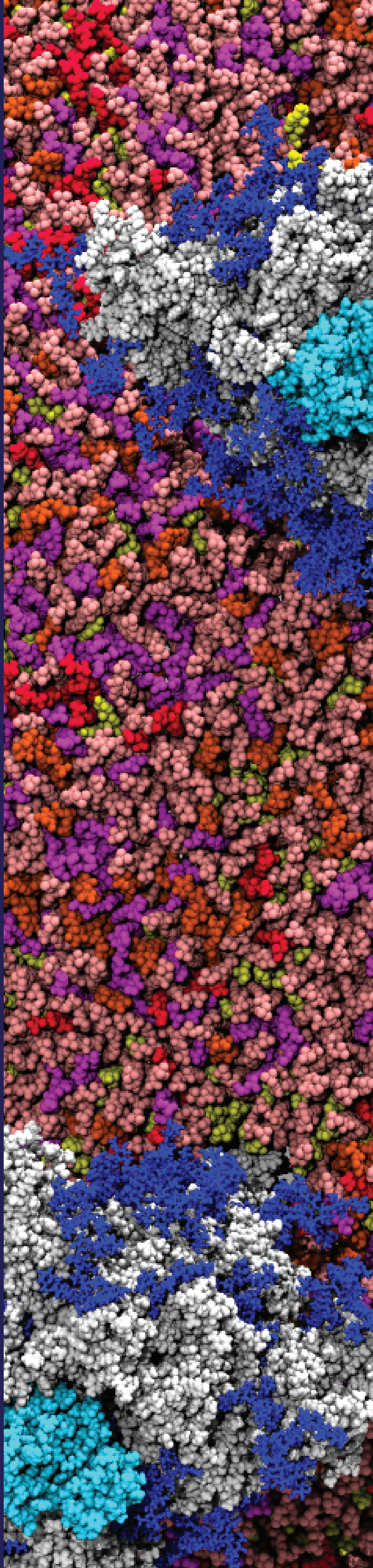
# 2020
## Operational Assessment Report



# Argonne Leadership Computing Facility

***On the cover:*** *A snapshot of a visualization of the SARS-CoV-2 viral envelope comprising 305 million atoms. A multi-institutional research team used multiple supercomputing resources, including the ALCF's Theta system, to optimize codes in preparation for large-scale simulations of the SARS-CoV-2 spike protein that were recognized with the ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research.*

***Image credit:*** *Rommie Amaro, Lorenzo Casalino, Abigail Dommer, and Zied Gaieb, University of California, San Diego*

# Contents

**Executive Summary/Background ..................................................... ES-1**

**Responses to Recommendations from the PY OAR ......................................... 1-1**

**1. User Support Results ........................................................................ 1-3**

ALCF Response ........................................................................................ 1-3
Survey Approach ..................................................................................... 1-4
Likert Scale and Numeric Mapping ........................................................... 1-4
1.1 User Support Metrics ........................................................................... 1-5
1.2 Problem Resolution Metrics .................................................................. 1-6
1.3 User Support and Outreach .................................................................. 1-7
1.3.1 Tier 1 Support .................................................................................. 1-7
1.3.1.1 Phone and E-mail Support ............................................................. 1-7
1.3.1.2 Improved Foreign Visit & Assignment Request (593) Form .................. 1-8
1.3.1.3 Multiple Resource Requests for Director's Discretionary Allocation ..... 1-8
1.3.1.4 Continuous Improvements ............................................................. 1-9
Centralized User Agreement and Acknowledgement Tracking ............. 1-9
Switch to Node-Hours ........................................................................ 1-9
Various Automation Scripts ................................................................ 1-9
Migration to ServiceNow ..................................................................... 1-9
1.3.2 Application Support ........................................................................... 1-9
1.3.2.1 Individual Projects ........................................................................ 1-9
Computational Physical Genomics: Exploring Potential Novel Cancer Therapies ........................................................................ 1-9
Predictive Modeling and Machine Learning for Functional Nanoporous Materials ......................................................... 1-10
Benchmarking Atomistic Simulations of Tribological Systems ........... 1-10
Multiscale Modeling of Soft Materials by Unifying Shell Elements and Anisotropic Coarse-Grained Force Fields in LAMMPS ........... 1-10
High-Speed Turbulence with Shocks over Non-adiabatic and Flexible Walls ........................................................................ 1-10
Toward Exascale Internal Combustion Engine Simulations ................. 1-10
The Energy Exascale Earth System Model (E3SM) Coupled Model Version 1: Description and Results at High Resolution .................. 1-11
1.3.3 Resource Support ............................................................................. 1-11
1.3.3.1 Continuous Integration Pipeline .................................................... 1-11
1.3.3.2 Spack at ALCF ............................................................................. 1-12
1.3.3.3 Theta Upgrade .............................................................................. 1-12
1.3.3.4 User Accounting Infrastructure Homogenization within the CELS Directorate (Userbase3) ............................................. 1-12
1.3.4 Outreach Efforts ................................................................................ 1-13
1.3.4.1 General Outreach ......................................................................... 1-13
User Advisory Council ........................................................................ 1-13
Connection to Science and Technology Partnerships and Outreach directorate ........................................................................ 1-13

# Contents (Cont.)


1.3.4.2 Training Activities ..... 1-13
Simulation, Data, and Learning Workshop (virtual) ..... 1-14
Aurora Early Science Program Dungeon (virtual) [funded by ALCF-3 and ECP] ..... 1-14
Intel COE for Aurora ESP Workshop 3 (virtual) [funded by ALCF-3 and ECP] ..... 1-14
2020 Performance, Portability, and Productivity at HPC (P3HPC) Forum (virtual) ..... 1-14
Argonne Training Program on Extreme-Scale Computing (virtual) ..... 1-14
ALCF-ECP CMake Workshop (virtual) ..... 1-14
Computational Performance Workshop (virtual) ..... 1-15
Intel COE for Aurora ESP Workshop 2 [funded by ALCF-3 and ECP] ..... 1-15
ALCF Webinars ..... 1-15
Aurora Early Science Program Hackathons (virtual) [funded by ALCF-3 and ECP] ..... 1-16
1.3.4.3 Community Outreach ..... 1-16
Science, Technology, Engineering and Math (STEM) Activities ..... 1-16
2020 Illinois Regional Middle School Science Bowl ..... 1-16
Introduce a Girl to Engineering Day ..... 1-16
Science Careers in Search of Women ..... 1-17
CodeGirls@Argonne Camp (virtual) ..... 1-17
Coding for Science Camp (virtual) ..... 1-17
Big Data Camp (virtual) ..... 1-17
Hour of Code (virtual) ..... 1-17
Facility Tours ..... 1-17
1.3.5 Communications ..... 1-18
1.3.5.1 Website Support Center Continuous Improvement ..... 1-18
1.3.5.2 Article Process Improvements ..... 1-18
1.3.5.3 Communications through Mailing Lists and Social Media ..... 1-18
1.3.5.4 Promotional Activities and Media Hits ..... 1-19
1.3.5.5 Other Publications ..... 1-20
Conclusion ..... 1-20

**2. Operational Performance ..... 2-1**

2.1 ALCF Response ..... 2-1
2.2 ALCF Production Resources Overview ..... 2-2
2.3 Definitions ..... 2-2
2.4 Theta ..... 2-3
2.4.1 Scheduled and Overall Availability ..... 2-3
2.4.1.1 Explanation of Significant Availability Losses ..... 2-3
2.4.2 System Mean Time to Interrupt (MTTI) and System Mean Time to Failure (MTTF) ..... 2-6
2.4.2.1 ALCF MTTI and MTTF Summary ..... 2-6

# Contents (Cont.)


2.4.3 Resource Utilization ..... 2-6
2.4.3.1 System Utilization ..... 2-6
2.4.3.2 Capability Utilization ..... 2-8
2.5 Storage ..... 2-10
2.5.1 Theta Lustre File System ..... 2-10
2.5.1.1 Scheduled and Overall Availability ..... 2-10
Explanation of Significant Availability Losses ..... 2-11
2.5.1.2 MTTI and MTTF ..... 2-12
MTTI and MTTF Summary ..... 2-12
2.5.2 Tape Storage ..... 2-13
2.5.2.1 Scheduled and Overall Availability ..... 2-13
Explanation of Significant Availability Losses ..... 2-13
2.5.2.2 MTTI and MTTF ..... 2-14
MTTI and MTTF Summary ..... 2-14
2.6 Center-Wide Operational Highlights ..... 2-15
2.6.1 Mira Decommission ..... 2-15
2.6.2 ThetaGPU Installation ..... 2-15
2.6.3 HDR Installation ..... 2-16
2.6.4 GFS Installation ..... 2-16
2.6.5 Shasta Testbed ..... 2-17
Conclusion ..... 2-17

**3. Allocation of Resources ..... 3-1**

ALCF Response ..... 3-1
3.1 Allocation of Resources ..... 3-1
3.1.1 Usage of the INCITE and ALCC Hours ..... 3-1
3.1.2 Facility Director's Discretionary Reserve Time ..... 3-4
Conclusion ..... 3-6

**4. Innovation ..... 4-1**

ALCF Response ..... 4-1
4.1 Operational Innovation ..... 4-1
4.1.1 Molecular Database for SARS-CoV-2 Research ..... 4-1
4.1.2 Collaboration with OpenPBS Community ..... 4-2
4.1.3 Hyper-Parameter Optimization (HPO) and Network Architecture Search (NAS) for AI-Driven Science ..... 4-3
4.1.4 Adaptive Routing Study ..... 4-3
4.1.5 Effective Virtual Meetings ..... 4-5
4.1.6 Aurora Readiness Assessments of INCITE Projects ..... 4-6
4.2 Research Activities for Future Operations ..... 4-7
4.2.1 Improved Software Environment for Emerging Technologies ..... 4-7
4.2.2 Tools for Improving Code Performance on Emerging Hardware ..... 4-9
4.2.3 Intel Discrete GPU Testbed ..... 4-10
4.2.4 Joint Laboratory for System Evaluation ..... 4-11

# Contents (Cont.)


4.2.5 AI Testbed ........ 4-13
4.3 Postdoctoral Fellows ........ 4-14

**5. Risk Management ........ 5-1**

ALCF Response ........ 5-1
5.1 Risk Management Process Overview ........ 5-1
5.1.1 Risk Review Board ........ 5-3
5.1.2 Risk Management in Day-to-Day Operations ........ 5-3
5.1.3 Continuation of the ALCF-3 Project ........ 5-4
5.2 Major Risks Tracked for the Review Year ........ 5-4
5.3 Risks Encountered in the Review Year and Their Mitigations ........ 5-5
5.3.1 Users Gain Access to Data That They Should Not Have Permission to Review ........ 5-5
5.3.2 Facility Power Interruptions ........ 5-6
5.4 Retired Risks ........ 5-6
5.5 New and Recharacterized Risks since the Last Review ........ 5-7
5.5.1 Rate of Hardware Failure Affects Availability of Key Production Resources Availability ........ 5-7
5.6 Top Operating Risks Monitored Closely for the Next Year ........ 5-7
Conclusion ........ 5-8

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

ALCF Response ........ 6-1

**7. Security ........ 7-1**

ALCF Response ........ 7-1
7.1 Continual Improvement in Cyber Security Practices ........ 7-1
7.2 Cyber Security Plan ........ 7-3
7.3 Foreign Visitor Access and Export Controls ........ 7-3

**8. Strategic Results ........ 8-1**

ALCF Response ........ 8-1
8.1 Scientific Accomplishments ........ 8-1
8.2 Scientific Highlights ........ 8-2
8.2.1 Predictive Modeling and Machine Learning for Functional Nanoporous Materials ........ 8-3
8.2.2 The Overarching Framework of Core-Collapse Supernova Explosions ........ 8-5
8.2.3 COVID-19 Spread and Effectiveness of Interventions ........ 8-7
8.2.4 Modeling Fusion Plasma-Facing Components ........ 8-9
8.2.5 Three Dimensions, Two Microscopes, One Code: Automatic Differentiation for X-ray Nanotomography Beyond the Depth of Focus Limit ........ 8-11
8.2.6 Shock-Induced Multi-Material Mixing ........ 8-13
8.2.7 The Last Journey ........ 8-15
8.3 Stakeholder Engagement ........ 8-17
8.3.1 Engagement Highlights ........ 8-18

# Contents (Cont.)


8.3.2 Summary of Engagements with the Exascale Computing Project ........ 8-19
ECP-Funded Positions in ALCF ........ 8-19
Continuous Integration Pipeline ........ 8-20
Communication between the ALCF and the ECP Resource Allocation Council ........ 8-20
8.3.3 COVID-19 Response ........ 8-20
Conclusion ........ 8-21

**Appendix A – Calculations ........ A-1**

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

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

**Appendix C – Factual Status Document ........ C-1**


# Figures


2.1 Theta Weekly Availability for CY 2020 ........ 2-4

2.2 System Utilization over Time by Program ........ 2-7

2.3 Theta INCITE Overall Capability ........ 2-9

2.4 Theta Capability Node-Hours by Program ........ 2-9

2.5 Theta Job Usage by Size ........ 2-10

2.6 Theta Lustre Weekly Availability for CY 2020 ........ 2-11

2.7 HPSS Weekly Availability for CY 2020 ........ 2-14

3.1 Theta INCITE 2020 Allocation Usage (Note: Projects are randomly ordered.) ........ 3-2

3.2 Theta ALCC 2019–2020 Allocation Usage (Note: Projects are randomly ordered.) ........ 3-3

3.3 Theta ALCC 2020–2021 Allocation Usage as of Dec. 31, 2020 (Note: Projects are randomly ordered.) ........ 3-4

3.4 Theta CY 2020 DD Allocations by Domain ........ 3-6

4.1 REST API to Access the nCOV Postgres Database ........ 4-2

4.2 Post-training results of DeepHyper with NAS using A3C on Combo Cancer problem with the large search space run on 512 and 1,024 nodes of Theta with agent scaling ........ 4-3

# Figures (Cont.)

4.3(a) MILC and MILCREORDER on three different job sizes ordered by the number of groups spanned with AD0 and AD3 routing modes. .... 4-4

4.3(b) System-wide counter metrics before (AD0) and after (AD3) changing the default routing on Theta. The color coding indicates information recorded on 4 different categories of ports on an Aries router. .... 4-5

4.4 SU(3)xSU(3) performance (GRID mini) with and without ALCF-developed LLVM/OpenMP optimizations. .... 4-8

4.5 Search space for a Monte Carlo Tree Search. Each node represents an experiment that can run and the execution time measured. In particular, each node is a different loop optimization. We start with the root node that represents the kernel without optimization. Each child is another loop transformation applied in addition to the ones in the parent. .... 4-9

4.6 THAPI Trace Output .... 4-9

4.7 Iprof Screen Output .... 4-10

4.8 Intel's Xe DG1 GPU in Yarrow .... 4-11

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

8.1 Big data from high-throughput simulations and a deep neural network are utilized to find optimal conditions for contracting a solution phase (green box) with a zeolite (orange box) to yield pure liquid solvent and a zeolite with enriched product, followed by partial desorption of the solvent into a vapor phase, leading to essentially pure product that can subsequently be desorbed. .... 8-3

8.2 This visualization depicts the neutrino-driven roiling convection of the nuclear material behind the newly re-energized supernova shock wave that surrounds the newly birthed neutron star, just hundreds of milliseconds after shock revival. The isosurfaces are of specific entropy, and the coloring follows the electron fraction. The pathlines trace infalling parcels of matter caught up in the turbulence of the explosion. The neutrino-heated turbulent bubbles help drive the supernova explosion shock wave outwards. .... 8-5

8.3 Weekly counts of the newly infected counties by ZIP code under differing individual behavior scenarios. .... 8-7

8.4 Images adapted from K.D. Hammond, et al., *Nucl. Fusion* 59 (2019) 066035. Snapshot from a simulation of helium atoms beneath a tungsten (001) surface exposed to 100 eV helium plasma for 840 ns (fluence of $8.4 \times 10^{18}$ m$^{-2}$): (left-top) top-down and (left-bottom) side views highlighting changes in surface height. Cluster size distributions for two fluxes are shown on the right. .... 8-9

8.5 Our automatic-differentiation-based software reconstructs beyond-depth-of-focus objects in both holography (A) and ptychography (B), providing results free of diffraction-induced artifacts (C). .... 8-11

# Figures (Cont.)

8.6 Image shows instantaneous planar (y=0) slices of transverse velocity v at final simulation time for low (left) and high (right) compressibility cases with unity density ratio in a limited region of the flow. Arrows indicate decorrelation length scales based on velocity fluctuations along the x and z axis. The Reynolds shear stress associated with turbulent mixing decreases with compressibility and reduces its growth rate. While this compressibility effect has been known since the mid-1970s, the present simulations have provided a novel physical explanation. Turbulence scales were found to be internally limited in this flow by acoustic communication. 8-13

8.7 Visualization of the Last Journey simulation. Shown are thin density slices for the full box (lower left corner) and zoom-ins at different levels. The panel on the lower right focuses on the largest cluster in the simulation. 8-15

8.8 Breakdown of some key activities by ALCF in CY 2020. The first pie chart (left) breaks down the 79 events by type. This is primarily derived from how the event self identifies. The second pie chart (right) breaks down the same 79 events by the community being engaged. 'Multi' refers to events such as Supercomputing (SC), where multiple communities are engaged. 8-17

# Tables

ES.1 Summary of the Target and Actual Data for the Previous Year (PY) (CY 2020) Metrics ES-2

1.1 All 2020 User Support Metrics and Results 1-3

1.2 2020 User Survey Results by Allocation Program 1-6

1.3 2019 and 2020 User Support Metrics 1-6

1.4 Tickets Addressed Metric 1-7

1.5 Ticket Categorization for 2019 and 2020 1-8

1.6 2020 Primary Communication Channels 1-19

1.7 2020 Target Audiences 1-19

1.8 Publications Designed for Print 1-20

2.1 Summary of All Metrics Reported in the Operational Performance Section 2-1

2.2 Availability Results for Theta 2-3

2.3 MTTI and MTTF Results 2-6

2.4 System Utilization Results 2-7

2.5 Node-Hours Allocated and Used by Program 2-8

2.6 Capability Results 2-8

# Tables (Cont.)


2.7 Availability Results ........ 2-11
2.8 MTTI and MTTF Results ........ 2-12
2.9 Availability Results ........ 2-13
2.10 MTTI and MTTF Results ........ 2-14
3.1 INCITE 2020 Time Allocated and Used on Theta ........ 3-2
3.2 ALCC Time Allocated and Used on Theta in CY 2020 ........ 3-4
3.3 DD Time Allocated and Used on Theta, 2020 ........ 3-6
5.2 Risk Retired during CY 2020 ........ 5-7
5.3 Top Operating Risks Monitored for CY 2021 ........ 5-8
8.1 Summary of Users' Peer-Reviewed Publications in CY 2020 ........ 8-2
8.2 Summary of Users' Peer-Reviewed Publications for 5-year Moving Window ........ 8-2
8.3 Summary of ALCF Participation in SC20 ........ 8-18
A.1 Capability Definitions for Theta ........ A-2
A.2 Total and Reportable Nodes for Theta ........ A-3

# Executive Summary/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 run around the clock every day of the year. From an operational standpoint, the high level of service these centers provide and the exceptional science they produce justify their existence to the DOE Office of Science and the U.S. Congress.

And while the COVID-19 pandemic has had a deep impact on many aspects of our lives in 2020, this Operational Assessment Report describes how the Argonne Leadership Computing Facility (ALCF) met or exceeded every one of its goals for calendar year (CY) 2020 as an advanced scientific computing center. More details about how ALCF responded to the pandemic can be found in its Factual Status Document response to the December 2020 Office of Science User Facilities Roundtable (Appendix C).

In 2020, the ALCF operated its production resource, Theta, an 11.7-petaflops Intel-based Cray XC40 system. In September 2020, Theta was augmented with 24 NVIDIA DGX A100-based nodes (3.9-petaflops) and has its own queue, known as ThetaGPU, set aside for the direct support of COVID-19 research in CY 2020. The new integrated Theta supercomputer supports diverse workloads, integrating data analytics with artificial intelligence (AI) training and learning into a single platform.

Last year, Theta delivered a total of 20.9 million node-hours to 14 Innovative and Novel Computational Impact on Theory and Experiment (INCITE) projects and 7.4 million node-hours to ASCR Leadership Computing Challenge (ALCC) projects (18 awards for the 2019–2020 ALCC year and 31 awards for the 2020–2021 ALCC year), as well as substantial support to Director's Discretionary (DD) projects (4.7 million node-hours). As Table ES.1 shows, Theta performed exceptionally in terms of overall availability (93.8 percent), scheduled availability (98.3 percent), and utilization (97.7 percent; Table 2.1).

Mira, the ALCF's previous production system, officially ceased production operations on December 31, 2019; however, it delivered an additional 67 million node-hours in support of 2 DD projects, until it terminated all operation on March 9, 2020.

As of the submission of this document, ALCF's user community has published 246 papers in high-quality, peer-reviewed journals and technical proceedings. ALCF also offered a comprehensive program of high-performance computing (HPC) support services to help its community make productive use of the facility's diverse and growing collection of resources.

As the LCF prepares to enter the exascale era in the coming years, ALCF researchers are already leading and contributing to several strategic activities that aim to push the boundaries of what's possible in computational science and engineering. When the exascale systems arrive, the Leadership Computing Facility will once again stand ready to deliver science on day one.

**Table ES.1 Summary of the Target and Actual Data for the Previous Year (PY) (CY 2020) Metrics**

| Area | Metric | 2020 Target | 2020 Actual |
|---|---|---|---|
| **User Results** | User Survey – **Overall Satisfaction** | 3.5/5.0 | 4.6/5.0 |
| | 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% | 42.6 % |
| | % User Problems Addressed within Three Working Days | 80.0% | 97.7% |
| **Business Results** | Theta Overall Availability | 80.0% | 93.8% |
| | Theta Scheduled Availability | 90.0% | 98.3% |
| | % of INCITE node hours from jobs run on 20.0% or more of Theta (800–4008 of the reportable nodes) | N/A | 84.8% |
| | % of INCITE node hours from jobs run on 60.0% or more of Theta (2400–4008 of the reportable) | N/A | 24.1% |
| | Lustre File System Overall Availability | N/A | 94.8% |
| | Lustre File System Scheduled Availability | N/A | 98.8% |
| | HPSS [a] Overall Availability | N/A | 95.2% |
| | HPSS Scheduled Availability | N/A | 98.9% |

[a] HPSS = high-performance storage system.

# Responses to Recommendations from the PY OAR

The Facility did not receive any recommendations from the previous OAR Committee.

*This page intentionally left blank.*

# 1. User Support Results

***Are the processes for supporting the 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 problems, and to conduct outreach. The 2020 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) 2020.

**Table 1.1 All 2020 User Support Metrics and Results [a]**

| | | 2019 Actual | 2020 Target | 2020 Actual |
|---|---|---|---|---|
| **Number Surveyed** | | 1,246 | N/A [c] | 1,174 |
| **Number of Respondents (Response Rate)** | | 522 (41.9%) | 25.0% | 500 (42.6%) |
| **Overall Satisfaction** | Mean | 4.5 | 3.5 | 4.6 |
| | Variance | 0.5 | N/A | 0.5 |
| | Standard Deviation | 0.7 | N/A | 0.7 |
| **Problem Resolution** | Mean | 4.5 | 3.5 | 4.6 |
| | Variance | 0.5 | N/A | 0.4 |
| | Standard Deviation | 0.7 | N/A | 0.7 |
| **User Support** | Mean | 4.6 | 3.5 | 4.6 |
| | Variance | 0.4 | N/A | 0.4 |
| | Standard Deviation | 0.7 | N/A | 0.7 |
| | | **2019 Actual** | **2020 Target** | **2020 Actual** |
| **% of Problems Addressed Within Three Working Days [b]** | | 93.3% | 80.0% | 97.7% |

[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 time 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 now polls workshop attendees separately.

The 2020 user survey closely resembled the 2019 survey, with a few minor modifications to the infrastructure, performance, and debugging tools/services listed in various questions. Three new questions were added: one that captured users' enthusiasm for an ALCF Slack channel; another that gauged their interest in participating in an annual meeting, and an open-ended question that asked how ALCF teams could better support their project needs; and two questions came out: one that captured user preferences on a dashboard, and another that asked about the use of continuous integration, containers, and workflows.

The 2020 survey was administered by a new vendor, Bargile Associates, and consisted of 7 required questions and 20 optional questions. The survey and associated e-mail campaign ran from November 19, 2020, through December 31, 2020. Each reminder e-mail included a unique 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 Likert Scale questions in the ALCF user survey use a six-point 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) |

Only 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 for the fiscal year 2020 was invited to complete a user survey. Of the 1,174 individuals asked, 500 responded, for a 42.6 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 results are not statistically significant. Other metrics are comparable, in that the variations are statistically insignificant.

**Table 1.2 2020 User Survey Results by Allocation Program**

| 2020 Metrics by Program | | INCITE | ALCC | INCITE + ALCC | DD | All |
|---|---|---|---|---|---|---|
| **Number Surveyed** | | 302 | 80 | 382 | 792 | 1,174 |
| **Number of Respondents** | | 119 | 39 | 158 | 342 | 500 |
| **Response Rate** | | 39.4% | 48.8% | 41.4% | 43.2% | 42.6% |
| **Overall Satisfaction** | Mean | 4.6 | 4.6 | 4.6 | 4.5 | 4.6 |
| | Variance | 0.5 | 0.4 | 0.5 | 0.5 | 0.5 |
| | Standard Deviation | 0.7 | 0.6 | 0.7 | 0.7 | 0.7 |
| **Problem Resolution** | Mean | 4.6 | 4.5 | 4.5 | 4.6 | 4.6 |
| | Variance | 0.5 | 0.5 | 0.5 | 0.4 | 0.4 |
| | Standard Deviation | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 |
| **User Support** | Mean | 4.5 | 4.5 | 4.5 | 4.6 | 4.6 |
| | Variance | 0.5 | 0.5 | 0.5 | 0.4 | 0.4 |
| | Standard Deviation | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 |
| **All Questions** | Mean | 4.5 | 4.5 | 4.5 | 4.6 | 4.6 |
| | Variance | 0.5 | 0.5 | 0.5 | 0.4 | 0.5 |
| | Standard Deviation | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 |

As Table 1.3 shows, in 2020, the ALCF again exceeded the targets for overall satisfaction and user support.

**Table 1.3 2019 and 2020 User Support Metrics**

| Survey Area | 2019 Target | 2019 Actual | 2020 Target | 2020 Actual |
|---|---|---|---|---|
| **Overall Satisfaction Rating** | 3.5/5.0 | 4.5/5.0 | 3.5/5.0 | 4.6/5.0 |
| **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 problem tickets addressed in three days or less, which ALCF exceeded. A ticket is defined as “addressed” once the following conditions have been met: (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 Tickets Addressed Metric**

| | 2019 Target | 2019 Actual | 2020 Target | 2020 Actual |
|---|---|---|---|---|
| **% of Problems Addressed Within Three Working Days [a]** | 80.0% | 93.3% | 80.0% | 97.7% |
| **Avg. of Problem Resolution Ratings** | 3.5/5.0 | 4.5/5.0 | 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 Outreach

### 1.3.1 Tier 1 Support

#### *1.3.1.1 Phone and E-mail Support*

In 2020, the ALCF answered and categorized 4,578 support tickets, a 17.9 percent decrease from the previous year. The biggest decreases were in Accounts, Access, and Automated E-mail Responses categories (Table 1.5). There are three possible explanations for this change: First, with the improvements made to the account and project management system (UB3), some of the tickets pertaining to access and accounts that we would normally see were eliminated. Second, the Blue Gene/Q (BG/Q) systems were retired at the end of 2019, which means we had very few support tickets related to the three BG/Q systems—Mira, Cetus, and Vesta—in 2020. Third, almost all of our users were provisioned mobile tokens in 2020. The self-service nature of activating a mobile token eliminated the need for users to reach out to Support to unlock the token, as is the case for physical tokens.

**Table 1.5 Ticket Categorization for 2019 and 2020**

| Category | 2019 | 2020 |
|---|---|---|
| Access | 1,010 (18%) | 806 (18%) |
| Accounts | 2,414 (43%) | 1,830 (40%) |
| Allocations | 629 (11%) | 629 (14%) |
| Applications Software | 197 (4%) | 195 (4%) |
| Automated E-mail Responses | 348 (6%) | 205 (4%) |
| Compilers | 23 (0%) | 61 (1%) |
| Data Transfer | 38 (1%) | 45 (1%) |
| Debugging and Debuggers | 11 (0%) | 9 (0%) |
| File System | 124 (2%) | 162 (4%) |
| HPSS [a] and Quota Management | 130 (2%) | 102 (2%) |
| Libraries | 36 (1%) | 24 (1%) |
| Miscellaneous | 87 (2%) | 148 (3%) |
| Network | 11 (0%) | 7 (0%) |
| Performance and Performance Tools | 42 (1%) | 15 (0%) |
| Reports | 278 (5%) | 181 (4%) |
| Scheduling | 196 (4%) | 154 (3%) |
| Visualization | 8 (0%) | 5 (0%) |
| TOTAL TICKETS | 5,582 (100%) | 4,578 (100%) |

[a] HPSS = high-performance storage system.

#### *1.3.1.2 Improved Foreign Visit & Assignment Request (593) Form*

ALCF worked to improve Argonne's visitor registration form, which is integrated into its account and project management application (UB3). As a result of this work, the form was updated to allow foreign national users to upload the required immigration documents. This change benefitted the entire lab, not just the ALCF, as it saves a lot of time and effort for our users. User details are passed from UB3 to the Argonne visitor registration form, eliminating the need for users to reenter some of the data. We reviewed and improved the front-end form validation and business rules for the visitor registration form.

#### *1.3.1.3 Multiple Resource Requests for Director's Discretionary Allocation*

The director's discretionary allocation request form was modified to make the proposal process easier for users who were interested in multiple resources. In its previous version, a user had to fill out the form separately for each resource, which was repetitive, time consuming, and redundant. Now, the resources are presented as checkboxes, and the user provides details and justification for each resource in only a single request.

#### 1.3.1.4 Continuous Improvements

##### Centralized User Agreement and Acknowledgement Tracking

The Master User Agreement and User Acknowledgement (MUA/Ack) database was a stand-alone application used to record and track the user institutions that have a user agreement with Argonne National Laboratory, and user acknowledgments of this agreement. This database was migrated into UB3 and integrated with the forms within UB3. This centralization has allowed the User Support group to shorten the time taken to look up and process requests that need an MUA or an Ack.

##### Switch to Node-Hours

ALCF switched the allocation unit on Theta from core-hour to node-hour. With this change, the allocation units for all of the resources at ALCF are uniform. Charging and reporting have also been changed to node-hours.

##### Various Automation Scripts

Many scripts used for internal workflows were automated using Jenkins, particularly audit scripts used to monitor and improve the overall health of UB3 data. Scripts include automatic assignment of login nodes to ThetaGPU users, status change of inactive users, and review of users' change in citizenship status.

##### Migration to ServiceNow

ALCF is in the process of migrating its service desk software from Request Tracker (RT) to ServiceNow. The project is in the testing phase and is expected to be deployed in early 2021. With this migration, ALCF will move away from an older version of RT to a platform that is up-to-date and in use lab-wide.

### 1.3.2 Application Support

#### 1.3.2.1 Individual Projects

##### Computational Physical Genomics: Exploring Potential Novel Cancer Therapies

This INCITE project aims to enable a high-throughput, label-free approach to quantifying macromolecular assembly structure in the nanometer-length scale regime to advance research in gene therapy. ALCF staff has provided comprehensive support for finite-difference time-domain (FDTD) code porting on Theta, and porting and profiling of supplementary analysis software. Significant support provided by ALCF included initial structure construction for molecular dynamics (MD) simulation and setup of nanoscale particle simulation. Additionally, staff assisted with usage monitoring and large reservation coordination on Theta driven by manuscript submission deadlines. This project has already resulted in two publications: (1) Eid, A., et al., "Characterizing Chromatin Packing Scaling in Whole Nuclei Using Interferometric Microscopy," *Opt. Lett.* **45**, (2020), 4810–4813; and (2) Li, Y., et al., "Nanoscale Chromatin Imaging and Analysis (Nano-ChIA) Platform Bridges 4-D Chromatin Organization with Molecular Function," *Microsc. Microanal.* **26**, no. S2 (2020), 1046–50.

### *Predictive Modeling and Machine Learning for Functional Nanoporous Materials*

This project targeted the use of hierarchical modeling and machine learning to accelerate discovery and design of materials for a variety of energy-related applications. ALCF support for this work included developing a scalable initialization scheme for multi-billion particle simulations of bubble collapse in water; addressing int32 overflow issues; and code optimization for a ~3x speedup on Theta—resulting in a 2020–2021 ALCC project award. A manuscript is in preparation. This work aligned with goals of the DOE-funded Nanoporous Materials Genome Center (NMGC) supported by the Computational Chemical Sciences award.

### *Benchmarking Atomistic Simulations of Tribological Systems*

An ALCF staff member assisted a research team from Freudenberg Technology Innovation to prepare for their DD work aimed at modeling tribological properties on Theta. The staff member supplied build scripts, executables, and job submission scripts for benchmarks early in the project phase, and provided hands-on technical assistance during the 2020 Virtual ALCF Computational Performance Workshop in May. Possible outcomes of this work include a future ALCC or INCITE award or an ALCF–industry collaboration.

### *Multiscale Modeling of Soft Materials by Unifying Shell Elements and Anisotropic Coarse-Grained Force Fields in LAMMPS*

This work addresses multiscale modeling of soft materials by coupling the finite-element method with anisotropic coarse-grained models to simulate large-scale systems, such as biological membranes. An ALCF staff member worked with the principal investigator (PI) to vectorize a smooth-particle hydrodynamics interaction potential in LAMMPS (2x speedup on Theta) as a proxy for the actual model of interest. This work yielded performance data for a 2021 INCITE proposal.

### *High-Speed Turbulence with Shocks over Non-adiabatic and Flexible Walls*

An ALCF Catalyst provided onboarding support for this INCITE work, first by setting up Kokkos on Cooley to allow the PIs to develop their uPDE code for heterogeneous computing platforms, then by building the Trilinos software stack on Theta following a planned operating system (OS) upgrade. He also provided input on parallel input/output (I/O) for the HybridX code, which was the main workhorse for the INCITE computations. This project used about 107% of its INCITE allocation. The PI was very appreciative of the support received from ALCF, directly from the project Catalyst, and indirectly by attending the relevant training opportunities hosted and/or promoted by the facility.

### *Toward Exascale Internal Combustion Engine Simulations*

The objective of this research was to understand the causes of cycle-to-cycle variability (CCV) in an engine. Nek5000, a leading, open source, high-order spectral element code, was used to simulate turbulent flow in the engine combustion chamber. Multi-cycle large-eddy simulations (LES) and closed-cycle direct numerical simulations (DNS) of the in-cylinder flow processes were performed for the General Motors (GM) Transparent Combustion Chamber (TCC-III) optical engine. These simulations, which were one of the most detailed simulations of an internal combustion engine (ICE), show some novel mechanisms of turbulence generation and CCV. An ALCF performance engineer assisted in the original porting of Nek5000 to Theta, including investigating the single-node performance using the Libxsmm library for small matrix multiplication operations. An ALCF postdoc was hired to develop, test, and port scalable algorithms in Nek5000, which would improve the performance of the ICE simulations on Theta. ALCF staff helped with reservation requests and special queue accommodations, and members of ALCF's visualization group provided in-situ and post-hoc visualization capabilities, which helped to improve understanding of the turbulent flow-field inside an ICE. This calculation ran on 802 nodes (51,328 cores) of Theta. The results from this work were used show how conventional wall models fall short of capturing the wall shear stress and heat flux. In collaboration with Stanford University, this data is being utilized to improve wall models for low-order techniques.

***The Energy Exascale Earth System Model (E3SM) Coupled Model Version 1: Description and Results at High Resolution***

The Energy Exascale Earth System Model (E3SM) is a relatively new, fully coupled Earth system and climate model used in major international model simulation projects and mission-defined efforts for the DOE. This INCITE study performed the first simulation of the model in its high-resolution configuration. This higher-resolution version is able to capture the most energetic motions in the ocean, which are poorly represented in standard-resolution, coupled climate models, as well as the largest of storms in the atmosphere. Evaluation of this simulation confirms the benefits of high resolution found by other models with a few notable exceptions. These discrepancies with other studies are interesting because they provide a richer understanding of how and why resolution affects model bias. Another key finding is that climate and aerosol sensitivity in E3SM is unaffected by resolution change. This study also confirmed the benefits of increased resolution for studying fine-scale features such as hurricanes and orographic precipitation. Finally, the high-resolution version of E3SM is shown to compare favorably to its low-resolution counterpart and to the models participating in Phase 5 of the Coupled Model Intercomparison Project (CMIP). The production runs for this work required 800 nodes on the Theta resource and extended into the first few months of the INCITE 2019 allocation. The project received support from ALCF staff, who improved job throughput in the queues and enabled the team to make use of the overburn policy. Project team members have generated the following publication: Caldwell, P.M., A. Mametjanov, Q. Tang, L.P. Van Roekel, J.-C. Golaz, W. Lin, et al., "The DOE E3SM coupled model version 1: Description and results at high resolution," *J. Adv. Model Earth Sy.,* **11**, 4095 (2019).

### 1.3.3 Resource Support

#### *1.3.3.1 Continuous Integration Pipeline*

Over the past few years, the ALCF has been working closely with the Exascale Computing Project (ECP) on a number of projects. One of these projects, ECP-CI, brings the ability to run Continuous Integration (CI) testing to high-performance computing (HPC) hardware at the facilities. Last year, we stood up one of the first instances supporting this capability on Theta. This project leverages a modified Gitlab runner to integrate with the batch schedulers and run as the given user.

Over the past year, we have continued to increase capabilities on the Theta CI instance by adding an additional 24 projects and 30 users, which brings us to a total of 35 projects. In addition to the Theta capabilities, we have expanded this functionality into the Joint Laboratory for System Evaluation (JLSE) to facilitate the use of CI testing on early access hardware. Having this functionality allows our users and developers to make critical strides in ensuring that their codes are ready for exascale systems when they arrive at DOE facilities.

ALCF is continuing to work with the ECP-CI project to expand the functionality of the software and make it easier and safer for projects to leverage it. The ECP-CI project is working closely with Gitlab to merge in changes and adapt in ways that will make this easier to deploy and manage at all DOE HPC facilities.

#### 1.3.3.2 Spack at ALCF

Highlights and accomplishments this past year with Spack at ALCF have been plentiful. The majority of ECP's E4S packages were able to be built using Spack for deployment to Theta. We also focused on the early hardware systems in JLSE, where enhancements were made to Spack, including the addition of a new oneAPI compiler class, which enables it to build packages using the Intel OneAPI compiler suite. A prioritized set of packages of interest to ALCF was built using the new compiler class. A complete four-stage, Spack-generated CI build pipeline was established and run to completion for PETSc using the OneAPI compiler. This pipeline was implemented on the Gitlab CI platform that was established in JLSE.

#### 1.3.3.3 Theta Upgrade

Theta, a Cray XC40 with Intel KNL processors, underwent a massive upgrade in July of 2020 that included upgrading the OS from SUSE Linux Enterprise Server (SLES) 12 to 15 and the Cray Linux Environment (CLE) from CLE 6 to 7; and the Cray Programming Environment was completely replaced with the latest compilers and libraries. On the storage side, the Sonexion software was upgraded from NEO 2 to 3.4, another large jump, and the Lustre software was upgraded from 2.7 to 2.12. Although temporarily a disruption, this upgrade brought Theta to the very latest in all aspects of the software for our users. We minimized outages for the last year by bypassing some upgrades, thereby avoiding some buggy releases and now capitalizing on needs for the latest software versions to be compatible with other systems that were installed in 2020. The upgrade and the resulting outage were delayed to July 2020 from an earlier time in the year to provide uninterrupted access to Theta to users doing COVID-19 research.

#### 1.3.3.4 User Accounting Infrastructure Homogenization within the CELS Directorate (Userbase3)

The Computing, Environment and Life Sciences (CELS) directorate is in the early stages of aligning and homogenizing infrastructure across the ALCF; the Laboratory Computing Resource Center (LCRC, an Argonne-only supercomputing resource); the JLSE, our shared testbed facility; and the CELS General Computing Environment (GCE, general-use servers for the directorate). One of the first steps in this process is unifying the user account management system. We use an internally developed system called Userbase, and all CELS user accounts are being migrated to this system. ALCF was the first to adopt the system, but we have now successfully migrated the GCE and LCRC over to this platform and are in the midst of moving JLSE over to it. In addition, when we were adding GCE, at their request we added support for integration with the lab-wide identity management system; all users are required to have an

Argonne account, and the local accounts are auto-generated from the lab accounts. GCE and LCRC are currently operating like this. JLSE will come online in that mode, and then we will migrate ALCF to this model. Besides managing the accounts, the system is also able to assist with some automated provisioning such as creating and managing Unix groups, setting user shells, creating home directories, and the like. They all share the same underlying core components in terms of authentication, authorization, identity, and data validation; each "domain" (ALCF, LCRC, etc.) can customize its site, set its own policies, and write events to customize the operation, etc., based on their needs.

## 1.3.4 Outreach Efforts

### *1.3.4.1 General Outreach*

#### *User Advisory Council*

ALCF's User Advisory Council advises the director on proposed changes to policies and services that impact users, and provides feedback on how the facility is doing. Members are appointed by the director and share an expert knowledge of the tasks and requirements of specific applications or domain areas. In 2020, the council met three times: June 15, September 22, and December 15, 2020, to hear updates on facility preparations for Aurora (training/workshops, early access to resources, outreach efforts with Intel) and other topics. Five new members were appointed to the 2020–2021 council to replace outgoing members.

#### *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 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 have the opportunity to connect with young women and introduce them to potential career paths in STEM. The 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. The ALCF also attends the Richard Tapia Celebration of Diversity in Computing Conference to recruit attendees from a diverse set of backgrounds and ethnicities.

### *1.3.4.2 Training Activities*

ALCF offers workshops and webinars on various tools, systems, and frameworks. These hands-on training programs are designed to help projects take advantage of leadership-class computers available at ALCF and enhance the performance and productivity of their research program. The facility 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 2020 training activities:

### *Simulation, Data, and Learning Workshop (virtual)*

This annual workshop helps researchers improve the performance and productivity of simulation, data science, and machine learning (ML) applications on ALCF systems. Participants learn how to scale data-centric science on ALCF systems; set up workflows and use containers; and test and debug codes. The workshop was held December 8–10, 2020, and attracted 139 attendees (with 80 people joining on the call on the first day).

### *Aurora Early Science Program Dungeon (virtual) [funded by ALCF-3 and ECP]*

The very first dungeon programming event for the ALCF Early Science Program (ESP) teams was held in December 2020 and was led by the Intel Center of Excellence (COE) for Aurora team. Four ESP project teams participated in this intense 3-day event where they worked on developing, porting, and profiling their codes with Intel and Argonne experts.

### *Intel COE for Aurora ESP Workshop 3 (virtual) [funded by ALCF-3 and ECP]*

The Intel COE for Aurora ESP workshop series helps researchers, particularly ESP and ECP project teams, prepare for Aurora. Participants get hands-on time to develop, test, and profile their codes using the latest Aurora software development kit (SDK) on Argonne's JLSE DG1 system and other systems, in consultation with Intel and Argonne experts. This invitation-only workshop took place October 21–22 (part 1) and November 5, 2020 (part 2), and attracted a total of 166 attendees.

### *2020 Performance, Portability, and Productivity at HPC (P3HPC) Forum (virtual)*

The annual P3HPC Forum is open to application developers, technology vendors, and software developers and is focused on problems of performance portability across current and future HPC platforms. The workshop explored best practices, identified major challenges, and worked with vendors and tool providers on potential implementations and solutions. The virtual workshop took place September 1–2, 2020, and attracted 350 registrants and 186 attendees, with 80 talks and 20 poster presentations.

### *Argonne Training Program on Extreme-Scale Computing (virtual)*

The Argonne Training Program on Extreme-Scale Computing (ATPESC) is a two-week summer program that teaches early career researchers about the tools and techniques needed to conduct scientific computing research on leading supercomputers. The 2020 program took place July 26–August 7, 2020, and attracted 73 attendees from 45 different institutions worldwide. The program featured talks by leading computer and computational scientists; included a track for machine learning and deep learning topics; and provided access to DOE leadership-class systems at Argonne, Oak Ridge National Laboratory (Oak Ridge or ORNL), and the National Energy Research Scientific Computing (NERSC) Center for hands-on training sessions. Video playlists from 2016–2020 ATPESC programs are available on Argonne's YouTube training channel. ATPESC is part of the Exascale Computing Project.

### *ALCF-ECP CMake Workshop (virtual)*

This workshop was designed to help exascale code developers advance their use of CMake on ALCF computing resources, including Aurora. CMake is a cross-platform, open-source build-system generator that is maintained and supported by Kitware, which hosted the event along with ALCF and the ECP. Attendees learned how to write a build system generator capable of seamlessly configuring for multiple unique architectures with a variety of compilers. The workshop, held July 15–17, 2020, was geared toward debugging and taking advantage of CMake capabilities. Registration was limited to 50 participants.

### *Computational Performance Workshop (virtual)*

This annual workshop provides code readiness exercises for prospective INCITE teams. Participants work with ALCF staff and vendor staff to debug/benchmark codes on Theta and Cooley; explore techniques for enhancing code performance; and expand their data science skills. The 2020 workshop was held May 5–7, 2020, and attracted 147 attendees.

### *Intel COE for Aurora ESP Workshop 2 [funded by ALCF-3 and ECP]*

This invitation-only workshop focused on helping ESP and ECP researchers prepare applications and software technologies for ALCF's Aurora exascale system. 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. The workshop took place at Argonne from February 25–27, 2020, and attracted 108 attendees.

### *ALCF Webinars*

The 2020 ALCF webinar program consisted of two tracks: *Developer Sessions*, attended by software and hardware developers and early users of a particular technology; and the *Aurora Early Adopter Series*, which focused on various topics aimed at preparing users for Aurora. These talks are archived on ALCF's YouTube channel with the associated training materials on the 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 2020 webinar program was as follows:

- Profiling Application Performance using Intel Vtune and Advisor (January 29, 2020)
- Automating Research Data Workflows with Globus (February 19, 2020)
- DAOS: Next-Generation Data Management for Exascale (March 25, 2020)
- Software Deployment with Spack (April 29, 2020)
- UPC++: An Asynchronous RMA/RPC Library for Distributed C++ Applications (May 20, 2020)
- Preparing Applications for Aurora using the Intel DPC++ Compatibility Tool (June 2, 2020) [funded by ALCF-3]
- OpenMP Offload Capabilities in the oneAPI HPC Toolkit (June 24, 2020)
- Best Practices for Queueing and Running Jobs on Theta (July 29, 2020)
- Data Analysis and Visualization at the ALCF (August 25, 2020)
- Overview of the Intel oneAPI Math Kernel Library (oneMKL) (September 30, 2020)
- Towards Interactive High-Performance Computing with ALCF JupyterHub (October 28, 2020)
- Profiling Deep Learning Performance with Intel VTune (December 16, 2020)

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

Hackathon events for projects in ALCF's Early Science Program are designed for early code preparation using systems available to ALCF. ALCF staff, in partnership with the Intel COE teams, held four complete hackathons in Fall 2020. Part of a fifth hackathon was held in December.

## *1.3.4.3 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.

### *Science, Technology, Engineering and Math (STEM) Activities*

#### *Summer Student Program*

Every summer, the ALCF opens its doors to a new class of student researchers who work alongside staff mentors to tackle research projects that address issues at the forefront of scientific computing. This year, the internship program went virtual. The facility hosted 33 students ranging from high school seniors to Ph.D. candidates. From enabling quantum computer simulations to applying machine learning to weather simulations, the interns had the opportunity to gain hands-on experience with some of the most advanced computing technologies in the world. The summer program culminated with a series of seminars in which the students presented their project results to the ALCF community.

#### *2020 Illinois Regional Middle School Science Bowl*

Argonne's Educational Programs and Outreach division hosted the DOE 2020 Regional Science Bowl at Oswego East High School, where teams of middle school students from 15 schools competed on a range of STEM topics. The winners of regional events across the country advanced to the Finals of the DOE National Science Bowl in Washington, D.C., in May (January 25, 2020).

#### *Introduce a Girl to Engineering Day*

Introduce a Girl to Engineering Day is an annual lab-sponsored event held in February that pairs approximately 100 local eighth graders with Argonne engineers and scientists, including ALCF staff members, for a day of presentations and hands-on activities focused on STEM careers (February 20, 2020).

### *Science Careers in Search of Women*

Science Careers in Search of Women (SCSW) is a long-standing lab-sponsored event held in April that is aimed at encouraging female high school students to pursue STEM research careers. Chicagoland suburban schools can register up to 6 high school girls and Chicago Public Schools can register up to 10. The day-long agenda includes a keynote presentation, Q&A panel discussions with Argonne women scientists, a facility tour, poster session, and career fair. ALCF staff members annually volunteer to fill various SCSW roles. (SCSW 2020 was canceled due to the COVID-19 pandemic.)

### *CodeGirls@Argonne Camp (virtual)*

CodeGirls@Argonne is a three-day STEM summer camp for 6th- and 7th-grade girls who have little or no experience in programming, and is taught by Argonne Learning Center staff and ALCF research staff. The campers meet women scientists who use code to solve problems, practice Python coding fundamentals, and experiment with robotics. The group also tours the ALCF machine room and visualization lab (June 22–26, 2020).

### *Coding for Science Camp (virtual)*

Coding for Science Camp is a week-long coding camp held in July for 30 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 (July 13–17, 2020).

### *Big Data Camp (virtual)*

Big Data Camp is a week-long summer camp held in July for high school juniors and seniors who have programming experience to teach them techniques for probing and analyzing massive scientific datasets, including data visualization methods. Organized by Argonne's Educational Programs Office and taught by Argonne computer scientists, the 2020 campers worked with data from the Array of Things (AoT) project and its follow-on Software-Defined Sensor Network (SAGE) project, an Argonne–University of Chicago urban sensor project, to experiment firsthand with investigative research (July 20–24, 2020).

### *Hour of Code (virtual)*

The Hour of Code takes place each year in December during Computer Science Education Week (CSEdWeek). More than 50 Argonne computer scientists, including many ALCF staff members, visit Chicago and area suburban schools to assist teachers in various Hour of Code activities. CSEdWeek was established by Congress in 2009 to raise awareness about the need to elevate computer science education at all levels (December 7–13, 2020).

## ***Facility Tours***

Visitors to Argonne can request a tour of the ALCF. ALCF visitors include student groups, members of Congress 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.

On March 17, 2020, ALCF suspended all in-person tours when the Laboratory restricted on-campus access to only the research and operations necessary to move vital research forward, and closed the campus to visitors and non-Argonne users. Staff transitioned to live virtual tours for several student groups, including ALCF summer interns and participants in the summer coding camps and attendees of ATPESC 2020. Members of Argonne leadership, including the ALCF director and deputy division director, also hosted several DOE VIP tours of the new facility, including Under Secretary for Science Paul Dabbar in July, Office of Science Director Chris Fall in September, and Counterintelligence Deputy Director Jose Delgado in November.

### 1.3.5 Communications

#### *1.3.5.1 Website Support Center Continuous Improvement*

ALCF's web documentation is an important aspect of its user support program and contains a wide range of how-to resources as well as video tutorials on how to run an effective science campaign. A new ALCF website was launched on December 9, 2019, with a focus on creating a Support Center to centralize guides and machine documentation, facility updates, and training materials. A committee of technical staff members meets monthly to review and make recommendations on the content found in the Support Center. The continuous improvement effort is a combination of creating new web features and redesigning existing webpages. Additional features include the ability to limit an advanced search to strictly user documentation and to filter and limit search content in the new library of training videos and slides.

Because the new site launched at the end of 2019, an assessment of the Support Center's performance was conducted at the end of CY 2020, and measured the number of views of both the new and redesigned webpages. ALCF's Communications team tracked substantial growth in pageviews across both areas, including the pageviews of the new library of training videos and slides of webinars (5,014) and an increase in pageviews the redesigned training and events listing (16,893, an increase of 10,542 pageviews over what was measured in CY 2019).

#### *1.3.5.2 Article Process Improvements*

In 2020, Argonne adopted the DivvyHQ workflow automation platform to manage the content process for all public-facing articles, including those originating in the ALCF. The DivvyHQ platform tracks stories from initiation through editorial review, including approvals, replacing a content process that relied on e-mail and the Box file-sharing service. The new platform provides a shared production schedule, automated e-mail reminders, a chat function, and an area where story production teams can upload related media and comment on drafts. Adoption of DivvyHQ laboratory-wide also ensures that all relevant stakeholders are aware of articles that involve multiple divisions and/or directorates.

#### *1.3.5.3 Communications through Mailing Lists and Social Media*

The facility maintains several communication channels, including direct e-mail campaigns, scriptable e-mail messages, social media postings, and website postings (Table 1.6; target audiences are identified in Table 1.7). Users can opt out of the system notify and newsletter mailing lists. In 2019, the ALCF communications team added two new social media channels, Twitter and LinkedIn, to help extend the reach of their campaigns.

**Table 1.6 2020 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 1.7 2020 Target Audiences**

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

The e-mail newsletter Newsbytes highlights ALCF-supported research or advancements, advertises upcoming training events and allocation program announcements, and provides links to relevant news stories. Special e-mails advertising training and fellowship opportunities are sent throughout the year, as needed.

### *1.3.5.4 Promotional Activities and Media Hits*

In 2020, the ALCF produced and placed 50 original science stories in various news outlets in coordination with Argonne's Communications & Public Affairs (CPA) Division and other ALCF direct relationships. The media team uses the social monitoring software Meltwater to track media hits. In 2020, Meltwater reported 411 unique ALCF media hits, 233 of which were chronicled on the ALCF website, and an audience reach of 426.7 million (an increase of 100.8 million).

#### 1.3.5.5 Other Publications

The ALCF produces several high-quality publications throughout of the year that describe aspects of the facility's mission and summarize its research achievements (Table 1.8). Most of these documents are available for download on the ALCF website.

**Table 1.8 Publications Designed for Print**

| Publication | Frequency | When |
|---|---|---|
| **Press and Visitor Packets** | As Needed | As Needed |
| **Industry Brochure** | As Needed | As Needed |
| **Annual Report** | Yearly | March |
| **Science Report** | Yearly | September |
| **Fact Sheets** | Yearly | November |
| **INCITE Poster** | Yearly | December |

## Conclusion

Our users are at the forefront of all of ALCF's interactions. As a user facility, the ALCF strives to improve our user experience processes and to help our customers make the most of their allocation time on our resources. In 2020, improvements included working with ALCC and INCITE project teams to transform and optimize their scientific codes and helping DD projects reach their scientific goals of obtaining ALCC and INCITE awards. The ALCF worked in partnership with other national laboratories and ECP to present sessions on our work at the annual meetings of major national scientific societies. The ALCF upgraded Theta to provide the latest compilers and libraries for our users, and augmented the system to provide an additional resource for COVID-19 research. The Support team adapted to the remote working environment with minimal disruption to the level of service we provide our users, and conducted a wide range of training and outreach events, quickly pivoting to a virtual format, to teach best practices and to help our users explore new technologies.

# 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.

Table 2.1 summarizes all metrics reported in this section.

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

| | Theta (Cray XC40): 4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 | | | | Filesystem (Sonexion 2000 Lustre): 9.2 PB | | Archive (HPSS tape) Three 10,000-slot Libraries LTO8/LTO6 Drives and Tapes 305 PB | |
|---|---|---|---|---|---|---|---|---|
| | CY 2019 | | CY 2020 | | CY 2020 | | CY 2020 | |
| | Target | Actual | Target | Actual | Target | Actual | Target | Actual |
| Scheduled Availability | 90.0% | 99.8% | 90.0% | 98.3% | N/A | 98.8% | N/A | 98.9% |
| Overall Availability | 90.0% | 95.0% | 90.0% | 93.8% | N/A | 94.8% | N/A | 95.2% |
| System MTTI | N/A | 12.39 days | N/A | 11.45 days | N/A | 18.26 days | N/A | 17.42 days |
| System MTTF | N/A | 72.86 days | N/A | 32.74 days | N/A | 180.95 days | N/A | 180.99 days |
| INCITE Usage | 17.8M | 18.5M | 17.8M | 20.9M | N/A | N/A | N/A | N/A |
| Total Usage | N/A | 32.8M | N/A | 33.1M | N/A | N/A | N/A | N/A |
| System Utilization | N/A | 97.0% | N/A | 97.7% | N/A | N/A | N/A | N/A |
| INCITE Overall Capability [a] | N/A | 70.6% | N/A | 84.8% | N/A | N/A | N/A | N/A |
| INCITE High Capability [b] | N/A | 10.5% | N/A | 24.1% | N/A | N/A | N/A | N/A |

[a] Theta Overall Capability = Jobs using ≥ 20.0 percent (800 nodes, 51,200 cores) of Theta.

[b] Theta High Capability = Jobs using ≥ 60.0 percent (2400 nodes, 153,600 cores) of Theta.

## 2.2 ALCF Production Resources Overview

During CY 2020, ALCF operated several production resources: Theta, Sonexion Lustre filesystem, and HPSS (high-performance storage system).

- Theta supported both ALCC and INCITE campaigns all year. Theta is a 4,392-node, ~281K-core, 11.69-PF Cray XC40 with 891 TB of RAM. Two cabinets (384 nodes) are not considered part of the allocated resources to help with redundancy. See section 2.6.2 for details on the GPU expansion of Theta, which went into production on January 1, 2021. Iota is an independent test and development resource and shares no resources with Theta. Iota is a 40-node, 2,650-core Cray XC40 with 8,320 GB of RAM.
- A Sonexion 2000 Lustre filesystem is mounted by Theta and the data and analysis resource with 9.2 PB of usable space.
- The facility-wide HPSS 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 Section A.7 in Appendix A.)

## 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 DOE HPC User 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. If any user can read and write to any portion of the disk space, the storage resources are considered to be available. The availability metric provides measures that are indicative of the stability of the systems and the quality of the maintenance procedures.
- *MTTI* is the time, on average, to any outage on the system, whether unscheduled or scheduled. It is also known as MTBI (Mean Time Between Interrupt).
- *MTTF* is the time, on average, to an unscheduled outage on the system.
- *Usage* is defined as resources consumed in units of node-hours.
- *System Utilization* is the percentage of time that the system's computational nodes run user jobs. No adjustment is made to exclude any user group, including staff and vendors.

## 2.4 Theta

### 2.4.1 Scheduled and Overall Availability

Theta entered full production on July 1, 2017. The GPU extension to Theta was entered into production on January 1, 2021. In consultation with the DOE Program Manager, ALCF has agreed to a target of 90 percent overall availability and a target of 90 percent scheduled availability (ASCR requested that all user facilities use a target of 90 percent for scheduled availability for the lifetime of the production resources). Table 2.2 summarizes the scheduled and overall availability results, respectively, for Theta.

**Table 2.2 Availability Results for Theta**

| Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 | | | | |
|---|---|---|---|---|
| | CY 2019 | | CY 2020 | |
| | Target (%) | Actual (%) | Target (%) | Actual (%) |
| **Scheduled Availability** | 90.0 | 99.8 | 90.0 | 98.3 |
| **Overall Availability** | 90.0 | 95.0 | 90.0 | 93.8 |

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

#### *2.4.1.1 Explanation of Significant Availability Losses*

This section briefly describes the causes of major losses of availability for the period January 1, 2020, through December 31, 2020, as shown in Figure 2.1.


**Figure 2.1 Theta Weekly Availability for CY 2020**

**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, but not all events.

**January 31, 2020: Scheduling impacted by configuration change**

One DVS node was removed from the pool of 60 for use with our installation of LDMS for additional monitoring. A configuration file was not updated properly, and this resulted in jobs not being able to be scheduled. This is a very rare configuration change, so the process was not well understood by the team. The process was documented so that if performed again in the future, the issue will not be repeated.

**February 2, 2020: OST crashed and scheduling paused**

An OST (object storage target) crashed causing I/O to hang on Sunday, February 2; and more than 100 compute nodes crashed at the same time. The hardware watchdog process caught an OST server error and initiated a reset; however, the reset caused problems. The remainder of the OSTs were unaffected. The root cause of the failure was linked to firmware. The firmware was upgraded since this happened, and no repeat cases have happened.

**March 5, 2020: MOM3 node crashed with uncorrectable errors**
The MOM3 node where the primary Cobalt process runs crashed with uncorrectable errors. The processor was replaced. All running jobs were lost.

**July 13 to 22, 2020: Theta and Sonexion upgrade**
A planned major software upgrade to Theta and the Sonexion Lustre file system that began July 13 required an additional four days to complete, thereby creating an unscheduled outage adjacent to the scheduled outage. Various problems were encountered during the Sonexion upgrade despite Cray having vetted our Sonexion configuration prior to the upgrade and ALCF having conducted testing on the T&D Sonexion. ALCF worked closely with Cray to resolve each issue; however, many problem fixes triggered new problems when the installation processes were rerun. Fixing the BIOS firmware activation failures was an especially time-consuming task that required resetting the BIOS by physically moving a jumper to drain the CMOS and rebooting the system one failure at a time. Despite the extended outage, this overdue Sonexion upgrade was successfully completed with no data loss or subsequent problems to fix. One further note, we jumped from version 2.7 to 3.4 on the Sonexion software and thus skipped several potential upgrades—this meant fewer outages for upgrades, but also meant that this upgrade was a bit more complicated. Also, because of that large difference in versions, the upgrade could not be put off any longer in order to receive proper support from Cray.

**October 18-20, 2020: Scheduled power outage**
A scheduled ALCF outage from October 18 to October 20 shut down all resources in the data center for electrical work that included moving emergency power off (EPO) buttons. The resources were shut down on Sunday, the electrical work was completed the following day, and the work needed to bring back the resources was completed the day after.

**November 24, 2020: Breaker replaced in substation**
A Theory and Computer Science (TCS) Building substation breaker tripped on November 23. This event did not affect Theta as the Theta logins, KNL computes, and management nodes were unaffected. An electrical contractor assessed the breaker the next day and recommended replacing it. This action required powering down the entire substation and Theta the morning of November 24. Theta was returned to production later that evening. Having a spare breaker on-hand was a lesson learned from a similar previous event.

**December 17, 2020: I/O nodes in bad state**
The Theta I/O service nodes got into a bad state basically preventing any further I/O operations. Although the administrators tried to fix the problem with scheduling halted, it was determined that the best course of action was to reboot the machine. The Theta Lustre filesystem itself was not affected.

### 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 2.3 summarizes the current MTTI and MTTF values, respectively, for Theta.

**Table 2.3 MTTI and MTTF Results**

| Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 | | | | |
|---|---|---|---|---|
| | CY 2019 | | CY 2020 | |
| | Target | Actual | Target | Actual |
| **System MTTI** | N/A | 12.39 days | N/A | 11.45 days |
| **System MTTF** | N/A | 72.86 days | N/A | 32.74 days |

Theta currently functions on a biweekly maintenance schedule to perform Cray driver upgrades, hardware replacements, OS upgrades, etc. Further, ALCF uses these preventative maintenance (PM) opportunities to schedule other potentially disruptive maintenance such as facilities power and cooling work, and storage systems upgrades and patching. That being said, ALCF canceled or postponed several PMs this year as COVID-19 research was a priority. Although Theta's biweekly maintenance schedule does not directly affect MTTF, it generally tends to cap MTTI at 14 days.

An increase in the number of unscheduled full-system outages in CY 2020 (10) versus CY 2019 (4), and more total unscheduled time due to the Theta software upgrade, accounts for the large drop in MTTF. Several, but not all of the MTTF events are described in the previous section.

### 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 2.4 summarizes ALCF system utilization results, and Figure 2.2 shows system utilization over time by program.

**Table 2.4 System Utilization Results**

| Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 | | | | |
|---|---|---|---|---|
| | CY 2019 | | CY 2020 | |
| | Target | Actual | Target | Actual |
| System Utilization | N/A | 97.0% | N/A | 97.7% |

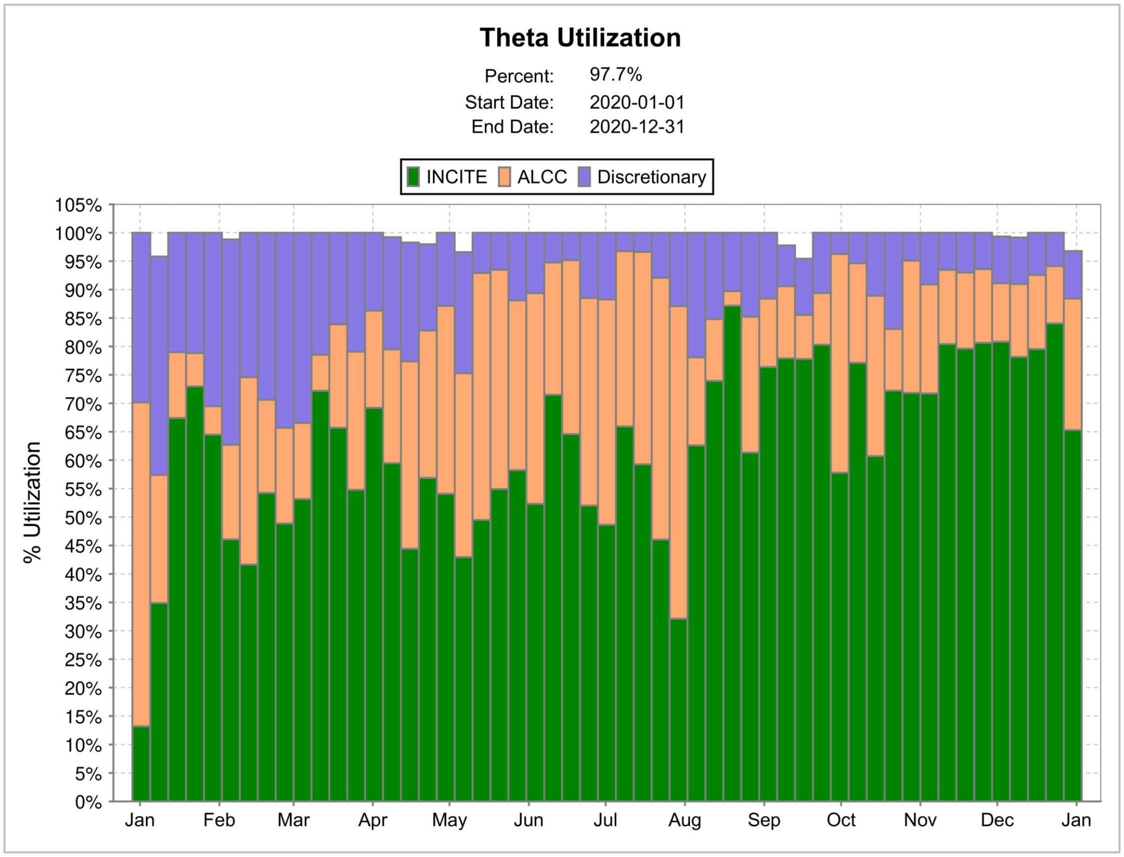



**Figure 2.2 System Utilization over Time by Program**

The system utilization for Theta was 97.7 percent for its 2020 production period of January 1, 2020, through December 31, 2020.

Table 2.5 shows how Theta's system hours were allocated and used by allocation source. Multiplying the theoretical hours by availability and system utilization values that were agreed upon with ALCF's DOE Program Manager determines the hours available. Of the hours available, 60 percent were allocated to the INCITE program, up to 20 percent to ALCC program allocations, and 20 percent to DD allocations. The allocated values for the DD allocations appear higher than expected because they represent a rolling allocation. A majority of DD projects are exploratory investigations, so the time allocations are often not used in full. DD allocations are discussed in detail in Section 3.1.2. In CY 2020, Theta delivered a total of 33.1 million node-hours.

**Table 2.5 Node-Hours Allocated and Used by Program**

| Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 | | | | | | |
|---|---|---|---|---|---|---|
| | CY 2019 | | | CY 2020 | | |
| | Allocated | Used | | Allocated | Used | |
| | Node-hours | Node-hours | % | Node-hours | Node-hours | % |
| **INCITE** | 17.8M | 18.5M | 56.5 | 17.8M | 20.9M | 63.2 |
| **ALCC** | 5.6M | 3.3M | 10.0 | 6.7M | 7.4M | 22.4 |
| **DD** | 16.9M | 11.0M | 33.5 | 9.1M | 4.7M | 14.4 |
| **Total** | 40.3M | 32.8M | 100.0 | 33.6M | 33.1M | 100.0 |

**Summary**: For CY 2020, 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. No targets were set for INCITE, nor were there any targets for ALCC And DD projects, but data is provided for all as reference. ALCC usage also encompasses COVID-19 research done on Theta per ASCR. Figure 2.4 shows the three programs' utilization of total core hours (from Table 2.6) over time, and Figure 2.5 shows the overall distribution of job sizes over time.

**Table 2.6 Capability Results**

| Theta (Cray XC40):4008-node, 251K-core 64 TB MCDRAM 770 TB DDR4 | | | | | | |
|---|---|---|---|---|---|---|
| | CY 2019 | | | CY 2020 | | |
| Capability Usage | Total Hours | Capability Hours | Percent Capability | Total Hours | Capability Hours | Percent Capability |
| **INCITE Overall [a]** | 18.5M | 13.1M | 70.6% | 20.9M | 17.7M | 84.8% |
| **INCITE High [b]** | 18.5M | 1.9M | 10.5% | 20.9M | 5.0M | 24.1% |
| **ALCC Overall** | 3.3M | 0.9M | 26.3% | 7.4M | 4.7M | 63.8% |
| **ALCC High** | 3.3M | 0.0M | 0.0% | 7.4M | 0.2M | 3.3% |
| **Director's Discretionary Overall** | 11.0M | 4.4M | 39.8% | 4.7M | 1.7M | 35.2% |
| **Director's Discretionary High** | 11.0M | 0.8M | 6.9% | 4.7M | 0.8M | 16.7% |
| **TOTAL Overall** | 32.8M | 18.3M | 55.8% | 33.1M | 24.1M | 72.9% |
| **TOTAL High** | 32.8M | 2.7M | 8.3% | 33.1M | 6.1M | 18.4% |

[a] Theta Overall Capability = Jobs using ≥ 20.0 percent (800 nodes, 51,200 cores) of Theta.

[b] Theta High Capability = Jobs using ≥ 60.0 percent (2400 nodes, 153,600 cores) of Theta.

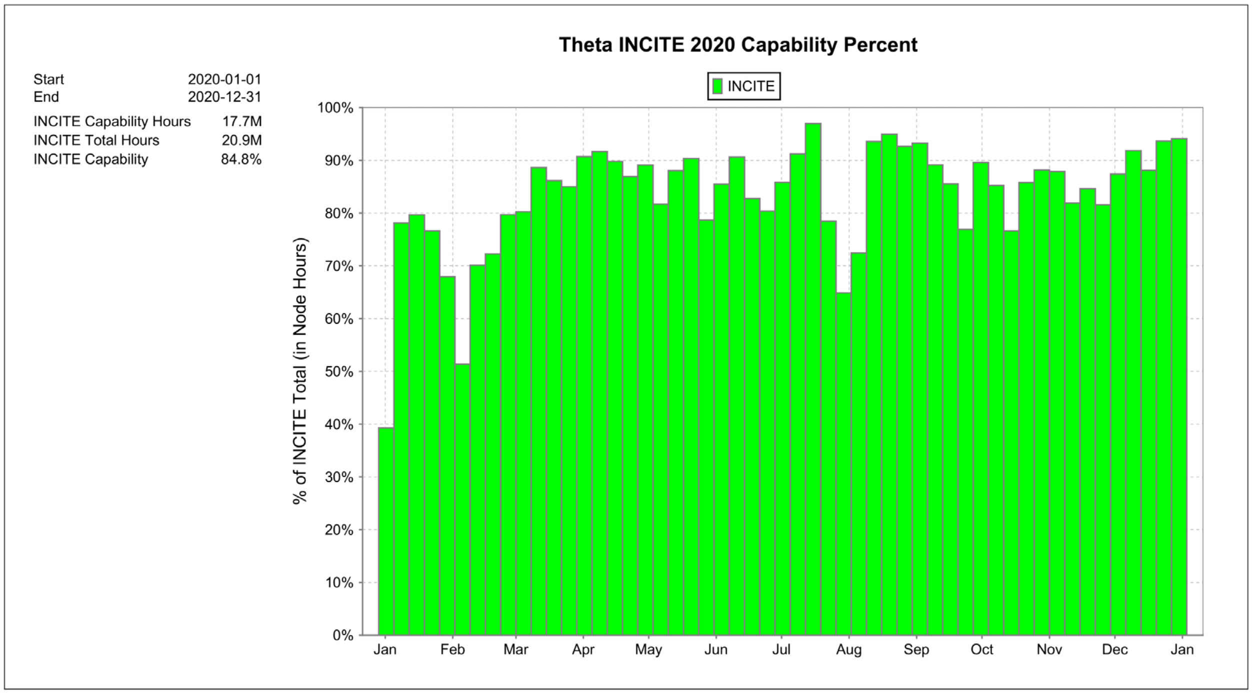



**Figure 2.3 Theta INCITE Overall Capability**

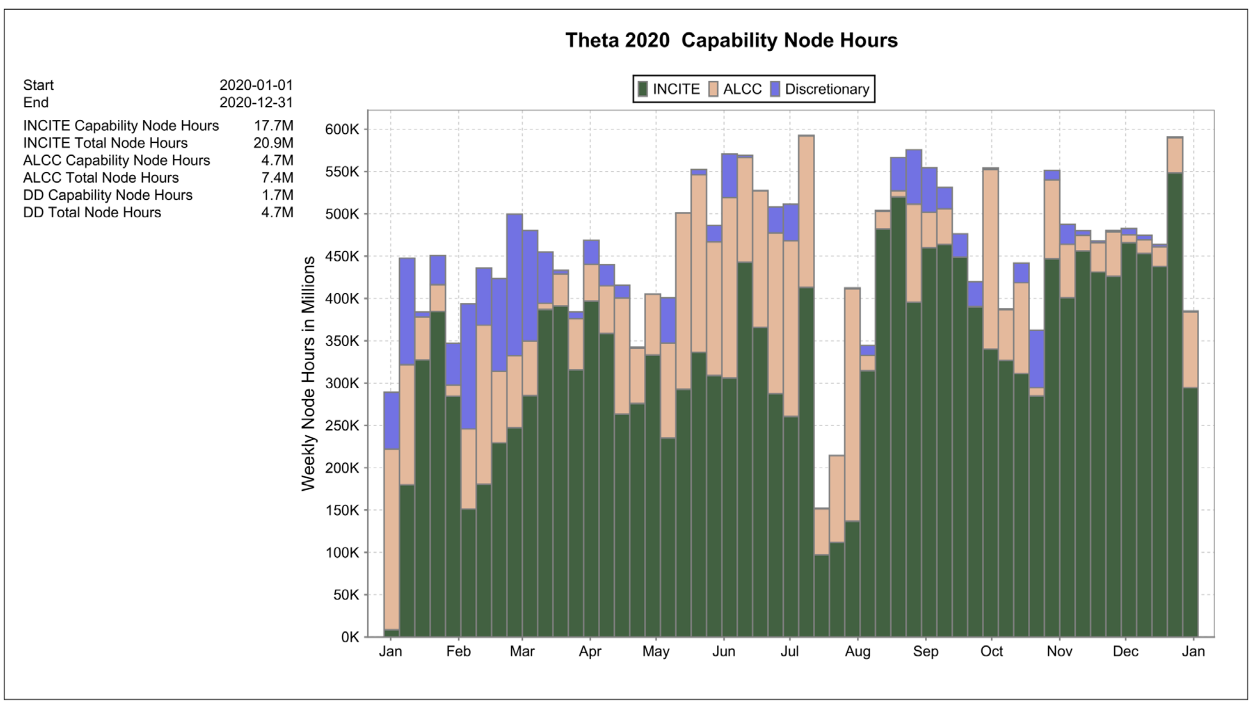



**Figure 2.4 Theta Capability Node-Hours by Program**

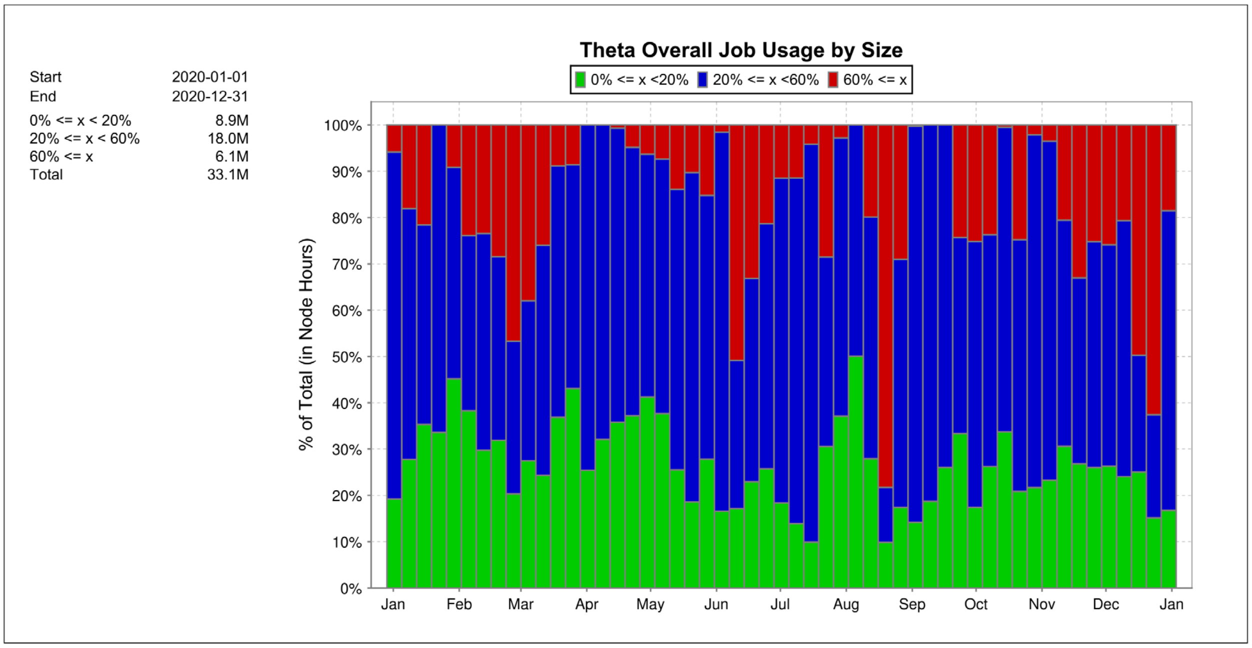



**Figure 2.5 Theta Job Usage by Size**

## 2.5 Storage

This section covers availability and MTTI/F metrics for the production storage resources. This is the first year that storage resources are being covered separately, and the availability sections will propose targets for the year.

### 2.5.1 Theta Lustre File System

Theta Lustre is a Cray Sonexion 2000 Lustre filesystem 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 at the same time, and Theta Lustre was in previous years considered part of Theta and not treated separately. Theta Lustre entered full production on July 1, 2017, at the same time as Theta.

#### *2.5.1.1 Scheduled and Overall Availability*

We are proposing target metrics of 90 percent overall and scheduled availability for CY 2021 since it is strongly integrated with Theta and in response to ASCR's request that all user facilities use a target of 90 percent for scheduled availability for the lifetime of the production resources. Table 2.7 summarizes the availability results.

**Table 2.7 Availability Results**

| Lustre File System | | |
|---|---|---|
| | CY 2020 | |
| | Target (%) | Actual (%) |
| Scheduled Availability | N/A | 98.8 |
| Overall Availability | N/A | 94.8 |

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

### *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, 2020, through December 31, 2020, as annotated in Figure 2.6.

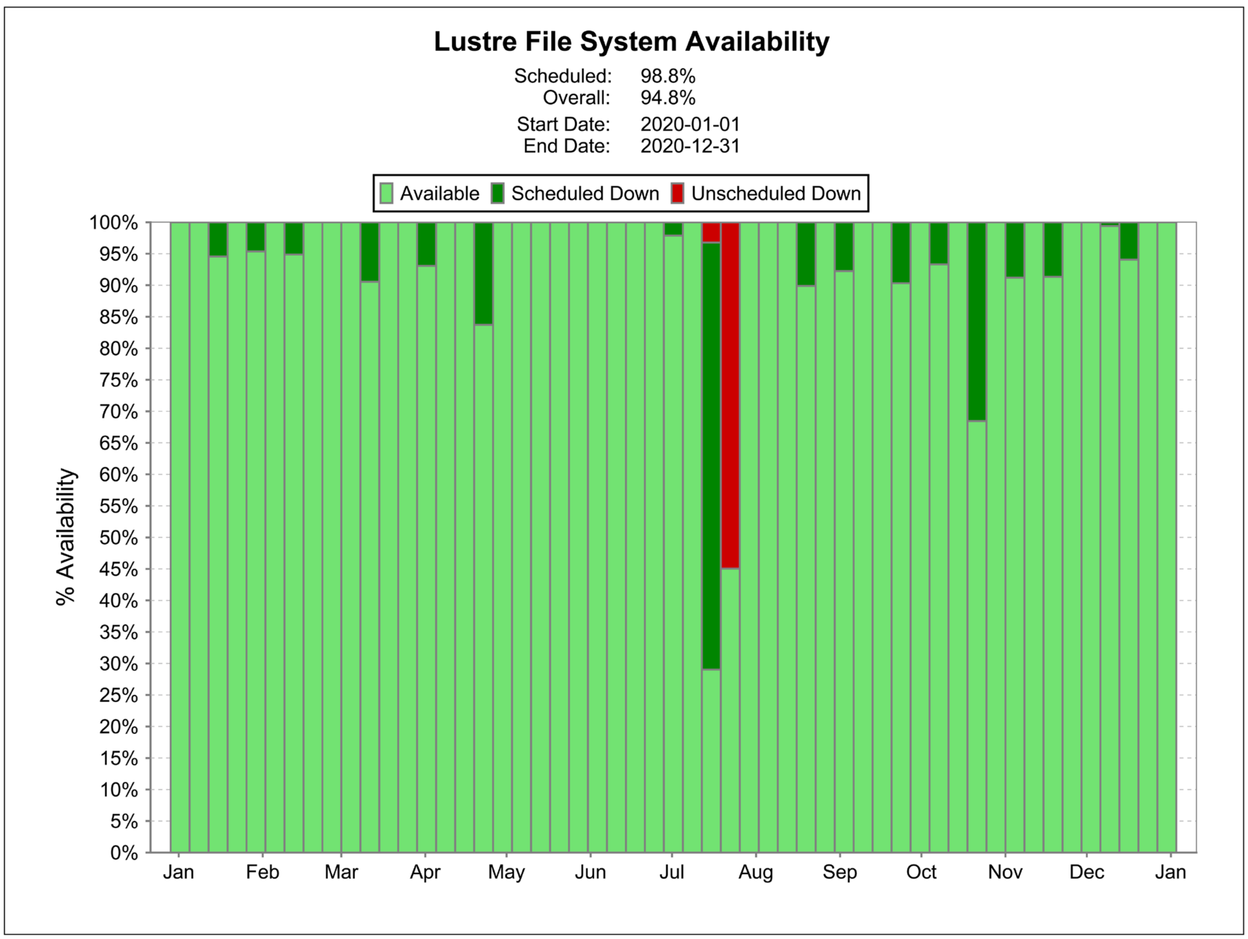



**Figure 2.6 Theta Lustre Weekly Availability for CY 2020**

**Graph Description:** Each bar in Figure 2.6 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. Each of the significant loss events is described in detail below (these also appeared in the Theta section).

**July 13 to 22, 2020: Theta and Sonexion upgrade**
A planned major software upgrade to Theta and the Sonexion Lustre file system that began July 13 required an additional four days to complete, thereby creating an unscheduled outage adjacent to the scheduled outage. Various problems were encountered during the Sonexion upgrade despite Cray having vetted our Sonexion configuration prior to the upgrade and ALCF having conducted testing on the T&D Sonexion. ALCF worked closely with Cray to resolve each issue; however, many problem fixes triggered new problems when the installation processes were rerun. Fixing the BIOS firmware activation failures was an especially time-consuming task that required resetting the BIOS by physically moving a jumper to drain the CMOS and rebooting the system one failure at a time. Despite the extended outage, this overdue Sonexion upgrade was successfully completed with no data loss or subsequent problems to fix. One further note, we jumped from version 2.7 to 3.4 on the Sonexion software and thus skipped several potential upgrades—this meant fewer outages for upgrades, but also meant that this upgrade was a bit more complicated. Also, because of that large difference in versions, the upgrade could not be put off any longer in order to receive proper support from Cray.

**October 18–20, 2020: Scheduled power outage**
There was a scheduled ALCF outage from October 18–20, 2020, to shut down all resources in the data center for electrical work that included moving EPO buttons. The LCF resources were shut down on a Sunday, the electrical work was completed Monday, and LCF brought back resources starting Monday evening and then completed the work Tuesday.

### 2.5.1.2 MTTI and MTTF

#### MTTI and MTTF Summary

MTTI and MTTF are reportable values with no specific targets. Table 2.8 summarizes the current MTTI and MTTF values.

**Table 2.8 MTTI and MTTF Results**

| Lustre File System | | |
|---|---|---|
| | CY 2020 | |
| | Target | Actual |
| System MTTI | N/A | 18.26 days |
| System MTTF | N/A | 180.95 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. Due to COVID-19 research being performed on Theta, several PMs were canceled or postponed to allow crucial research to take place.

### 2.5.2 Tape Storage

The facility-wide HPSS tape archive was available to all ALCF users from all compute resources in 2020 as it has been in previous years. The tape storage is comprised of three 10,000-slot libraries with LTO8 tape drives and LTO8 tapes, with some legacy LTO6 drives and tapes remaining. The first tape library went into production in 2009 in the old Interim Supercomputing Support 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 ALCF also permanently vacated 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.

#### *2.5.2.1 Scheduled and Overall Availability*

As this is the first time we are reporting HPSS availability, we propose target metrics of 95 percent scheduled availability and 90 percent overall availability going forward. Table 2.9 summarizes the availability results.

**Table 2.9 Availability Results**

| HPSS Archive | | |
|---|---|---|
| | CY 2020 | |
| | Target (%) | Actual (%) |
| Scheduled Availability | N/A | 98.9 |
| Overall Availability | N/A | 95.2 |

Note that HPSS is considered unavailable when users are not able to 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 January 1, 2020, through December 31, 2020, as annotated in Figure 2.7.

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

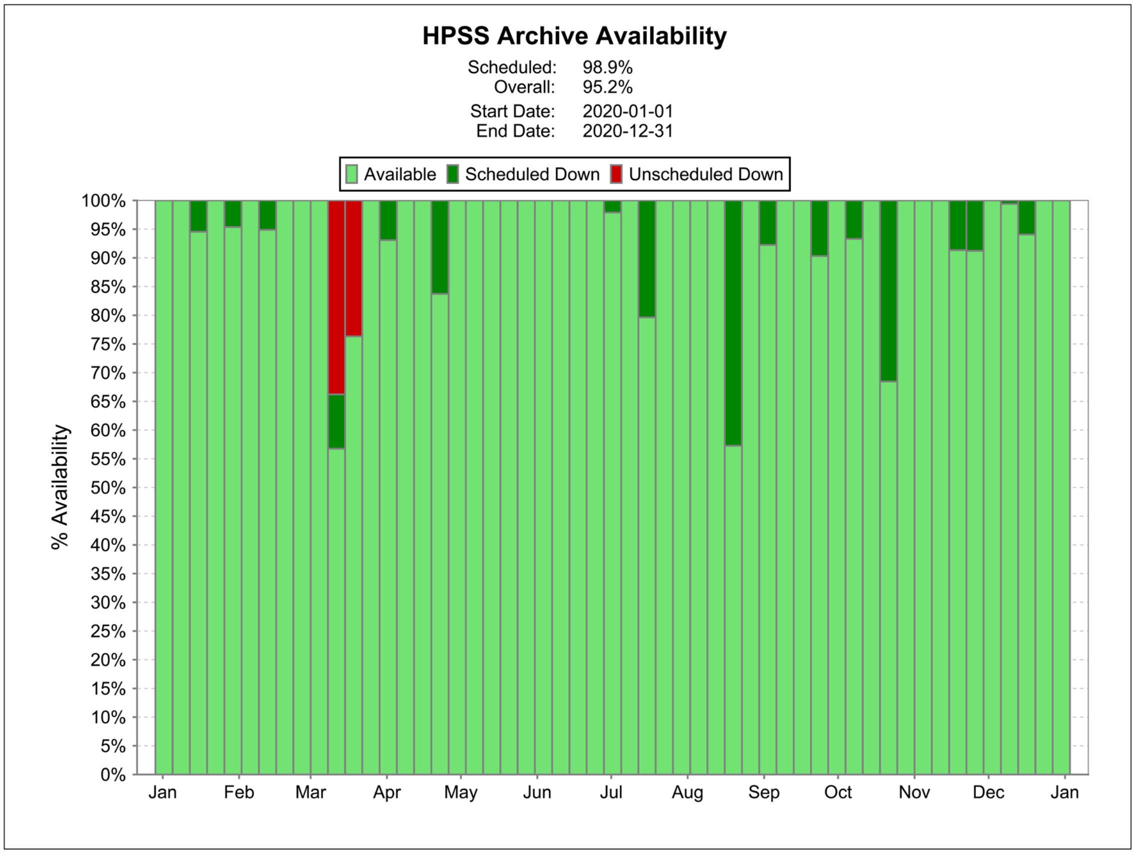



**Figure 2.7 HPSS Weekly Availability for CY 2020**

**March 12–14, 2020: Single library under repair**
One library suffered a damaged tractor cable and required replacement. The tractor cable was replaced, but more parts were required to stabilize the cable. The library was finally fixed on March 16. Although this library was not in service, the other two libraries remained in production and were accessible.

### 2.5.2.2 MTTI and MTTF

#### MTTI and MTTF Summary
MTTI and MTTF are reportable values with no specific targets. Table 2.10 summarizes the current MTTI and MTTF values.

**Table 2.10 MTTI and MTTF Results**

| HPSS Archive | | |
|---|---|---|
| | CY 2020 | |
| | Target | Actual |
| **System MTTI** | N/A | 17.42 days |
| **System MTTF** | N/A | 180.99 days |

HPSS maintenance is not regularly scheduled, but when it is scheduled, it is opportunistically lined up with Theta maintenance. HPSS is often available even though other resources may be in preventative maintenance.

## 2.6 Center-Wide Operational Highlights

### 2.6.1 Mira Decommission

Mira, a 10-petaflops IBM Blue Gene/Q system with 49,152 nodes, was decommissioned on March 9, 2020. It was number 3 on the Top 500 when it debuted, and was still in the top 25 when it was retired. It delivered 39.6 billion core-hours for more than 800 projects in its seven-plus years of production. Over that time, more than 700 full-machine jobs were run—a remarkable feat. Mira supported two special projects until March 9, after which time it and all BG/Q systems were decommissioned.

In March and April 2020, ALCF Operations staff worked to safely decommission the 54 Blue Gene/Q cabinets of Mira, Vesta, and Cetus. ALCF worked with Argonne's Procurement and the datacenter's facility management, as well as engaged recycling contractor Elgin Recycling to remove each of the 4,400-pound Blue Gene/Q racks, as well as additional legacy support racks. This task was more made challenging because of a deadline driven by building construction. ALCF prepared work planning documents for the various steps of work as noted in Section 6.

The effort included working with facility management to safely disconnect power and drain water from the Blue Gene/Q cooling loop. This step was performed with minimal impact to Theta and other water-cooled systems in the datacenter. After draining and establishing an electrically safe work environment, ALCF staff disconnected power cables, networking cables, and cooling lines from the racks.

Once disconnection was complete, Elgin Recycling came in and followed procedures to remove all node boards from each cabinet, splitting the roughly 4,400 pounds and distributing it to two 2,200-pound packages: a cabinet and stack of node boards. This approach was followed for safe transport. Elgin Recycling took 5 business days to fully remove the 54 Blue Gene/Q racks and 9 additional support racks. ALCF Operations staff followed this up with 3 business days of effort to remove over 40 miles of optical cable (approximately 5,000 pounds) and roughly 500 pounds of copper cable.

All of the physical work was performed on-site during the first months of Argonne's COVID-19 pandemic response. All Argonne staff and contractors followed normal safety procedures in addition to the heightened controls included in response to the pandemic. There were no reported safety or COVID-19 exposures among any of our teams.

### 2.6.2 ThetaGPU Installation

In support of COVID-19 research, ALCF executed a plan to expand the Theta supercomputer with NVIDIA DGX-A100 GPUs. The expansion fulfills DOE ASCR's high-priority request to provide an essential resource to support COVID-19 research as soon as possible. The DGX-A100 nodes are composed of two AMD Rome CPUs (64-core) and eight A100 GPUs. There are

1 TB of DDR4 memory and 320 GB of GPU memory per node, and each node has eight HDR200 ports and 2 HDR200 storage ports connected in a basic fat tree with full bisection.

ALCF received the first three DGX-A100 nodes in May and installed them into the Theta test and development machine (Iota is an air-cooled XC40). This placement allowed ALCF to quickly install and have them available for use by COVID-19 researchers within 2 days of delivery.

The second phase was the delivery of the remaining 21 nodes, which were delivered on July 20 after the vendor experienced delays due to COVID-19. These nodes arrived in 7 racks, which were immediately installed and had power within 2 days. The vendor installation took 2 weeks longer than they had planned. ALCF identified some missing components and configuration errors by the vendor as well, further delaying acceptance, which happened on September 8. The first friendly users were put on within a few days, and the machine was opened up in production to Argonne computing on September 22. General user availability and production began on January 1, 2021.

### 2.6.3 HDR Installation

In anticipation of a new global filesystem and the Aurora supercomputer coupled with retirement of the BG/Q hardware and aging QDR InfiniBand network, ALCF installed a new InfiniBand HDR200 core. The HDR200 core consists of a three-level fat tree built with two 800–port, water-cooled core chassis feeding 80 40-port, air-cooled edge switches, for a total of 1600 ports of 200 Gbps HDR InfiniBand with a non-blocking aggregate bandwidth of 40 TB/s.

The new HDR200 core was installed in June and fully operational by July. It was subsequently fully integrated with the aging QDR, FDR, and EDR networks, with management nodes of the HDR core controlling routing. The integration was completed very successfully with no issues

### 2.6.4 GFS Installation

The ALCF has been working toward a new global filesystem (GFS) since the RFP award in late 2019. The Cray E1000 submission was chosen to provide one 150 PB filesystem (Grand) at 1 TB/s and one 50 PB filesystem (Eagle) at 300 GB/s. The latter was designated to be a community filesystem where data sharing with non-ALCF users would be possible. By July 2020, the hardware had been fully delivered, and by September 20, 2020, the GFS was accepted.

After award, ALCF decided to change the configurations of the filesystems into two equivalent filesystems, each with a capacity of 100 PB and peak I/O speeds of 650 GB/s. The change was made to minimize the likelihood of duplicates of large data given that the faster filesystem would be the most attractive place for large I/O, and users would want to share some of these same very large data sets with non-ALCF communities. Thus, data sets from those projects where sharing is desired will be put on the sharing filesystem from the start to avoid the need for duplicates. Because the sharing capability is actually orthogonal to the filesystem implementation using Globus, it would be possible to implement sharing on the non-sharing filesystem at a later date if desired.

In summary, ALCF has a GFS configured as two identically sized filesystems in Grand and Eagle, and Eagle will have the additional capability of sharing via Globus. The filesystems started being in production on January 1, 2021.

### 2.6.5 Shasta Testbed

A Shasta testbed consisting of 256 AMD Rome processors assembled in a single cabinet was delivered to ALCF on March 30. The testbed, named Crux, is one of many machines that ALCF is using to prepare for Aurora. Installing and running the Shasta software on Crux, while initially a buggy process, was instructive, and it was accepted on July 30 after successfully running HPL several times on version 1.2 of Shasta.

After acceptance, the machine was studied carefully and prepared for an install of the Shasta v1.3 software. However, the plans changed and HPE’s HPCM system management software was installed instead so ALCF could obtain experience with a completely different software stack.

## Conclusion

ALCF is maximizing the use of its HPC systems and other resources consistent with its mission. We have exceeded the 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 2020, this regular failure analysis has continued to drive code changes to software infrastructure at ALCF (like UB3 and Cobalt) and has provided details to support debugging of compute and storage system problems.

*This page intentionally left blank.*

# 3. Allocation of Resources

***Is the allocation of resources reasonable and effective?***

## ALCF Response

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

## 3.1 Allocation of Resources

### 3.1.1 Usage of the INCITE and ALCC Hours

The INCITE 2020 program allocated 17.8M node-hours on Theta. The allocation usage for Theta is shown in Figure 3.1. A total of 20.9M node-hours was delivered to INCITE on Theta (Table 3.1). Of the 14 INCITE projects, only 1 project used less than 50 percent of its allocation. The other 13 projects used more than 75 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. The overuse of Theta was made possible through the use of the backfill queue (low priority) and an “overburn” policy that permitted projects to continue running capability-sized jobs after their allocations were completely exhausted.

For the 2019–2020 ALCC year, 18 projects had allocations on Theta for a total of 6.7M node-hours. The allocation usage is shown in Figure 3.2. Three of these projects used under 50 percent of their allocations, 5 projects used between 50 and 90 percent of their allocations, and 10 projects used their entire allocations or more.

The 2020–2021 ALCC year is approximately halfway through its allocation cycle. So far, the 31 projects have received allocations of 6.5M node-hours on Theta. The projects have used a total of 1.8M node-hours from July 1, 2020, through December 31, 2020. The allocation usage is shown in Figure 3.3.

Table 3.2 summarizes the ALCC node-hours allocated and used in CY 2020, and includes COVID-19 research projects done on Theta per ASCR. The total ALCC node-hours allocated, 6.7M, are calculated by adjusting the 2019–2020 and 2020–2021 Theta ALCC year allocations by the percentage of their award cycle occurring in CY 2020, then summing these two values. The total ALCC node-hours used, 7.4M, is the sum of all node-hours charged against any Theta ALCC project in CY 2020.

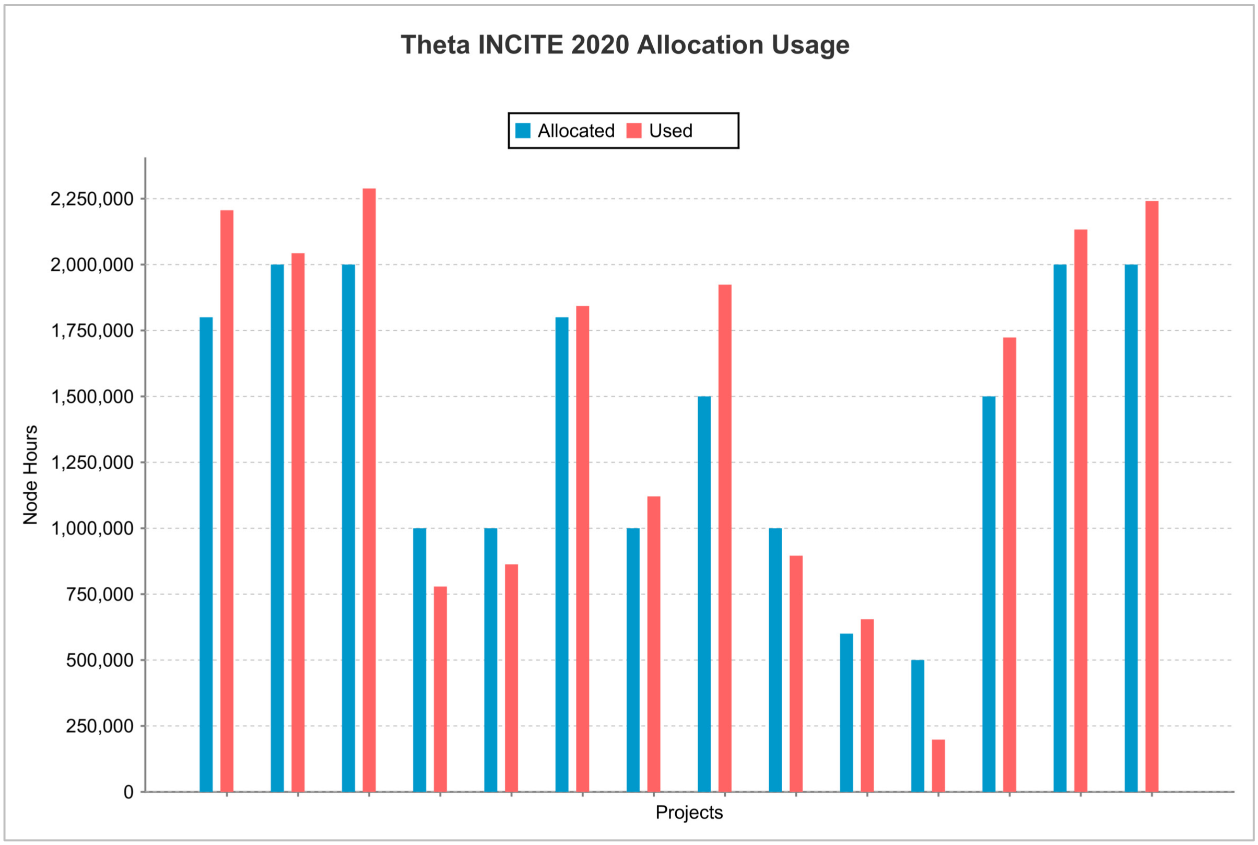



**Figure 3.1 Theta INCITE 2020 Allocation Usage (Note: Projects are randomly ordered.)**

**Table 3.1 INCITE 2020 Time Allocated and Used on Theta**

| Projects | Theta |
|---|---|
| **Allocated Node-Hours** | 17.8M |
| **Used Node-Hours** | 20.9M |

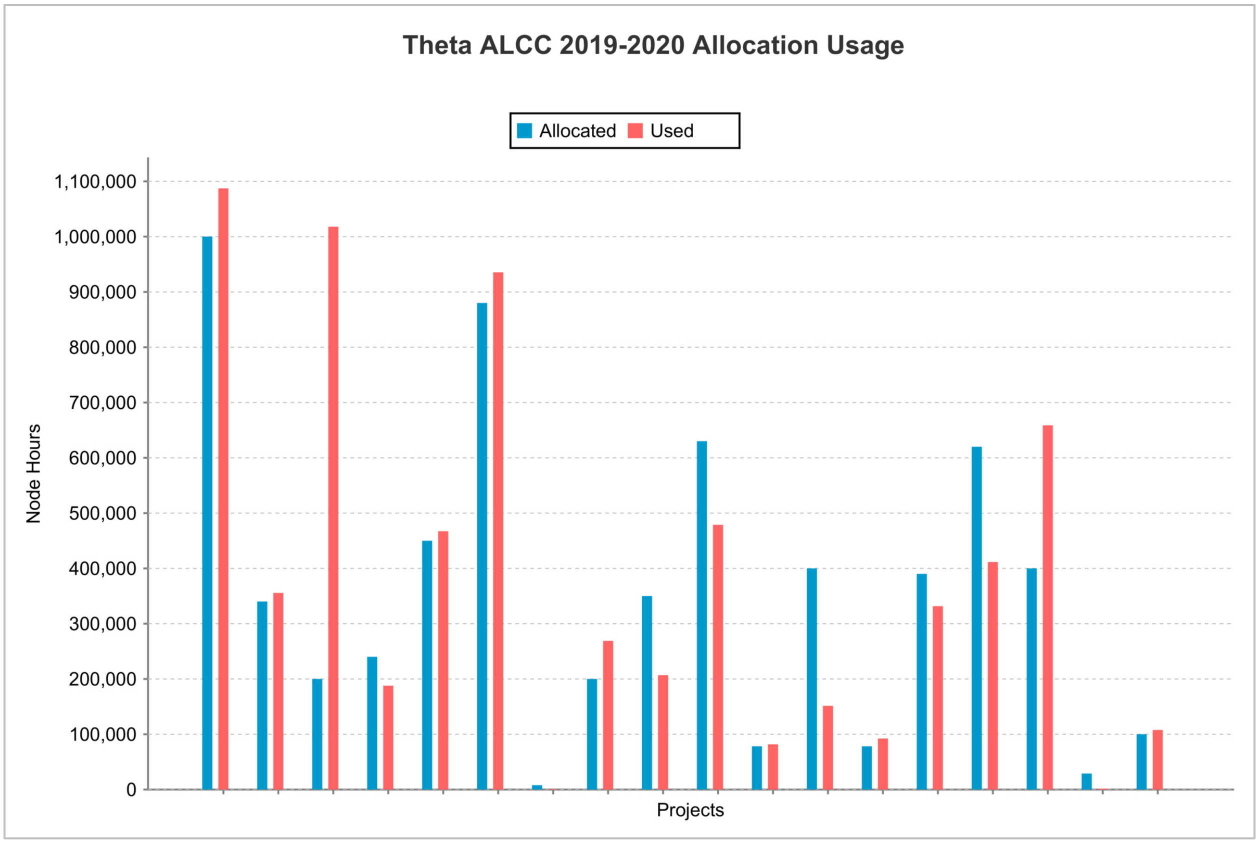



**Figure 3.2 Theta ALCC 2019–2020 Allocation Usage (Note: Projects are randomly ordered.)**

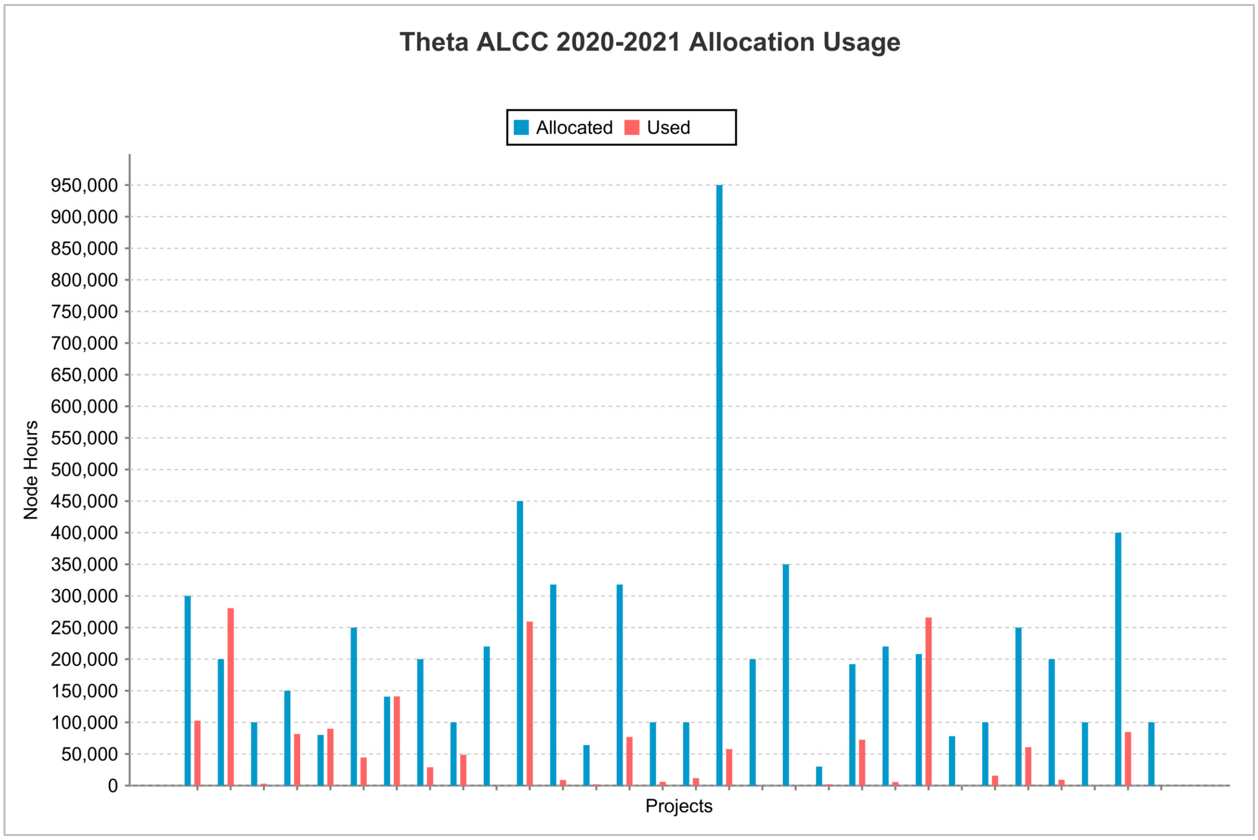



**Figure 3.3 Theta ALCC 2020–2021 Allocation Usage as of Dec. 31, 2020 (Note: Projects are randomly ordered.)**

**Table 3.2 ALCC Time Allocated and Used on Theta in CY 2020**

| Projects | Theta |
|---|---|
| **Allocated Node-Hours** | 6.7M [a] |
| **Used Node-Hours** | 7.4M [b] |

[a] Allocation total is obtained by adjusting each ALCC cycle allocation (2019–2020, 2020–2021) to prorate it for the amount of time allocated in CY 2020, then summing.

[b] Usage total is the number of node-hours charged for jobs run against any ALCC allocation in CY 2020.

### 3.1.2 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 proposing for and receiving an INCITE award, 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 a program 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 division activities. All activities come 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.

As of the mid-point of 2018, the discretionary pool grew to 20 percent of the system to support the needs of the ECP program. ECP and the computing facilities run a Resource Allocations Council (RAC) that meets on a monthly basis 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 2020. 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; rather, the time is provided on a first-come, first-served basis. 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, 2020**

| Projects | Theta |
|---|---|
| **Allocated Node-Hours** | 9.1M |
| **Used Node-Hours** | 4.7M |

A list of the CY 2020 DD projects on Theta, including title, PI, institution, and hours allocated, is provided in Appendix B.

Figure 3.4 provides a breakdown of the CY 2020 DD allocations by domain for Theta.

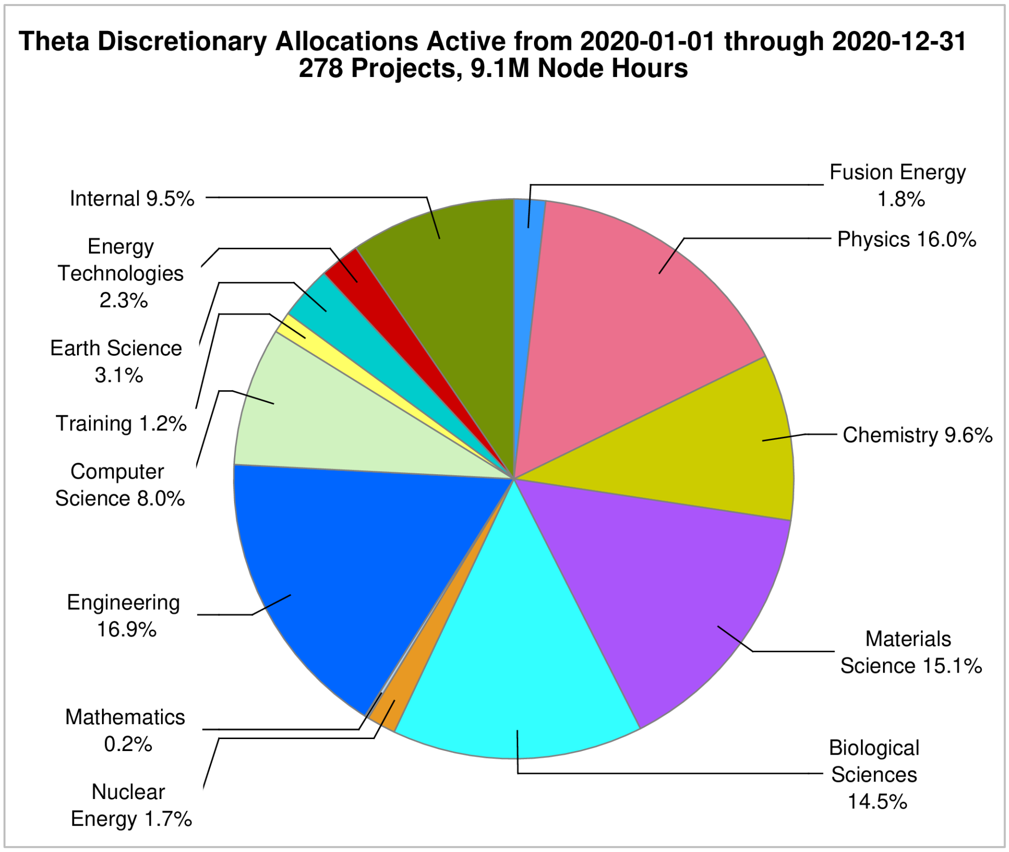



**Figure 3.4 Theta CY 2020 DD Allocations by Domain**

## Conclusion

The ALCF delivered the following node-hours to the allocation programs in CY 2020: 20.9 million to INCITE, 7.4 million to ALCC, and 4.7 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 on HPC machines and to achieve science goals that could not otherwise have been reached.

# 4. Innovation

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

***(b) Is the facility advancing research, either intramurally or through external collaborations, that will impact next generation high performance computing platforms?***

***(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 2020. 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

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

### 4.1.1 Molecular Database for SARS-CoV-2 Research

**Challenge**: Argonne was part of collaborative project that used high-performance computing and artificial intelligence to discover novel small molecules that can inhibit various virus proteins and inhibit the replication of SARS-CoV-2. However, the database initially developed was unable to return results quickly enough for researchers to do broad queries.

**Approach:** ALCF staff worked with the project to stand up a new COVID Research Molecule Database by first standing up a PostgreSQL database server for the 2.5 TB of data representing 4.2 billion molecules from 24 different sources. However, this database took too long to respond to queries. ALCF staff then implemented the database using IBM's DB2 Database with BLU Acceleration. The database was tuned and optimized so that researchers' queries were returned in a timely fashion. ALCF explored three ways to store the data: regular database tables, compressed database tables, and IBM's BLU columnar tables.

Because the data needed to be shared outside of Argonne, ALCF staff deployed a REST API server that communicated with the database (Figure 4.1) and implemented API services that allowed researchers to query the database without knowing the database structure.

**Impact/Status:** The proof-of-concept implementation showed researchers were able to query 4.2 billion molecules, with a 15–200 ms response time versus the hundreds of seconds with the Postgres version. ALCF can apply this technology to other scientific processes and data.

Testing showed significant opportunity for improvement on similar research in the future if the workflow were moved to IBM's DB2 BLU. Since many of the rows have the same values for some of the columns, a BLU solution would be faster and require less disk space since it utilizes compression as well as taking advantage of a SIMD architecture to scan the data in parallel. Implementing all of this functionality on an IBM DB2 database with one billion rows of data can show significantly improved performance, with queries experiencing 500x speedups over the initial Postgres installation.

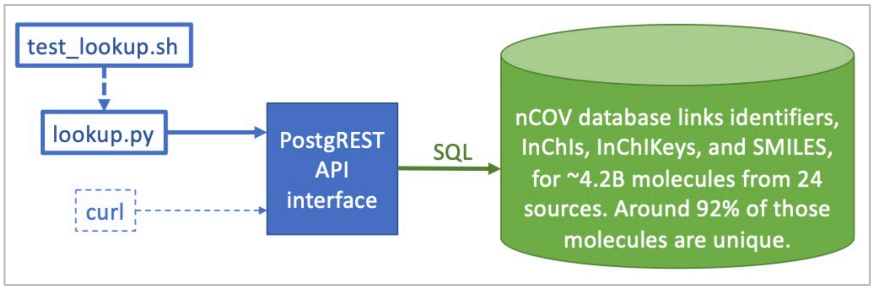



**Figure 4.1 REST API to Access the nCOV Postgres Database**

### 4.1.2 Collaboration with OpenPBS Community

**Challenge:** ALCF must ensure that it has a scheduler capable of meeting the demands of an exascale workload that includes not only the traditional large-scale MPI simulations, but data-intensive and AI workloads, as well. As noted last year, ALCF has discontinued new development on the Argonne-developed Cobalt scheduler and is now contributing to the OpenPBS open source community.

**Approach:** ALCF collaborates with the OpenPBS community to develop features needed by ALCF, focused on contributing the functionality that exists in Cobalt, particularly around instrumentation, which is needed to drive reporting tools. In 2020, ALCF developed three feature additions that were merged into the OpenPBS master. These are additional event hooks for the following events:

- Execute admin command event – allows PBS sites to track when a node is taken on/offline by an administrator, as well as scheduler configuration changes.
- Node state change event – allows PBS sites to track when the server determines that a node has gone on/offline.
- Modifications that allow a PBS developer to do a standard Python import of the pbs and pbs_ifl modules, useful in development and testing.

Additionally, ALCF submitted five bug reports. For two of these, ALCF's fixes have been merged. Two others are going through review. In the final case, Altair said the bug is complex enough and invasive enough that they will fix it themselves.

ALCF is also involved in community discussions about the future of OpenPBS. ALCF has provided significant feedback on the multi-server design for additional resiliency and scalability. ALCF suggested the use of Cython to ease development on the server and improve performance. Altair has issued a design document for public comment and plans on implementing it. ALCF has also been actively encouraging the development of standardized APIs, such as REST. Altair is reviewing this internally, but has not yet decided to actively work on it.

**Impact/Status:** Collaborating with the OpenPBS community allows ALCF to leverage work and expertise from a wide variety of sources, producing better software at a lower cost to ALCF while increasing the impact. In 2021, ALCF will deploy the first production machines running OpenPBS. ALCF will continue to pursue standardized interfaces, improvements in the reservation system, and significant interaction with the OpenPBS community and other stakeholders.

### 4.1.3 Hyper-Parameter Optimization (HPO) and Network Architecture Search (NAS) for AI-Driven Science

**Challenge:** ALCF supports many projects using deep neural networks, but development of models can be time consuming for domain scientists.

**Approach:** ALCF, in collaboration with Argonne's Mathematics and Computer Science (MCS) Division, has enhanced the DeepHyper AutoML tool for AI-driven science. DeepHyper has been augmented to help automate the development of deep learning-based predictive models using network architecture search (NAS).

**Impact/Results:** The NAS approach has identified deep neural network architectures in cancer research with fewer trainable parameters, shorter training time, and accuracy matching or surpassing that of their manually designed counterparts (Figure 4.2). DeepHyper has scaled on ALCF supercomputers and is being used by Aurora Early Science application teams, as well as with application domains spanning cancer research, weather forecasting, and fluid dynamics.

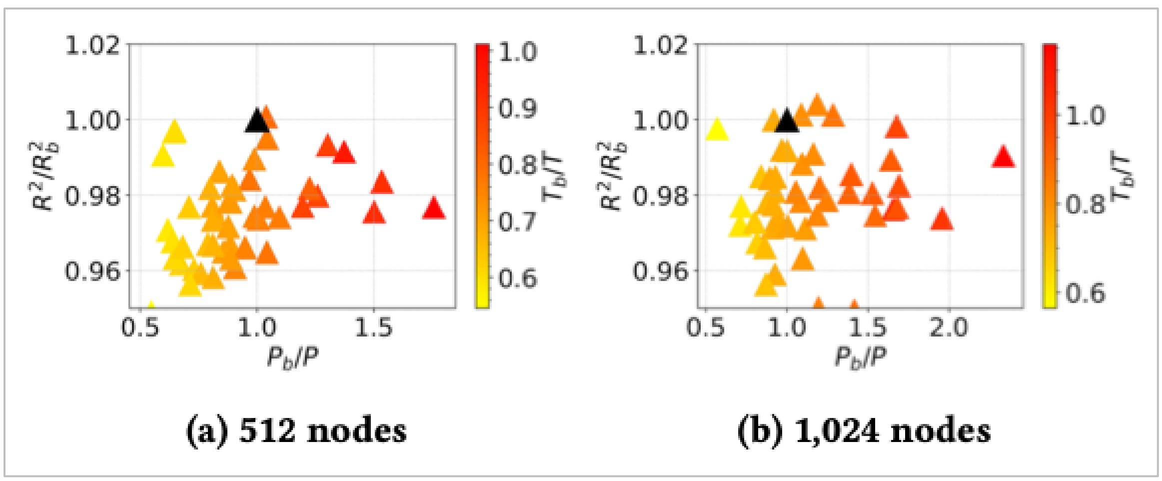



**Figure 4.2 Post-training results of DeepHyper with NAS using A3C on Combo Cancer problem with the large search space run on 512 and 1,024 nodes of Theta with agent scaling.**

### 4.1.4 Adaptive Routing Study

**Challenge:** The Cray Aries Network interconnects the nodes on Theta using a three-level dragonfly topography. The network performance of the dragonfly-based systems is determined by the efficient operation of the adaptive routing. The Aries adaptive routing protocol selects the best path through the network on a packet-by-packet basis based on the congestion information. The routing method attempts to avoid congested routes by allowing for packets to traverse over non-minimal routes if required. Aries supports four different adaptive routing schemes to balance between minimal and non-minimal route selection to avoid congestion, however, the default scheme recommended by the vendor was found to be suboptimal as per our detailed experimental analysis.

**Approach:** ALCF staff studied the workload and performance of applications running on Theta under the vendor-recommended default (adaptive 0) routing scheme and compared that with the adaptive 3 routing scheme. Adaptive 3 routing scheme has more bias toward minimal routes. The evaluation demonstrated between a 0–10% improvement in production application runtimes. This was enabled by better path selection, as preference to minimal routes helped reduce overall application-level congestion (Figure 4.3(a)), as well as system-level congestion (Figure 4.3(b)).

**Results/Impact:** Based on this study, ALCF management decided to make the adaptive 3 routing scheme the default routing scheme on Theta. ALCF staff changed the default routing on Theta to adaptive 3 in July 2020. Based on this work and resultant publication ("Performance Evaluation of Adaptive Routing on Dragonfly-based Production Systems," Sudheer Chunduri, Kevin Harms et al. IPDPS 2021), NERSC made adaptive 3 the default routing scheme on Cori.

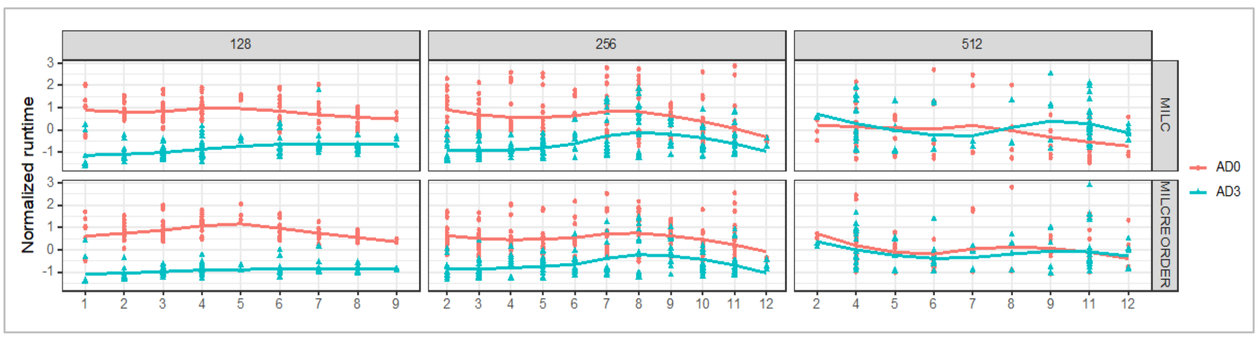



**Figure 4.3(a) MILC and MILCREORDER on three different job sizes ordered by the number of groups spanned with AD0 and AD3 routing modes.**

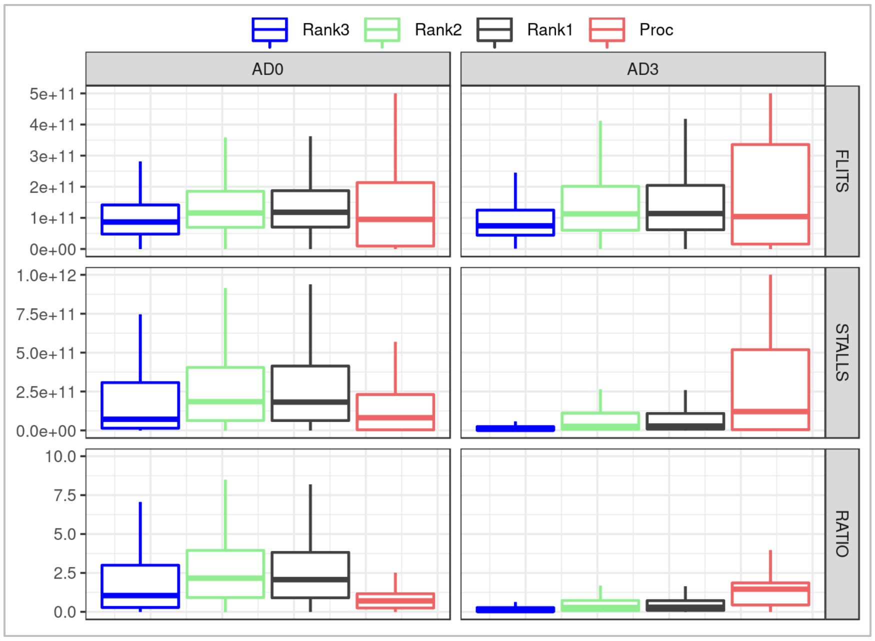



**Figure 4.3(b) System-wide counter metrics before (AD0) and after (AD3) changing the default routing on Theta. The color coding indicates information recorded on 4 different categories of ports on an Aries router.**

### 4.1.5 Effective Virtual Meetings

**Challenge:** Due to the pandemic, the ALCF training team decided to move all workshops to a virtual format. Because most of our workshops have a strong hands-on component, we had to rework the in-person experience such that the virtual format still provides one-on-one interaction with ALCF and industry experts. Time-zone difference was another factor in reworking the format, length, and hours of each workshop. We had to consider problems that could occur due to internet connections and also working-from-home challenges.

**Approach:** To enhance the hands-on experience, there were at least two subject matter experts available at instant chat platforms, as well as during talks to answer questions. There were general channels, as well as specific topics to allow users to discuss their questions with experts and other fellow users experienced in the matter. The start and end times of each workshop were changed to allow for reasonable start times across different time zones. To prevent problems associated with the internet and working from home, we worked with speakers and pre-recorded all of the talks. Then the organizing committee replayed those videos allowing speakers to answer questions in real time.

**Impacts/Status:** During CY 2020, ALCF held five public workshops and three invitation-only workshops, as follows (by date):

*Public*

- Computational Performance Workshop, May 5–7, with 147 attendees.
- CMake Workshop, July 15–17, with 50 attendees.

- ATPESC, July 26–August 7, with 73 attendees.
- 2020 Performance, Portability and Productivity at HPC Forum, Sept 1–2, with 80 talks and 20 poster sessions, and with 186 attendees.
- Science, Data and Learning Workshop, December 8–10, with 139 attendees.

*Invitation-only*

- COE Aurora Workshop #2 Argonne from February 25–27, with 108 attendees.
- COE Aurora Workshop #3, October 21–22 (part 1) and November 5, 2020 (part 2), with a total of 166 attendees.

Participation increased over similar in-person events, and attendees were very positive about the experience.

### 4.1.6 Aurora Readiness Assessments of INCITE Projects

**Challenge:** Future ALCF machines will have a significantly different architectures than the previous Mira and the current Theta system. ALCF needed a deeper engagement with teams to understand if INCITE projects were on a path to supporting the new architecture.

**Approach:** With the 2020 INCITE projects, the catalyst team added a new component to their initial and ongoing engagement with the project teams: assessing the state of readiness for INCITE applications to transition to the GPU-accelerated architecture of Aurora. As part of the established welcome calls between project catalysts and INCITE project PIs, ALCF introduced a new set of questions to guide the discussion on Aurora. Key questions included:

- Are the project's application codes already GPU-enabled, and, if so, what programming model is used?
- What plans exist for porting to GPU-accelerated architectures?
- What key library dependencies might need Aurora-optimized implementations?
- Is the team interested in collaborating with ALCF to port code to Aurora and exascale?

Part of the motivation for this was to identify applications and development teams with an interest in developing for Aurora, but that were not already engaged with ALCF on the effort through avenues such as the Aurora ESP and ECP's Application Integration work at the DOE facilities. This might reveal opportunities for ALCF to boost Aurora development.

**Impacts/Status:** Answers to these questions went into project-specific documents that the catalyst team uses to track the status of the project throughout the INCITE year. For projects with an interest in collaboration on Aurora development, science team management evaluates the merit and likelihood of success, including the prospect for future INCITE projects on Aurora. For projects ALCF chooses to collaborate with on Aurora development, the document provides sections to record plans and progress on plans enacted.

## 4.2 Research Activities for Future Operations

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 Improved Software Environment for Emerging Technologies

**Challenge:** The software environment (compilers, frameworks, APIs, libraries, etc.) for HPC must evolve as new hardware is introduced and new usage patterns emerge.

**Approach:** ALCF staff have led in a number of large community efforts to define and develop the capabilities needed for future systems.

ALCF staff led the effort to add several new capabilities into the OpenMP 5.1 specification. These features will be supported in LLVM/Clang, but also in vendors' compilers across the HPC ecosystem.

- Loop tiling and unrolling – These loop transformations are critical to the performance of many kinds of algorithms, and now users will be able to instruct the compiler to perform these transformations instead of implementing them *by hand*.
- Begin/end variant declarations – This feature is essential to support math functions (e.g., sin, cos), complex types, and more on accelerators.
- Assumptions – With this feature, developers can provide hints to the compiler to enable optimizations that could not be performed without it.

ALCF staff developed and integrated implementations of these and other new OpenMP features into LLVM/Clang. In addition, optimization capabilities of LLVM have been advanced specifically to increase the performance of programs using OpenMP constructs (Figure 4.4). In particular, ALCF staff have focused on the improvements needed for exascale computing.

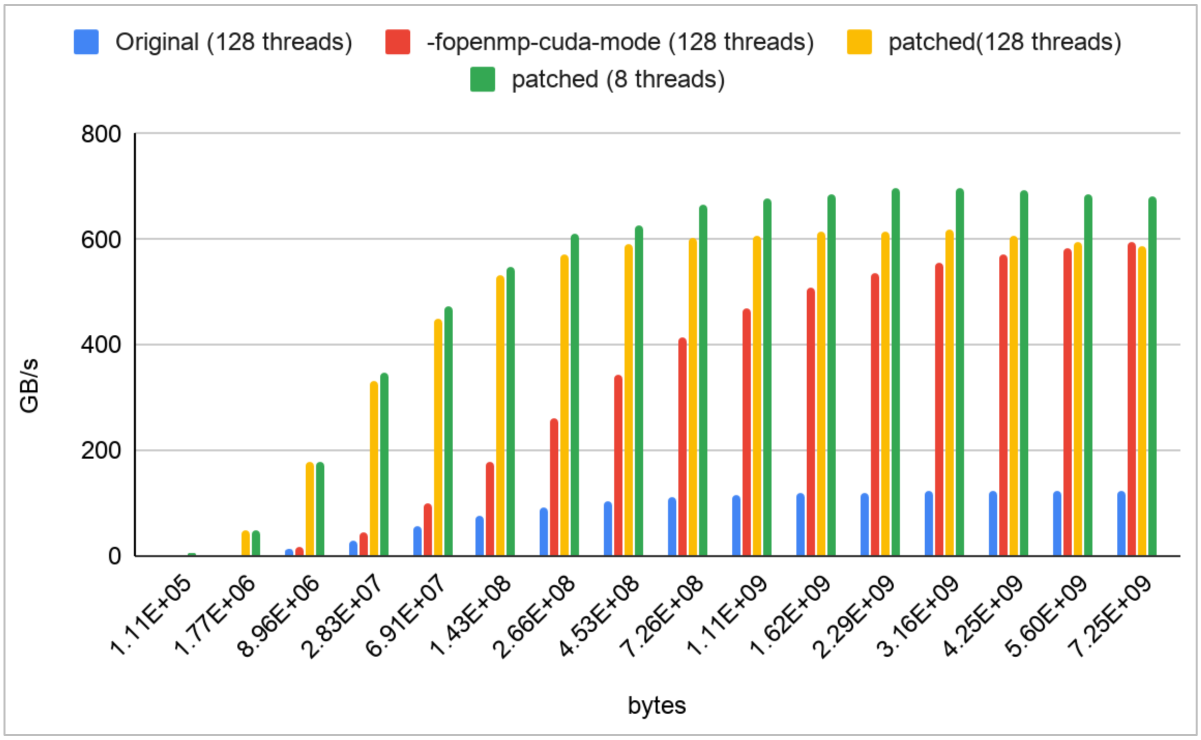



**Figure 4.4 SU(3)xSU(3) performance (GRID mini) with and without ALCF-developed LLVM/OpenMP optimizations.**

ALCF staff developed a new infrastructure for the Clang/LLVM compilers to support user-directed loop transformations (Figure 4.5), and new directives for users to control those optimizations. ALCF staff also connected these new directives to the MCS-developed ytopt autotuner and demonstrated substantial performance enhancements on several proxy applications.

**Results/Status:** All of the work has been contributed to the appropriate standards or open source communities. The OpenMP features were contributed to the OpenMP ARB and incorporated into OpenMP 5.1, which was released in November 2020. The LLVM/Clang work has been contributed to the LLVM OpenMP open source project. Some features are in LLVM 11.0, released in October 2020. Others will be in LLVM 11.1 in early 2021.

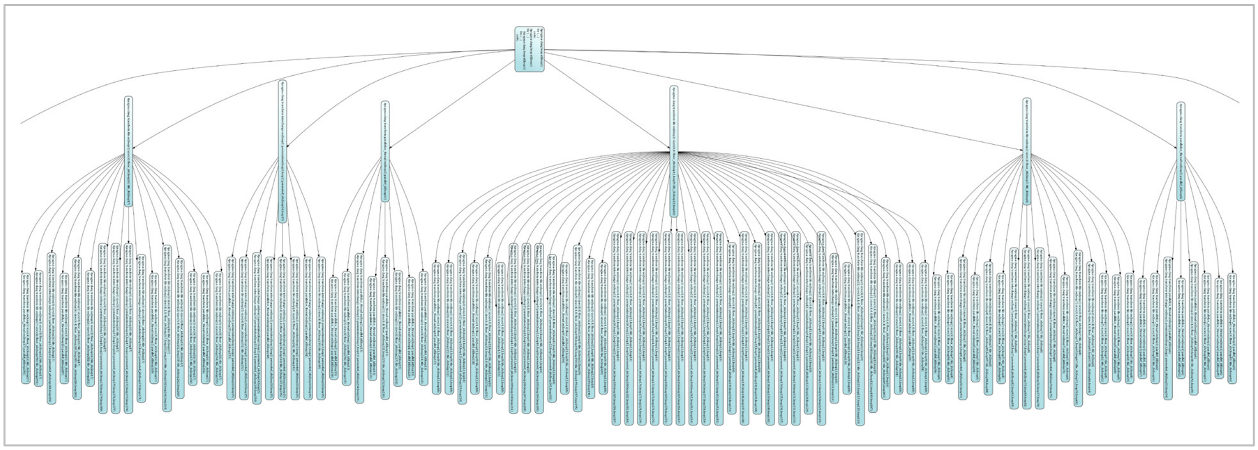


**Figure 4.5 Search space for a Monte Carlo Tree Search. Each node represents an experiment that can run and the execution time measured. In particular, each node is a different loop optimization. We start with the root node that represents the kernel without optimization. Each child is another loop transformation applied in addition to the ones in the parent.**

### 4.2.2 Tools for Improving Code Performance on Emerging Hardware

**Challenge:** It is challenging for programmers to understand the impact and performance of routines offloaded to GPUs. Many traditional tools do not support GPUs, and many GPU-specific tools do not support Intel GPUs.

**Approach:** ALCF staff deployed a number of software tools and evaluated them for utility with Intel GPUs.

With Intel's new accelerator technologies, ALCF required a method to understand what actions applications were taking with respect to accelerator offload without investigating every line of code. THAPI provides a trace of all offload API usage for OpenCL or Level 0 (Figure 4.6).

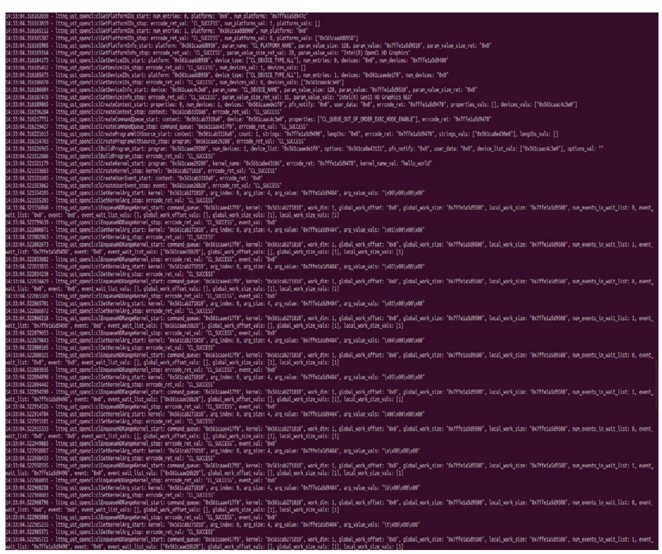
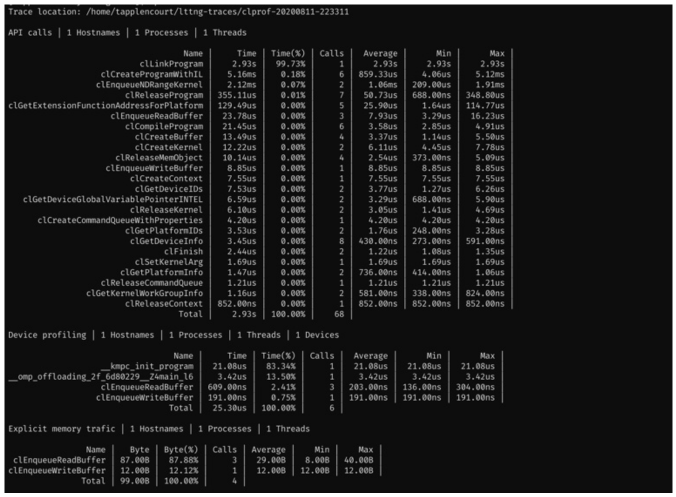


**Figure 4.6 THAPI Trace Output**

ALCF constructed a simple, lightweight performance profiler for Intel's integrated and DG1 GPUs. Iprof is built on the THAPI tracer infrastructure (Figure 4.7). It provides a simple tool for ALCF, ESP, and ECP users to analyze data movement, kernel offload timing, and kernel launch parameters. Users are not required to rebuild code, and the profiling overhead enables analysis of full applications.



Trace location: /home/tapplencourt/lttng-traces/clprof-20200811-223311

API calls | 1 Hostnames | 1 Processes | 1 Threads

| Name | Time | Time(%) | Calls | Average | Min | Max |
|---|---|---|---|---|---|---|
| clLinkProgram | 2.93s | 99.73% | 1 | 2.93s | 2.93s | 2.93s |
| clCreateProgramWithIL | 5.16ms | 0.18% | 6 | 859.33us | 4.06us | 5.12ms |
| clEnqueueNDRangeKernel | 2.12ms | 0.07% | 2 | 1.06ms | 209.00us | 1.91ms |
| clReleaseProgram | 355.11us | 0.01% | 7 | 50.73us | 688.00ns | 348.80us |
| clGetExtensionFunctionAddressForPlatform | 129.49us | 0.00% | 5 | 25.90us | 1.64us | 114.77us |
| clEnqueueReadBuffer | 23.78us | 0.00% | 3 | 7.93us | 3.29us | 16.23us |
| clCompileProgram | 21.45us | 0.00% | 6 | 3.58us | 2.85us | 4.91us |
| clCreateBuffer | 13.49us | 0.00% | 4 | 3.37us | 1.14us | 5.50us |
| clCreateKernel | 12.22us | 0.00% | 2 | 6.11us | 4.45us | 7.78us |
| clReleaseMemObject | 10.14us | 0.00% | 4 | 2.54us | 373.00ns | 5.09us |
| clEnqueueWriteBuffer | 8.85us | 0.00% | 1 | 8.85us | 8.85us | 8.85us |
| clCreateContext | 7.55us | 0.00% | 1 | 7.55us | 7.55us | 7.55us |
| clGetDeviceIDs | 7.53us | 0.00% | 2 | 3.77us | 1.27us | 6.26us |
| clGetDeviceGlobalVariablePointerINTEL | 6.59us | 0.00% | 2 | 3.29us | 688.00ns | 5.90us |
| clReleaseKernel | 6.10us | 0.00% | 2 | 3.05us | 1.41us | 4.69us |
| clCreateCommandQueueWithProperties | 4.20us | 0.00% | 1 | 4.20us | 4.20us | 4.20us |
| clGetPlatformIDs | 3.53us | 0.00% | 2 | 1.76us | 248.00ns | 3.28us |
| clGetDeviceInfo | 3.45us | 0.00% | 8 | 430.00ns | 273.00ns | 591.00ns |
| clFinish | 2.44us | 0.00% | 2 | 1.22us | 1.08us | 1.35us |
| clSetKernelArg | 1.69us | 0.00% | 1 | 1.69us | 1.69us | 1.69us |
| clGetPlatformInfo | 1.47us | 0.00% | 2 | 736.00ns | 414.00ns | 1.06us |
| clReleaseCommandQueue | 1.21us | 0.00% | 1 | 1.21us | 1.21us | 1.21us |
| clGetKernelWorkGroupInfo | 1.16us | 0.00% | 2 | 581.00ns | 338.00ns | 824.00ns |
| clReleaseContext | 852.00ns | 0.00% | 1 | 852.00ns | 852.00ns | 852.00ns |
| Total | 2.93s | 100.00% | 68 | | | |

Device profiling | 1 Hostnames | 1 Processes | 1 Threads | 1 Devices

| Name | Time | Time(%) | Calls | Average | Min | Max |
|---|---|---|---|---|---|---|
| __kmpc_init_program | 21.08us | 83.34% | 1 | 21.08us | 21.08us | 21.08us |
| __omp_offloading_2f_6d80229__Z4main_l6 | 3.42us | 13.50% | 1 | 3.42us | 3.42us | 3.42us |
| clEnqueueReadBuffer | 609.00ns | 2.41% | 3 | 203.00ns | 136.00ns | 304.00ns |
| clEnqueueWriteBuffer | 191.00ns | 0.75% | 1 | 191.00ns | 191.00ns | 191.00ns |
| Total | 25.30us | 100.00% | 6 | | | |

Explicit memory trafic | 1 Hostnames | 1 Processes | 1 Threads

| Name | Byte | Byte(%) | Calls | Average | Min | Max |
|---|---|---|---|---|---|---|
| clEnqueueReadBuffer | 87.00B | 87.88% | 3 | 29.00B | 8.00B | 40.00B |
| clEnqueueWriteBuffer | 12.00B | 12.12% | 1 | 12.00B | 12.00B | 12.00B |
| Total | 99.00B | 100.00% | 4 | | | |

**Figure 4.7 Iprof Screen Output**

To evaluate the offloading capabilities of compilers, ALCF designed an OpenMP 5.0-based offload test suite named OvO, which stands for "OpenMP Validator for Offload." OvO contains about 2,700 OpenMP offload tests for C++ and Fortran, which are procedurally generated. It contains a series of scripts to compile, run, and check the correctness of all of the tests in a single run. It's also used internally by some vendors.

**Results/Status:** THAPI, Iprof, and OvO are deployed on ALCF's discrete GPU testbed and are available to all users. The tools have allowed ALCF staff and early users to better understand how their code runs on Intel's discrete GPUs.

### 4.2.3 Intel Discrete GPU Testbed

**Challenge:** ALCF needs to test a variety of software on Intel's discrete GPUs in anticipation of Aurora.

**Approach:** ALCF staff worked with Intel to deploy a testbed of twenty nodes, each with two Intel Xeon CPUs and four Intel DG1 GPUs (80 GPUs in total), connected by a Mellanox EDR InfiniBand. This testbed allows the testing of Intel Data Parallel C++ and OpenMP offload runtimes using multiple discrete GPUs within a node. The DG1s are a good stand-in for eventual Aurora GPUs. The system called Yarrow is deployed in JLSE (Figure 4.8).

**Impact/Status:** The testbed is available for ALCF staff, Aurora ESP project members, and ECP teams for testing and code development. The testbed allows ALCF staff and users to better prepare for Aurora.


**Figure 4.8 Intel's Xe DG1 GPU in Yarrow**

### 4.2.4 Joint Laboratory for System Evaluation

Argonne's JLSE enables researchers to assess and improve next-generation computing platforms of interest to the DOE. Established by the CELS computing divisions and run by ALCF, the 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.

JLSE testbeds 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))
- **Yarrow:** Supermicro SYS-1029GQ-TNRT with Intel development GPUs (code name DG1) and Xeon Gold 6226R CPUs (Intel's first discrete GPU)
- **Iris:** SuperMicro X11SSH-GF-1585 Server Motherboard with Intel Xeon E3-1585 v5 CPU and Iris Pro Graphics P580 GPU (Intel integrated Gen9 GPUs)

- **Presque:** Intel DAOS nodes (DCPMM and NVMe storage) with Intel DAOS file system

Other JLSE testbeds include:

- **CooperLake:** Supermicro SYS-240P-TNRT with Intel Xeon 8354H CPUs (first Xeon to support for BF16 datatype)
- Intel Xeon Phi Knights Landing (KNL) Cluster
- NVIDIA DGX–1 (V100 GPUs)
- Intel Xeon Skylake Cluster
- Intel oneAPI
- HPE Comanche Prototype ARM64 Cluster
- IBM Power System AC922 (Power9 CPU, V100 GPU)
- Supermicro NVIDIA V100 and P100 cluster
- Atos Quantum Learning Machine

In 2020, the JLSE supported nearly 700 users spanning more than 80 projects. These projects ranged from application portability to software development to tools and compiler development for an ALCF ESP project. Teams from within the ECP's Application Development and Software Technology groups, developing applications and software for Aurora, have had access to the JLSE Gen9 nodes and the Aurora SDK for their development work. 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:** Application teams from the ALCF Early Science Program 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.

**PetrelKube:** This small Kubernetes testbed is managed with Rancher and hosts back-end services supporting web portals running outside of the JLSE and HPC workflows. The testbed utilizes Petrel for its primary data storage, initially staging data over the network and eventually having direct access to Petrel via a POSIX interface.

**Argo:** Argo is a new 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 the laboratory's testbeds to develop and optimize a lightweight, low-level threading and task framework for OpenMP and other programming models (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 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:** A number of 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.

### 4.2.5 AI Testbed

The AI Testbed at ALCF, established in 2020, provides an infrastructure of next-generation machines with hardware accelerators customized for AI applications. The testbed provides a platform to evaluate usability and performance of ML-based HPC applications running on these accelerators. The goal is to better understand how to integrate AI accelerators with ALCF's existing and upcoming supercomputers to accelerate science insights.

Key activities include:

- Maintain a range of hardware and software environments for AI accelerators.
- Provide a platform to benchmark applications, programming models, and ML frameworks.
- Support science application teams in the porting and evaluation of their applications.

- Coordinate with vendors during their product development.

Currently, the AI Testbed has Cerebras, SambaNova and Graphcore systems set up and running. ALCF is also working closely with other vendors.

- **Cerebras** is a wafer-scale, deep learning accelerator architected on a 462 $cm^2$ silicon wafer with approximately 400,000 processor cores and on-die interconnect. The current configuration is a single node.
- **SambaNova** is architected with two reconfigurable dataflow units (RDUs) on a PCIe Gen4 accelerator card. The current configuration is a half-rack SN10-8R system consisting of two nodes, each with four accelerators interconnected with a custom RDU-direct link.
- **Graphcore** is a Colossus GC2 intelligent processing unit (IPU) that uses 1,216 IPU tiles, each with an independent core and memory. The current configuration uses a Dell DSS8440 featuring eight dual-IPU GC2 PCIe cards connected with IPU-Link technology in a standard 4U server.

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, 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 these are installed and supported. The environment also provides support for Singularity-based containers and for executing Jupyter notebooks.

The AI Testbed is currently open to Argonne users. Several application teams, including nuclear physics, materials science, biosciences, astrophysics, cosmology, imaging sciences, and precision medicine, are active users. ALCF expects to open the system up to academia, other DOE laboratories, industry, and others in the near future.

## 4.3 Postdoctoral Fellows

ALCF supports a steady-state postdoctoral fellowship program. Within the purview of this program, ALCF supports one named postdoctoral fellow, the ALCF's Margaret Butler for Computational Science Fellow. Postdocs are awarded 1-year appointments that have 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 that these postdocs continue to be 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 2020, ALCF hired four new postdoctoral researchers (fellows) and supported a total of 17, representing a range of scientific domain backgrounds. Of those, 3 postdocs were terminated and 2 were transferred to another division. Of the 12 active postdocs, 5 are ESP, 4 are ECP, and 3 are steady-state. The ALCF postdocs work on various research topics, including computational chemistry; ML/AI for particle physics; exascale interconnect performance evaluation; ATLAS, simulation, and deep learning; fusion and deep learning; Lattice QCD, ML, generative modeling/simulation-based inference, and Markov chain Monte Carlo methods; electronic structure; climate-scale simulation; materials science; cosmological n-body simulations; neuroscience and synchrotron X-ray tomography; dark matter halo modeling and clustering analysis and cosmological parameter estimation.

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 could focus on over the next year; opportunities for development; and, if entering the third year, what will be most beneficial to enabling a smooth career transition. The division director also meets on a monthly basis 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 postdoctoral researchers in CY 2020:

**Abhishek Bagusetty** (Ph.D., chemical engineering, University of Pittsburgh). **Hired:** January 2020. **Research area:** computational chemistry. **Current projects:** Aurora ESP NWChemEx. **Accomplishments include**: (1) made code improvements in the NWChemEx software ecosystem and the ECP ecosystem; (2) has two manuscripts in review for ECP Special Issue in Elsevier *Parallel Comput.*; (3) submitted one conference poster (IWOCL'21); and (4) is a 2020 Argonne Postdoctoral Performance nominee for outstanding research efforts.

**Denis Boyda** (Ph.D., theoretical physics, National Research Center Kurchatov Institute). **Hired:** September 2020. **Research area:** ML/AI for particle physics using HPC. **Current projects:** Aurora ESP: ML for Lattice QCD (PI: William Detmold). **Accomplishments include**: (1) implemented novel mathematical approach for speeding up Hamiltonian Monte Carlo integration and tested on small problems; (2) ported code to test on early Intel hardware and early DL software; created mini apps, wrote a test suite, and reported bugs to Intel; (3) implemented and tested several ways to improve memory usage within this project's DL code; (4) implemented model parallelism and tested scalability on current systems, comparing model and data parallelism; and (5) currently writing two papers.

**Kevin A. Brown** (Ph.D., mathematical and computer science, Tokyo Institute of Technology). **Hired:** October 2019. **Research area:** exascale interconnect performance evaluation. **Current projects:** CODES model validation (completed); QoS model in CODES (ongoing). This work will guide the configuration of Aurora and other Slingshot machines at DOE facilities. **Accomplishments include**: (1) validated the 1D Dragonfly model in the CODES simulation toolkit and upgraded the model's QoS implementation to support dual rate limiters; (2) participated in the *2020 Argonne Postdoctoral Research and Career Symposium*; (3) completed one submission to *2021 ISC High Performance Conference*; and (4) presented two invited talks.

**Tyler Burch** (Ph.D., physics, Northern Illinois University). **Hired:** May 2020. **Research area:** ATLAS, simulation, deep learning. **Current projects:** Aurora ESP: ATLAS (PI: Walter Hopkins). **Accomplishments include:** (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 (under development).

**Kyle Felker** (Ph.D., physics, Yale University). **Hired:** September 2019. **Research area:** fusion, deep learning. **Current projects:** Aurora ESP: Fusion RNN (PI: Bill Tang)—project involves scaling-up studies of Fusion Tokamak reactors using deep learning to predict operational failures. **Accomplishments include:** (1) deployment of the previous successful LSTM (long short-term memory) network on Theta and running hyperparameter searches with DeepHyper; (2) integration of additional detector data into the training and rerunning hyperparameter scans; (3) contributions to ALCF's Data Science group; and (4) the providing of support to attendees of the SDL workshop. **Future work:** porting the LSTM to PyTorch for testing on JLSE.

**Sam Foreman** (Ph.D., physics, University of Iowa). **Hired:** August 2019. **Research area:** Lattice QCD, ML, generative modeling/simulation-based inference, Markov chain Monte Carlo methods. **Current projects:** Application of L2HMC algorithm to simulation in Lattice QCD. Developed a mini-app, called l2hmc-qcd, that uses ML to build more efficient sampling techniques, thereby reducing one of the major bottlenecks in the HEP/Lattice QCD workflow. **Accomplishments include**: (1) contributed extensive code development of l2hmc-qcd; built and tested a proof-of-concept model (manuscript in progress); (2) presented one invited talk at the University of Iowa; (3) contributed regular presentations with the "Critical Slowing Down" subgroup of the Lattice QCD ECP project; (4) presented tutorial at ALCF's SDL Workshop for AI; and (5) helped to build and test ML frameworks on ThetaGPU as a member of the Data Science team.

**Kevin Gasperich** (Ph.D., chemistry, University of Pittsburgh). **Hired:** November 2019. **Research area:** electronic structure. **Current projects:** ECP-HI QMCPACK; Quantum Package solids, and initial work for exascale porting. **Accomplishments include**: (1) contributed to the 20X speedup of the multi-determinants branch of QMCPACK, leading to a 3X application speed up; (2) developed a Python mini-app to prototype new algorithms that will enable development of exascale sCI code; (3) has one publication in *J. Chem. Phys.*; and (4) presented one invited talk (an international collaboration).

**Andrea Orton** (Ph.D., geophysics/atmospheric sciences, Purdue University). **Hired:** October 2020. **Research area:** climate scale simulation. **Current projects:** DD: Earth System Modeling project. **Accomplishments include**: (1) is 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; and (2) participated in the *2020 Argonne Postdoctoral Research and Career Symposium*.

**Pankaj Rajak** (Ph.D., computational materials science, University of Southern California). **Hired:** October 2018. **Research area:** materials science, HPC, DL, reinforcement learning (RL). **Current projects:** (1) development and optimization of the massively parallel nonadiabatic ab-initio quantum molecular dynamics (NAQMD) and reactive molecular dynamics (RMD) software code for Aurora; and (2) working on applied ML research in materials science domain, with emphasis on designing autonomous AI agents using RL that have in-build capability of decision making and are able to finish wide range assigned tasks with little human supervision. **Accomplishments include**: (1) has two conference publications (one best paper) and two journal publications (one in *J. Phys: Conf. Series 1461* and one in *Software X*; (2) has two manuscripts in review for *npj Comput. Mater.; (3)* attended five major conferences.

**Esteban (Steve) Rangel** (Ph.D., computer science, Northwestern University). **Hired:** July 2018. **Research area:** computational science, cosmological n-body simulations, HPC. **Current projects:** ESP/HACC (100%, January–July, ECP Application Integration (50%, August–December), ECP/ExaSky (25%, August-December), and Threadwork (25%, August–December). **Accomplishments include**: (1) made significant code improvements to HACC; (2) has one manuscript in review for *ApJ* and had one manuscript accepted by *ApJS*; (3) presented a talk at the Aurora COE ESP Workshop 2; and (4) is a recipient of the Argonne Director's Award: Contributions to HACC for the Last Journey (ALCC) project.

**Rafael Vescovi** (Ph.D., physics, University of Campinas). **Hired:** May 2019. **Research area:** neuroscience and synchrotron X-ray tomography. **Current projects:** Aurora ESP project to enable connectomics at exascale. **Accomplishments include**: (1) development of a Python package that aims to wrap existing tools for electron microscope image processing and establish pipelines on HPC facilities; (2) establishment of a faster iteration process that enabled several large-scale samples to be processed at ALCF; (3) establishment of new collaborations with the UChicago Advanced Electron Microscope Core; (4) has one publication in *2020 IEEE/ACM 2nd Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)* and one publication in *Sci Data.*

**Antonio Villarreal** (Ph.D., astrophysics, University of Pittsburgh). **Hired:** August 2018. **Research area:** dark matter halo modeling and clustering analysis and cosmological parameter estimation for Large Synoptic Survey Telescope (LSST)-Dark Energy Science Collaboration (DESC); HPC. **Current projects:** Works with 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 (management of the image simulation workflow, taking in community input catalogs and providing raw images for processing); (2) designed a workflow that would work in multiple HPC facilities; (3) has one manuscript in review and has one manuscript in progress; (4) developed an algorithm to assist in analysis of large-scale cosmology simulations.

*This page intentionally left blank.*

# 5. Risk Management

***Is the Facility effectively managing operational risks?***

## ALCF Response

The overview of the risk management process that ALCF follows laid out in Section 5.1 below clearly demonstrated that ALCF successfully managed both its project risks and operational risks in CY 2020. 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. The risks that occurred and the mitigations for those risks are described in greater detail in Section 5.3. Section 5.6 and Table 5.2 provide details on the major risks that will be tracked in CY 2021.

*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 the 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.

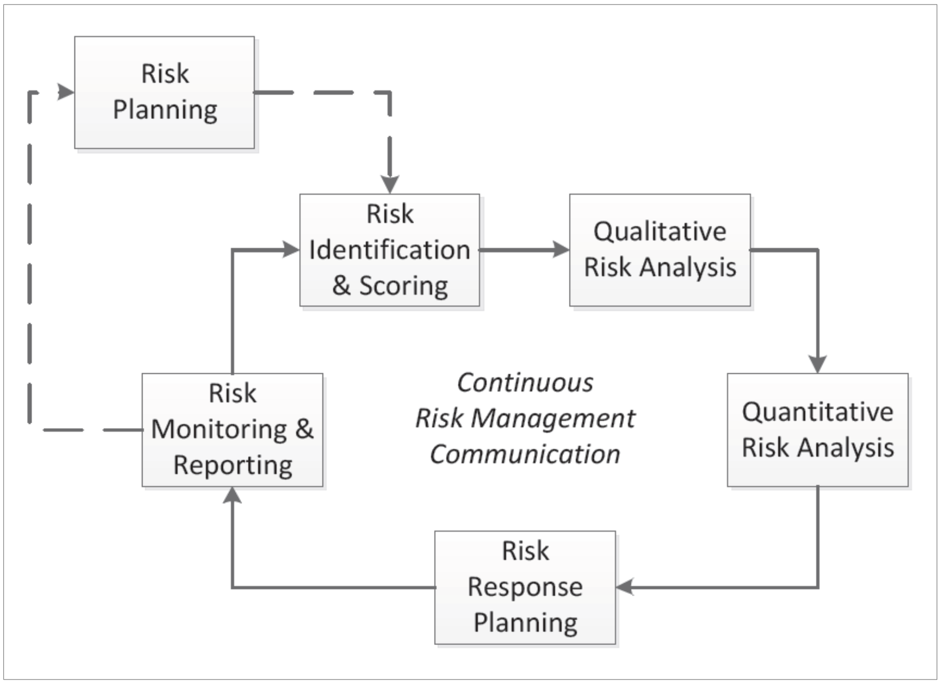



**Figure 5.1 Illustration of the Iterative Risk Management Process from the ALCF's Risk Management Plan**

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 to 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 as a result 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). On the operations side, subject matter experts estimate risk mitigation costs and use them to inform management reserves.

In addition to formal and individual risk meetings and the Risk Review Board meetings, 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. 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.

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 cover 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 in 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 2020. 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 utilizing the OPRA integration with the Primavera project management tool. 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 was operating both Mira and Theta during CY 2020, and planned the growth of both the staff and the budget in order to bring the facility to full strength. As such, ALCF continues to monitor a large number of major risks for the facility. One risk was retired in CY 2020 and that is covered in section 5.4.

**Four** major operations risks were tracked for CY 2020, **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 2021, 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 2021. |
| **25** | Staff Recruitment Challenges | No | High | The ALCF added 11 new staff members and 1 reclassified staff member in CY 2020. One of the new hires is funded by ECP. 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 7 staff members during CY 2020, one of them funded by ECP. Three staff members transferred internally to other divisions. 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 **two** risks in CY 2020. 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, while another has a rating of Very Low.

### 5.3.1 Users Gain Access to Data That They Should Not Have Permission to Review

| **1046: Users gain access to data that they should not have permission to review** | |
|---|---|
| **Risk Owner** | HPC Infrastructure Team Lead (John Reddy) |
| **Probability** | Very Low |
| **Impact** | Cost: Very Low; Technical Scope: Very Low |
| **Risk Rating** | Very Low |
| **Primary Management Strategies** | Follow cyber security best practices.<br>Perform periodic reviews and scans of file system looking for issues.<br>User education. |
| **Triggers** | Scan of files shows bad permissions on files.<br>User notifies support of an issue. |

### Description

A directory containing source code and some shared secrets for ALCF management software was identified in world-readable folders.

### Evaluation

After investigation we found that the issue occurred due to scope creep on the use of a shared network file system (NFS) mount.

### Management

The Operations team is instituting more strict reviews of configuration changes, including security review of changes previously deemed too minor for such reviews.

## 5.3.2 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** | The ALCF's director of operations participates in the Data Center management group. ALCF pays part of the cost of an Argonne Data Center liaison.<br>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 November 23, 2020, a breaker on a substation feeding the datacenter tripped causing ThetaGPU compute nodes to lose power.

### Evaluation

A similar event took place over a year ago, and a spare breaker was procured as a result of that event. On November 24, 2020, the tripped breaker was replaced with the spare. This required the entire substation to be powered down, which required Theta to be powered down as well.

### Management

The faulty breaker was replaced with a spare that was available immediately.

## 5.4 Retired Risks

One risk was retired during CY 2020. This risk is summarized in Table 5.2.

**Table 5.2 Risk Retired during CY 2020**

| ID | Title | Rating | Management Strategies | Notes |
|---|---|---|---|---|
| **1047** | Compiler bugs do not get fixed | Very Low | Manually fix code with a workaround so that compiler bug is not triggered.<br>Use open source compilers.<br>Support development of open source compilers. | This risk was no longer relevant after the retirement of Mira. |

## 5.5 New and Recharacterized Risks since the Last Review

There are no new risks and one recharacterized risk to report since the last review. The details of the recharacterized risk appear below, including the risk owner, an assessment of the risk's probability and impacts, a description of the changes, and what management strategies have been developed to address the risk. This risk has a rating of Very Low.

### 5.5.1 Rate of Hardware Failure Affects Availability of Key Production Resources Availability

| 1084: Rate of hardware failure affects key production resources availability | |
|---|---|
| **Risk Owner** | Director of Operations (Mark Fahey) |
| **Probability** | Very Low |
| **Impact** | Cost: Very Low; Technical Scope: Very Low |
| **Risk Rating** | Very Low |
| **Primary Management Strategies** | Work with vendor to resolve issues.<br>Monitor hardware performance.<br>Have full support on hardware from vendor.<br>Purchase additional racks of hardware to serve as tested hot-spare hardware supply. |
| **Triggers** | Component failure rates increase. |

The risk was originally titled: Rate of hardware failure exceeds projections.

The original risk was created when the BG/Q was no longer supported by IBM, and replacement parts were not available through them. The ALCF Operations team then began projecting the rate of hardware failures based on historical data. The original risk would be triggered if the failure rate ever exceeded the projection as that would mean the size of the machine would start decreasing accordingly. Because the Intel/Cray system Theta is currently within the window of support by the vendor, this concern changed from failure exceeding available replacement parts to broadly apply to failures that affect the availability of resources.

## 5.6 Top Operating Risks Monitored Closely for the Next Year

Table 5.3 lists the current top operating risks being closely monitored for CY 2021, 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 5.3 Top Operating Risks Monitored for CY 2021**

| 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 document is regularly reviewed and updated to reflect the dynamic nature of risk management, as well as 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 the past year.

# 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

The ALCF is a leader in health and safety at Argonne National Laboratory. The facility has never experienced a lost time incident. In 2020, ALCF experienced no recordable injuries, near misses, first aid incidents, or pause/stop work events. The division's commitment to safety also extends to office work safety and to psychological safety. Two office ergonomic evaluations and 17 at-home ergonomic evaluations were performed on ALCF employees in 2020. Additional communication efforts were made to inform employees of available resources such as the Employee Assistance Program to help address COVID-related stress.

In March 2020, the Laboratory was placed in a minimum safe operations posture. During this time, ALCF quickly adapted to, developed, and implemented COVID-19 controls into its work practices and emerged as a top safety leader. During Argonne's "Min Safe" mode of operation, ALCF was able to decommission the Mira machine while utilizing COVID-19 controls to ensure our workers' and contract workers' safety. The decommissioning of Mira was completed in May 2020 without incident. As the laboratory transitioned to a limited operation posture later in the summer, ALCF continued to work safely while following all COVID controls recommended by the Laboratory. Workers successfully completed several large installations, including Crux and ThetaGPU, without incident.

ALCF continues to utilize its local work planning and control manual for executing work according to Integrated Safety Management guiding principles and various approaches from DOE-HDBK-1211-2014: *Activity-Level Work Planning and Control Implementation*. The division uses Aware, the laboratory's work planning and control online application. ALCF analyzes hazard, using laboratory subject matter experts (SMEs) when necessary, and control measures to ensure that work is performed while mitigating risks to the safety, environment, mission, security, and health of the users, public, and workers. All ALCF work control documents (WCDs) were reviewed by the work planners and updated in 2020 to either retire obsolete documents or update with COVID-19 controls.

Formal task-based WCDs are developed for complex tasks such as the decommissioning of Mira. Management actively monitors work scope and updates hazard analyses and revises work planning and control documents for emergency work or whenever there is an unexpected change to previously planned work. During the decommissioning of Mira, the WCD was reviewed and communicated daily, and adjustments were made as needed to implement the most current State and Laboratory COVID-19 recommendations. ALCF also utilized a Hazard Assessment and Control (HAC) document to assess the effectiveness and feasibility of needed COVID-19 controls while executing the planned work. All supplementary WCD documentation is housed within Aware. Documentation such as the Job Safety Analysis (JSA), that covered the work completed by contractors, is retained within the Laboratory's official cloud resource, Box. ALCF is a leader at the laboratory, being the first to implemented COVID-19 controls in their contractor JSAs. This best practice led to a laboratory-wide requirement that COVID-19 controls be addressed in JSAs for contractors.

ALCF continues to use formal management assessments, such as the 2020 Noise in the Data Center Assessment, to seek continuous improvement within the division. The 2020 Noise assessment included two additional noise surveys in the data center performed by industrial hygienists after noisy equipment was installed. The noise assessment concluded that anyone working in the data center for more than 2 hours at a time at any point during a single-year period should be included in the Hearing Conservation Program (HCP). Subsequently, with the assistance of the noise SME, 21 ALCF workers were enrolled into the HCP to assess and better protect the workers' health and safety. In addition to formal management assessments, ALCF's division director and ESH coordinator conduct biannual health and safety inspections. These inspections serve as an opportunity to observe work, seek worker feedback, and identify areas for improvement. The 2020 inspections took place with COVID-19 controls maintained during the walk-throughs.

# 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 Argonne's internal security personnel, and between ALCF and the Argonne Cybersecurity Program Office (CSPO).

## 7.1 Continual Improvement in Cyber Security Practices

ALCF continuously works to maintain a strong cyber security posture. In 2019, ALCF deployed its custom UserBase3 (UB3) user management application that, among its features, supports rapid development and enhanced usability and security. Throughout 2020, we have continued to improve developer workflows for UB3 and other applications, including integrating continuous integration (CI) workflows to monitor Python and JavaScript libraries for security vulnerabilities, and integrating static code analysis for identifying potential security issues in the code. Integrating CI capabilities with cyber workflows has allowed our developers and security team to collaboratively identify certain vulnerabilities before code is pushed to production systems or picked up by our security scanners—giving our developers the maximum amount of time to test and verify the fixes.

ALCF is vigilant in our efforts to stay informed about active and evolving threats, vulnerabilities, and potential exploitations. Earlier this year, ALCF's Security team partnered with the OLCF, NERSC, and other facilities to quickly assess the risks of an active security incident affecting several HPC facilities across Europe. Working together, the facilities identified possible attack vectors and talked to the affected sites to obtain advanced knowledge of the attacks. The ALCF and others were then able to open tickets with our vendors and test the vulnerabilities on our systems—all within 24 hours of the attacks going public. In addition, ALCF staff were able to verify that no users had attempted to run this exploiting program on our system.

In CY 2020, 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 include these taken by our security personnel, who:

1. Conduct privileged access reviews across the environment to ensure that everyone has the appropriate level of access.
2. Conduct reviews of how and where data is stored to ensure that data is 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. Update the National Institute of Standards and Technology (NIST) Certification package, including NIST 800-53, 800-34, 800-30, and 800-18 compliance documents.
7. Archive and delete obsolete data and systems.
8. Sets 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.
11. Deployed Elastic Beats agent to collect and store machine logs in Elasticsearch for short-term analysis.
12. Deployed and evaluated the integration of new security scanners.

Some of these activities revealed security vulnerabilities that were promptly addressed and fixed, usually within days of their discovery. Immediately upon detection, ALCF staff would also investigate all relevant logs to determine whether the security vulnerability had been exploited. In CY 2020, 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:** Identification of a large number of failed login attempts made by one specific user. **Mitigation:** Log analysis showed that user error, not a brute force attack, was the cause. ALCF staff worked with the user to remediate the issue on their system.
2. **Issue:** The CSPO identified new security issues with public-facing ALCF services. This is typically considered a low severity issue. **Mitigation:** ALCF staff reviewed the CSPO information and worked to address the issues in a timely manner.
3. **Issue:** Passwords in some applications were found to be stored insecurely. **Mitigation:** Evaluation of the application data integrity showed that no unauthorized access to the associated databases had occurred. Passwords were rotated as a precaution against potential exploit.
4. **Issue:** Security risks were identified in pre-production applications under development. **Mitigation:** The development instances were locked down while the developers addressed the issues. The fixes were verified by cyber security personnel prior to application deployment.

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 needs.
3. Improving real-time log analysis techniques.

The CSPO performs an annual internal assessment with all of the divisions, called the Division Site Assist Visit (DSAV). The DSAV is an audit of each division's compliance with NIST-800-53 controls. Each year, the DSAV assessment covers roughly one-third of the controls. The CSPO did not perform the 2020 DSAV due to prioritization of major cyber projects and the transition to working from home due to the COVID-19 global pandemic. Although the DSAV was not performed, Argonne complied with all audits and continued to manage the security posture of the lab as a whole. In the 2019 DSAV, the CSPO identified three opportunities for improvement that ALCF addressed during CY 2020:

1. Update our disk-shredding exemption.
    a. ALCF's disk-shredding exemption is now in line with current Argonne cyber policies.
2. Automatically deactivate ALCF staff accounts when their relationship with the laboratory and the facility is terminated.
    a. ALCF developed a plan to deactivate ALCF accounts as soon as the Argonne account is deactivated. This plan is expected to take 1–2 years to implement.
3. Verify with CSPO that they continue to receive all logs that they require.
    a. The CSPO's logging project work is ongoing. ALCF is working with the office to provide all of the logs they require.

The CSPO has already started the 2021 DSAV. ALCF will continue to work with the CSPO to verify that all Argonne security standards and practices are met.

## 7.2 Cyber Security Plan

The Argonne 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 National Laboratory maintains robust, continuous monitoring of the Cyber Security Program as detailed in the ATO letter, included at the end of this section.

## 7.3 Foreign Visitor Access and Export Controls

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

To apply for an ALCF account, the user fills out a secure webform in the ALCF Account and Project Management system (UB3) where they provide details such as legal name, a valid e-mail address, work address, phone number, and country of citizenship. They also identify the ALCF project they are associated with. In addition, all foreign nationals (non-U.S. citizens) are required to fill out their personal, employer, demographic, and immigration/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 has to 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 our 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 would like 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

Dr. Paul K. Kearns -2- JAN 2 2 2018

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

# 8. Strategic Results

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

***(b) Has the Facility demonstrated effective engagements with strategic stakeholders (i.e., beyond the user population)?***

***(c) Is the Facility operating in a manner that enables delivery of facility mission and Department of Energy mission including maintaining a vibrant US effort in science and engineering?***

## 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 accomplishments;
- Scientific highlights; and
- Stakeholder engagement.

## 8.1 Scientific 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 are performed 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.

The Facility also 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 collect 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 Materials*, one in *Nature Methods*, two in *Nature Physics*, three in *Nature Communications,* two in *npj Computational Materials,* two in *Scientific Reports,* and two in *Scientific Data* (combined in the *Nature* journals category in the table below); three in *Science Advances (*listed under the *Science* journals category in the table below*)*; one in the *Proceedings of the National Academy of Sciences (PNAS)*; seven in *Physical Review Letters*; and three 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 after the prior year's OAR deadline.

In addition to publications, ALCF projects earned multiple awards in 2020, including HPCwire awards. The HPCwire Readers' Choice Award for Best Use of HPC in Energy was given to ALCF INCITE PI Muhsin Ameen for work on engine design (covered in the Science Highlights Section 8). The HPCwire Readers' Choice Award for Best Use of HPC in Industry was given to ALCF DD PI Roberto Torrelli for designing cleaner propulsion systems for engines. The team combined experimental results from the Advanced Photon Source with simulations on Theta to improve injector design in engines.

**Table 8.1 Summary of Users' Peer-Reviewed Publications in CY 2020**

| Nature Journals | Science | PNAS | Physical Review Letters | SC | Total 2020 Publications |
|---|---|---|---|---|---|
| 13 | 3 | 1 | 7 | 3 | 246 |

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

| OAR Year | CY 2016 | CY 2017 | CY 2018 | CY 2019 | CY 2020 |
|---|---|---|---|---|---|
| Total Publications | 199 | 225 | 276 | 286 | 246 |

## 8.2 Scientific 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 is what the project used. The individual bars represent the percentage of time used on the fraction of the machine shown below the bar, which are "no capability," "low capability," and "high capability" from left to right.

## 8.2.1 Predictive Modeling and Machine Learning for Functional Nanoporous Materials

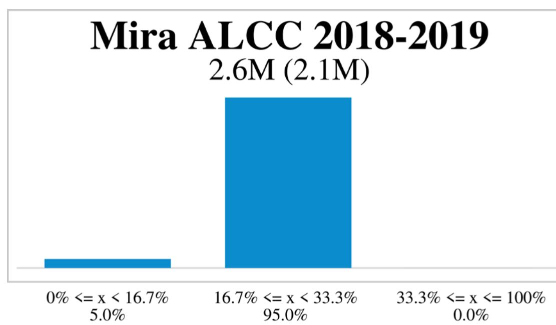
### The Science

Molecular simulations have been shown to predict accurate sorption equilibria, but the number and computational expense of the required simulations can be exceedingly large to optimize conditions for separating complex mixtures. A deep neural network (NN), called *SorbNet*, was developed that can predict multi-component adsorption isotherms over a wide range of temperatures and pressures and allows for finding optimal conditions for challenging separations. The deep NN developed here serves as a surrogate for molecular simulations of complex sorption equilibria that accurately reproduce simulation results; it can thus be used to generate continuous isotherm functions. Exercising the transfer learning capability of *SorbNet*, the NN was tuned with a small amount of additional data to successfully predict properties for similar sorption systems.

**Figure 8.1 Big data from high-throughput simulations and a deep neural network are utilized to find optimal conditions for contracting a solution phase (green box) with a zeolite (orange box) to yield pure liquid solvent and a zeolite with enriched product, followed by partial desorption of the solvent into a vapor phase, leading to essentially pure product that can subsequently be desorbed.**

Image courtesy of J. Ilja Siepmann.

### The Impact




More energy-efficient separation processes are needed to reduce the energy footprint of the chemical industries, but a large amount of the required pressure and temperature data is typically unavailable; furthermore, modeling complex sorption equilibria can be computationally expensive. The *SorbNet* network enables optimization of phase and sorption equilibrium properties in complex thermodynamic systems (Figure 8.1) for engineering and industrial processes without need for a laborious iterative search based on molecular simulations. The workflow created here provides a new avenue for modeling sorption equilibria with the coupling of machine learning and molecular simulations.

### Summary

The simulation training data was generated from multicomponent Gibbs ensemble Monte Carlo (GEMC) simulations to compute the equilibrium loading of binary systems. Two types of diols—butane-1,4-diol (C4) and pentane-1,5-diol (C5)—were used with water (W) or ethanol (E) as the solvent to create the binary mixtures to be separated within one of two all-silica zeolite adsorbents (MFI and LTA). Four combinations of adsorbent-diol-solvent were simulated at various loadings within the zeolite phase: MFI-C5-W, MFI-C5-E, MFI-C4-W, and LTA-C5-W. For each system, 256 temperature/volume combinations were investigated with 32 independent simulations at each state point. Adsorption data was generated on Mira using an ensemble workflow consisting of 131,072 simulations. The desorptive process was modeled as a three-stage process: (1) adsorption from a solution phase, (2) partial desorption of the solvent into a vapor phase, and (3) desorption of the desirable product. *SorbNet* accurately predicted the adsorbent and solvent loadings compared to simulation results and with similar variance in results as compared to simulations. Performance was compared against a shallow NN with the same number of parameters; and it was found that, while the shallow NN had some predictive power, the deep NN had superior performance. *SorbNet* was then applied to optimize process conditions for the MFI-C5-W system readily providing observables, such as the optimal temperature as a function of operation pressure for representative initial loadings, at a fraction of the computational cost of the corresponding molecular simulation workflow. *SorbNet* can utilize transfer learning to tackle different product compounds, solvents, and porous materials. The hidden-layer features in *SorbNet* were characterized to understand the predictive power of the network, identifying that higher-branched layers were playing a role in predicting sorption loading from features extracted by the lower layers. The network was found to be transferable to similar systems, although information on the new system is required for retraining the network if the new system is too different (i.e., changing adsorbates).

**ALCF Contribution:** ALCF staff have aided the project over the years, originally helping to port the ensemble computing workflow to Mira. ALCF staff worked with the lead author at the ALCF Simulation, Data, and Learning Workshop in October 2018, onboarding him to ALCF resources, and previously worked with the second author at the May 2016 ALCF Workshop.

**Contact**

J. Ilja Siepmann
University of Minnesota
siepmann@umn.edu

**Publication**

Y. Sun, R. F. DeJaco, and J. Ilja Siepmann, "Deep neural network learning of complex binary sorption equilibria from molecular simulation data," *Chem. Sci.* **10**, 4377–4388 (2019). **doi**: 10.1039/c8sc05340e

**Highlight Categories**

Performer/Facility: ASCR-ALCF

**Date Submitted to ASCR:** January 2020

## 8.2.2 The Overarching Framework 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 to inform the interpretation of data from large-scale experiments.

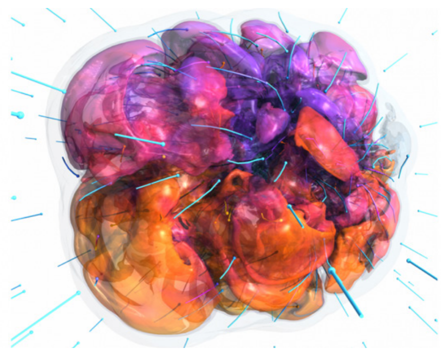
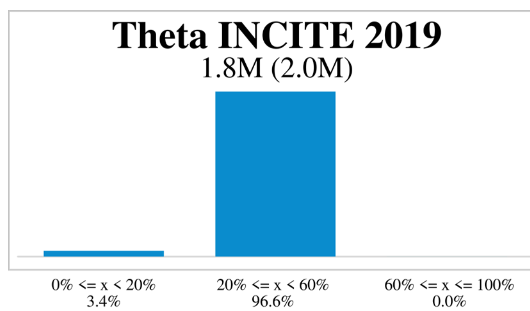


**Figure 8.2 This visualization depicts the neutrino-driven roiling convection of the nuclear material behind the newly re-energized supernova shock wave that surrounds the newly birthed neutron star, just hundreds of milliseconds after shock revival. The isosurfaces are of specific entropy, and the coloring follows the electron fraction. The pathlines trace infalling parcels of matter caught up in the turbulence of the explosion. The neutrino-heated turbulent bubbles help drive the supernova explosion shock wave outwards.**

Image courtesy of Joseph A. Insley and Silvio Rizzi, Argonne National Laboratory; Adam Burrows and Hiroki Nagakura, Princeton University.

### 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 team's efforts to advance the fundamental theoretical understanding of supernova explosions will benefit ongoing research efforts 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.




### Summary

Since the 1960s, there has been an agonizingly slow march toward demonstrating a robust mechanism of supernovae explosion. Although 2D simulations of supernovae have supported the theory that capturing a small fraction of the neutrinos emitted during collapse powers the explosions, detailed 3D calculations proving this paradigm are lacking. With the power of leadership-class supercomputers and continued advances in software, researchers now have the capabilities to tackle this long-standing challenge in nuclear astrophysics. To carry out this study, the team is using FORNAX, their new highly scalable, 3D radiation-hydrodynamics code. By addressing the transport operator with an explicit method, the code significantly reduces the computational complexity and communication overhead of traditional multidimensional radiative transfer solutions by bypassing the need for global iterative solvers. The team is running large-scale simulations on Theta aimed at determining whether the neutrino mechanism (Figure 8.2) is a robust explanation for supernova explosions and the stellar progenitor dependence of their observables.

Building upon previous work, the team is using Theta to complete a suite of non-rotating 3D supernova models. The suite includes simulations of the collapse, bounce, and explosion of progenitor massive stars at several different solar masses. The researchers are using these simulations for several studies, including an exploration of the impact of spatial resolution on the outcome and character of 3D supernova simulations. In another study, they are exploring the correlations between the time series and angular distributions of the neutrino and gravitational wave emissions and the real-time dynamics of the shock and the proto-neutron star core.

One of the most important conclusions is that the most massive progenitor models need to be continued for longer physical times, perhaps to many seconds, to asymptote to a final state, in particular vis à vis explosion energy. Moreover, while the majority of the inner ejecta have electron-fraction of 0.5, there is a substantial proton-rich tail. Those models that explode more lethargically and a bit later after bounce tend not to include much neutron-rich ejecta, while those that explode more quickly, such as the lowest-mass progenitors, can eject some more neutron-rich matter. However, in all of their 3D models, the inner ejecta has a net proton-richness. If true, this systematic result has important consequences for the nucleosynthesis yields as a function of progenitor.

**ALCF Contribution:** ALCF staff from the Visualization Team developed workflows to visualize the simulation data, overcoming a number of ParaView's limitations on performing large-scale particle tracing, and helped produce images and movies of the 3D entropy and electron-fraction distributions of exploding and non-exploding models for publications and presentations. The PI is in negotiations with the American Museum of Natural History in New York City to display some of the graphics created by ALCF staff there. The team's ALCF Catalyst provided support on Theta with ensemble job submissions and dependencies.

**Contact**

Adam Burrows
Professor of Astrophysical Sciences, Princeton University
burrows@astro.princeton.edu

**Publication**

A. Burrows, D. Radice, D. Vartanyan, H. Nagakura, M. Aaron Skinner, J.C. Dolence, "The overarching framework of core-collapse supernova explosions as revealed by 3D FORNAX simulations," *Mon. Not. R. Astron. Soc.*, **491**, 2, 2715–2735 (2020). **doi**: 10.1093/mnras/stz3223

**Highlight Categories**

Performer/Facility: ASCR-ALCF

**Date Submitted to ASCR:** May 2020

### 8.2.3 COVID-19 Spread and Effectiveness of Interventions

**The Science**

Argonne researchers have developed CityCOVID, an agent-based model capable of tracking detailed COVID-19 transmission (Figure 8.3). Agent-based modeling is an approach for capturing the dynamics of heterogeneous, interacting, adaptive agents at an individual, granular level of detail. When applied to a city like Chicago, CityCOVID includes a synthetic population representing the 2.7 million residents of Chicago and the 1.2 million geo-located locations where they can co-locate, including households, schools, workplaces, hospitals, nursing homes, dormitories, and jails. Throughout a simulated day, each individual, or agent, moves from place to place, hour by hour, engaging in social activities and interactions with co-located agents, where COVID-19 exposure events can occur. The COVID-19 disease progression is modeled within each individual, including differing symptom severities, hospitalizations, and age-dependent probabilities of transitions between disease stages.

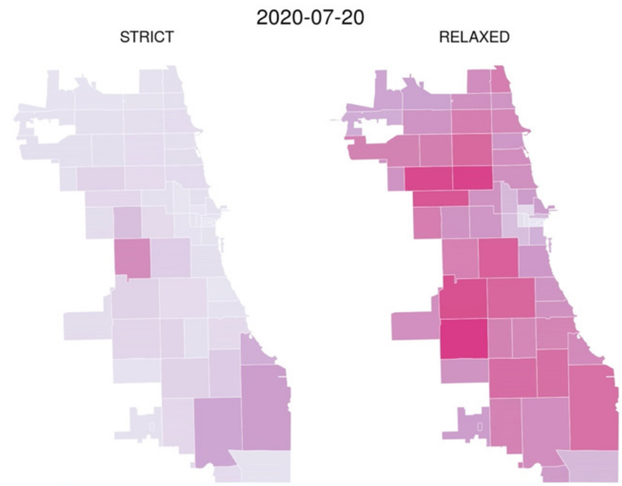
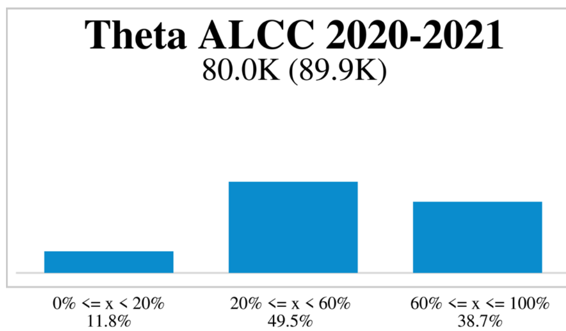



**Figure 8.3 Weekly counts of the newly infected counties by ZIP code under differing individual behavior scenarios.**

Image courtesy of Jonathan Ozik, Argonne National Laboratory.




**The Impact**

Using ALCF resources, CityCOVID is being used to calibrate unobserved model parameters, such as the time-varying degree of individual self-protective behaviors across the population, and to simulate a variety of interventions and future scenarios. Model results are being provided to the City of Chicago and Cook County Public Health Departments as well as the Illinois Governor's COVID-19 Modeling Task Force. While the Argonne team has been using Chicago as a testbed for developing these capabilities, CityCOVID is being extended to other regions as well.

**Summary**

With COVID-19 drastically altering daily life for people across the planet, DOE's Argonne National Laboratory has moved quickly to join the global fight against the pandemic. Among the laboratory's most powerful resources for scientific research is the supercomputer Theta, housed at ALCF, a DOE Office of Science User Facility. More than 250 nodes on the machine were immediately reserved for multipronged research into the disease.

Led by Argonne computational scientist Jonathan Ozik and Argonne Distinguished Fellow Charles (Chick) Macal, one of these branches of research oversees the development of epidemiological models to simulate the spread of COVID-19 throughout the population.

The significant computational demands of the project result from the models' stochastic components, which encapsulate the underlying uncertainties and parameters of the simulation, and the complexity of the population-level outcomes that result from the interactions of millions of individual software agents. The models pursue lines of inquiry that will be familiar to anyone following the virus in the news media—for example, the difference in outcomes yielded by implementing different nonpharmaceutical interventions (NPIs) and how to safely ease off of the NPIs. The Argonne-developed technologies Repast HPC (https://repast.github.io), ChiSIM (https://github.com/Repast/chiSIM), EMEWS (https://emews.github.io), and Swift/T (http://swift-lang.org/Swift-T/) are used to develop the model and to run the large-scale parameter estimation and NPI scenario workflows on Theta. ALCF Petrel and Globus services are being used for data archiving, distribution, and automation.

**ALCF Contribution:** ALCF was actively engaged with ensuring that the CityCOVID work completed in a timely manner. This task was quite manual as staff assisted with dynamically adjusting reservations based on surge needs of the project. ALCF installed software requirements for the CityCOVID software stack and worked with the team on the initial porting and enablement on Theta.

**Contact**

Jonathan Ozik
Argonne National Laboratory
jozik@anl.gov

**Publications**

J. Ozik, J. M. Wozniak, N. Collier, C. M. Macal, and M. Binois, "A Population Data-Driven Workflow for COVID-19 Modeling and Learning," Finalist 2020 ACM Gordon Bell Special Prize for HPC-Based COVID-19 Research, to appear in the International Journal of High-performance Computing Applications (IJHPCA).

**Highlight Categories**

Performer/Facility: ASCR-ALCF

**Date Submitted to ASCR:** July 2020

### 8.2.4 Modeling Fusion Plasma-Facing Components

**The Science**

Fusion as a practical energy source requires improved knowledge of plasma surface interactions and materials engineering design of components to survive the extreme heat and particle flux exposure conditions of a fusion power plant. One complication that has been identified is the creation of "fuzz" on the surface of tungsten—the primary plasma-facing divertor material in ITER—after a few hours of exposure to energetic helium ions; this fuzz degrades material stability. The fuzz formation mechanism is complicated and difficult to probe either experimentally or computationally. Large-scale molecular dynamics simulations were used to generate sufficiently long simulations that are relatively free of finite-size effects, which help to elucidate the mechanisms of helium transport and tungsten surface deformation (Figure 8.4).

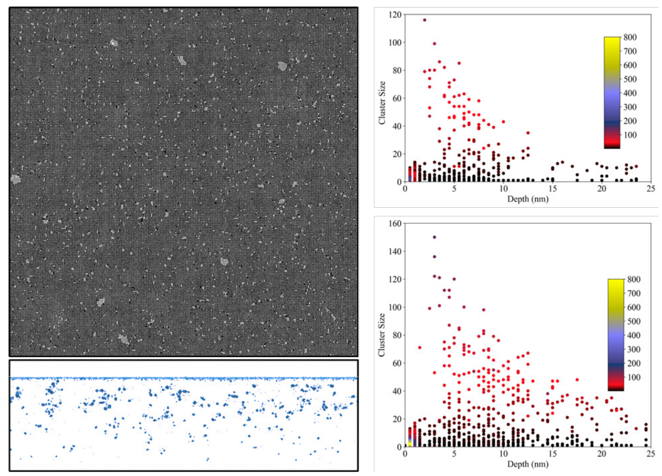
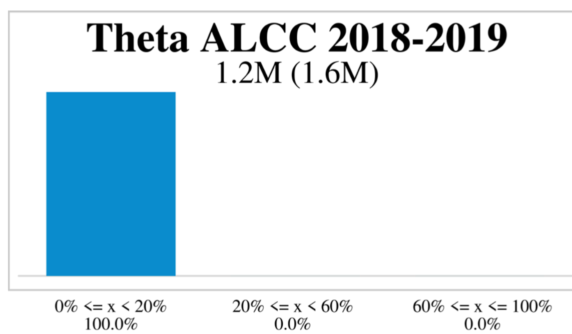



**Figure 8.4 Images adapted from K.D. Hammond, et al., *Nucl. Fusion* 59 (2019) 066035. Snapshot from a simulation of helium atoms beneath a tungsten (001) surface exposed to 100 eV helium plasma for 840 ns (fluence of 8.4 × $10^{18}$ $m^{-2}$): (left-top) top-down and (left-bottom) side views highlighting changes in surface height. Cluster size distributions for two fluxes are shown on the right.**

**The Impact**

Pronounced changes across three orders of magnitude of helium flux were observed for helium retention and helium bubble depth distribution, with deep bubbles tending to grow larger and create more prominent surface features over time. Results suggest that nearly all prior molecular simulations reported to date suffer from finite-size effects, and that helium flux is a very important parameter in determining the behavior of helium in plasma-facing components. The simulations also provide much-needed benchmarks for larger-scale models, which are necessary to better understand the mechanisms of helium transport, retention, and agglomeration. This, in turn, will help to identify material design strategies and potentially mitigate issues resulting from the extreme conditions of a fusion energy environment.




**Summary**

Molecular dynamics simulations with LAMMPS were used to simulate helium flux effects for two surface orientations. Lateral dimensions of up to 100 nanometers were considered to reduce finite-size effects, such as the formation of bubbles spanning the full width of the supercell. Four distinct helium fluxes were considered, ranging from $10^{25}$ to $10^{28}$ $m^{-2}$ $s^{-1}$, with simulation times ranging from 500 nanoseconds up to 2.5 microseconds, depending on the system size.

A temperature of 933 K was chosen so as to be at the low end of the fuzz-forming regime. The helium insertion depth was sampled from a distribution consistent with 100-eV incident energy, which results in 90% of helium atoms being implanted within 5 nm of the surface.

The depth distribution of helium ions was found to depend strongly on helium flux, with self-trapping at the surface being more probable at the highest fluxes due to the resulting higher density of helium near the surface. The result is a layer of near-surface bubbles that tended to coalesce. At high fluence, bubble-bursting events can result in significant and sudden changes in surface morphology. Bubble-bursting events result in abrupt changes in helium retention, causing retention to fall off at high fluence. Deep bubbles that rupture may serve to create pathways all the way to the plasma, allowing helium to penetrate deeper into the material and potentially form even larger bubbles that subsequently burst, causing further damage. Results suggest four distinct regimes that need to be considered: initial exposure, bubble growth, small bubbles near the surface beginning to burst, and larger bubbles beginning to burst that result in significant damage. Larger simulations are needed to test whether an additional regime exists involving bubbles further below the surface, although it is likely (because of the length of time that would be necessary) that these will need to be performed using techniques other than molecular dynamics. ITER-relevant fluxes are expected to generate significant numbers of helium clusters relatively deep in material before a substantial bubble population is built up.

The simulations reported in this highlight were generated in part using ALCF resources on Mira from several consecutive ALCC awards starting in 2013. Computing resources at NERSC were also utilized. Total usage for this and related projects at ALCF and NERSC was in excess of 300 million core-hours. All data for images in the figure except those for the upper right image were obtained using Mira.

**ALCF Contribution:** The team made good use of ALCF's overburn policy over the years, exceeding 100% of allocated hours every year since 2013. More recently, ALCF staff provided an optimized Makefile for compiling LAMMPS on Theta.

**Contact**

Karl Hammond
University of Missouri
hammondkd@missouri.edu

**ASCR Allocation PI:** Brian Wirth, University of Tennessee

**Publication**

K.D. Hammond, et al., "Helium flux effects on bubble growth and surface morphology in plasma-facing tungsten from large-scale molecular dynamics simulations," *Nucl. Fusion* **59**, 066035 (2019). DOI: 10.1088/1741-4326/ab12f6.

**Highlight Categories**

Performer/Facility: ASCR-ALCF

**Date Submitted to ASCR:** July 2020

## 8.2.5 Three Dimensions, Two Microscopes, One Code: Automatic Differentiation for X-ray Nanotomography Beyond the Depth of Focus Limit

### The Science

The APS Upgrade (APS-U) project will eventually deliver X-ray beams with much higher brightness to image centimeter-thick samples with sub-100-nm resolution. However, the multiple scattering problem of X-rays inside such a thick specimen must be addressed before researchers can fully harness the power of the upgraded light source. By leveraging the powerful computing facilities at ALCF, the team has devised a novel method to provide high-quality 3D reconstructions of heavily scattering samples measured under these complicated imaging scenarios.

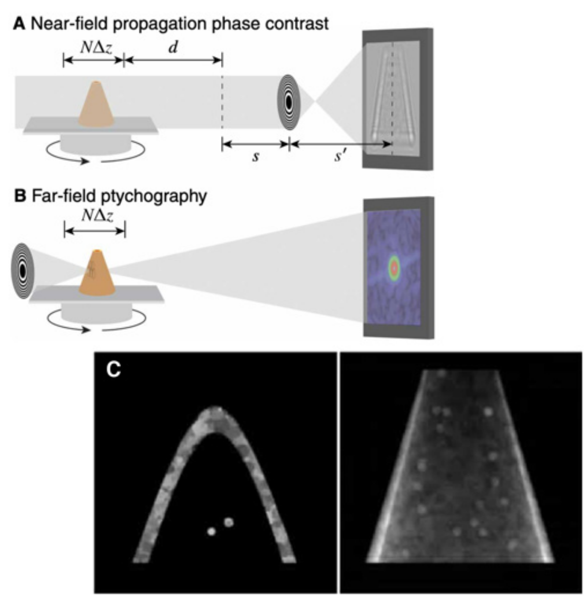
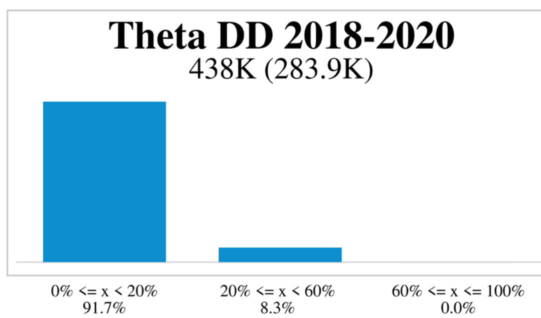



**Figure 8.5 Our automatic-differentiation-based software reconstructs beyond-depth-of-focus objects in both holography (A) and ptychography (B), providing results free of diffraction-induced artifacts (C).**

Image courtesy of Chris Jacobsen, Argonne National Laboratory.

### The Impact

Our proposed method enables the 3D structures of thick samples to be retrieved faithfully, which would otherwise be pestered by artifacts and distortions due to multiple scattering of X-rays (Figure 8.5). Moreover, we exploited the power of automatic differentiation that comes with most unsupervised machine learning tools, which makes our software highly versatile and adaptable to many variants and setups of X-ray microscopes.




### Summary

As the APS-U project is about to deliver sufficiently high beam brightness to image thick samples with high resolution, the diffraction of X-rays inside a thick specimen's volume becomes increasingly problematic, as this is not accounted for by most conventional X-ray 3D reconstruction algorithms. We propose a model-based method to address this issue, where a precise physics model is used to predict the detected wave intensity for a guessed object function, and the object is constantly adjusted to reduce the mismatch between the predicted intensity and the actual measurements. The method is shown to provide good reconstruction results that are free of diffraction-induced artifacts as observed in traditional reconstruction methods. In addition, we used the automatic differentiation (AD) capability of machine learning tools as the solver for our inversion problem, which eliminates the need of manually deriving the update equation, and makes the program very flexible and expandable.

With very minimal adaptation, our code can work for both holography and ptychography. In other words, we can now reconstruct three-dimensional objects beyond the depth of focus for two microscopy techniques, and with AD, we can perform these tasks with one code!

**ALCF Contribution:** ALCF staff members helped with parallelization of the software on ALCF supercomputers, provided guidance on best practices, and assisted with the setup and testing of the application code on ALCF systems.

**Contact**

Chris Jacobsen
Argonne National Laboratory
jcjacobsen@anl.gov

**Publication**

M. Du, Y.S.G. Nashed, S. Kandel, D. Gürsoy, and C. Jacobsen, "Three dimensions, two microscopes, one code: Automatic differentiation for x-ray nanotomography beyond the depth of focus limit," *Sci Adv* **6**, eaay3700 (2020). DOI: https://doi.org/10.1126/sciadv.aay3700

**Highlight Categories**

Performer/Facility: ASCR-ALCF

**Date Submitted to ASCR:** August 2020

### 8.2.6 Shock-Induced Multi-Material Mixing

**The Science**

Turbulent mixing is relevant in high-speed propulsion and other compressible flows for aerospace applications, and in high-energy density applications such as inertial confinement fusion and supernova dynamics; however, turbulence under these conditions remains poorly understood. This project focused on answering some of the fundamental (open) questions regarding the physics of mixing in turbulent shear layers, such as the effect of compressibility on the turbulent mixing process, for which there is, as yet, no consensus on the physical reason for growth rate reduction in supersonic mixing layers, and to quantify variable density effects across the compressibility range.

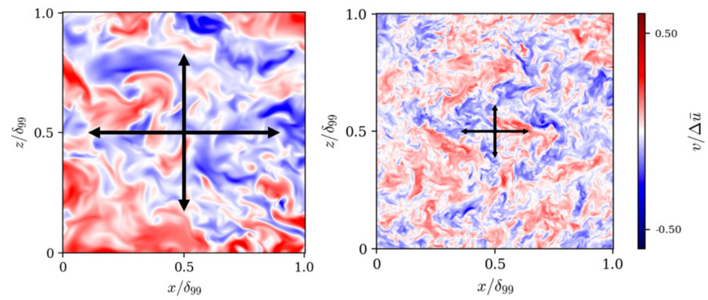
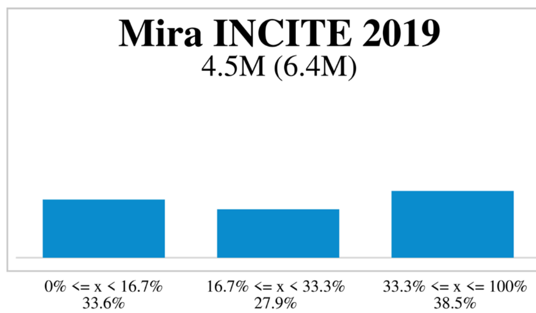



**Figure 8.6 Image shows instantaneous planar (y=0) slices of transverse velocity v at final simulation time for low (left) and high (right) compressibility cases with unity density ratio in a limited region of the flow. Arrows indicate decorrelation length scales based on velocity fluctuations along the x and z axis. The Reynolds shear stress associated with turbulent mixing decreases with compressibility and reduces its growth rate. While this compressibility effect has been known since the mid-1970s, the present simulations have provided a novel physical explanation. Turbulence scales were found to be internally limited in this flow by acoustic communication.**

Image courtesy of Sanjiva Lele, Stanford University.

Compared to shock-induced turbulent mixing resulting from the Richtmyer-Meshkov instability, studied in the PI's group previously, this study on the physics of compressible mixing with shear and variable density effects (Figure 8.6) provides invaluable insights on a wider range of problems within the engineering and science community, and complements difficulties in experimental measurements of highly compressible mixing layers. As a result, highly resolved numerical simulations play a critical role in bridging the gap between theory and experiments by allowing exploration of important aspects of turbulent flow dynamics that are inaccessible to experiments.




**The Impact**

Using up to 32,768 nodes of Mira, the simulations focused on the effects of compressibility and density variation on turbulent structures, and are the first to quantify these effects on turbulent scales across a comprehensive range of parameters. Published experimental results at lower compressibility levels are used to validate simulation data on crucial data points of stresses and growth rates. The examination of turbulent scales from the high-resolution simulations brought to light an internal regulation mechanism. This result was obtained by quantifying decorrelation, or "internal," time and length scales of energy-containing eddies, and analyzing the scaling of key turbulent stress magnitudes and budget terms in the context of these newly quantified scales. The results of these simulations and the models of turbulence-mixing length scales that are subsequently developed are expected to improve engineering models of highly compressible and variable density turbulence. It is also expected that the improved models will enable innovations in new combustion devices and propulsion systems.

**Summary**

The aim of this INCITE project was to improve our understanding of compressible mixing flows by performing high-resolution simulations of temporally evolving mixing layers across a large range of compressibility levels, and to separately evaluate the effects of density variations on mixing. There are several distinguishing features of both the methodology and the specific goals of this project as enumerated below:

1. The simulations employ high-order, minimally dissipative numerical schemes to demonstrate the state of the art in simulations of compressible turbulent mixing.
2. We will compare the results obtained from the highly resolved simulations with existing (lower compressibility and resolution) data turbulent mixing layers to establish confidence in the results obtained.
3. We explicitly quantified and accounted for the effects of numerics, and performed a budget analyses of turbulent energies and stresses in the compressible turbulent mixing flow phenomenon, where the decrease in growth of the transverse velocity fluctuations (i.e., orthogonal to the streamwise direction), and Reynold stresses, with increasing compressibility (i.e., higher convective Mach number), is now better understood.

In this project, the simulations were performed at higher compressibility levels, quantified in this project by the convective Mach number Mc, and also using sufficiently large domains in streamwise, spanwise, and transverse directions to allow full decorrelation of the turbulent scales. Present simulations used three times larger domains than in previously published simulations. These larger-span simulations have made it possible to extract reliable results from the self-similar growth period of the mixing layer, which is often difficult to achieve without outgrowing the computation domain. The high-resolution data makes it possible to quantify the decorrelation scales of the transverse velocity fluctuations; and using the present simulation data, the PI's team has been able to demonstrate that these decorrelation scales are integral to explaining mixing layer behavior in the high Mach number regime. The database that has been generated by using the INCITE allocation has yielded novel scientific insights into the behavior of highly compressible turbulence. Turbulence under these conditions was found to internally regulate its spatial scale and intensity (Figure 8.6). These scientific discoveries would not have been possible without the large-scale simulations.

**ALCF Contribution:** ALCF support staff provided timely responses on questions pertaining to the project team's jobs in the queue. Before the APS-DFD 2019 meeting, during Q3, the team's ALCF Catalyst assisted in obtaining additional overburn time to complete the study's objectives.

**Contact**

Sanjiva Lele
Stanford University
lele@stanford.edu

**Publication**

K. Matsuno and S.K. Lele, “Compressibility effects in high speed turbulent shear layers – revisited,” AIAA 2020-0573, Published Online: January 5, 2020. https://doi.org/10.2514/6.2020-0573.

**Highlight Categories**

Performer/Facility: ASCR-ALCF

**Date Submitted to ASCR:** October 2020

### 8.2.7 The Last Journey

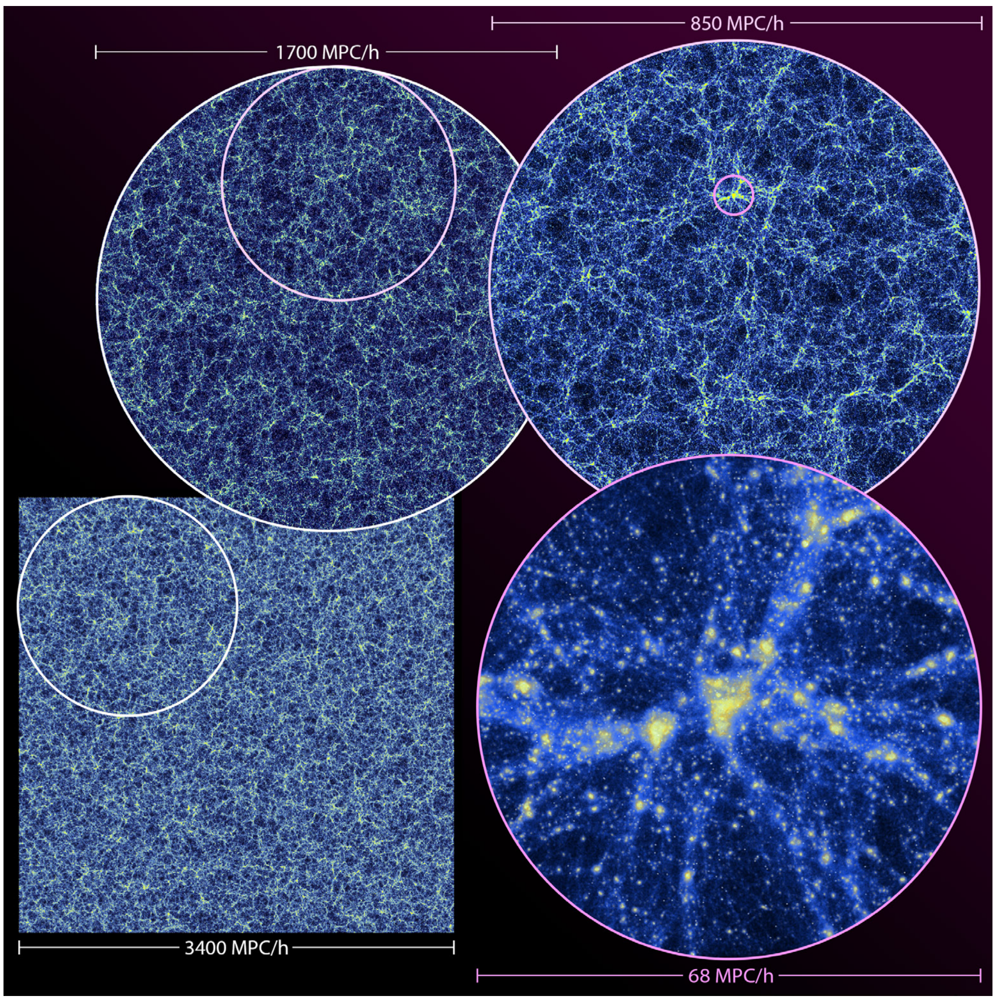


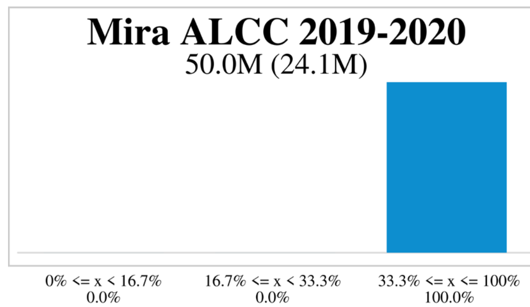

**Figure 8.7 Visualization of the Last Journey simulation. Shown are thin density slices for the full box (lower left corner) and zoom-ins at different levels. The panel on the lower right focuses on the largest cluster in the simulation.**

Image courtesy of Katrin Heitmann, et al.

**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 to generate and analyze the Last Journey simulation; one of the largest cosmological structure formation simulations ever performed. 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.

**Summary**

The team used the Hardware/Hybrid Accelerated Cosmology Code (HACC) to simulate the formation of large-scale structures in the universe over cosmic time by tracing the motions of 1.24 trillion mass-tracing particles under the influence of gravity in an expanding universe. 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.

**ALCF Contribution**

ALCF support for this work was comprehensive and spanned several years, and included performance engineering and code development support for HACC; and catalyst support, tool refinements, and images and movies of the Last Journey simulation.

**Contact**

Katrin Heitmann
Argonne National Laboratory
heitmann@anl.gov

**Publication**
K. Heitmann et al., "The Last Journey. I: An Extreme-Scale Simulation on the Mira Supercomputer," accepted by ApJS, https://arxiv.org/abs/2006.01697

**Highlight Categories**
Performer/Facility: ASCR-ALCF

**Date Submitted to ASCR:** January 2021

## 8.3 Stakeholder Engagement

To help ensure that the ALCF delivers on its mission of delivering breakthrough science, staff outreach needs to closely engage with domain science and keep a close eye on directions for supercomputing technologies. The ALCF provides a crucial balance of understanding how production science applications and computer science technologies can move into new and exciting machine architectures in the near term and 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 conference, workshops, reviews, and meetings. In 2020, staff participated in over 79 events. The figure below (Figure 8.8) breaks down these events by both type and community. Staff members support DOE mission needs by serving on review committees, advisory boards, and 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.

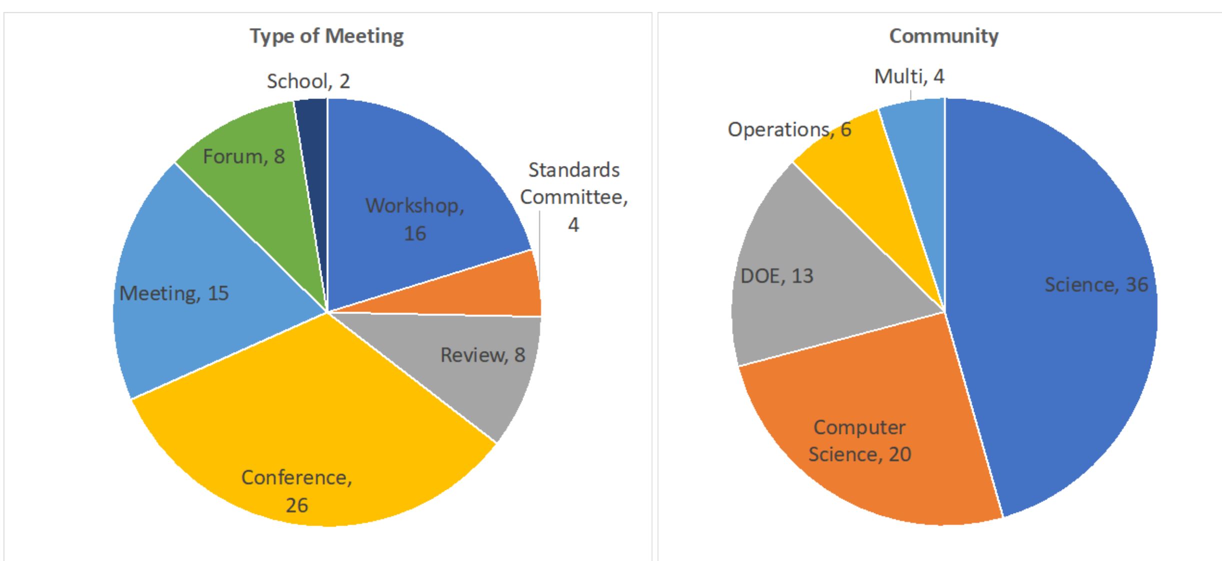



**Figure 8.8 Breakdown of some key activities by ALCF in CY 2020. The first pie chart (left) breaks down the 79 events by type. This is primarily derived from how the event self identifies. The second pie chart (right) breaks down the same 79 events by the community being engaged. 'Multi' refers to events such as Supercomputing (SC), where multiple communities are engaged.**

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

### 8.3.1 Engagement Highlights

#### *Supercomputing 2020*

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 a large participation in the event.

**Table 8.3 Summary of ALCF Participation in SC20**

| Program | Total |
|---|---|
| **Tutorials** | 2 |
| **Workshops** | 17 |
| **Papers** | 3 |
| **State of Practice** | 1 |
| **Poster** | 3 |
| **Birds of a Feather** | 3 |
| **Exhibitor Forum** | 1 |
| **Booth Talks** | 4 |

#### *Engagement in Standards and Community Groups*

The ALCF participates in a number of HPC standards and community groups in order 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 Xeon Phi User Group (IXPUG); C++ Standards Committee; OpenMP Architectural Review Board; Khronos OpenCL and SYCL working groups; OpenSFS; and MPI Forum.

On Khronos, ALCF participates in both the OpenCL and SYCL working groups. In 2020, the SYCL working group was focused on the new SYCL2020 specification published this year. In 2020, the OpenCL working group was focused on the OpenCL 3.0 specification. ALCF staff also maintain the Linux open source OpenCL loader.

On the C++ Standards Committee, ALCF votes on C++ proposals and participates in the Evolution Working Group (EWG) and Library Evolution WG (LEWG). An ALCF staff member is vice chair of the LEWG Incubator which looks at proposals not yet ready for the LEWG.

OpenSFS promotes the Lustre parallel file system and organizes the Lustre User Group conference. ALCF has 3 major Lustre file systems (theta-fs0, Grand, and Eagle). ALCF helped organize a public webinar series in place of the Lustre Users Group for 2020. An ALCF member is vice president of OpenSFS.

An ALCF staff member serves on the OpenMP Architectural Review Board. Three other staff members attend the language committee meetings. ALCF participates in reviewing, forming, and voting on new OpenMP specifications including the OpenMP5.1 specification released in 2020.

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

The P3HPC Forum, described in the User Support Section 1.3.4.2, is more than User Training. Originally focused on sharing definitions, best practices, and ideas for performance portability across DOE, this event has grown to a public event open to the public. The impact of this event has grown as the landscape of HPC platforms has made performance portability a viable target for supercomputing applications. In 2020, ALCF organized P3HPC as a key opportunity for the HPC community to discuss progress in performance portability.

#### *Summer Student Research Programs*

As mentioned in Outreach section 1.3.3.4, 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 2020, 33 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. Katherine Riley is a level-4 lead for the Application Integration area. 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 Industry Council. Haritha Siddabathuni Som is the level-3 lead for Facility Resource Utilization and Scott Parker is the level-3 lead for Application Integration. 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, as well as PathForward and CI efforts.

In 2020, more than 26 ALCF staff attended the Virtual ECP Annual Meeting held February 3–7, 2020, 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 the ECP training lead to promote ECP training activities to ALCF users.

#### *ECP-Funded Positions in ALCF*

To ensure that ECP goals are successful at the ALCF, ECP funds ALCF staff. The ALCF ECP/HI Applications Integration effort made great strides in 2020 porting and testing ECP applications across many GPUs including Intel GPUs. There are 17 staff members funded at various levels to work with ECP Application Development projects. In addition, 1 staff member is in place to focus on planning training, and the Intel COE for Aurora is staffed with 2 FTEs.

Additionally, two staff members have been funded to develop and deploy the continuous integration capabilities, and one staff member for supporting software technologies, and a half an FTE focused on ECP containers.

#### *Continuous Integration Pipeline*

Section 1.3.3.1, Resource Support, covers the deployment and growth of the Continuous Integration Pipeline through the ECP-CI project. This resource provides a key tool for ECP projects to have regular, automated testing on ALCF resources.

#### *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 RAC (discussed in Section 3.1.2) 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.

### 8.3.3 COVID-19 Response

At the beginning on March 2020, with DOE guidance, the ALCF began prioritizing COVID-19 research on facility resources. Projects that were contributing to fighting COVID-19 were given priority access to resources and staff. ALCF reached out to the user community to inform researchers that work on COVID-19 would be supported. Initially, all COVID-19 work was through the discretionary pool and has been now transitioned to ALCC time.

In total, 10 COVID-19 projects have compute time at ALCF, and an additional project served as a public database of information for the community. Over 1.7M node-hours were used on Theta, and two projects were finalists in the SC 2020 special COVID-19 Gordon Bell competition. The ALCF staff put significant effort into collaborating with most of these projects to ensure a quick start, and good performance. Some key activities regarding COVID-19 in 2020 include:

1. COVID-19 jobs ran with the highest priority on compute resources. Significant ALCF effort went into supporting the scheduling of COVID-19 projects for very rapid turn-around and near-real time while still supporting regular services.
2. A special queue of committed hardware was created on Theta. This ranged in size over the year, most often between 128 and 256 nodes.
3. Expertise of ALCF staff was made available to all projects, from computational and data science expertise to systems staff.
4. ALCF participated in the HPC Consortium for COVID-19 research including reviewing science merit and helping to match to appropriate resources. Through the HPC Consortium, the ALCF is supporting 3 projects as well as other COVID-19 projects already in the discretionary pool.

5. Monthly reporting on project status has been implemented.
6. ALCF expanded Theta with ThetaGPU to support compute needs for COVID-19 projects.

## 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 facility machines. In CY 2020, researchers participating in projects using ALCF resources published 246 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. A number 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 Calculation Details

**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 will be counted as an adjacent unscheduled outage. Typically, this designation would be assigned for a return to service four or more hours later than the scheduled end time. The centers have not yet agreed on a specific definition for this rare scenario.

***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 Calculation Details

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) Calculation Details

**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***:

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

## A.4 System Mean Time to Failure (MTTF) Calculation Details

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

***Formula:***

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

## A.5 Utilization Calculation Details

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***:

$$\textbf{Utilization} = \left(\frac{\textit{nodehours used in period}}{\textit{nodehours available in period}}\right) * \textbf{100}$$

## 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*:**

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

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

## 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.3.

**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, 2020 – December 31, 2020**
**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 | 19,385 |
| 3DMPCK | Harold Schock | Michigan State University | Three-Dimensional Multiphase Methodology to Study Piston Cylinder-kit Tribology | Energy Technologies | 31,250 |
| Acoustic_bubbles | Parviz Moin | Stanford University | Acoustic characterization of oceanic bubbles in turbulent environments using direct computation and modeling methods | Engineering | 9,738 |
| AIASMAAR | Rui Hu | Argonne National Laboratory | Artificial Intelligence Assisted Safety Modeling and Analysis of Advanced Reactors | Nuclear Energy | 25,868 |
| AIElectrolytes | Logan Timothy Ward | Argonne National Laboratory | Reinforcement Learning Driven Automated Optimization of Battery Electrolytes | Materials Science | 16,447 |
| aimaterialsADSP | Marco Govoni | The University of Chicago (UChicago) | Advanced Materials Characterization with AI-informed Computation | Materials Science | 46,789 |
| aimaterialsADSP | Marco Govoni | The University of Chicago (UChicago) | Advanced Materials Characterization with AI-informed Computation | Materials Science | 35,636 |
| Allinea | Raymond M. Loy,Kalyan Kumaran | Argonne National Laboratory | Improved debugging memory usage for BG/Q | Internal | 744 |
| AMASE | Rajkumar Kettimuthu | Argonne National Laboratory | Architecture and Management for Autonomous Science Ecosystems | Computer Science | 410 |
| AMASE | Rajkumar Kettimuthu | Argonne National Laboratory | Architecture and Management for Autonomous Science Ecosystems | Computer Science | 609 |
| AMEWS_DD | Giulia Galli | The University of Chicago (UChicago) | Advanced Materials for Energy-Water Systems Center | Materials Science | 234,375 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| AM_modelling | Lianghua Xiong | Argonne National Laboratory | Advanced modelling of phase transformation in metal additive manufacturing | Materials Science | 781 |
| APS-Linac | Hairong Shang, Yine Sun | Argonne National Laboratory | APS Linac Operation Optimization with AI | Physics | 1,634 |
| APSDataAnalysis | Nicholas Schwarz | Argonne National Laboratory | APS Beamline Data Processing and Analysis | Computer Science | 3,906 |
| APSDataAnalysis | Nicholas Schwarz | Argonne National Laboratory | APS Beamline Data Processing and Analysis | Computer Science | 1,450 |
| 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 | 3,617 |
| 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 | 34,141 |
| 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 | 15,711 |
| AstroAccel_Post | Dmitri Uzdensky, Gregory R Werner | University of Colorado-Denver | Astrophysical particle accelerators: magnetic reconnection and turbulence | Physics | 200 |
| atlasMLbjets | Rui Wang | Argonne National Laboratory | Using ML in b-jet identification at ATLAS | Physics | 2,343 |
| atlas_aesp | Walter Howard Hopkins | Argonne National Laboratory | Simulating and Learning in the ATLAS detector at the Exascale | Physics | 78,125 |
| atlas_aesp | Walter Howard Hopkins | Argonne National Laboratory | Simulating and Learning in the ATLAS detector at the Exascale | Physics | 14,734 |
| atlas_aesp | Walter Howard Hopkins | Argonne National Laboratory | Simulating and Learning in the ATLAS detector at the Exascale | Physics | 0 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| ATPESC2020 | Raymond M. Loy | Argonne National Laboratory | Argonne Training Program on Extreme-Scale Computing 2020 | Training | 15 |
| ATPESC2020 | Raymond M. Loy | Argonne National Laboratory | Argonne Training Program on Extreme-Scale Computing 2020 | Training | 7,812 |
| ATPESC_Instructors | Raymond M. Loy | Argonne National Laboratory | Argonne Training Program on Extreme-Scale Computing for ALL Instructors | Training | 15 |
| AutoBEM | Joshua Ryan New | Oak Ridge National Laboratory (ORNL) | Automatic Building Energy Modeling and analysis | Energy Technologies | 15,625 |
| AutoBEM | Joshua Ryan New | Oak Ridge National Laboratory (ORNL) | Automatic Building Energy Modeling and analysis | Energy Technologies | 4,601 |
| AutomatedBench | Kevin Harms | Argonne National Laboratory | Automated Benchmarking | Computer Science | 5,000 |
| AutomatedBench | Kevin Harms | Argonne National Laboratory | Automated Benchmarking | Computer Science | 8,749 |
| BIP167 | Philip Kurian | Howard University | Computing van der Waals many-body dispersion effects for MD simulations and THz spectroscopy of biomacromolecules in water | Physics | 13,939 |
| BIP167 | Philip Kurian | Howard University | Computing van der Waals many-body dispersion effects for MD simulations and THz spectroscopy of biomacromolecules in water | Physics | 8,182 |
| bloodflow_dd | Jifu Tan | Northern Illinois University (NIU) | Microfluidic design and optimization for cell separation | Engineering | 11,794 |
| bloodflow_dd | Jifu Tan | Northern Illinois University (NIU) | Microfluidic design and optimization for cell separation | Engineering | 7,855 |
| BrainImagingML | Thomas David Uram | Argonne National Laboratory | Large-scale Brain Imaging and Reconstruction | Biological Sciences | 6,250 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| BS-SOLCTRA | Esteban Meneses | Costa Rica National High Technology Center | Plasma Physics Simulations for SCR-1 Stellarator | Physics | 21,014 |
| Bubblecollapse | Eric Johnsen | University of Michigan | Dynamics and heat transfer in bubble collapse near solid surfaces | Engineering | 2,400 |
| Bubblecollapse | Eric Johnsen | University of Michigan | Dynamics and heat transfer in bubble collapse near solid surfaces | Engineering | 23,437 |
| BubbleOcean | Parviz Moin | Stanford University | Multiscale bubble breakup and gas transfer in turbulent oceanic environments | Engineering | 10,666 |
| Bubble_Collapse | Charlotte Noemie Barbier | Oak Ridge National Laboratory (ORNL) | High Definition Simulation of cavitating bubble near a wall with a shear flow | Engineering | 14,815 |
| Bubble_Collapse | Charlotte Noemie Barbier | Oak Ridge National Laboratory (ORNL) | High Definition Simulation of cavitating bubble near a wall with a shear flow | Engineering | 2,666 |
| candle_aesp | Rick Lyndon Stevens, Thomas Scott Brettin, Venkatram Vishwanath | Argonne National Laboratory | Virtual Drug Response Prediction | Biological Sciences | 156,250 |
| candle_aesp | Rick Lyndon Stevens, Thomas Scott Brettin, Venkatram Vishwanath | Argonne National Laboratory | Virtual Drug Response Prediction | Biological Sciences | 36,835 |
| candle_aesp | Rick Lyndon Stevens, Thomas Scott Brettin, Venkatram Vishwanath | Argonne National Laboratory | Virtual Drug Response Prediction | Biological Sciences | 0 |
| Carbon_composites | Hendrik Heinz | University of Colorado-Boulder | Rational design of ultrastrong composites | Materials Science | 15,798 |
| Carbon_composites | Hendrik Heinz | University of Colorado-Boulder | Rational design of ultrastrong composites | Materials Science | 31,250 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| Carbon_composites | Hendrik Heinz | University of Colorado-Boulder | Rational design of ultrastrong composites | Materials Science | 31,063 |
| CASPER | John Paul Nelson Walters | University of Southern California (USC) | CASPER: Award: CASPER: Compiler Abstractions Supporting high Performance on Extreme-scale Resources | Computer Science | 239 |
| CASPER | John Paul Nelson Walters | University of Southern California (USC) | CASPER: Award: CASPER: Compiler Abstractions Supporting high Performance on Extreme-scale Resources | Computer Science | 3,840 |
| catalysis_aesp | David Hamilton Bross | Argonne National Laboratory | Exascale Computational Catalysis | Chemistry | 100,000 |
| catalysis_aesp | David Hamilton Bross | Argonne National Laboratory | Exascale Computational Catalysis | Chemistry | 0 |
| catalysis_aesp | David Hamilton Bross | Argonne National Laboratory | Exascale Computational Catalysis | Chemistry | 35,362 |
| Catalyst | Katherine M Riley, Christopher James Knight, James Clifton Osborn, Timothy Joe Williams | Argonne National Laboratory | Catalyst | Internal | 78,231 |
| CFDML | Himanshu Sharma | Argonne National Laboratory | Embedding Machine Learning Models with Numerical Solvers for Scientific Applications | Engineering | 3,229 |
| CFDML | Himanshu Sharma | Argonne National Laboratory | Embedding Machine Learning Models with Numerical Solvers for Scientific Applications | Engineering | 3,265 |
| cfdml_aesp | Kenneth Edward Jansen | University of Colorado-Boulder | Data Analytics and Machine Learning for Exascale CFD | Engineering | 156,250 |
| cfdml_aesp | Kenneth Edward Jansen | University of Colorado-Boulder | Data Analytics and Machine Learning for Exascale CFD | Engineering | 36,835 |
| cfdml_aesp | Kenneth Edward Jansen | University of Colorado-Boulder | Data Analytics and Machine Learning for Exascale CFD | Engineering | 0 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CharmRTS | Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia | University of Illinois at Urbana-Champaign | Charm++ and its applications | Computer Science | 7,812 |
| CharmRTS | Laxmikant Kale, Abhinav Bhatele, Juan Jose Galvez-Garcia | University of Illinois at Urbana-Champaign | Charm++ and its applications | Computer Science | 3,817 |
| climate_severe | Vittorio Angelo Gensini | Northern Illinois University (NIU) | Anticipating Severe Weather Events via Dynamical Downscaling | Earth Science | 1,562 |
| climate_severe | Vittorio Angelo Gensini | Northern Illinois University (NIU) | Anticipating Severe Weather Events via Dynamical Downscaling | Earth Science | 1,184 |
| Climate_Water | Jiali Wang | Argonne National Laboratory | Linking climate to water: implementing a 4km regional climate model with hydrologic model coupling (WRF-Hydro) using Argonne's H | Earth Science | 31,250 |
| cmake | Robert James Maynard | Kitware Inc. | CMake Training | Computer Science | 156 |
| CobaltDevel | Paul Michael Rich, William Edward Allcock | Argonne National Laboratory | Cobalt Development | Internal | 9,468 |
| Comp_Perf_Workshop | Raymond M. Loy, Yasaman Ghadar | Argonne National Laboratory | ALCF Computational Performance Workshop | Training | 140,625 |
| connectomics_aesp | Nicola Joy Ferrier, Thomas David Uram | Argonne National Laboratory | Enabling Connectomics at Exascale to Facilitate Discoveries in Neuroscience | Biological Sciences | 312,500 |
| connectomics_aesp | Nicola Joy Ferrier, Thomas David Uram | Argonne National Laboratory | Enabling Connectomics at Exascale to Facilitate Discoveries in Neuroscience | Biological Sciences | 73,671 |
| connectomics_aesp | Nicola Joy Ferrier, Thomas David Uram | Argonne National Laboratory | Enabling Connectomics at Exascale to Facilitate Discoveries in Neuroscience | Biological Sciences | 0 |
| CONUS-Carbon | Jinxun Liu | U.S. Geological Survey (USGS) | Terrestrial ecosystem carbon cycle of the conterminous U.S. | Earth Science | 483 |
| CONUS-Carbon | Jinxun Liu | U.S. Geological Survey (USGS) | Terrestrial ecosystem carbon cycle of the conterminous U.S. | Earth Science | 8,949 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| Cray | Torrance Ivan Leggett, Mark R Fahey, Susan Marie Coghlan, Timothy Joe Williams | Cray Inc. | Cray Installation | Internal | 78,231 |
| cray-hpo | Michael Adnan Salim | Argonne National Laboratory | Scaling Studies of CrayAI Hyperparameter Optimization | Computer Science | 18,666 |
| CRRCS | Veerabhadra Kotamarthi | Argonne National Laboratory | Convection Resolved Regional Scale Climate Simulations | Earth Science | 3,200 |
| CSC249ADCD01 | Ian Foster | Argonne National Laboratory | 2.2.6.03 ADCD01-CODAR | Computer Science | 10,000 |
| CSC249ADCD01 | Ian Foster | Argonne National Laboratory | 2.2.6.03 ADCD01-CODAR | Computer Science | 3,000 |
| CSC249ADCD01 | Ian Foster | Argonne National Laboratory | 2.2.6.03 ADCD01-CODAR | Computer Science | 10,000 |
| CSC249ADCD01 | Ian Foster | Argonne National Laboratory | 2.2.6.03 ADCD01-CODAR | Computer Science | 3,000 |
| CSC249ADCD02 | Susan Marie Mniszewski, Timothy C. Germann | Los Alamos National Laboratory (LANL) | 2.2.6.04 ADCD02-COPA: Co-Design Center for Particle Applications | Physics | 20,760 |
| 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 | 1,000 |
| CSC249ADCD02 | Susan Marie Mniszewski, Timothy C. Germann | Los Alamos National Laboratory (LANL) | 2.2.6.04 ADCD02-COPA: Co-Design Center for Particle Applications | Physics | 1,000 |
| CSC249ADCD02 | Susan Marie Mniszewski, Timothy C. Germann | Los Alamos National Laboratory (LANL) | 2.2.6.04 ADCD02-COPA: Co-Design Center for Particle Applications | Physics | 1,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADCD04 | Tzanio Kolev, Misun Min, Paul Fischer | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.06 CEED: Center for Efficient Exascale Discretizations | Computer Science | 1,000 |
| CSC249ADCD04 | Tzanio Kolev, Misun Min, Paul Fischer | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.06 CEED: Center for Efficient Exascale Discretizations | Computer Science | 4,400 |
| CSC249ADCD04 | Tzanio Kolev, Misun Min, Paul Fischer | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.06 CEED: Center for Efficient Exascale Discretizations | Computer Science | 4,392 |
| CSC249ADCD04 | Tzanio Kolev, Misun Min, Paul Fischer | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.06 CEED: Center for Efficient Exascale Discretizations | Computer Science | 4,400 |
| CSC249ADCD04 | Tzanio Kolev, Misun Min, Paul Fischer | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.06 CEED: Center for Efficient Exascale Discretizations | Computer Science | 3,392 |
| CSC249ADCD05 | Mahantesh Halappanavar | Pacific Northwest National Laboratory (PNNL) | 2.2.6.07 ADCD05-ExaGraph | Computer Science | 10,000 |
| CSC249ADCD05 | Mahantesh Halappanavar | Pacific Northwest National Laboratory (PNNL) | 2.2.6.07 ADCD05-ExaGraph | Computer Science | 1,000 |
| CSC249ADCD05 | Mahantesh Halappanavar | Pacific Northwest National Laboratory (PNNL) | 2.2.6.07 ADCD05-ExaGraph | Computer Science | 7,812 |
| CSC249ADCD05 | Mahantesh Halappanavar | Pacific Northwest National Laboratory (PNNL) | 2.2.6.07 ADCD05-ExaGraph | Computer Science | 2,500 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADCD05 | Mahantesh Halappanavar | Pacific Northwest National Laboratory (PNNL) | 2.2.6.07 ADCD05-ExaGraph | Computer Science | 10,000 |
| CSC249ADCD08 | Francis Joseph Alexander | Brookhaven National Laboratory (BNL) | 2.2.6.08 ADCD08-ExaLearn | Physics | 71,000 |
| CSC249ADCD08 | Francis Joseph Alexander | Brookhaven National Laboratory (BNL) | 2.2.6.08 ADCD08-ExaLearn | Physics | 1,000 |
| CSC249ADCD08 | Francis Joseph Alexander | Brookhaven National Laboratory (BNL) | 2.2.6.08 ADCD08-ExaLearn | Physics | 1,000 |
| CSC249ADCD08 | Francis Joseph Alexander | Brookhaven National Laboratory (BNL) | 2.2.6.08 ADCD08-ExaLearn | Physics | 10,000 |
| CSC249ADCD502 | Kenneth John Roche | Pacific Northwest National Laboratory (PNNL) | 2.2.6.02 ADCD502 Application Assessment | Computer Science | 1,000 |
| CSC249ADCD502 | Kenneth John Roche | Pacific Northwest National Laboratory (PNNL) | 2.2.6.02 ADCD502 Application Assessment | Computer Science | 1,000 |
| CSC249ADCD502 | Kenneth John Roche | Pacific Northwest National Laboratory (PNNL) | 2.2.6.02 ADCD502 Application Assessment | Computer Science | 1,000 |
| CSC249ADCD502 | Kenneth John Roche | Pacific Northwest National Laboratory (PNNL) | 2.2.6.02 ADCD502 Application Assessment | Computer Science | 1,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADCD504 | David Richards, Shirley Victoria Moore | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.01 ADCD504-Proxy Applications | Computer Science | 1,000 |
| CSC249ADCD504 | David Richards, Shirley Victoria Moore | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.01 ADCD504-Proxy Applications | Computer Science | 1,000 |
| CSC249ADCD504 | David Richards, Shirley Victoria Moore | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.01 ADCD504-Proxy Applications | Computer Science | 1,000 |
| CSC249ADCD504 | David Richards, Shirley Victoria Moore | Lawrence Livermore National Laboratory (LLNL) | 2.2.6.01 ADCD504-Proxy Applications | Computer Science | 1,000 |
| CSC249ADOA01 | Rick Lyndon Stevens, Thomas Scott Brettin | Argonne National Laboratory | 2.2.4.03 ADOA01 CANDLE | Biological Sciences | 1,000 |
| CSC249ADOA01 | Rick Lyndon Stevens, Thomas Scott Brettin | Argonne National Laboratory | 2.2.4.03 ADOA01 CANDLE | Biological Sciences | 50,000 |
| CSC249ADOA01 | Rick Lyndon Stevens, Thomas Scott Brettin | Argonne National Laboratory | 2.2.4.03 ADOA01 CANDLE | Biological Sciences | 1,000 |
| CSC249ADOA01 | Rick Lyndon Stevens, Thomas Scott Brettin | Argonne National Laboratory | 2.2.4.03 ADOA01 CANDLE | Biological Sciences | 1,000 |
| CSC249ADSE03 | Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie | Fermi National Accelerator Laboratory (Fermilab) | 2.2.1.01 ADSE03-LatticeQCD | Physics | 100,000 |
| CSC249ADSE03 | Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie | Fermi National Accelerator Laboratory (Fermilab) | 2.2.1.01 ADSE03-LatticeQCD | Physics | 100,000 |
| CSC249ADSE03 | Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie | Fermi National Accelerator Laboratory (Fermilab) | 2.2.1.01 ADSE03-LatticeQCD | Physics | 50,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADSE03 | Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie | Fermi National Accelerator Laboratory (Fermilab) | 2.2.1.01 ADSE03-LatticeQCD | Physics | 20,000 |
| CSC249ADSE03 | Andreas Samuel Kronfeld, Norman Howard Christ, Paul Mackenzie | Fermi National Accelerator Laboratory (Fermilab) | 2.2.1.01 ADSE03-LatticeQCD | Physics | 100,000 |
| CSC249ADSE04 | Danny Perez | Los Alamos National Laboratory (LANL) | 2.2.1.04 ADSE04-EXAALT - Molecular dynamics at the exascale | Nuclear Energy | 15,000 |
| CSC249ADSE04 | Danny Perez | Los Alamos National Laboratory (LANL) | 2.2.1.04 ADSE04-EXAALT - Molecular dynamics at the exascale | Nuclear Energy | 1,000 |
| CSC249ADSE04 | Danny Perez | Los Alamos National Laboratory (LANL) | 2.2.1.04 ADSE04-EXAALT - Molecular dynamics at the exascale | Nuclear Energy | 1,000 |
| CSC249ADSE04 | Danny Perez | Los Alamos National Laboratory (LANL) | 2.2.1.04 ADSE04-EXAALT - Molecular dynamics at the exascale | Nuclear Energy | 470 |
| CSC249ADSE05 | David Trebotich | Lawrence Berkeley National Laboratory (LBNL) | 2.2.3.04 ADSE05-Subsurface | Earth Science | 1,000 |
| CSC249ADSE05 | David Trebotich | Lawrence Berkeley National Laboratory (LBNL) | 2.2.3.04 ADSE05-Subsurface | Earth Science | 1,000 |
| CSC249ADSE05 | David Trebotich | Lawrence Berkeley National Laboratory (LBNL) | 2.2.3.04 ADSE05-Subsurface | Earth Science | 1,000 |
| CSC249ADSE05 | David Trebotich | Lawrence Berkeley National Laboratory (LBNL) | 2.2.3.04 ADSE05-Subsurface | Earth Science | 1,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADSE08 | Steven Hamilton, Paul Kollath Romano | Oak Ridge National Laboratory (ORNL) | 2.2.2.03 ADSE08 ExaSMR | Nuclear Energy | 1,000 |
| CSC249ADSE08 | Steven Hamilton, Paul Kollath Romano | Oak Ridge National Laboratory (ORNL) | 2.2.2.03 ADSE08 ExaSMR | Nuclear Energy | 10,000 |
| CSC249ADSE08 | Steven Hamilton, Paul Kollath Romano | Oak Ridge National Laboratory (ORNL) | 2.2.2.03 ADSE08 ExaSMR | Nuclear Energy | 10,000 |
| CSC249ADSE08 | Steven Hamilton, Paul Kollath Romano | Oak Ridge National Laboratory (ORNL) | 2.2.2.03 ADSE08 ExaSMR | Nuclear Energy | 10,000 |
| CSC249ADSE09 | Paul Richard Charles Kent, Anouar Benali | Oak Ridge National Laboratory (ORNL) | 2.2.1.06 QMCPACK: Predictive and Improvable Quantum-mechanics Based Simulations | Materials Science | 1,000 |
| CSC249ADSE09 | Paul Richard Charles Kent, Anouar Benali | Oak Ridge National Laboratory (ORNL) | 2.2.1.06 QMCPACK: Predictive and Improvable Quantum-mechanics Based Simulations | Materials Science | 1,000 |
| CSC249ADSE09 | Paul Richard Charles Kent, Anouar Benali | Oak Ridge National Laboratory (ORNL) | 2.2.1.06 QMCPACK: Predictive and Improvable Quantum-mechanics Based Simulations | Materials Science | 1,000 |
| CSC249ADSE09 | Paul Richard Charles Kent, Anouar Benali | Oak Ridge National Laboratory (ORNL) | 2.2.1.06 QMCPACK: Predictive and Improvable Quantum-mechanics Based Simulations | Materials Science | 1,000 |
| CSC249ADSE11 | Theresa Windus | University of Washington | 2.2.1.02 ADSE11-NWChemEx: Tackling Chemical, Materials, & Biomolecular Challenges in Exascale | Chemistry | 4,000 |
| CSC249ADSE11 | Theresa Windus | University of Washington | 2.2.1.02 ADSE11-NWChemEx: Tackling Chemical, Materials, & Biomolecular Challenges in Exascale | Chemistry | 4,000 |
| CSC249ADSE11 | Theresa Windus | University of Washington | 2.2.1.02 ADSE11-NWChemEx: Tackling Chemical, Materials, & Biomolecular Challenges in Exascale | Chemistry | 4,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADSE12 | Amitava Bhattacharjee | Princeton Plasma Physics Laboratory (PPPL) | 2.2.2.05 ADSE12 WDMAPP | Computer Science | 40,000 |
| CSC249ADSE14 | Jacqueline Chen | Sandia National Laboratories, California | 2.2.2.02 ADSE14-Combustion-Pele: Transforming Combustion Science & Technology with Exascale Simulations | Engineering | 100,000 |
| CSC249ADSE14 | Jacqueline Chen | Sandia National Laboratories, California | 2.2.2.02 ADSE14-Combustion-Pele: Transforming Combustion Science & Technology with Exascale Simulations | Engineering | 40,000 |
| CSC249ADSE14 | Jacqueline Chen | Sandia National Laboratories, California | 2.2.2.02 ADSE14-Combustion-Pele: Transforming Combustion Science & Technology with Exascale Simulations | Engineering | 80,000 |
| CSC249ADSE14 | Jacqueline Chen | Sandia National Laboratories, California | 2.2.2.02 ADSE14-Combustion-Pele: Transforming Combustion Science & Technology with Exascale Simulations | Engineering | 30,000 |
| CSC249ADSE16 | Mark S Gordon | Ames Laboratory | 2.2.1.03 ADSE16-GAMESS | Chemistry | 11,000 |
| CSC249ADSE16 | Mark S Gordon | Ames Laboratory | 2.2.1.03 ADSE16-GAMESS | Chemistry | 10,200 |
| CSC249ADSE16 | Mark S Gordon | Ames Laboratory | 2.2.1.03 ADSE16-GAMESS | Chemistry | 11,000 |
| CSC249ADSE16 | Mark S Gordon | Ames Laboratory | 2.2.1.03 ADSE16-GAMESS | Chemistry | 34,750 |
| CSC249ADSE16 | Mark S Gordon | Ames Laboratory | 2.2.1.03 ADSE16-GAMESS | Chemistry | 11,000 |
| CSC249ADSE16 | Mark S Gordon | Ames Laboratory | 2.2.1.03 ADSE16-GAMESS | Chemistry | 13,100 |
| CSC249ADSE17 | Charlie Catlett, Melissa Ree Allen, Rajeev Jain, Scott A Ehling | Argonne National Laboratory | 2.2.4.01 Urban: Multiscale Coupled Urban Systems | Computer Science | 1,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADSE17 | Charlie Catlett, Melissa Ree Allen, Rajeev Jain, Scott A Ehling | Argonne National Laboratory | 2.2.4.01 Urban: Multiscale Coupled Urban Systems | Computer Science | 1,000 |
| CSC249ADSE17 | Charlie Catlett, Melissa Ree Allen, Rajeev Jain, Scott A Ehling | Argonne National Laboratory | 2.2.4.01 Urban: Multiscale Coupled Urban Systems | Computer Science | 1,000 |
| CSC249ADSE18 | Daniel Kasen | Lawrence Berkeley National Laboratory (LBNL) | 2.2.3.01 ADSE18 Exastar | Physics | 1,000 |
| CSC249ADSE18 | Daniel Kasen | Lawrence Berkeley National Laboratory (LBNL) | 2.2.3.01 ADSE18 Exastar | Physics | 1,000 |
| CSC249ADSE22 | Slaven Peles, Andres Marquez, Zhenyu Huang | Pacific Northwest National Laboratory (PNNL) | 2.2.4.02 ADSE22-ExaSGD | Energy Technologies | 1,000 |
| CSC249ADSE22 | Slaven Peles, Andres Marquez, Zhenyu Huang | Pacific Northwest National Laboratory (PNNL) | 2.2.4.02 ADSE22-ExaSGD | Energy Technologies | 1,000 |
| CSC249ADSE22 | Slaven Peles, Andres Marquez, Zhenyu Huang | Pacific Northwest National Laboratory (PNNL) | 2.2.4.02 ADSE22-ExaSGD | Energy Technologies | 1,000 |
| CSC249ADSE22 | Slaven Peles, Andres Marquez, Zhenyu Huang | Pacific Northwest National Laboratory (PNNL) | 2.2.4.02 ADSE22-ExaSGD | Energy Technologies | 1,000 |
| CSC249ADTR01 | Daniel Edward Laney | Lawrence Livermore National Laboratory (LLNL) | 2.3.5.10 ADTR01-ExaWorks | Computer Science | 1,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC249ADTR02 | Ashley Barker | Oak Ridge National Laboratory (ORNL) | 2.4.6.02 ADTR02- Productivity | Computer Science | 1,000 |
| CSC249ADTR02 | Ashley Barker | Oak Ridge National Laboratory (ORNL) | 2.4.6.02 ADTR02- Productivity | Computer Science | 1,500 |
| CSC249ADTR02 | Ashley Barker | Oak Ridge National Laboratory (ORNL) | 2.4.6.02 ADTR02- Productivity | Computer Science | 1,000 |
| CSC249ADTR02 | Ashley Barker | Oak Ridge National Laboratory (ORNL) | 2.4.6.02 ADTR02- Productivity | Computer Science | 1,000 |
| CSC250STDA05 | Kenneth Dean Moreland | Sandia National Laboratories, New Mexico | 2.3.4.13 STDA05-ECP/VTK-m | Computer Science | 1,000 |
| CSC250STDA05 | Kenneth Dean Moreland | Sandia National Laboratories, New Mexico | 2.3.4.13 STDA05-ECP/VTK-m | Computer Science | 1,000 |
| CSC250STDA05 | Kenneth Dean Moreland | Sandia National Laboratories, New Mexico | 2.3.4.13 STDA05-ECP/VTK-m | Computer Science | 1,000 |
| CSC250STDA05 | Kenneth Dean Moreland | Sandia National Laboratories, New Mexico | 2.3.4.13 STDA05-ECP/VTK-m | Computer Science | 1,000 |
| CSC250STDA05 | Kenneth Dean Moreland | Sandia National Laboratories, New Mexico | 2.3.4.13 STDA05-ECP/VTK-m | Computer Science | 1,000 |
| CSC250STDM10 | Surendra Byna, Venkatram Vishwanath | Lawrence Berkeley National Laboratory (LBNL) | 2.3.4.15 ExaIO - Delivering Efficient Parallel I/O on Exascale Computing Systems with HDF5 and Unify | Computer Science | 4,500 |
| 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 | 3,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| 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 | 5,000 |
| CSC250STDM10 | Surendra Byna, Venkatram Vishwanath | Lawrence Berkeley National Laboratory (LBNL) | 2.3.4.15 ExaIO - Delivering Efficient Parallel I/O on Exascale Computing Systems with HDF5 and Unify | Computer Science | 4,500 |
| 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 | 1,000 |
| 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 | 1,000 |
| 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 | 1,000 |
| 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 | 1,000 |
| 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 | 1,000 |
| 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 | 1,000 |
| 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 | 1,000 |
| 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 | 10,000 |

| 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 | 8,000 |
| 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 | 1,000 |
| 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 | 10,000 |
| CSC250STDM16 | James Paul Ahrens, Terece Louise Turton | Lawrence Livermore National Laboratory (LLNL) | 2.3.4.16 STDM16-ALPINE/ZFP | Computer Science | 1,000 |
| CSC250STDM16 | James Paul Ahrens, Terece Louise Turton | Lawrence Livermore National Laboratory (LLNL) | 2.3.4.16 STDM16-ALPINE/ZFP | Computer Science | 1,000 |
| CSC250STDM16 | James Paul Ahrens, Terece Louise Turton | Lawrence Livermore National Laboratory (LLNL) | 2.3.4.16 STDM16-ALPINE/ZFP | Computer Science | 1,000 |
| CSC250STDT10 | Jeffrey S Vetter | Oak Ridge National Laboratory (ORNL) | 2.3.2.10 STDT10 PPROTEAS-TUNE | Computer Science | 1,000 |
| CSC250STDT10 | Jeffrey S Vetter | Oak Ridge National Laboratory (ORNL) | 2.3.2.10 STDT10 PPROTEAS-TUNE | Computer Science | 1,000 |
| CSC250STDT10 | Jeffrey S Vetter | Oak Ridge National Laboratory (ORNL) | 2.3.2.10 STDT10 PPROTEAS-TUNE | Computer Science | 1,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC250STDT10 | Jeffrey S Vetter | Oak Ridge National Laboratory (ORNL) | 2.3.2.10 STDT10 PPROTEAS-TUNE | Computer Science | 1,000 |
| CSC250STDT11 | Barbara Mary Chapman | Stony Brook University | 2.3.2.11 SOLLVE: Scaling OpenMP with LLVm for Exascale | Computer Science | 6,000 |
| CSC250STDV01 | Charles Vernon Atkins | Kitware Inc. | 2.3.4.01 STDV01-Data and Visualization Software Development Kit | Computer Science | 1,000 |
| CSC250STDV01 | Charles Vernon Atkins | Kitware Inc. | 2.3.4.01 STDV01-Data and Visualization Software Development Kit | Computer Science | 1,000 |
| CSC250STML13 | Jack Dongarra | The University of Tennessee at Knoxville | 2.3.3.13 STML13 - CLOVER | Computer Science | 1,000 |
| CSC250STML13 | Jack Dongarra | The University of Tennessee at Knoxville | 2.3.3.13 STML13 - CLOVER | Computer Science | 1,000 |
| CSC250STML13 | Jack Dongarra | The University of Tennessee at Knoxville | 2.3.3.13 STML13 - CLOVER | Computer Science | 1,000 |
| CSC250STMS05 | Ulrike Meier Yang, Satish Balay | Argonne National Laboratory | 2.3.3.01 STMS05-Extreme-scale Scientific xSDK for ECP | Mathematics | 1,000 |
| CSC250STMS05 | Ulrike Meier Yang, Satish Balay | Argonne National Laboratory | 2.3.3.01 STMS05-Extreme-scale Scientific xSDK for ECP | Mathematics | 1,000 |
| CSC250STMS05 | Ulrike Meier Yang,Satish Balay | Argonne National Laboratory | 2.3.3.01 STMS05-Extreme-scale Scientific xSDK for ECP | Mathematics | 1,000 |
| CSC250STMS05 | Ulrike Meier Yang, Satish Balay | Argonne National Laboratory | 2.3.3.01 STMS05-Extreme-scale Scientific xSDK for ECP | Mathematics | 1,000 |
| CSC250STMS07 | Todd S. Munson, Hong Zhang, Richard Tran Mills, Satish Balay | Argonne National Laboratory | 2.3.3.06 STMS07-PETSc/TAO for Exascale | Mathematics | 1,500 |
| CSC250STMS07 | Todd S. Munson, Hong Zhang, Richard Tran Mills, Satish Balay | Argonne National Laboratory | 2.3.3.06 STMS07-PETSc/TAO for Exascale | Mathematics | 1,500 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC250STMS07 | Todd S. Munson, Hong Zhang, Richard Tran Mills, Satish Balay | Argonne National Laboratory | 2.3.3.06 STMS07-PETSc/TAO for Exascale | Mathematics | 1,500 |
| CSC250STMS07 | Todd S. Munson, Hong Zhang, Richard Tran Mills, Satish Balay | Argonne National Laboratory | 2.3.3.06 STMS07-PETSc/TAO for Exascale | Mathematics | 1,500 |
| CSC250STNS01 | Michael Lang, Terece Louise Turton | Los Alamos National Laboratory (LANL) | 2.3.6.01 - STNS01 -LANL ATDM ST Projects | Computer Science | 1,000 |
| CSC250STNS01 | Michael Lang, Terece Louise Turton | Los Alamos National Laboratory (LANL) | 2.3.6.01 - STNS01 -LANL ATDM ST Projects | Computer Science | 1,000 |
| CSC250STNS01 | Michael Lang, Terece Louise Turton | Los Alamos National Laboratory (LANL) | 2.3.6.01 - STNS01 -LANL ATDM ST Projects | Computer Science | 1,000 |
| CSC250STNS01 | Michael Lang, Terece Louise Turton | Los Alamos National Laboratory (LANL) | 2.3.6.01 - STNS01 -LANL ATDM ST Projects | Computer Science | 1,000 |
| CSC250STPM09 | Pavan Balaji | Argonne National Laboratory | 2.3.1.07 STPM09-Exascale MPI | Computer Science | 8,000 |
| CSC250STPM11 | George Bosilca, Earl Luther Carr, Jack Dongarra, Thomas Herault | The University of Tennessee at Knoxville | 2.3.1.09 STPM11 ParSEC: Distributed Tasking | Computer Science | 1,000 |
| CSC250STPM11 | George Bosilca, Earl Luther Carr, Jack Dongarra, Thomas Herault | The University of Tennessee at Knoxville | 2.3.1.09 STPM11 ParSEC: Distributed Tasking | Computer Science | 1,000 |
| CSC250STPM11 | George Bosilca, Earl Luther Carr, Jack Dongarra, Thomas Herault | The University of Tennessee at Knoxville | 2.3.1.09 STPM11 ParSEC: Distributed Tasking | Computer Science | 1,000 |
| CSC250STPM11 | George Bosilca, Earl Luther Carr, Jack Dongarra, Thomas Herault | The University of Tennessee at Knoxville | 2.3.1.09 STPM11 ParSEC: Distributed Tasking | Computer Science | 1,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| 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 | 2,000 |
| CSC250STPM17 | Paul Hamilton Hargrove, Erich Strohmaier | Lawrence Berkeley National Laboratory (LBNL) | 2.3.1.14 STPM17-UPC++ & GASNet | Computer Science | 1,000 |
| CSC250STPM17 | Paul Hamilton Hargrove, Erich Strohmaier | Lawrence Berkeley National Laboratory (LBNL) | 2.3.1.14 STPM17-UPC++ & GASNet | Computer Science | 1,000 |
| CSC250STPM17 | Paul Hamilton Hargrove, Erich Strohmaier | Lawrence Berkeley National Laboratory (LBNL) | 2.3.1.14 STPM17-UPC++ & GASNet | Computer Science | 1,000 |
| CSC250STPM17 | Paul Hamilton Hargrove, Erich Strohmaier | Lawrence Berkeley National Laboratory (LBNL) | 2.3.1.14 STPM17-UPC++ & GASNet | Computer Science | 1,000 |
| CSC250STPR19 | Peter Hugh Beckman | Argonne National Laboratory | 2.3.1.19 STPR19 Argo: Argo/Power Steering | Computer Science | 1,000 |
| CSC250STPR19 | Peter Hugh Beckman | Argonne National Laboratory | 2.3.1.19 STPR19 Argo: Argo/Power Steering | Computer Science | 1,000 |
| CSC250STPR19 | Peter Hugh Beckman | Argonne National Laboratory | 2.3.1.19 STPR19 Argo: Argo/Power Steering | Computer Science | 1,000 |
| CSC250STPR19 | Peter Hugh Beckman | Argonne National Laboratory | 2.3.1.19 STPR19 Argo: Argo/Power Steering | Computer Science | 1,000 |
| CSC250STPR27 | David Edward Bernholdt | Oak Ridge National Laboratory (ORNL) | 2.3.1.17 STPR27-OMPI-X: Open MPI for Exascale | Materials Science | 1,000 |
| CSC250STPR27 | David Edward Bernholdt | Oak Ridge National Laboratory (ORNL) | 2.3.1.17 STPR27-OMPI-X: Open MPI for Exascale | Materials Science | 3,000 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| CSC250STPR27 | David Edward Bernholdt | Oak Ridge National Laboratory (ORNL) | 2.3.1.17 STPR27-OMPI-X: Open MPI for Exascale | Materials Science | 1,000 |
| CSC250STPR27 | David Edward Bernholdt | Oak Ridge National Laboratory (ORNL) | 2.3.1.17 STPR27-OMPI-X: Open MPI for Exascale | Materials Science | 1,000 |
| CSC250STTO09 | Jack Dongarra,Anthony Danalis,Earl Luther Carr,Heike Jagode | The University of Tennessee at Knoxville | 2.3.2.06 STTO09 EXAPAPI | Computer Science | 1,000 |
| CSC250STTO09 | Jack Dongarra, Anthony Danalis, Earl Luther Carr, Heike Jagode | The University of Tennessee at Knoxville | 2.3.2.06 STTO09 EXAPAPI | Computer Science | 1,000 |
| CSC250STTO09 | Jack Dongarra, Anthony Danalis, Earl Luther Carr, Heike Jagode | The University of Tennessee at Knoxville | 2.3.2.06 STTO09 EXAPAPI | Computer Science | 1,000 |
| CSC250STTO09 | Jack Dongarra, Anthony Danalis, Earl Luther Carr, Heike Jagode | The University of Tennessee at Knoxville | 2.3.2.06 STTO09 EXAPAPI | Computer Science | 1,000 |
| CSC250STTO11 | John Michael Mellor-Crummey | Rice University | 2.3.2.08 STTO11 HPCToolkit | Computer Science | 1,000 |
| CSC250STTO11 | John Michael Mellor-Crummey | Rice University | 2.3.2.08 STTO11 HPCToolkit | Computer Science | 1,000 |
| CSC250STTO11 | John Michael Mellor-Crummey | Rice University | 2.3.2.08 STTO11 HPCToolkit | Computer Science | 1,000 |
| CSC250STTO11 | John Michael Mellor-Crummey | Rice University | 2.3.2.08 STTO11 HPCToolkit | Computer Science | 1,000 |
| CSC251HIHE05 | Scott Dov Pakin, Simon David Hammond | Los Alamos National Laboratory (LANL) | 2.4.2.01 HIHE05-Analytical Modeling - Hardware Evaluation Working Groups | Computer Science | 1,000 |
| CSC251HIHE05 | Scott Dov Pakin, Simon David Hammond | Los Alamos National Laboratory (LANL) | 2.4.2.01 HIHE05-Analytical Modeling - Hardware Evaluation Working Groups | Computer Science | 1,000 |

| 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 | 1,000 |
| CSC251HIHE05 | Scott Dov Pakin, Simon David Hammond | Los Alamos National Laboratory (LANL) | 2.4.2.01 HIHE05-Analytical Modeling - Hardware Evaluation Working Groups | Computer Science | 1,000 |
| CSC251HISD01 | Ryan Adamson | Los Alamos National Laboratory (LANL) | 2.4.4.01 HISD01-Software Integration | Computer Science | 1,000 |
| CSC251HISD01 | Ryan Adamson | Los Alamos National Laboratory (LANL) | 2.4.4.01 HISD01-Software Integration | Computer Science | 1,000 |
| CSC251HISD01 | Ryan Adamson | Los Alamos National Laboratory (LANL) | 2.4.4.01 HISD01-Software Integration | Computer Science | 1,000 |
| CSC251HISD01 | Ryan Adamson | Los Alamos National Laboratory (LANL) | 2.4.4.01 HISD01-Software Integration | Computer Science | 1,000 |
| CSCSTDT12345 | Patrick McCormick | Los Alamos National Laboratory (LANL) | 2.3.2.12 Flang: open-source Fortran front end for the LLVM infrastructure | Computer Science | 1,000 |
| CSCSTDT12345 | Patrick McCormick | Los Alamos National Laboratory (LANL) | 2.3.2.12 Flang: open-source Fortran front end for the LLVM infrastructure | Computer Science | 1,000 |
| CTOP | Devesh Tiwari | Northeastern University | Cost-Efficient and Throughput-Oriented Power Capping on Production HPC Systems | Computer Science | 11,469 |
| CTOP | Devesh Tiwari | Northeastern University | Cost-Efficient and Throughput-Oriented Power Capping on Production HPC Systems | Computer Science | 31,250 |
| CTSegNet | Christopher Powell | Argonne National Laboratory | AI-based segmentation of X-ray Tomographic Data | Energy Technologies | 781 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| Cu_CO2_ethanol | Ping Liu | Brookhaven National Laboratory (BNL) | Theoretical Study of Cu-Based Catalysts Ethanol Synthesis | Chemistry | 3,025 |
| darkskyml_aesp | Salman Habib | Argonne National Laboratory | Dark Sky Mining | Physics | 14,734 |
| darkskyml_aesp | Salman Habib | Argonne National Laboratory | Dark Sky Mining | Physics | 0 |
| darkskyml_aesp | Salman Habib | Argonne National Laboratory | Dark Sky Mining | Physics | 62,500 |
| datascience | Venkatram Vishwanath | Argonne National Laboratory | ALCF Data Science and Workflows Allocation | Internal | 78,231 |
| DDES_JET | Roberto Paoli | University of Illinois at Chicago | DDES of compressible jet flows with OpenFoam | Engineering | 31,250 |
| DDES_JET | Roberto Paoli | University of Illinois at Chicago | DDES of compressible jet flows with OpenFoam | Engineering | 14,754 |
| DDICF-Dev | Duc Minh Cao | University of Rochester | Direct-Drive Inertial Confinement Fusion Code Porting and Proposal Preparation | Fusion Energy | 390 |
| DDICF-Dev | Duc Minh Cao | University of Rochester | Direct-Drive Inertial Confinement Fusion Code Porting and Proposal Preparation | Fusion Energy | 390 |
| DDICF-Dev | Duc Minh Cao | University of Rochester | Direct-Drive Inertial Confinement Fusion Code Porting and Proposal Preparation | Fusion Energy | 1,562 |
| deMon2k | Andreas Koster | Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV) | First-Principle Simulation of the Melting of Aluminum Nanoclusters | Chemistry | 15,625 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| DetonationEngines | Venkatramanan Raman | University of Michigan | High-fidelity Simulations of Detonation Engines | Engineering | 7,812 |
| 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 | 16,157 |
| 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 | 31,250 |
| 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 | 15,711 |
| Diagnostics-CNN | Lander Ibarra | Argonne National Laboratory | Inverse Problem CNN Computation for Diagnostics | Energy Technologies | 189 |
| dist_relational_alg | Sidharth Kumar | The University of Alabama at Birmingham | Distributed relational algebra at scale | Computer Science | 2,812 |
| dist_relational_alg | Sidharth Kumar | The University of Alabama at Birmingham | Distributed relational algebra at scale | Computer Science | 2,812 |
| DLHMC | James Clifton Osborn | Argonne National Laboratory | Deep Learning HMC | Physics | 789 |
| DLHMC | James Clifton Osborn | Argonne National Laboratory | Deep Learning HMC | Physics | 312 |
| DLILT | Krishnan Raghavan | Argonne National Laboratory | Deep learning for inverse Laplace transform | Computer Science | 15,615 |
| DLILT | Krishnan Raghavan | Argonne National Laboratory | Deep learning for inverse Laplace transform | Computer Science | 10,084 |
| dl_am | Satish Karra | Los Alamos National Laboratory (LANL) | Scalable Deep Learning Workflow for Additive Manufacturing | Engineering | 1,373 |

| 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 | 837 |
| DMplex_Petsc | Oana Marin | Argonne National Laboratory | Unstructured domain support at scale in PETSc | Engineering | 7 |
| DNA-silver | Hannu Hakkinen | University of California-Irvine | Computational Investigations of Optical and Chiroptical Properties of DNA-Stabilized Silver Nanoclusters | Chemistry | 31,250 |
| DNS3D | Ramesh Balakrishnan | Argonne National Laboratory | Direct Numerical Simulation of Three Dimensional Turbulence | Engineering | 521 |
| DNS3D | Ramesh Balakrishnan | Argonne National Laboratory | Direct Numerical Simulation of Three Dimensional Turbulence | Engineering | 78,125 |
| DNS3D | Ramesh Balakrishnan | Argonne National Laboratory | Direct Numerical Simulation of Three Dimensional Turbulence | Engineering | 15,625 |
| DNSforLIS | Antonino Ferrante | University of Washington | DNS investigation for the law of incipient separation | Engineering | 23,501 |
| Drug_FEP_Data | Wei Jiang | Argonne National Laboratory | Machine elarning of drug binding and toxicity based on high throughput free energy computations | Biological Sciences | 78,125 |
| DynamicCS | Jonathan Tyler Schwartz, Huihuo Zheng | Argonne National Laboratory | Dynamic Compressed Sensing for Real-time Tomographic Reconstruction | Materials Science | 9,375 |
| DynamicCS | Jonathan Tyler Schwartz, Huihuo Zheng | Argonne National Laboratory | Dynamic Compressed Sensing for Real-time Tomographic Reconstruction | Materials Science | 5,575 |
| DynStall | Anupam Sharma | Iowa State University (ISU) | Analysis and Mitigation of Dyna mic Stall in Energy Machines | Engineering | 1,579 |
| Dynstall_DD_1 | Anupam Sharma | Iowa State University (ISU) | Characterization of Dynamic Stall Onset | Engineering | 718 |
| ecp-ci | Mark R Fahey | Argonne National Laboratory | ECP Operations Testing | Computer Science | 78 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| ECP_SDK | Sameer Suresh Shende | University of Oregon | Deploying the ECP SDK software stack at ALCF | Computer Science | 1,040 |
| EE-ECP | Xingfu Wu, Valerie Taylor | Argonne National Laboratory | Energy efficient tradeoff among execution time, power, and resilience of two ECP applications | Computer Science | 10,000 |
| EE-ECP | Xingfu Wu, Valerie Taylor | Argonne National Laboratory | Energy efficient tradeoff among execution time, power, and resilience of two ECP applications | Computer Science | 4,000 |
| EE-ECP | Xingfu Wu, Valerie Taylor | Argonne National Laboratory | Energy efficient tradeoff among execution time, power, and resilience of two ECP applications | Computer Science | 15,625 |
| EE-ECP | Xingfu Wu, Valerie Taylor | Argonne National Laboratory | Energy efficient tradeoff among execution time, power, and resilience of two ECP applications | Computer Science | 14,435 |
| EIC | Whitney Richard Armstrong | Argonne National Laboratory | EIC Detector Simulations | Physics | 7,812 |
| ELSI_dev | Alvaro Vazquez Mayagoitia | Duke University | ELSI Developers | Computer Science | 15 |
| ELSI_dev | Alvaro Vazquez Mayagoitia | Duke University | ELSI Developers | Computer Science | 169 |
| EngineDNS | Christos Frouzakis | Eidgenössische Technische Hochschule Zürich (ETH Zurich) | Towards reactive DNS in complex internal combustion engine geometries | Engineering | 17,908 |
| EPW_Scaling_Tests | Feliciano Giustino | The University of Texas at Austin | Engineering the electron-phonon interaction for next generation materials design | Materials Science | 781 |
| extrap-noise-ai | Marcus Ritter | Technische Universität Darmstadt | Noise Resilient Performance Modeling with Deep Learning | Computer Science | 2,352 |
| FDTD_Cancer_2a | Allen Taflove | Northwestern University | Computational Physical Genomics: Exploring Potential Novel Cancer Therapies | Biological Sciences | 5,445 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| 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 | 9,516 |
| 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 | 13,671 |
| FireballPlasma | Deborah Levin | University of Illinois at Urbana-Champaign | "Validated Particle Agglomeration Models for Fireball Plasma and Fallout", Defense Threat Reduction Agency, Project No. AWARD# HDTRA11710 | Engineering | 781 |
| FTI_Tribo_AS_DD | Matthias Baldofski | Aalto University | Benchmarking atomistic simulations of tribological systems | Materials Science | 11,914 |
| fusiondl_aesp | William Tang | Princeton University | Accelerated Deep Learning Discovery in Fusion Energy Science | Fusion Energy | 7,812 |
| fusiondl_aesp | William Tang | Princeton University | Accelerated Deep Learning Discovery in Fusion Energy Science | Fusion Energy | 3,683 |
| fusiondl_aesp | William Tang | Princeton University | Accelerated Deep Learning Discovery in Fusion Energy Science | Fusion Energy | 0 |
| GA-bfg3d | Mark Kostuk | General Atomics | General Atomics Next Generation Skew-Symmetric Fluid Solver | Physics | 11,718 |
| GA-bfg3d | Mark Kostuk | General Atomics | General Atomics Next Generation Skew-Symmetric Fluid Solver | Physics | 3,993 |
| GA-bfg3d | Mark Kostuk | General Atomics | General Atomics Next Generation Skew-Symmetric Fluid Solver | Physics | 11,718 |
| gas-turbine | Pinaki Pal | Argonne National Laboratory | Deep Learning-Augmented Flow Solver to Improve the Design of Gas-Turbine Engines | Engineering | 23,191 |
| gasturbine_opt | Prithwish Kundu | Argonne National Laboratory | Optimization of Ultra-Clean Transient Turbine Combustor | Energy Technologies | 31,250 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| GFDL_Ensemble | Thomas Edward Robinson | Geophysical Fluid Dynamics Laboratory (GFDL) | Ensemble-based Regression Tests for Climate Model Reproducibility | Earth Science | 5,013 |
| GNPMem | Tarak K Patra | Indian Institute of Technology Madras | Computational design of polymer grafted nanoparticle membrane | Materials Science | 6,841 |
| GNPMem | Tarak K Patra | Indian Institute of Technology Madras | Computational design of polymer grafted nanoparticle membrane | Materials Science | 15,625 |
| GrainBoundaries | Wissam A Saidi | University of Pittsburgh | Structure and Properties of Grain Boundaries in Materials for Energy Applications | Materials Science | 31,250 |
| GrainBoundaries | Wissam A Saidi | University of Pittsburgh | Structure and Properties of Grain Boundaries in Materials for Energy Applications | Materials Science | 16,447 |
| GraphNeuralNetwork | Qi Yu | University of California-San Diego | Accelerating Traffic Simulation via Graph Neural Networks | Computer Science | 30,625 |
| HACC_aesp | Katrin Heitmann | Argonne National Laboratory | Extreme-Scale Cosmological Hydrodynamics | Physics | 0 |
| HACC_aesp | Katrin Heitmann | Argonne National Laboratory | Extreme-Scale Cosmological Hydrodynamics | Physics | 3,118 |
| HACC_aesp | Katrin Heitmann | Argonne National Laboratory | Extreme-Scale Cosmological Hydrodynamics | Physics | 31,250 |
| HED_Flows | Hussein Aluie | University of Rochester | Scale-Aware Modeling of Instabilities and Mixing in HED Flows | Fusion Energy | 9,503 |
| HED_Flows | Hussein Aluie | University of Rochester | Scale-Aware Modeling of Instabilities and Mixing in HED Flows | Fusion Energy | 15,625 |
| HEPCloud-FNAL | Burt Holzman | Fermi National Accelerator Laboratory (Fermilab) | High Energy Physics Computing for Fermilab experiments via HEPCloud | Physics | 1,813 |
| HEPCloud-FNAL | Burt Holzman | Fermi National Accelerator Laboratory (Fermilab) | High Energy Physics Computing for Fermilab experiments via HEPCloud | Physics | 7,812 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| HEPCloud-FNAL | Burt Holzman | Fermi National Accelerator Laboratory (Fermilab) | High Energy Physics Computing for Fermilab experiments via HEPCloud | Physics | 2,554 |
| Hephaestus | Andrey Beresnyak | Naval Research Laboratory | Hephaestus heterogeneous mode | Physics | 15,625 |
| HEP_on_HPC | Jim B Kowalkowski, Marc Francis Paterno | Fermi National Accelerator Laboratory (Fermilab) | HEP analysis workflows on HPC | Physics | 3,125 |
| HEP_on_HPC | Jim B Kowalkowski, Marc Francis Paterno | Fermi National Accelerator Laboratory (Fermilab) | HEP analysis workflows on HPC | Physics | 6,250 |
| HEP_on_HPC | Jim B Kowalkowski, Marc Francis Paterno | Fermi National Accelerator Laboratory (Fermilab) | HEP analysis workflows on HPC | Physics | 5,000 |
| hifiturbfsi | Ivan Bermejo Moreno | University of Southern California (USC) | High-fidelity simulation of supersonic turbulent flow-structure interaction and mixing | Engineering | 26,562 |
| HighMachTurbulence | Sanjiva K Lele | Stanford University | Simulations of Turbulence at High Compressibility | Engineering | 7,798 |
| HighMachTurbulence | Sanjiva K Lele | Stanford University | Simulations of Turbulence at High Compressibility | Engineering | 31,250 |
| HighMachTurbulence | Sanjiva K Lele | Stanford University | Simulations of Turbulence at High Compressibility | Engineering | 15,625 |
| HighReyTurb_PostProc | Robert D. Moser, Myoungkyu Lee | The University of Texas at Austin | Data analysis of turbulent Channel Flow at High Reynolds number | Engineering | 184 |
| HighReyTurb_PostProc | Robert D. Moser, Myoungkyu Lee | The University of Texas at Austin | Data analysis of turbulent Channel Flow at High Reynolds number | Engineering | 10,407 |
| HiMB_Beamline | Eremey Vladimirovich Valetov | Michigan State University | Development and Optimization of a Novel High-Intensity Muon Beamline | Physics | 10,355 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| 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 | 16,085 |
| 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 | 25,390 |
| HNPballistics | Sinan Keten | Northwestern University | Engineering Nanocellulose based Hairy Nanoparticle Assemblies for High Ballistic Impact Performance | Engineering | 7,852 |
| HNPballistics | Sinan Keten | Northwestern University | Engineering Nanocellulose based Hairy Nanoparticle Assemblies for High Ballistic Impact Performance | Engineering | 15,625 |
| hpcbdsm | Tanwi Mallick | Argonne National Laboratory | High-Performance Computing and Big Data Solutions for Mobility Design and Planning | Computer Science | 15,625 |
| HybridTreecodePME | Henry Asamoah Boateng | Bates College | A hybrid tree code particle mesh Ewald method for electrostatics | Mathematics | 2,000 |
| hydro | Eugene Yan | Argonne National Laboratory | Deep learning solution for flood forecasting | Earth Science | 7,254 |
| IBR2M | Alberto Talamo | Argonne National Laboratory | Transient Simulation of IBR2M | Nuclear Energy | 7,913 |
| ICEPIC | Andrey Diomidovich Andreev | University of New Mexico | High-fidelity particle-in-cell simulations of high-power microwave and millimeter-wavelength vacuum electronic devices | Physics | 265 |
| ICE_InSitu | Muhsin Mohammed Ameen, Saumil Sudhir Patel | Argonne National Laboratory | Towards Exascale Internal Combustion Engine Simulations with In-Situ Analysis | Chemistry | 506 |
| ICE_InSitu | Muhsin Mohammed Ameen, Saumil Sudhir Patel | Argonne National Laboratory | Towards Exascale Internal Combustion Engine Simulations with In-Situ Analysis | Chemistry | 58,739 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| IL_elec | Zhengcheng Zhang | Argonne National Laboratory | Ionic liquid as a potential electrolyte of high performance lithium ion battery | Chemistry | 15,625 |
| IL_elec | Zhengcheng Zhang | Argonne National Laboratory | Ionic liquid as a potential electrolyte of high performance lithium ion battery | Chemistry | 15,798 |
| IL_elec | Zhengcheng Zhang | Argonne National Laboratory | Ionic liquid as a potential electrolyte of high performance lithium ion battery | Chemistry | 62,500 |
| ImageNanoX_DD | Phay J Ho | Argonne National Laboratory | X-ray Imaging of Transient Structure and Dynamics of Nanoparticles | Physics | 10,192 |
| ImageNanoX_DD | Phay J Ho | Argonne National Laboratory | X-ray Imaging of Transient Structure and Dynamics of Nanoparticles | Physics | 8,123 |
| INSPIRE | Peter Coveney | University College London (UCL) | INtegrated and Scalable PredictIon of REsistanace | Chemistry | 15,625 |
| Intel | Kalyan Kumaran, Scott Parker, Timothy Joe Williams, Venkatram Vishwanath | Argonne National Laboratory | Intel employees in support of Theta | Internal | 78,231 |
| interconnect_bench | Devi Sudheer Kumar Chunduri | Argonne National Laboratory | Collaboration with Cray on Interconnect related studies | Computer Science | 1,443 |
| inversedesign_oerorr | Rafael Gomez-Bombarelli | Massachusetts Institute of Technology (MIT) | Inverse design of oxide catalysts for OER/ORR | Materials Science | 1,468 |
| iRF-LOOP | Daniel A Jacobson | Oak Ridge National Laboratory (ORNL) | Explainable-AI Applications for Systems Biology | Biological Sciences | 7,812 |
| iRF-LOOP | Daniel A Jacobson | Oak Ridge National Laboratory (ORNL) | Explainable-AI Applications for Systems Biology | Biological Sciences | 4,531 |
| iRF-LOOP | Daniel A Jacobson | Oak Ridge National Laboratory (ORNL) | Explainable-AI Applications for Systems Biology | Biological Sciences | 4,803 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| Job_Interference | Zhiling Lan | Illinois Institute of Technology (IIT) | Workload Interference Analysis on Theta | Computer Science | 7,631 |
| Job_Interference | Zhiling Lan | Illinois Institute of Technology (IIT) | Workload Interference Analysis on Theta | Computer Science | 7,367 |
| Large-ScalePF | Jiamian Hu | University of Wisconsin-Madison | Phase-field Simulations and Atomistic Modeling of Large-Scale Spin-Lattice Dynamics | Materials Science | 12,197 |
| large3dxrayADSP | Chris Johnson Jacobsen | Northwestern University | X-ray microscopy of extended 3D objects: scaling towards the future | Biological Sciences | 187,157 |
| LatticeQCD_aesp | Paul Mackenzie, Norman Howard Christ | Fermi National Accelerator Laboratory (Fermilab) | Lattice Quantum Chromodynamics Calculations for Particle and Nuclear Physics | Physics | 156,250 |
| LatticeQCD_aesp | Paul Mackenzie, Norman Howard Christ | Fermi National Accelerator Laboratory (Fermilab) | Lattice Quantum Chromodynamics Calculations for Particle and Nuclear Physics | Physics | 0 |
| LatticeQCD_aesp | Paul Mackenzie, Norman Howard Christ | Fermi National Accelerator Laboratory (Fermilab) | Lattice Quantum Chromodynamics Calculations for Particle and Nuclear Physics | Physics | 3,118 |
| LES_CommTransAir | Parviz Moin | Stanford University | Large-Eddy Simulation of a Commercial Transport Aircraft Model | Engineering | 1,675 |
| LES_CommTransAir_DD | Parviz Moin | Stanford University | Large Eddy Simulation of Commercial Transport Aircraft | Engineering | 31,250 |
| LIGHTCONTROL | Sandra Gail Biedron | University of New Mexico | Light sources and their control using AI techniques | Physics | 1,562 |
| LIGHTCONTROL | Sandra Gail Biedron | University of New Mexico | Light sources and their control using AI techniques | Physics | 1,977 |
| LLVM | Hal Joshua Finkel | Argonne National Laboratory | LLVM Compiler Tools for ALCF Systems | Internal | 1,562 |
| LQCDdev | James Clifton Osborn | Argonne National Laboratory | Lattice QCD development | Physics | 2,042 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| lqcdml_aesp | William Detmold | Massachusetts Institute of Technology (MIT) | Machine Learning for Lattice Quantum Chromodynamics | Physics | 156,250 |
| lqcdml_aesp | William Detmold | Massachusetts Institute of Technology (MIT) | Machine Learning for Lattice Quantum Chromodynamics | Physics | 29,468 |
| lqcdml_aesp | William Detmold | Massachusetts Institute of Technology (MIT) | Machine Learning for Lattice Quantum Chromodynamics | Physics | 0 |
| LQCD_VeloC | Chulwoo Jung | Brookhaven National Laboratory (BNL) | Lattice QCD with VeloC | Physics | 781 |
| 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 | 25,000 |
| LTP-Opt | Vyaas Gururajan | Argonne National Laboratory | Low Temperature Plasma Optimization | Chemistry | 16,634 |
| magstructsADSP | Trevor David Rhone, Prof. Kaxiras | Rensselaer Polytechnic Institute (RPI) | Machine learning magnetic properties of van der Waals heterostructures | Materials Science | 304,130 |
| Maintenance | Mark R Fahey | Argonne National Laboratory | LCF Operations System Maintenance | Internal | 78,231 |
| marine-twin | Flavio Dal Forno Chuahy | Oak Ridge National Laboratory (ORNL) | Marine Digital-Twin Full-Scale Simulation | Energy Technologies | 16,347 |
| matml_aesp | Noa Marom | Carnegie Mellon University | Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials | Materials Science | 140,625 |
| matml_aesp | Noa Marom | Carnegie Mellon University | Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials | Materials Science | 38,085 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| matml_aesp | Noa Marom | Carnegie Mellon University | Many-Body Perturbation Theory Meets Machine Learning to Discover Singlet Fission Materials | Materials Science | 0 |
| MCPSI_DD | Ahren Ward Jasper | Argonne National Laboratory | High temperature kinetics and dynamics of fluxional molecules | Chemistry | 2,925 |
| mep | Whitney Richard Armstrong | Argonne National Laboratory | Medium Energy Physics Data Processing | Physics | 7,623 |
| MERF_Data_Layer | Santanu Chaudhuri, Jakob Ryan Elias | Argonne National Laboratory | MERF Data Layer | Materials Science | 8,033 |
| MERF_Data_Layer | Santanu Chaudhuri, Jakob Ryan Elias | Argonne National Laboratory | MERF Data Layer | Materials Science | 15,625 |
| mesoscale-QM | Long-Qing Chen | The Pennsylvania State University (Penn State/PSU) | Computational Mesoscale Science and Open Software for Quantum Materials | Materials Science | 78 |
| mesoscale-QM | Long-Qing Chen | The Pennsylvania State University (Penn State/PSU) | Computational Mesoscale Science and Open Software for Quantum Materials | Materials Science | 156 |
| metastable | Subramanian Sankaranarayanan | Argonne National Laboratory | Metastable phase diagram of material | Materials Science | 31,250 |
| metastable | Subramanian Sankaranarayanan | Argonne National Laboratory | Metastable phase diagram of material | Materials Science | 12,374 |
| MI2Dmaterials | Trevor David Rhone | Rensselaer Polytechnic Institute (RPI) | Materials informatics study of two-dimensional magnetic materials and their heterostructures | Materials Science | 9,772 |
| MICCoM-train | Marco Govoni | Argonne National Laboratory | Training for MICCoM codes | Materials Science | 1,406 |
| microwave_catalysis | fanglin che | University of Massachusetts-Lowell | Characterizing Non-Thermal Effects of Microwave Accelerated Heterogeneous Catalysis | Chemistry | 5,533 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| MIT-MSEI-Partnership | Santanu Chaudhuri | Argonne National Laboratory | MIT-Argonne Manufacturing Collaboration | Materials Science | 8,494 |
| ML-GA | Pinaki Pal | Argonne National Laboratory | Machine Learning Workflow Tools for Rapid Optimization of Product Designs and Manufacturing Processes | Energy Technologies | 27,212 |
| ML4MPF | Gina Maureen Sforzo | Argonne National Laboratory | Enabling predictive simulations of reacting multiphase flows via data-driven emulation | Engineering | 22,966 |
| MLPerfHPC | Murali Krishna Emani | Argonne National Laboratory | Evaluation of Scientific ML Benchmarks on HPC Systems | Computer Science | 14,077 |
| MLPerfHPC | Murali Krishna Emani | Argonne National Laboratory | Evaluation of Scientific ML Benchmarks on HPC Systems | Computer Science | 31,250 |
| mmaADSP | Eliu Antonio Huerta Escudero | University of Illinois at Urbana-Champaign | Deep Learning at Scale for Multimessenger Astrophysics through the NCSA-Argonne Collaboration | Physics | 104,356 |
| MOAB_App | Vijay Subramaniam Mahadevan | Argonne National Laboratory | MOAB Algorithmic Performance Portability | Mathematics | 8,140 |
| ModelingSoftMatter | Zhangli Peng | University of Illinois at Chicago | Multiscale Modeling of Soft Materials by Unifying Shell Elements and Anisotropic Coarse-Grained Force Fields in LAMMPS | Biological Sciences | 9,805 |
| MoltenSalts | Nicholas Everett Jackson | University of Illinois at Urbana-Champaign | Automated Active Learning on ALCF for Machine Learning Forcefield Automation | Nuclear Energy | 17,203 |
| MPICH_MCS | Kenneth James Raffenetti, Pavan Balaji | Argonne National Laboratory | MPICH - A high performance and widely portable MPI implementation | Computer Science | 39,062 |
| MPICH_MCS | Kenneth James Raffenetti, Pavan Balaji | Argonne National Laboratory | MPICH - A high performance and widely portable MPI implementation | Computer Science | 13,125 |
| MPICH_MCS | Kenneth James Raffenetti, Pavan Balaji | Argonne National Laboratory | MPICH - A high performance and widely portable MPI implementation | Computer Science | 15,798 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| MPI_Aurora_Intel | Devi Sudheer Kumar Chunduri | Argonne National Laboratory | MPI development for Aurora | Computer Science | 1,529 |
| MSEMNeuro | Thomas David Uram | Argonne National Laboratory | Reconstruction of neuronal connectivity from MSEM imaged tissue | Biological Sciences | 7,538 |
| MSEMNeuro | Thomas David Uram | Argonne National Laboratory | Reconstruction of neuronal connectivity from MSEM imaged tissue | Biological Sciences | 15,625 |
| Mu2e_HEPCloud | Iuri Artur Oksuzian | Argonne National Laboratory | Mu2e simulations through the HEPCloud Project | Physics | 10,666 |
| multimode_comb | Pinaki Pal | Argonne National Laboratory | High-Fidelity CFD Simulations of Multi-Mode Combustion | Energy Technologies | 1,520 |
| multimode_comb | Pinaki Pal | Argonne National Laboratory | High-Fidelity CFD Simulations of Multi-Mode Combustion | Energy Technologies | 46,875 |
| multiphysics_aesp | Amanda Randles | Duke University | Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations | Engineering | 156,250 |
| multiphysics_aesp | Amanda Randles | Duke University | Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations | Engineering | 22,450 |
| multiphysics_aesp | Amanda Randles | Duke University | Extreme-scale In Situ Visualization and Analysis of Fluid-Structure-Interaction Simulations | Engineering | 0 |
| NAMD_aesp | Benoit Roux, James C. Phillips | The University of Chicago (UChicago) | Free Energy Landscapes of Membrane Transport Proteins | Biological Sciences | 0 |
| NAMD_aesp | Benoit Roux, James C. Phillips | The University of Chicago (UChicago) | Free Energy Landscapes of Membrane Transport Proteins | Biological Sciences | 3,118 |
| NAMD_aesp | Benoit Roux, James C. Phillips | The University of Chicago (UChicago) | Free Energy Landscapes of Membrane Transport Proteins | Biological Sciences | 156,250 |
| NanoCluster_DD | Anastassia Alexandrova | University of California-Los Angeles | Heterogeneous Catalysis on Dynamic Nanocluster-Decorated Interfaces | Chemistry | 31,250 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| NanoCluster_DD | Anastassia Alexandrova | University of California-Los Angeles | Heterogeneous Catalysis on Dynamic Nanocluster-Decorated Interfaces | Chemistry | 46,875 |
| NanoCluster_DD | Anastassia Alexandrova | University of California-Los Angeles | Heterogeneous Catalysis on Dynamic Nanocluster-Decorated Interfaces | Chemistry | 31,250 |
| NanoCluster_DD | Anastassia Alexandrova | University of California-Los Angeles | Heterogeneous Catalysis on Dynamic Nanocluster-Decorated Interfaces | Chemistry | 23,437 |
| NAQMC_RMD_aesp | Aiichiro Nakano | University of Southern California (USC) | Metascalable Layered Materials Genome | Materials Science | 78,125 |
| NAQMC_RMD_aesp | Aiichiro Nakano | University of Southern California (USC) | Metascalable Layered Materials Genome | Materials Science | 0 |
| NAQMC_RMD_aesp | Aiichiro Nakano | University of Southern California (USC) | Metascalable Layered Materials Genome | Materials Science | 3,118 |
| 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 | 1,108 |
| 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 | 31,250 |
| networkbench | Devi Sudheer Kumar Chunduri, Elise Jennings, Kevin Harms, Misbah Mubarak | Argonne National Laboratory | network benchmarking and modeling | Computer Science | 17,230 |
| networkbench | Devi Sudheer Kumar Chunduri, Elise Jennings, Kevin Harms, Misbah Mubarak | Argonne National Laboratory | network benchmarking and modeling | Computer Science | 7,812 |

| 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 | 1,443 |
| NetworkScience | Mehmet H Gunes | University of Nevada, Reno | Analysis of complex networks on HPC: A case study of the Internet | Computer Science | 15,625 |
| neutrinoADSP | Andrzej Michał Szelc, Corey J Adams, Diego Garcia-Gamez | The University of Manchester | Developing High Performance Computing Applications for Liquid Argon Neutrino Detectors | Physics | 35,092 |
| neutrinoADSP | Andrzej Michał Szelc, Corey J Adams, Diego Garcia-Gamez | The University of Manchester | Developing High Performance Computing Applications for Liquid Argon Neutrino Detectors | Physics | 9,026 |
| Next | Paola Ferrario | Donostia International Physics Center | NEXT | Physics | 10,689 |
| NWChemEx_aesp | Theresa Windus, Alvaro Vazquez Mayagoitia | Pacific Northwest National Laboratory (PNNL) | NWChemEx: Tackling Chemical, Materials & Biochemical Challenges in the Exascale Era | Chemistry | 15,625 |
| NWChemEx_aesp | Theresa Windus, Alvaro Vazquez Mayagoitia | Pacific Northwest National Laboratory (PNNL) | NWChemEx: Tackling Chemical, Materials & Biochemical Challenges in the Exascale Era | Chemistry | 0 |
| NWChemEx_aesp | Theresa Windus, Alvaro Vazquez Mayagoitia | Pacific Northwest National Laboratory (PNNL) | NWChemEx: Tackling Chemical, Materials & Biochemical Challenges in the Exascale Era | Chemistry | 3,118 |
| OctopusNeuro | Thomas David Uram | Argonne National Laboratory | Octopus Neuroscience | Biological Sciences | 6,250 |
| OctopusNeuro | Thomas David Uram | Argonne National Laboratory | Octopus Neuroscience | Biological Sciences | 9,607 |
| OnSite_Workshop | Richard Michael Coffey, Avanthi Madduri | Argonne National Laboratory | Project for ALCF Workshops Onsite | Internal | 0 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| Operations | Mark R Fahey | Argonne National Laboratory | Systems administration tasks | Internal | 78,231 |
| OptADDN | Sandeep Madireddy | Argonne National Laboratory | Optimal Architecture discovery for deep probabilistic models and neuromorphic systems | Computer Science | 30,000 |
| pearlnecklace | Michael David Schneider | Lawrence Livermore National Laboratory (LLNL) | Pearl Necklace - LSST project exploration | Physics | 15,625 |
| PENA | Gabriel Serge Staffelbach | Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS) | Performance evaluation of new algorithms in AVBP | Chemistry | 25,277 |
| Performance | Scott Parker, Raymond M. Loy | Argonne National Laboratory | Performance | Internal | 78,231 |
| perf_research | Devi Sudheer Kumar Chunduri | Argonne National Laboratory | Performance group external facing research | Computer Science | 1,562 |
| perf_research | Devi Sudheer Kumar Chunduri | Argonne National Laboratory | Performance group external facing research | Computer Science | 7,812 |
| perf_research | Devi Sudheer Kumar Chunduri | Argonne National Laboratory | Performance group external facing research | Computer Science | 1,443 |
| PerovskiteMachine | Volker Wolfgang Blum | Duke University | High-Precision Dynamical Properties of Complex Perovskites by Interpolative Machine Learning | Materials Science | 20,312 |
| PHASTA_aesp | Kenneth Edward Jansen | University of Colorado-Boulder | Extreme Scale Unstructured Adaptive CFD: From Multiphase Flow to Aerodynamic Flow Control | Engineering | 117,187 |
| PHASTA_aesp | Kenneth Edward Jansen | University of Colorado-Boulder | Extreme Scale Unstructured Adaptive CFD: From Multiphase Flow to Aerodynamic Flow Control | Engineering | 3,118 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| PHASTA_aesp | Kenneth Edward Jansen | University of Colorado-Boulder | Extreme Scale Unstructured Adaptive CFD: From Multiphase Flow to Aerodynamic Flow Control | Engineering | 0 |
| PHASTA_NCSU | Igor A Bolotnov | North Carolina State University (NCSU) | Multiphase Simulations of Nuclear Reactor Thermal Hydraulics | Engineering | 19,214 |
| PHASTA_NCSU | Igor A Bolotnov | North Carolina State University (NCSU) | Multiphase Simulations of Nuclear Reactor Thermal Hydraulics | Engineering | 10,937 |
| poly-ion-dd | Juan Pablo | The University of Chicago (UChicago) | Ion Conduction Through Polymeric Interfaces | Chemistry | 710 |
| PPE-CV-NTM | Sandra Jean Bittner | | Computational Study and Visualization Models for Non-Traditional Materials for Personal Protective Equipment | Computer Science | 100 |
| PrincetonConnectome | Sebastian Seung | Princeton University | Petascale neural circuit reconstruction | Biological Sciences | 7 |
| psr001 | Ronald Otis Grover | General Motors Company | Electric Motor Thermal Management Analysis | Engineering | 53,655 |
| PtychoHPC | Junjing Deng | Argonne National Laboratory | High performance computation for ptychography reconstruction | Materials Science | 8,171 |
| ptychotomo | Yi Jiang | Argonne National Laboratory | High-throughput ptycho-tomography | Physics | 6,250 |
| PUR-IRL | Nicholas Lee-Ping Chia | Mayo Clinic-Minnesota | Inferring the Reward Function of Cancer | Biological Sciences | 148 |
| Q-Pix | Jonathan Asaadi | The University of Texas at Arlington | QPix: Achieving kiloton scale pixelated readout for Liquid Argon Time Projection Chambers | Physics | 624 |
| Q-Pix | Jonathan Asaadi | The University of Texas at Arlington | QPix: Achieving kiloton scale pixelated readout for Liquid Argon Time Projection Chambers | Physics | 1,562 |
| QCSim | Hal Joshua Finkel | Argonne National Laboratory | Simulating Realistic Quantum Computers | Computer Science | 23,437 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| QCSim | Hal Joshua Finkel | Argonne National Laboratory | Simulating Realistic Quantum Computers | Computer Science | 23,437 |
| QMC-EFMO | Federico Zahariev | Iowa State University (ISU) | Linearly scaling QMC through EFMO fragmentation | Chemistry | 15,625 |
| QMCPACK_aesp | Anouar Benali | Argonne National Laboratory | Extending Moore's Law computing with Quantum Monte Carlo | Materials Science | 3,118 |
| QMCPACK_aesp | Anouar Benali | Argonne National Laboratory | Extending Moore's Law computing with Quantum Monte Carlo | Materials Science | 0 |
| QMCPACK_aesp | Anouar Benali | Argonne National Laboratory | Extending Moore's Law computing with Quantum Monte Carlo | Materials Science | 78,125 |
| QSim | Yuri Alexeev | Argonne National Laboratory | Quantum Simulations | Computer Science | 10,851 |
| QuantumDS | Alvaro Vazquez Mayagoitia | Argonne National Laboratory | Quantum mechanics and Data Science | Chemistry | 15,636 |
| radix-io | Philip Hutchinson Carns | Argonne National Laboratory | System software to enable data-intensive science | Computer Science | 6,189 |
| RaptorX | Jinbo Xu | Toyota Technological Institute at Chicago (TTIC) | Protein Folding through Deep Learning and Energy Minimization | Biological Sciences | 10,958 |
| RCM_4km | Jiali Wang | Argonne National Laboratory | Generation of a next level dataset for regional scale climate modeling: convective resolving spatial scales | Earth Science | 14,754 |
| Regional-climate | Veerabhadra Kotamarthi | Argonne National Laboratory | Convective Resolved Regional Climate Simulations | Earth Science | 117,187 |
| REI_Flares | Marc Cremer | Reaction Engineering International | Leveraging the UCF for Simulation of Industrial Flares | Chemistry | 8,831 |
| REI_Flares | Marc Cremer | Reaction Engineering International | Leveraging the UCF for Simulation of Industrial Flares | Chemistry | 4,723 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| REI_Flares | Marc Cremer | Reaction Engineering International | Leveraging the UCF for Simulation of Industrial Flares | Chemistry | 15,625 |
| ReservoirQuality | James Edward Guilkey | The University of Utah | Numerical prediction of sandstone reservoir quality | Earth Science | 21,484 |
| RL-fold | Arvind Ramanathan | Argonne National Laboratory | Targeting intrinsically disordered proteins using artificial intelligence driven molecular simulations | Biological Sciences | 23,050 |
| RL-fold | Arvind Ramanathan | Argonne National Laboratory | Targeting intrinsically disordered proteins using artificial intelligence driven molecular simulations | Biological Sciences | 4,231 |
| SAC | Shaama Mallikarjun Sharada | University of Southern California (USC) | Dynamical evolution, surface chemistry, and stability of atomically dispersed catalysts | Chemistry | 781 |
| scalablemoose | Fande Kong | Idaho National Laboratory (INL) | MOOSE scaling study | Nuclear Energy | 5,169 |
| SCRAMJET_VULCAN | Roberto Paoli | Argonne National Laboratory | Scramjet simulations on NASA VULCAN | Engineering | 15,623 |
| SCRAMJET_VULCAN | Roberto Paoli | Argonne National Laboratory | Scramjet simulations on NASA VULCAN | Engineering | 4,005 |
| SDL | Min Si | Argonne National Laboratory | Data analytics for HPC | Computer Science | 7,727 |
| SDL_Workshop | Raymond M. Loy | Argonne National Laboratory | ALCF Simulation, Data, and Learning Workshop | Training | 2,000 |
| SENSEI | Silvio Humberto Rafael Rizzi, Joseph Insley, Nicola Joy Ferrier, Venkatram Vishwanath | Argonne National Laboratory | Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery | Computer Science | 22,400 |
| SENSEI | Silvio Humberto Rafael Rizzi, Joseph Insley, Nicola Joy Ferrier, Venkatram Vishwanath | Argonne National Laboratory | Scalable Analysis Methods and In Situ Infrastructure for Extreme Scale Knowledge Discovery | Computer Science | 15,710 |

| 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 | 4,243 |
| SI_IBM | Saurabh Chawdhary | Argonne National Laboratory | Scalable, efficient sharp-interface immersed boundary method for fluid-structure interaction problems | Engineering | 3,590 |
| SI_IBM | Saurabh Chawdhary | Argonne National Laboratory | Scalable, efficient sharp-interface immersed boundary method for fluid-structure interaction problems | Engineering | 12,500 |
| SI_IBM | Saurabh Chawdhary | Argonne National Laboratory | Scalable, efficient sharp-interface immersed boundary method for fluid-structure interaction problems | Engineering | 3,089 |
| SolarWindowsADSP | Jacqueline Manina Cole | University of Cambridge | Data-Driven Molecular Engineering of Solar-powered Windows | Materials Science | 21,462 |
| SolarWindowsADSP | Jacqueline Manina Cole | University of Cambridge | Data-Driven Molecular Engineering of Solar-powered Windows | Materials Science | 19,061 |
| SolarWindowsADSP | Jacqueline Manina Cole | University of Cambridge | Data-Driven Molecular Engineering of Solar-powered Windows | Materials Science | 39,062 |
| SolarWindTurb | Jean Carlos Perez | Florida Institute of Technology | Understanding Solar Wind Turbulence in the inner Heliosphere | Physics | 15,625 |
| Solv_extraction | Lynda Soderholm | Argonne National Laboratory | Liquid-liquid extraction from solvent extraction | Chemistry | 31,250 |
| Solv_extraction | Lynda Soderholm | Argonne National Laboratory | Liquid-liquid extraction from solvent extraction | Chemistry | 31,250 |
| spentFuel | Angela Di Fulvio | University of Illinois at Urbana-Champaign | Cask Mis-loads Evaluation Techniques | Nuclear Energy | 16,088 |
| Stable-MABs | Deniz Cakir | University of North Dakota | The Stable MAB and MBenes for Spintronics and Battery Applications | Physics | 6,128 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| 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 | 7,812 |
| 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 | 0 |
| 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 | 3,118 |
| SweptWing_DD | Lian Duan | Missouri University of Science and Technology | Laminar-Turbulent Transition in Swept-Wing Boundary Layers | Engineering | 23,982 |
| TNContract | James Clifton Osborn | Argonne National Laboratory | Tensor Network Contractions for QIS | Physics | 16,262 |
| TNContract | James Clifton Osborn | Argonne National Laboratory | Tensor Network Contractions for QIS | Physics | 15,625 |
| Tools | Scott Parker | Argonne National Laboratory | ALCF Performance Tools | Internal | 15,775 |
| TotalView | Peter Michael Thompson, Raymond M. Loy | Rogue Wave Software, Inc. | TotalView Debugger on Blue Gene P | Internal | 649 |
| TotalView | Peter Michael Thompson, Raymond M. Loy | Rogue Wave Software, Inc. | TotalView Debugger on Blue Gene P | Internal | 643 |
| TurbLowBeta | Luca Franci | Queen Mary University of London | Hybrid simulations of kinetic plasma turbulence at low electron beta | Physics | 15,625 |
| TurbNet | Romit Maulik | Argonne National Laboratory | TurbNet: Scaleable physics-informed deep learning for turbulence model development | Engineering | 17,983 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| UA_TRANSITION | Hermann Fasel | The University of Arizona | Direct Numerical Simulations of Boundary-Layer Transition | Engineering | 7,812 |
| uboone | Bonnie Fleming | Fermi National Accelerator Laboratory (Fermilab) | MicroBooNE physics data processing | Physics | 14,644 |
| 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 | 3,118 |
| 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 | 0 |
| 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 | 125,000 |
| ultrafast | Jin Wang | Argonne National Laboratory | Developing High-Fidelity Simulation and Ultrafast X-Ray Imaging Tools for the APS-Upgrade | Engineering | 26,112 |
| ultrafast | Jin Wang | Argonne National Laboratory | Developing High-Fidelity Simulation and Ultrafast X-Ray Imaging Tools for the APS-Upgrade | Engineering | 19,214 |
| Ultrafast_X-ray | Jin Wang | Argonne National Laboratory | Ultrafast_X-ray | Engineering | 11,574 |
| 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 | 812 |
| VeloC | Bogdan Florin Nicolae | Argonne National Laboratory | VeloC: Very Low Overhead Checkpointing System | Computer Science | 1,562 |
| visualization | Joseph Insley, Michael E. Papka | Argonne National Laboratory | Visualization and Analysis Research and Development for ALCF | Internal | 15,646 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| Viz_Support | Joseph Insley, William Edward Allcock | Argonne National Laboratory | Visualization Support | Computer Science | 3,911 |
| VST-sode | Olaseni Sode | California State University (CSU), Northridge | Development of vibrational structure techniques for van der Waals and hydrogen-bonded complexes | Chemistry | 7,812 |
| wall_turb_dd | Ramesh Balakrishnan | Argonne National Laboratory | Wall Resolved Simulations of Canonical Wall Bounded Flows | Engineering | 31,250 |
| wall_turb_dd | Ramesh Balakrishnan | Argonne National Laboratory | Wall Resolved Simulations of Canonical Wall Bounded Flows | Engineering | 14,385 |
| wall_turb_dd | Ramesh Balakrishnan | Argonne National Laboratory | Wall Resolved Simulations of Canonical Wall Bounded Flows | Engineering | 15,625 |
| WaterHammer | Hong Zhang, Hong Zhang | Argonne National Laboratory | Water Hammer Simulation | Mathematics | 150 |
| 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 | 3,116 |
| WindScenarios | Ignas Vasilijus Satkauskas | National Renewable Energy Laboratory (NREL) | wind scenario generation for two-stage stochastic optimization | Energy Technologies | 5,227 |
| wisc_atlas | Sau Lan Yu Wu | University of Wisconsin-Madison | Apply quantum machine learing to High Energy Physics | Physics | 12,500 |
| WRLCMF | Krishnan Mahesh | University of Minnesota-Twin Cities | Wall-Resolved LES of complex maneuvering flows | Engineering | 14,111 |
| XFELO | Henry Freund | University of New Mexico | Simulation of x-ray FEL Oscillators | Physics | 11,881 |
| XGC_aesp | Choongseok Chang, Timothy Joe Williams | Princeton Plasma Physics Laboratory (PPPL) | High fidelity simulation of fusion reactor boundary plasmas | Fusion Energy | 156,250 |

| Project Name | PI Name | PI Institution | Project Title | Science Field (short) | Allocation Amount in Node-Hours |
|---|---|---|---|---|---|
| XGC_aesp | Choongseok Chang, Timothy Joe Williams | Princeton Plasma Physics Laboratory (PPPL) | High fidelity simulation of fusion reactor boundary plasmas | Fusion Energy | 0 |
| XGC_aesp | Choongseok Chang, Timothy Joe Williams | Princeton Plasma Physics Laboratory (PPPL) | High fidelity simulation of fusion reactor boundary plasmas | Fusion Energy | 3,118 |
| Xin-ZhongLiang | Xin-Zhong Liang, chao sun | University of Maryland | Xin-Zhong Liang | Earth Science | 31,422 |
| 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 | 523 |
| | | | | **Total DD** | **9,087,429** |

*This page intentionally left blank.*

# Appendix C – Factual Status Document

**To facilitate planning and conversation at the Office of Science User Facilities Roundtable planned for December 2020, please provide brief responses to the questions below. A paragraph or two per question is appropriate. Please send responses to Natalia Melcer and Ben Brown by COB Friday, November 20, 2020.**

1. How has the COVID-19 pandemic affected your facility? At what fraction of facility capacity are you currently operating? What is the distribution between virtual access and physical on-site access? How have facility operations been impacted? Have there been disproportionate impacts on any staff? Has any action been taken to minimize the impact on early career staff or other vulnerable staff groups?

   The COVID-19 pandemic has had little to no effect on the facility regarding the facility's user-facing side. We are operating at full capacity since all user access to our facility is 100% remote. The operations team has had full access to the facility to perform the required hands-on maintenance when needed. The operations team removed our old supercomputer Mira, installed a new 200PB filesystem, and stood up two new machines (Crux and ThetaGPU) during the pandemic.

   The facility, with respect to the staff, has seen many challenges. Face-to-face interactions have been replaced by what seems like an increase in video-based meeting. Informal brainstorming sessions are no longer possible and extremely challenging in remote mode. Those with caregiving responsibilities who lost standard support networks (e.g., schools, daycare, family networks) by the pandemic took on additional duties in addition to their regular work commitments. We have emphasized flexibility to all supervisors and staff during this period. We have encouraged frequent conversations with staff about how things are going and facilitate flexible work schedules. Argonne has started numerous initiatives to alleviate the issues faced by caregivers. The lab has rolled out a caregiving vacation program the allows for vacation donations by members of the Argonne community to a pool of days that other staff members can draw from as needed. Argonne has modified part-time work policies to adjust to the shifting needs of staff. Early career staff have lost out face to face mentoring and engagement in conferences. We have extended postdoctoral appointments as needed to accommodate challenges in the job market.

2. What approaches are you taking to provide productive user research experiences and services in a virtual context?

   The user research experience has not changed for the ALCF during the pandemic except for face-to-face interaction at workshops. We are usually a remote-only access facility. We have moved a few of our larger, typically in-person workshops and training events to an entirely virtual. We have not needed to cancel any of them, although we have shifted a few dates during the early stay-at-home days of the pandemic. We have seen increased participation in a few of the workshops and training sessions during the pandemic.

3. What, if any, changes to physical access have you implemented or considered for users, vendors, or maintenance personnel?

   There have been no changes for users, given the way we typically operate. There were a few delays in new user access due to the way certain portions of the user application are processed in the early days of the stay-at-home order. User account processing is now back to expected timelines.

   As mentioned earlier, the operations team has had full access to the facility during the pandemic. When needed, we have accommodated vendor access to the site. The challenges have been travel and the vendor's operating policies during this period.

4. Have you identified any lessons learned as a result of the COVID-19 pandemic?

   Lessons learned during the COVID-19 pandemic have been many. Starting with the need to be collocated - the facility and staff have continued to be productive. This experience has demonstrated that all the teams do not need to be in the same place simultaneously. This will likely allow for increased flexibility in remote working moving forward. Flexible hours and work schedules have turned out not to be an issue. Our staff was not prepared for an extended work period at home, and accommodations were needed to help provide productive and ergonomic environments.

5. What has the COVID-19 pandemic revealed about the needs for virtual access to computational, data management/analysis, software, and network resources or other types of instrumentation?

   Since our user community has historically been full remote access, there has been little to no changes needed. As mentioned earlier - the changes have mainly been in how we deliver training and hold workshops.

6. User Project Administration: Have project award periods been extended and if so, does this impact future proposal calls and facility operating costs (including budget carryover across the fiscal year)? Have there been any changes to allocation policies?

   We have not had the need to change allocation policies and no awards were extended as a result of COVID.

7. What have been the impacts on annual user facility meetings, strategic planning workshops, and short courses and schools (to train students, early career researchers, or future users)?

   Involvement and ability to hold user meetings, strategic planning workshops, and short courses are the one area we have needed to adapt due to the pandemic. We have transitioned all of our training and workshops to online virtual meetings. We have not had to cancel any of them. Given the rest of the world has also transitioned into virtual events, we have let our staff participate in a larger number of these activities than would not have been possible in the past.

8. Please provide any other comments or ideas as input for the roundtable.

   What policies at the DOE level need to be able to accommodate a more distributed workforce?

**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.