GPUs	Computer Science	114
Nucleant	John J Low	Argonne National Laboratory	Multi-scale modeling of nucleant evolution and alloy solidification for manufacturing of high-strength parts from aluminum alloys	Materials Science	6,926
NuQMC	Alessandro Lovato	Argonne National Laboratory	Nuclear quantum Monte Carlo	Physics	2,000
Operations	Mark R Fahey	Argonne National Laboratory	Systems administration tasks	Internal	1,566
OptADDN	Sandeep Madireddy	Argonne National Laboratory	Optimal Architecture discovery for deep probabilistic models and neuromorphic systems	Computer Science	4,000
oqs_int	David Beratan	Duke University	Efficient Simulation of Open Quantum Systems with Matrix Product States in the Interaction Picture	Chemistry	2,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
Pagoda-GPU	Paul Hamilton Hargrove	Lawrence Berkeley National Laboratory (LBNL)	Pagoda GPU development for ECP WBS 2.3.1.14	Computer Science	1,926
PARTURB3D	Ramesh Balakrishnan	Argonne National Laboratory	Simulating turbulent particulate flows inside enclosures	Engineering	2,000
PDE_ML	Ramin Baghgar Bostanabad	University of California-Irvine	Self-supervised Coupling of Deep Operator Surrogates for Scalable and Transferable Learning	Engineering	2,941
Performance	Scott Parker, Raymond M. Loy	Argonne National Laboratory	Performance	Internal	1,566
poly-ion-dd	Juan Pablo	The University of Chicago (UChicago)	Ion Conduction Through Polymeric Interfaces	Chemistry	1,472
PPE-CV-NTM	Sandra Jean Bittner	Individual	Computational Study and Visualization Models for Non-Traditional Materials for Personal Protective Equipment	Computer Science	8,900
ptychopy	Ke Yue	Argonne National Laboratory	Ptychopy: a framework for ptychographic data processing	Computer Science	2,000
PUR-IRL	Nicholas Lee-Ping Chia	Mayo Clinic-Minnesota	Inferring the Reward Function of Cancer	Biological Sciences	3,967
QCProxyApps	Graham Donald Fletcher	Argonne National Laboratory	Quantum Chemistry Proxy Applications	Chemistry	558
qsars_qm_vae	Brad Reisfeld	Colorado State University	Discovering quantitative structure activity relationships using quantum chemical descriptors and variational autoencoding	Biological Sciences	263
QSim	Yuri Alexeev	Argonne National Laboratory	Quantum Simulations	Computer Science	1,276

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
QuantumDS	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Quantum mechanics and Data Science	Chemistry	951
radix-io	Philip Hutchinson Carns	Argonne National Laboratory	System software to enable data-intensive science	Computer Science	112
RAPINS	Eliu Antonio Huerta Escudero	Argonne National Laboratory	Reproducible and Accelerated Physics-inspired Neural Networks	Physics	1,017
RaptorX	Jinbo Xu	Toyota Technological Institute at Chicago (TTIC)	Protein Folding through Deep Learning and Energy Minimization	Biological Sciences	5,394
rdesim	Venkatramanan Raman	University of Michigan	GPU-based Simulation of Shock-containing Flows	Engineering	4,000
Redox_ADSP	Logan Timothy Ward	Argonne National Laboratory	Autonomous Molecular Design for Redox Flow Batteries	Materials Science	234
RL-fold	Arvind Ramanathan	Argonne National Laboratory	Targeting intrinsically disordered proteins using artificial intelligence driven molecular simulations	Biological Sciences	12,238
rnn-robustness	Liam Benjamin Johnston	University of Wisconsin-Madison	Large-scale Factorial Experiment on RNN Robustness	Computer Science	217
Rxn_Diff	Neeraj Rai	Mississippi State University	Modeling reaction-diffusion processes for liquid-phase heterogeneous catalysis	Chemistry	2,000
sbi-fair	Pete Beckman,Kamil Antoni Iskra	Argonne National Laboratory	FAIR Surrogate Benchmarks Supporting AI and Simulation Research	Computer Science	950
SDL2021x	Wai Nim Alfred Tang	Phd Tutor Hub	SDL2021 Extension	Computer Science	47
SDL_Workshop	Raymond M. Loy	Argonne National Laboratory	ALCF Simulation, Data, and Learning Workshop	Training	2,000

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
SEEr-planning	Zhiling Lan	Illinois Institute of Technology (IIT)	Performance and Power Tradeoff Analysis of AI-Enabled Science on CPU-GPU System	Computer Science	32,400
SI_IBM	Saurabh Chawdhary	Argonne National Laboratory	Scalable, efficient sharp-interface immersed boundary method for fluid-structure interaction problems	Engineering	2,808
skysurvey_adsp	George Frazer Stein	Lawrence Berkeley National Laboratory (LBNL)	Learning Optimal Image Representations for Current and Future Sky Surveys	Physics	1,342
smarthpc	Zhengchun Liu	Argonne National Laboratory	System level approach to optimize neural architecture search in HPC environments	Computer Science	3,769
SolarWindowsADSP	Jacqueline Manina Cole	University of Cambridge	Data-Driven Molecular Engineering of Solar-powered Windows	Materials Science	2,094
SOLLVE	Sunita Chandrasekaran	Brookhaven National Laboratory (BNL)	Scaling OpenMP With LLVm for Exascale Performance and Portability	Computer Science	243
SuperBERT	Ian Foster	Argonne National Laboratory	Training of language models on large quantities of scientific text	Computer Science	19,408
SWIFT_CRADA	Venkatram Vishwanath	Argonne National Laboratory	Anomaly Detection for Swift Transaction Streams	Computer Science	171
THGSupport	Kevin Harms	Argonne National Laboratory	The HDF Group Support	Computer Science	245
TMEM_DEL	Diomedes Elias Logothetis	Northeastern University	Molecular Dynamics simulation on TMEM16A chloride channel	Biological Sciences	1,154
TomoADSP	Robert Hovden	University of Michigan	Dynamic Compressed Sensing for Real-time Tomographic Reconstruction	Materials Science	4,483

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
TomoEncoders	Aniket Tekawade	Argonne National Laboratory	TomoEncoders: Computer Vision Framework for 4D X-ray Tomography	Engineering	1,154
Tools	Scott Parker	Argonne National Laboratory	ALCF Performance Tools	Internal	1,000
TORCHANIA100	Alvaro Vazquez Mayagoitia	Argonne National Laboratory	Distributed training of Deep Neural Networks of the ANI Neural Network Potentials	Chemistry	896
Turbulent_Transport	Paul Fischer	Argonne National Laboratory	Turbulent Transport for Multiphysics Applications	Engineering	476
Uintah_GPU_benchmark	Marta Garcia Martinez	Argonne National Laboratory	Uintah benchmarks on GPU	Engineering	2,000
Ultrafast_X-ray	Jin Wang	Argonne National Laboratory	Ultrafast_X-ray	Engineering	3,276
User_Services	Haritha Siddabathuni Som, Sreeranjani Ramprakash	Argonne National Laboratory	User Services	Internal	54
VASPDEK	David Eugene Keller	University of Rochester	Vasp 2.0.1 GPU timing Simulations	Physics	0
VeloC	Bogdan Florin Nicolae	Argonne National Laboratory	VeloC: Very Low Overhead Checkpointing System	Computer Science	2,416
visualization	Joseph A Insley, Michael E. Papka	Argonne National Laboratory	Visualization and Analysis Research and Development for ALCF	Internal	1,566
wall_turb_dd	Ramesh Balakrishnan	Argonne National Laboratory	Wall Resolved Simulations of Canonical Wall Bounded Flows	Engineering	1,305
WaterHammer	Hong Zhang, Hong Zhang	Argonne National Laboratory	Water Hammer Simulation	Mathematics	1,846

Project Name	PI Name	PI Institution	Project Title	Science Field (Short)	Allocation Amount in Node-Hours
WGSanalysis	Elizabeth McNally	Northwestern University	Large scale alignment and analysis of whole human and mouse genomes, with focus on realigning to HG 38 and harmonized workflows	Biological Sciences	2,000
WRFGPU	Veerabhadra Rao Kotamarthi	Argonne National Laboratory	WRF GPU testing	Earth Science	1,498
WRLCMF	Krishnan Mahesh	University of Minnesota-Twin Cities	Wall-Resolved LES of complex maneuvering flows	Engineering	3,118
				Total DD	450,546

Appendix C – ALCF CY2021 Science Highlights

The following Science Highlights were submitted to ASCR for the 2021 OAR performance period.

Atomistic Insights Into Metal Hardening

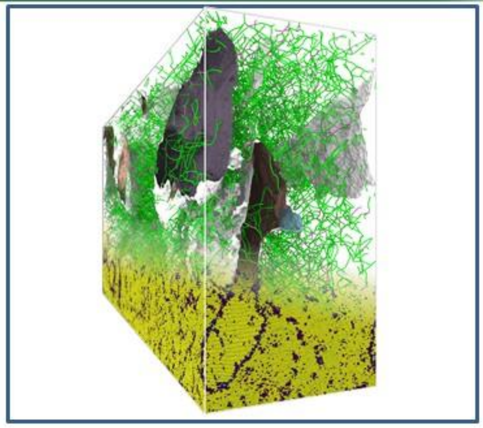
The Science

Although known since the earliest days of metallurgy that metals get stronger or harden when mechanically deformed, mechanisms of metal hardening continue to be investigated with tenacity to develop a predictive theory based on intrinsic properties of a material. While known that the motion of crystal lattice dislocations where layers of atoms slide over each other play an important role in the process, debates on mechanisms continued because of an inability to observe dislocations in situ during straining. Large-scale molecular simulations enable examining atomic level details of metal hardening and demonstrated staged hardening of metals is direct consequence of crystal rotation under uniaxial straining.

The Impact

While known for years of the connection between crystal plasticity and dislocation defects, no quantitative theory currently exists to predict metal hardening directly from the intrinsic properties of a material. From the detailed atomistic data collected from large-scale molecular simulations, the team observed that the basic mechanism is the same across all stages of metal hardening. This is at odds with widely divergent and contradictory views in literature, but the simulations clarify that three-stage hardening is not an intrinsic material property. The ability to simulate crystal reorientation at very large strains and identify the most likely slip deformation is important for improving the accuracy of engineering models used to predict the evolution of polycrystalline textures in several industrial processes, such as rolling, forging, extrusion, and other metal-forming operations.

LLNL, TU Darmstadt, UC Berkeley, Stanford



A data reduction workflow in which a detailed all-atom representation of compressed crystal (bottom) gradually morphs into a more economical representation (top) in terms of its lattice defects - dislocation lines and interfaces of twin particles. (Image courtesy of Alexander Stukowski).

Contact PI: Vasily V. Bulatov
ASCR Allocation PI: Vasily V. Bulatov
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: January 27, 2021
Publication(s) for this work: L.A. Zepeda-Ruiz, et. al., *Nat. Mater.* (2020).

U.S. DEPARTMENT OF ENERGY | Office of Science

NNSA ASC & TU Darmstadt

The Last Journey

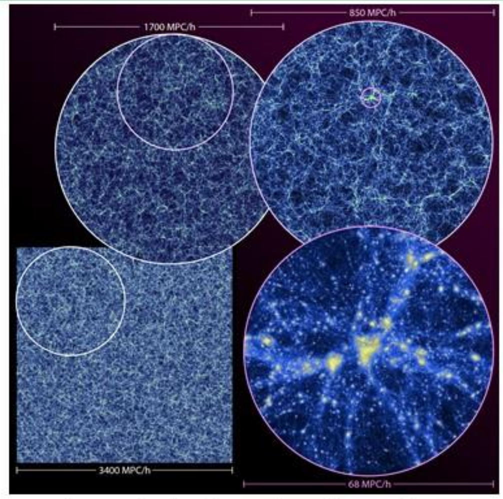
An Extreme-Scale Simulation on the Mira Supercomputer

The Science

Some of the most important observational tools for cosmologists trying to understand the content and history of the universe are measurements of the growth and distribution of structures on the largest scales, including how galaxies group into clusters and how those clusters are distributed. These measurements are used to determine how the content of the universe is divided between normal (baryonic) matter, dark matter, and dark energy, and also to try to understand more about the physical properties of dark matter and dark energy. Enormous current and near future astronomical sky surveys can measure positions of millions or billions of galaxies, and massive simulations of cosmological structure formation have become indispensable tools for interpreting the measurements, controlling systematic uncertainties well enough to use the statistical power of the data, and looking for tensions between our current theoretical understanding of cosmology and the signals written on the sky.

The Impact

The team used almost 400 million core-hours over the last six months of operation on the Mira supercomputer at ALCF to generate and analyze the Last Journey simulation, one of the largest cosmological structure formation simulations ever performed. The Hardware/Hybrid Accelerated Cosmology Code (HACC) was used to simulate the movements of 1.24 trillion particles tracing the mass distribution in the universe over cosmic time. HACC's CosmoTools infrastructure was used to perform the largest data analysis and reduction operations in-situ, dramatically reducing the amount of data storage required while still producing a rich set of outputs that will be used to support a wide variety of cosmological measurements. Outputs from the Last Journey simulation will be used to help with planning and analysis of current and upcoming sky surveys including the Dark Energy Spectroscopic Instrument (DESI), the Legacy Survey of Space and Time (LSST), the Stage 4 Cosmic Microwave Background experiment (CMB-S4), and the SPHEREx space telescope.



Zoomed views of a thin density slice of the simulation volume, ending with the largest cluster in the simulation. (Heitmann et al 2021)

Contact PI: Katrin Heitmann
ASCR Allocation PI: Katrin Heitmann
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: February 2021
Publication(s) for this work: Heitmann et al., ApJS, 252, 19 (2021)

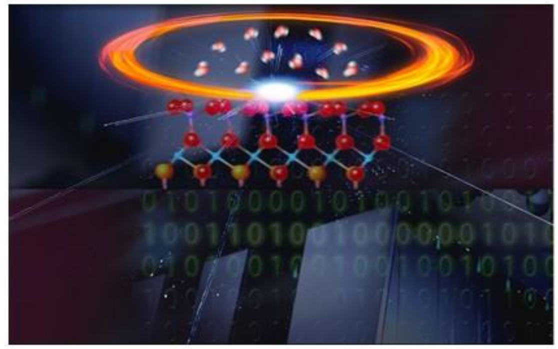
Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

Validating first-principles molecular dynamics calculations of oxide/water interfaces with x-ray reflectivity data

The Science

First-principles electronic structure calculations can reveal unique insights into molecular, structural, and reactive processes when combined with molecular simulation algorithms. However, such simulations rely on several design choices, whose validation is often a challenging and complex task requiring a means to confirm that they do indeed describe the real world. Properly validating simulation protocols is important to predict accurate structural models, and in turn requires careful experimental validation. Building upon a systematic testing of theoretical and numerical approximations and assessing the impact on accuracy, a general, quantitative validation protocol was developed for first-principal molecular simulations of oxide/aqueous interfaces using x-ray reflectivity (XR) data.



Pictorial representation of joint experimental and computational study of the atomic structure of interfaces. The study utilized the Advanced Photon Source (upper panel and Argonne Leadership Computing Facility (lower panel). (Emmanuel Gygi, UCSD).

The Impact

Metal oxide/water interfaces play a crucial role in electrochemical and photocatalytic processes, including water splitting and CO2 reduction to create fuel from sunlight. The process of developing a validation protocol highlighted the need for choosing appropriate numerical algorithms and for accurate first principles electronic structure theories to properly describe interfacial interactions; indeed, the XR data are sensitive not only to atomic structure, but also to the electron-density distributions. It was demonstrated that electronic structure simulations can accurately reproduce XR data. In order to validate the interactions between atoms and electrons at the interface, using XR data, accurate predictions not only of the surface structure but also of the bulk substrate and bulk water are required, since bulk and interfacial regions are probed simultaneously.

Contact PI: Giulia Galli
ASCR Allocation PI: Giulia Galli
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: March 15, 2021
Publication(s) for this work: K.J. Harmon et. al., *Phys. Rev. Materials* **4**, 113805 (2020).

U. Chicago, Argonne, UC Davis, James Madison University, Northwestern University

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE BES, ANL LDRD, DOD NDSEG, plus others

Reliable and practical computational description of molecular crystal polymorphs

The Science

Individual molecular properties are just one fundamental problem in computer-aided drug design. How those molecules are arranged when forming solids highly determine their resultant properties, such as their bioavailability. Further complicating the matter is the potential for molecules to form two or more crystal structures known as polymorphs. The prediction of crystal polymorphs is indeed very challenging due to the high dimensionality of the problem and the relatively small free energy differences (within 1 kJ/mol) between crystal polymorphs. By coupling two methods proven successful in blind test competitions, a hierarchical workflow was created providing a robust and computationally feasible procedure to accurately determine structures and stabilities of molecular crystals.

The Impact

The determination of molecular crystal structures is of high interest for drug design and nanotechnological applications, such as semiconductor synthesis, hybrid perovskites for solar energy production, and nano scale machines. As polymorphs often have different, potentially unwanted, properties, confirming the most thermodynamically stable solid structure of a pharmaceutical drug candidate enables an informed and critical assessment of the potential risk associated with the assumed stable form transforming in undesired ways. The large number of thermodynamically relevant conformations found with this capability also provides an excellent source for accurate datasets that can be used to further accelerate materials discovery with data-driven and machine learning techniques.

Crystal structure prediction workflow based on the GRACE software for sampling crystal structures and the first principles-based DFT+MBD framework. (Alexandre Tkatchenko, University of Luxembourg).

Contact PI: Alexandre Tkatchenko
ASCR Allocation PI: Alexandre Tkatchenko
ASCR Program/Facility: DD/ALCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: April 19, 2021
Publication(s) for this work: J. Hoja et. al., *Sci. Adv.* **5**, eaau3338 (2019).

U. Luxembourg, Princeton, Avant-garde Materials Simulation Deutschland GmbH, Cornell

U.S. DEPARTMENT OF ENERGY | Office of Science

DFG, ERC, DOE BES, Cornell University

High-fidelity gyrokinetic simulation of tokamak and ITER edge physics

The Science

Successful operation of ITER, the world's largest magnetic fusion experiment under construction in France, to prove the feasibility of fusion as a virtually inexhaustible source of carbon-free energy for generating electricity, requires avoiding damage to the material surface of the divertor plates from deposition of excessively narrow exhaust plasma heat loads. A simple extrapolation from present experimental results to ITER predicts a dangerously narrow heat-load width for ITER's production of 10-times more energy than the input energy. Since ITER is much larger in size and far stronger in magnetic field strength, its heat-load width may not be subject to the same physics as present tokamaks and the simple data extrapolation may be subject to large uncertainty. A first-principles-based prediction for the heat-load width could reduce this dangerous uncertainty. Such a study is now possible due to the availability of DOE resources.

The Impact

Establishing an accurate predictive formula for the exhaust heat-load width of future doughnut-shaped tokamak fusion reactors can allow scientists to operate ITER in a more comfortable and cost-effective way toward the goal of 0.5 GW of fusion power production from 50 MW of input power. A more accurate formula can also allow for more reliable design of future fusion reactors, which suffer from the limitation imposed by exhaust heat-load width on the divertor plates. The team simulated three ITER plasmas shown in the figure. The 10MA case has been obtained using all of Theta. Without this new critical data point, there would have been a large gap between the lowest and highest data-points and hence it would not have been possible to verify the correctness of the machine learning formula.

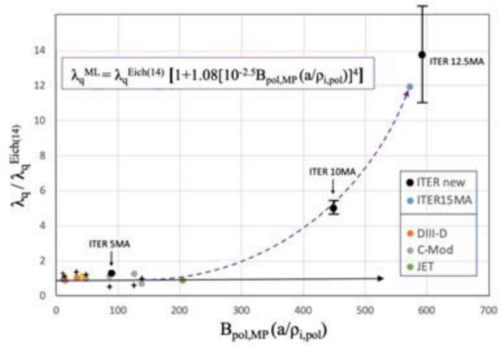


Image caption (credit: CS. Chang). A supervised machine-learning program finds the hidden kinetic parameter $a/\rho_{i,pol}$ and a simple predictive formula for $\lambda_q/\lambda_q^{Eich}$ where λ_q is the divertor heat-load width normalized using the Eich regression formula. Three more ITER simulations (black symbols) have been performed to verify the validity of the new ML-found formula (black dots). The 10MA case has been obtained from Theta, utilizing 4,352 nodes.

Contact PI: CS Chang (Princeton Plasma Physics Laboratory)
ASCR Allocation PI: CS Chang
ASCR Program/Facility: INCITE/ALCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: June 2021
Publication: Chang et al, *Physics of Plasma*, 2021,
doi: https://doi.org/10.1063/5.0027637.

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE-FES, DOE-ASCR

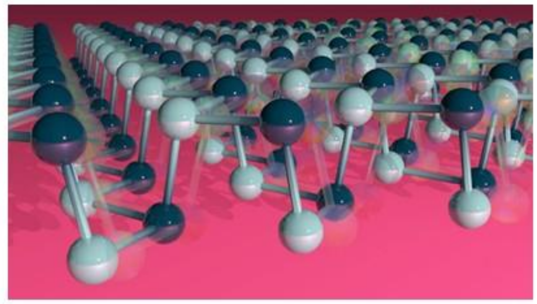
Optimized structure and electronic band gap of monolayer GeSe from quantum Monte Carlo methods

The Science

2D semiconductor nanomaterials are good candidates for use in electronic and optical sensors given their unique properties and tunability via altering their structure, such as pulling, pushing, or twisting their geometries. Quantum Monte Carlo (QMC) methods were used to find the structure and electronic band gap of 2D GeSe, determining that the gap and its nature are indeed highly tunable by strain. The geometry of the material was undetermined experimentally and out of reach of standard/common theoretical methods. Using a newly developed computational technique, the detailed atomic structure was determined, from which accurate optical properties of the 2D nanomaterial GeSe were obtained.

The Impact

GeSe is considered a promising material for devices used to detect the presence of light, such as solar cells and photodetectors. Highly accurate QMC methods were used to obtain the full geometry of a complex 2D nanomaterial for the first time. The newly developed QMC-based algorithm can accurately determine the structure of a material without calculating the atomic forces. The high tunability with strain of the band gaps indicates potential optical applications in this class of materials. This work also clearly demonstrated the need for highly accurate structural and electronic structure methods to reliably assess the properties of these materials in order to exploit them in future applications.



Optimized geometry of GeSe monolayer using the newly developed structural optimization algorithm within Quantum Monte Carlo (colored structure) compared to the initial Density Functional Theory optimized structure (clear structure). (Janet Knowles, Joseph Insley, Silvio Rizzi, and Victor Mateevitsi)

Contact PI: Olle Heinonen
ASCR Allocation PI: Paul R.C. Kent
ASCR Program/Facility: ALCF / OLCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: June 1, 2021
Publication(s) for this work: H. Shin, et. al., *Phys. Rev. Mater.* 5, 024002 (2021).

Argonne National Laboratory and Oak Ridge National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE BES CMS Center

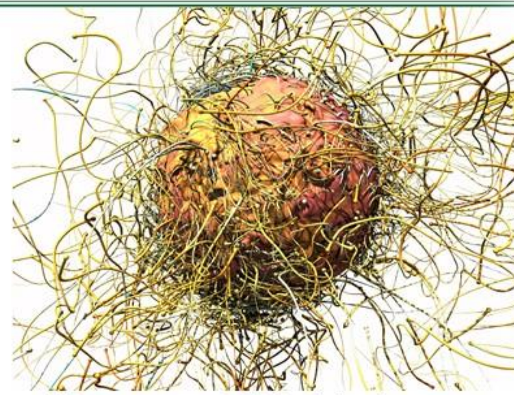
Towards a Definitive Model of Core-Collapse Supernova Explosions

The Science

Core-collapse supernovae dramatically announce the death of massive stars and the birth of neutron stars. These violent explosions, which produce the highest densities of matter and energy in the universe, are responsible for creating most of the elements in nature. The neutrino signal they emit carries information about the nuclear equation of state and the strength of their explosion is sensitive to how both the neutrinos and ultra-dense matter are treated. A fundamental theoretical understanding of such explosions is needed to advance research in nuclear and particle physics, and inform the interpretation of data from large-scale experiments.

The Impact

To shed light on this mysterious cosmological phenomenon, a research team led by Princeton University is using ALCF supercomputers to complete and analyze a suite of 19 simulations spanning a broad range of progenitor masses and structures, the largest such collection of 3D investigation in the history of core-collapse theory. The researchers are using these simulations for several studies, including exploring the impact of spatial resolution on the outcome and character of 3D supernova simulations, the strength and persistence of proto-neutron star convection, the neutrino and gravitational-wave signatures of core-collapse, and the origin of pulsar spins and kicks. The team's efforts to advance the fundamental understanding of supernova explosions will benefit ongoing research to determine the origin of the elements in the universe, measure gravitational waves, and interpret laboratory nuclear reaction rate measurements in light of stellar nucleosynthesis.



Newly-born neutron star left behind when the inner regions of a 25 solar-mass star explodes. The tangled "hair" are matter trajectories. Inner "sphere" is an isodensity surface colored by electron fraction. Image: Joseph Insley (ANL)

Contact PI: Adam Burrows
ASCR Allocation PI: Adam Burrows
ASCR Program/Facility: INCITE 2019/ALCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: July 2021
Publication for this work: A. Burrows and D. Vartanyan, Core-collapse supernova explosion theory. *Nature*, **589,** 29-39 (2021)

Princeton University, LANL, LLNL, LBNL

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE, NSF, Princeton IAS

Dielectric Constant of Liquid Water Determined with Neural Network Quantum Molecular Dynamics

The Science

The large dielectric constant of water is responsible for several physical properties, such as its excellent solvation properties and importance in biological systems. It is a challenging quantity to accurately predict in fully quantum mechanical calculations in part due to the need of very long trajectories to reach convergence. Based on training data from a highly accurate quantum-mechanical method, DFT-SCAN, neutral network (NN) molecular dynamics simulations coupled with an NN model to compute polarization yields an excellent description of liquid water. The computed dielectric constants are in good agreement with experiment over a range of temperatures.

The Impact

The accurate prediction of electrical polarization is essential for understanding the physical and chemical behavior in liquids and materials. The developed scalable scheme coupling multiple NNs enables modeling polarization in quantum accuracy. It is applicable to a range of systems, including other polar molecular liquids in biomedical applications, as well as nanoelectronics with potential applications in information and energy technologies. The physics-based integrated approach combines the accuracy of quantum molecular dynamics with the computational efficiency of empirical forcefield models and can be further improved by systematically increasing the quality of the training data for NNs. Computational speedups of ~5,000x were observed for small systems of water in this work and increase with system size.

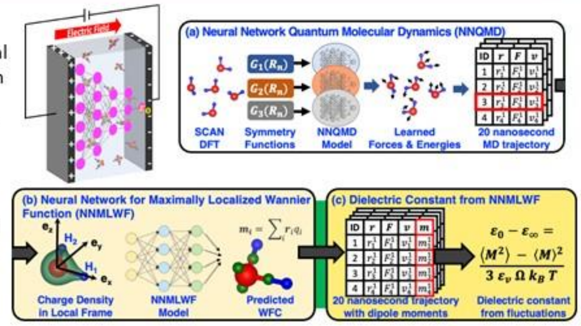


Illustration of scalable workflow for large-scale neural network molecular dynamics simulations of polarizable liquids and materials.
(Image created by Aravind Krishnamoorthy, courtesy of Priya Vashishta)

Contact PI: Priya Vashishta
ASCR Allocation PI: Priya Vashishta
ASCR Program/Facility: ALCF
ASCR PM: Sonia Sachs
Date submitted to ASCR: July 26, 2021
Publication(s) for this work: A. Krishnamoorthy, et. al., *Phys. Rev. Lett.* 126, 216403 (2021).

USC, Kumamoto University, ANL

U.S. DEPARTMENT OF ENERGY Office of Science

DOE BES CMS Center, ANL

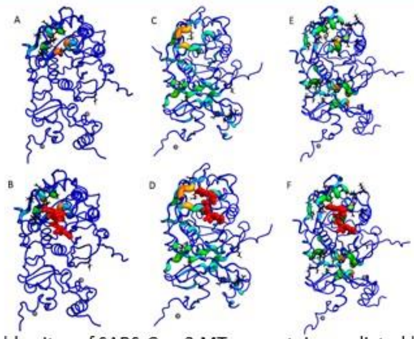
Finding druggable sites in SARS-Cov-2 proteins using molecular dynamics and machine learning

The Science

Coronavirus disease 2019 (COVID-19) is caused by a novel coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and was declared a pandemic by the World Health Organization on March 11, 2020. To date, only a few repurposed drugs have shown limited benefits in critically ill patients. The challenge here is the ability to accurately and efficiently determine where the drug-binding sites are located on target proteins. A molecular biophysics group at Johns Hopkins School of Medicine developed a joint computational and experimental approach using machine learning to accelerate novel drug discovery, which was able to locate known small-molecule binding sites as well as predict locations of sites not previously observed in experimental structures of SARS-Cov-2 and other proteins.

The Impact

General oral medications are expected to present significant usage flexibility and reduced manufacturing/transportation/storage cost than antibodies for COVID-19 control. Development of a high fidelity, high-resolution all-atom simulation and modeling methodology that can predict all drug binding sites as well as their local conformations is a key step towards rational drug design. The TACTICS workflow developed here, which is capable of detecting "cryptic" binding sites that are difficult to detect without a binding ligand, opens the door for identifying potential druggable sites. TACTICS does not require specialized simulations, can analyze both single structures and pre-computed trajectories, is self-contained within a single software package, and freely available for the research community to find druggable sites in other proteins.



Druggable sites of SARS-Cov-2 MTase protein predicted by TACTICS workflow: open pockets (top) and ligand-bound pockets (bottom) with residues of drug binding sites and ligand emphasized in bold color.
(Image reproduced from Figure 3 of *J. Chem. Inf. Model*, **6**1, 2897 (2021))

Contact PI: Albert Lau
ASCR Allocation PI: Albert Lau
ASCR Program/Facility: ALCF
ASCR PM: Benjamin Brown
Date submitted to ASCR: August 16, 2021
Publication(s) for this work: D. Evans, et. al., *J. Chem. Inf. Model* **61**, 2897 (2021).

John Hopkins School of Medicine

U.S. DEPARTMENT OF ENERGY | Office of Science

NIH

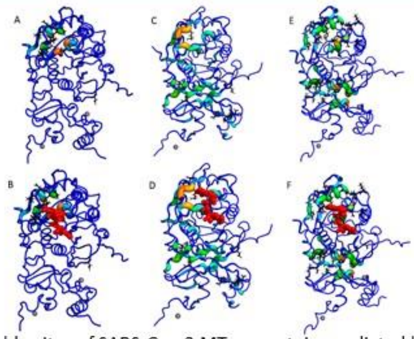
DLIO: A Data-Centric Benchmark for Scientific Deep Learning Applications

The Science

Coronavirus disease 2019 (COVID-19) is caused by a novel coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and was declared a pandemic by the World Health Organization on March 11, 2020. To date, only a few repurposed drugs have shown limited benefits in critically ill patients. The challenge here is the ability to accurately and efficiently determine where the drug-binding sites are located on target proteins. A molecular biophysics group at Johns Hopkins School of Medicine developed a joint computational and experimental approach using machine learning to accelerate novel drug discovery, which was able to locate known small-molecule binding sites as well as predict locations of sites not previously observed in experimental structures of SARS-Cov-2 and other proteins.

The Impact

General oral medications are expected to present significant usage flexibility and reduced manufacturing/transportation/storage cost than antibodies for COVID-19 control. Development of a high fidelity, high-resolution all-atom simulation and modeling methodology that can predict all drug binding sites as well as their local conformations is a key step towards rational drug design. The TACTICS workflow developed here, which is capable of detecting "cryptic" binding sites that are difficult to detect without a binding ligand, opens the door for identifying potential druggable sites. TACTICS does not require specialized simulations, can analyze both single structures and pre-computed trajectories, is self-contained within a single software package, and freely available for the research community to find druggable sites in other proteins.



Druggable sites of SARS-Cov-2 MTase protein predicted by TACTICS workflow: open pockets (top) and ligand-bound pockets (bottom) with residues of drug binding sites and ligand emphasized in bold color.
(Image reproduced from Figure 3 of *J. Chem. Inf. Model*, **6**1, 2897 (2021))

Contact PI: Venkatram Vishwanath
ASCR Allocation PI: Venkatram Vishwanath
ASCR Program/Facility: ALCF
ASCR PM: Benjamin Brown
Date submitted to ASCR: August 16, 2021
Publication(s) for this work: H. Devarajan, et. al., *JEEE/ACM Symposium on Cluster, Cloud and Internet computing* (2021).

Argonne National Laboratory and Illinois Institute of Technology

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE, NSF

Expanding the Scale of Urban Building Energy Modeling

The Science

The energy consumption of buildings can be modeled and simulated to gain insights into possible energy efficiencies that would help to reduce carbon emissions. However, the scale of the problem is vast. The simulation of a typical urban area involves 10,000's of buildings and each building model involves approximately 3,000 parameters. Researchers have now harnessed supercomputers using their software suite, Automatic Building Energy Modeling (AutoBEM), to rapidly create building energy models and study them in unprecedented detail providing utilities and industry with the information they need to improve energy efficiency.

A digital twin of Chattanooga, Tennessee processed with the AutoBEM software. All 178,337 buildings and simulation results can be searched, selected, and visualized using flexible regular expressions at bit.ly/virtual_epb (Image courtesy of Joshua New).

The Impact

In the United States, residential and commercial buildings account for 40% of energy consumption, 73% of electrical consumption, and 39% of emissions. In 2019 alone, the energy bill for the 125 million buildings in the US was $412 billion. As the threat of climate change looms, the US Department of Energy has set aggressive goals for energy-efficiency — a 30% reduction in energy use intensity (in units of kilowatt-hours per square foot) of all US buildings by 2030 compared to 2010. In their studies, researchers found that 99% of Chattanooga buildings would benefit from the efficiency technologies evaluated with an average of $28,500 saved per building per year. If implemented, such efficiencies could reduce greenhouse gas emissions and replace the need for new power plants.

Contact PI: Joshua New
ASCR Allocation PI: Joshua New
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Benjamin Brown
Date submitted to ASCR: September 13, 2021
Publication(s) for this work: B. Bass, J. New, and W. Copeland, *Energies* **14**, 132 (2021).

Oak Ridge National Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE BTO & OE

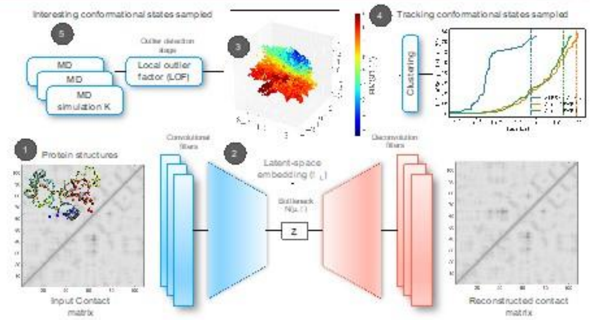
Stream-AI-MD: Streaming AI-driven Adaptive Molecular Simulations for Heterogeneous Computing Platforms

The Science

The research integrates scientific simulations with AI wherein near real-time feedback between the two approaches are used to understand how two proteins in the SARS-CoV-2 viral genome, nsp10 and nsp16, interact to help the virus replicate and elude the host's immune system. The team developed Stream-AI-MD, a novel application of deep learning to drive adaptive molecular dynamics (MD) simulations in a streaming manner from the ThetaGPU onto the Cerebras CS-1 AI accelerator platform to simultaneously analyze how the two proteins interact.

The Impact

Two scientific use-cases were used to evaluate the efficacy of Stream-AI-MD: (1) folding a small prototypical protein, namely $\beta\beta\alpha$-fold (BBA) FSD-EY and (2) understanding protein-protein interaction (PPI) within the SARS-CoV-2 proteome between two proteins, nsp16 and nsp10. By coupling simulations with AI/ML approaches, a speedup of at least 50X (i.e., time-to-solution) was achieved as well as enhanced sampling of complex conformational landscapes as quantified by the fraction of conformational states sampled with respect to equilibrium simulations (offering an additional order-of-magnitude speedup). Given the importance of the ongoing Coronavirus disease (COVID-19) pandemic, this approach can provide the necessary infrastructure to drive novel therapeutic discovery. Stream-AI-MD also has implications for how AI and high-performance computing (HPC) workflows will intersect in the future and on the design of future systems for AI-driven science.



Stream AI-MD simulations. (1) Protein trajectories are fed as contact map representations which are passed through (2) convolutional filters and then compressed using a dense layer to obtain (3) a latent representation of the protein conformational landscape. Reconstruction of the contact maps is constructed via deconvolution filters. The latent representation is then used to (4) track the conformational states sampled thus far, and (5) outlier conformations detected from these states using LOF are fed back to continue/restart productive simulations and kill unproductive simulations.

Contact PI: Arvind Ramanathan
ASCR Allocation PI: Arvind Ramanathan
ASCR Program/Facility: ALCC/ALCF
ASCR PM: Benjamin Brown
Date submitted to ASCR: September 14, 2021
Publication(s) for this work: A. Brace, et. al., *ACM PASC* (2021).

Argonne National Laboratory, Cerebras Systems Inc., Univ. of Chicago, Univ. of Illinois

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE

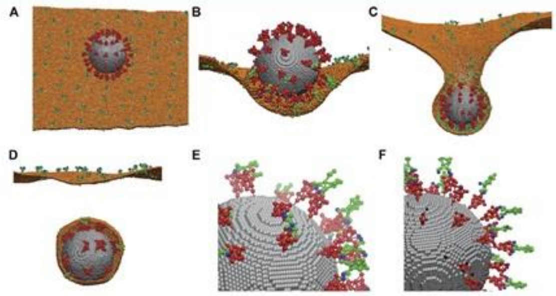
Modeling endocytosis of the SARS-CoV-2 coronavirus

The Science

Novel coronavirus Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) that causes Coronavirus disease 2019 (COVID-19) set in motion intensive experimental and computational research. To date, the exact infection mechanism of SARS-CoV-2 through cell membrane proteins remains unclear. The challenge here is the ability to accurately model and simulate the process of coronavirus entering human cells, namely, endocytosis. A bioengineering group at University of Illinois at Chicago developed a computational approach using coarse-grained (CG) models to simulate the endocytosis of SARS-CoV-2. The model was able to simulate a complete endocytosis as well as predict the binding mode of coronavirus to the receptor protein, not previously observed in computer simulations with all-atom models.

The Impact

A complete microscopic picture of how SARS-CoV-2 interacts with its receptor protein (ACE2) on cell membranes is key to understanding the infection mechanism of SARS-CoV-2 infection. However, development of a set of molecular models that can describe the coronavirus endocytosis in physiologically relevant spatial and temporal scales as well as its detailed binding mode with ACE2 is a difficult task. The CG model developed here, which takes both global conformation changes and local interaction details into account, opens the door to picture a complete coronavirus endocytosis. The CG model is built based on robust all-atom simulations of underlying molecular complexes, and self-contained within a single software package, and freely available for the research community to tune model resolution, if needed.



Simulations of the endocytosis (A-D) of a whole coronavirus by incorporating the coarse-grained models of the spike protein (red), the ACE2 receptor (green), the virus envelope (grey) and the membrane (brown), and a close view (E-F) of interactions between spike protein and ACE2 receptor. (Image reproduced from Figure 4 of *Frontiers in Physics*, **9**, 680983 (2021))

Contact PI: Zhangli Peng (UIC)
ASCR Allocation PI: Zhangli Peng (UIC)
ASCR Program/Facility: DD/ALCF
ASCR PM: Benjamin Brown
Date submitted to ASCR: Oct. 18, 2021
Publication(s) for this work: T. Leong, et. al., *Frontiers in Physics*, **9**, 680983 (2021)

University of Illinois at Chicago

U.S. DEPARTMENT OF ENERGY | Office of Science

NSF, UIC

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, Lemont, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

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

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

www.anl.gov

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