AUGUST 27, 2024

ALCF-4 PROJECT OVERVIEW

JINI RAMPRAKASH
ALCF-4 Project Director
Argonne National Laboratory
Leadership Computing Facility

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

Argonne NATIONAL LABORATORY
Argonne Leadership Computing Facility

AGENDA

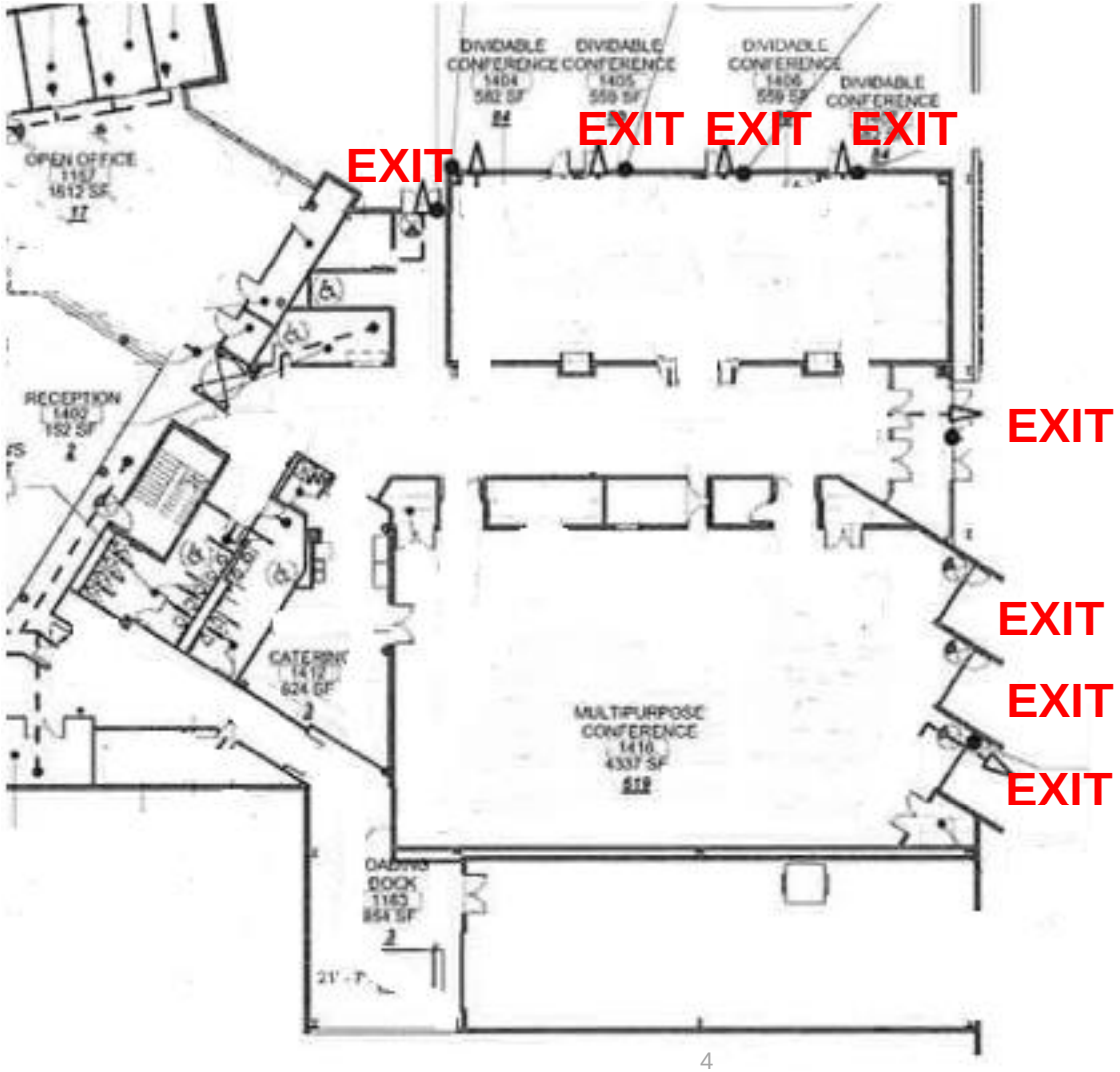
Times	Item	Owner
8:30	Executive Session	Review Chair
9:00	Welcome	Mike Papka
9:10	Project Overview	Jini Ramprakash
9:40	Technical Overview and Early Science	Kevin Harms Chris Knight
10:15	Break	
10:30	Technical Requirements	Taylor Childers
11:30	Benchmarks	Chris
12:15	(Working Lunch) Discussion & Questions from the committee	ALCF-4 Team
12:30	(Working Lunch) Executive Session	Review Chair
13:30	Facilities	Jon Cisek
14:15	ALCF-4 Risks Review	Noah / Jini
15:00	Break	
15:15	Executive Committee Q&A with ALCF-4 team	Review Chair
15:45	Executive Writing Session	Review Chair
17:00	Adjourn / Tour of Aurora	Susan Coghlan
18:00	Dinner	

IN CASE OF EMERGENCY

Dial 9-1-1 on an Argonne phone or 630-252-1911 on your cell phone and follow operator instructions

EMERGENCY EVACUATION FROM THE CONFERENCE CENTER

In case of fire or other emergencies follow the exit signs

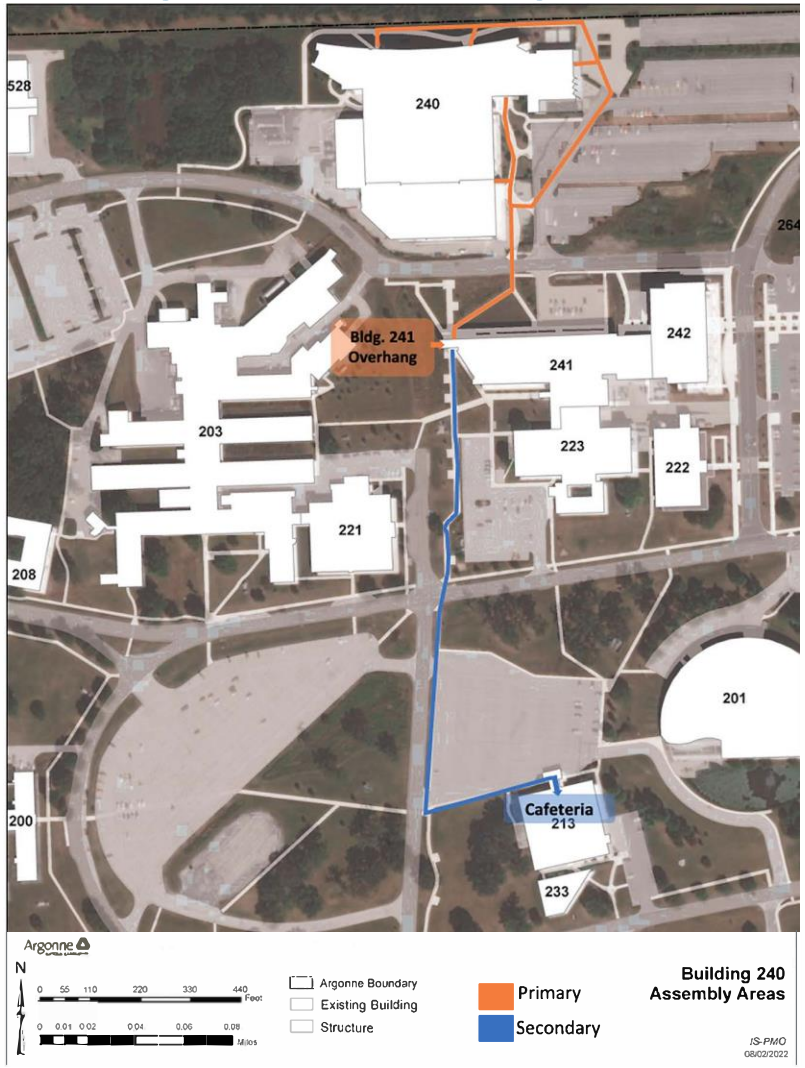


U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY | Argonne Leadership Computing Facility

ASSEMBLY AREA FOLLOWING EVACUATION

Follow the orange path below to Bldg. 241 overhang area



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

Argonne NATIONAL LABORATORY | Argonne Leadership Computing Facility

EMERGENCY SHELTERS

Proceed to rest rooms, catering area, loading dock, and stairwells

OTHER

REST ROOMS & DESIGNATED SMOKING AREA

- Men's and women's rest rooms are located just outside the entrance to the main 7 story section of the building on the east side of the Conference Center.
- There is a dedicated smoking area just outside the Conference Center around the corner of the main entrance on the north side of the building.
- Please use the smokers' receptacle located within the area.

AED LOCATION

- Near the restrooms

OUTLINE

- Origin of LCF @ DOE
- Mission Need Statement
- Key Project personnel and their roles
- ALCF-4 Project Goals
  - Key Performance Parameters
- Mapping charge questions to presentations

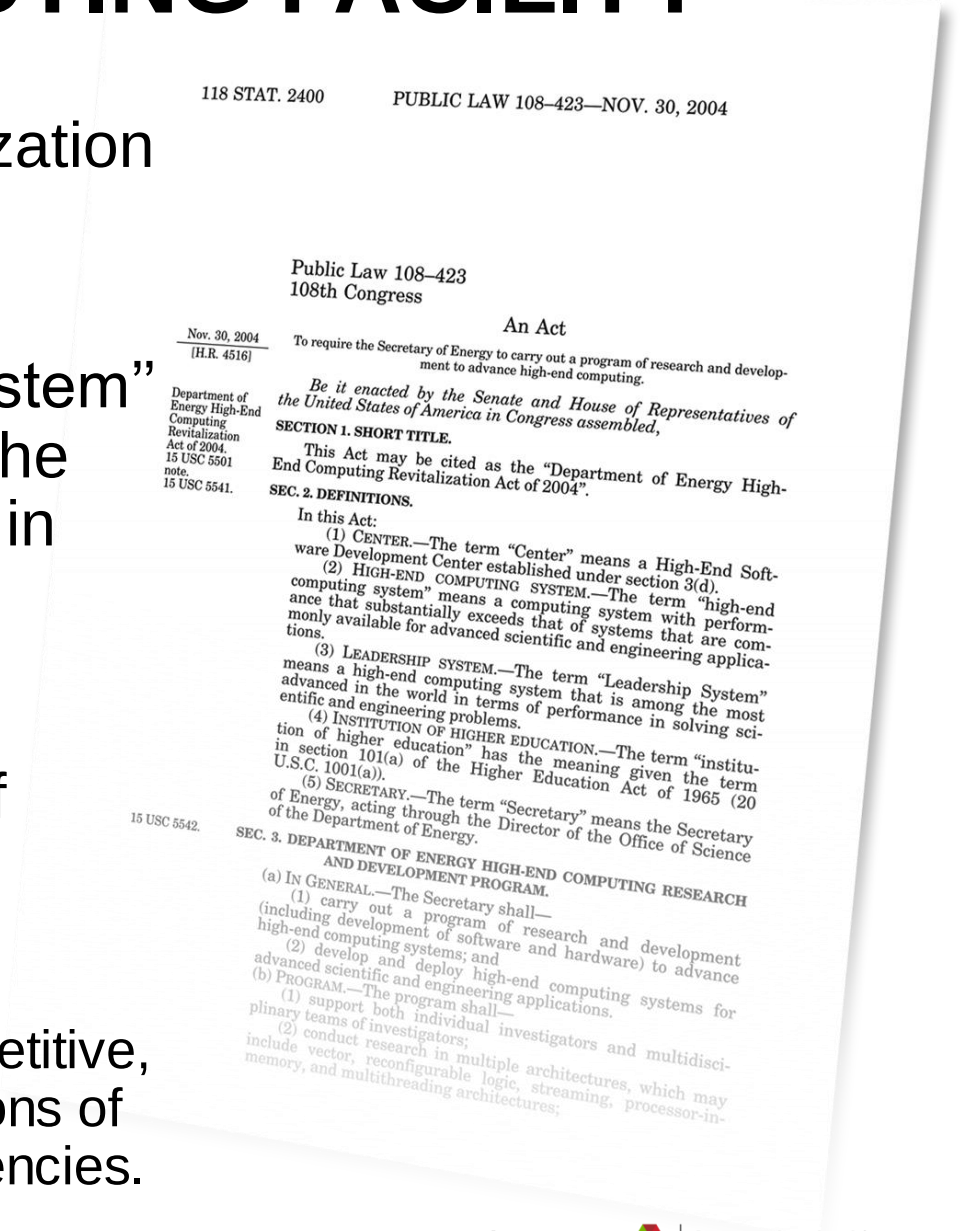
ORIGIN OF LEADERSHIP COMPUTING FACILITY

Department of Energy High-End Computing Revitalization Act of 2004 (Public Law 108-423):

LEADERSHIP SYSTEM.—The term “Leadership System” means a high-end computing system that is among the most advanced in the world in terms of performance in solving scientific and engineering problems.

The Secretary of Energy, acting through the Office of Science, shall

- Establish and operate Leadership Systems Facilities
- Provide access [to Leadership Systems Facilities] on a competitive, merit-reviewed basis to researchers in U.S. industry, institutions of higher education, national laboratories and other Federal agencies.



118 STAT. 2400 PUBLIC LAW 108–423—NOV. 30, 2004

Public Law 108–423
108th Congress

An Act

Nov. 30, 2004
[H.R. 4516]

To require the Secretary of Energy to carry out a program of research and development to advance high-end computing.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

Department of Energy High-End Computing Revitalization Act of 2004.
15 USC 5501 note.

SECTION 1. SHORT TITLE.

This Act may be cited as the “Department of Energy High-End Computing Revitalization Act of 2004”.

15 USC 5541.

SEC. 2. DEFINITIONS.

In this Act:

(1) CENTER.—The term “Center” means a High-End Software Development Center established under section 3(d).

(2) HIGH-END COMPUTING SYSTEM.—The term “high-end computing system” means a computing system with performance that substantially exceeds that of systems that are commonly available for advanced scientific and engineering applications.

(3) LEADERSHIP SYSTEM.—The term “Leadership System” means a high-end computing system that is among the most advanced in the world in terms of performance in solving scientific and engineering problems.

(4) INSTITUTION OF HIGHER EDUCATION.—The term “institution of higher education” has the meaning given the term in section 101(a) of the Higher Education Act of 1965 (20 U.S.C. 1001(a)).

(5) SECRETARY.—The term “Secretary” means the Secretary of Energy, acting through the Director of the Office of Science of the Department of Energy.

15 USC 5542.

SEC. 3. DEPARTMENT OF ENERGY HIGH-END COMPUTING RESEARCH AND DEVELOPMENT PROGRAM.

(a) IN GENERAL.—The Secretary shall—

(1) carry out a program of research and development (including development of software and hardware) to advance high-end computing systems; and

(2) develop and deploy high-end computing systems for advanced scientific and engineering applications.

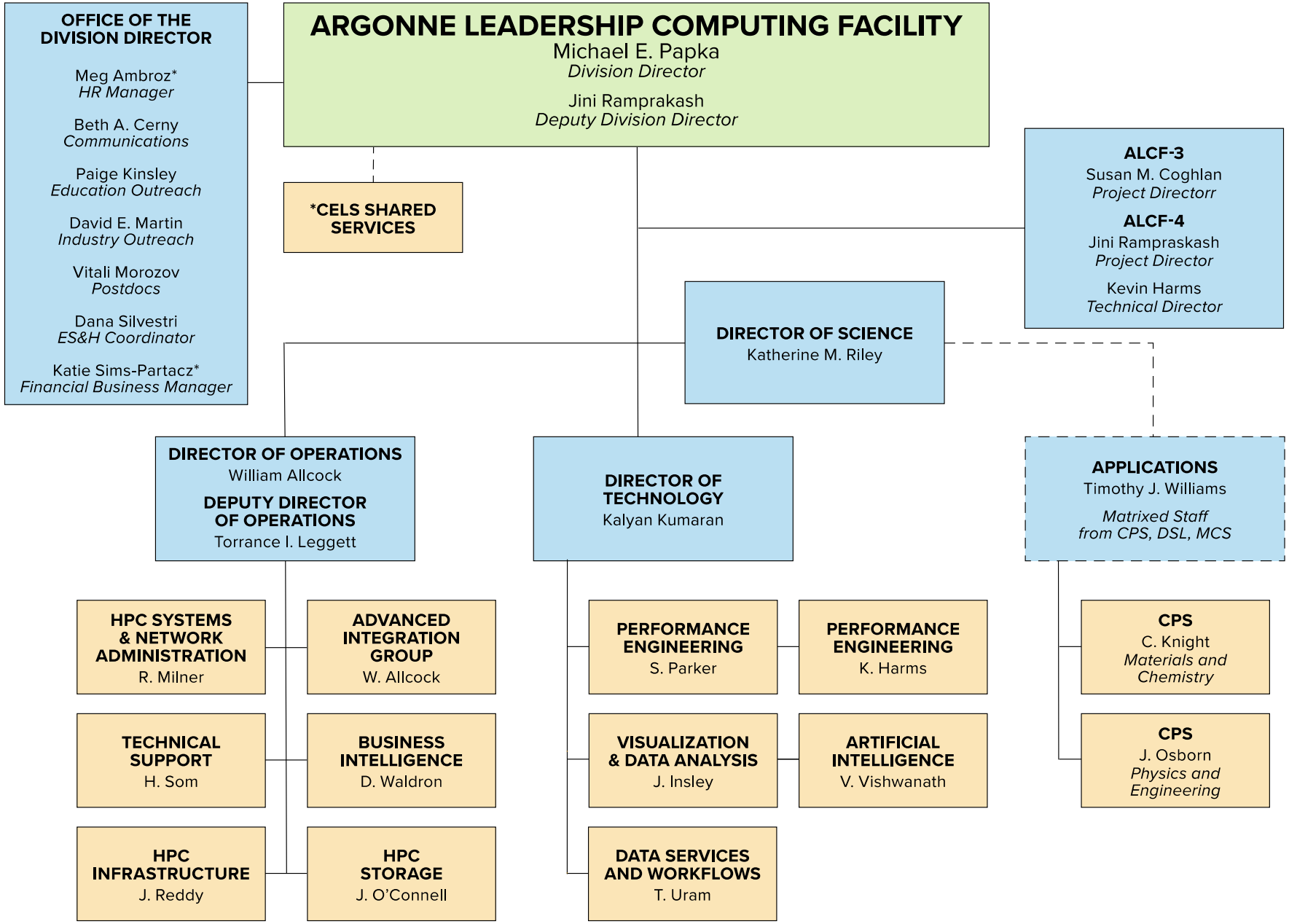
(b) PROGRAM.—The program shall—

(1) support both individual investigators and multidisciplinary teams of investigators;

(2) conduct research in multiple architectures, which may include vector, reconfigurable logic, streaming, processor-in-memory, and multithreading architectures;

DOE Leadership Computing Facility

- Established in 2004 as a collaborative, multi-lab initiative funded by DOE's ***Advanced Scientific Computing Research*** program
- Operates as **one facility** with two centers, at Argonne and at Oak Ridge National Laboratory
- Deploys and operates at least two advanced architectures that are **10-100 times more powerful** than systems typically available for open scientific research
- **Fully dedicated** to open science to address the ever-growing needs of the scientific community



U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
Argonne NATIONAL LABORATORY | Argonne Leadership Computing Facility


EXIT
55
50
Argonne
Aurora
A brand-new
class of system

ALCF AT A GLANCE IN 2023

- Users pursue scientific challenges
- In-house experts to help maximize results
- Resources fully dedicated to open science

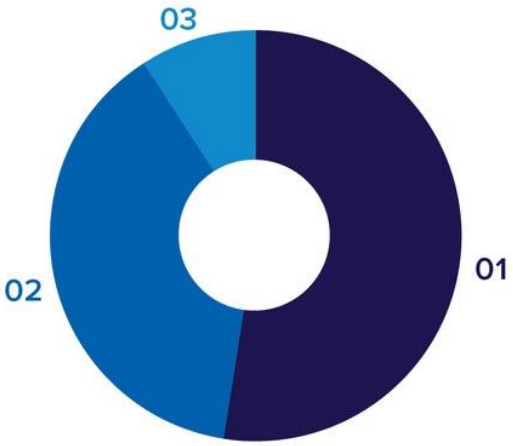
35.7M node-hours of compute time

417 active projects

1,624 facility users

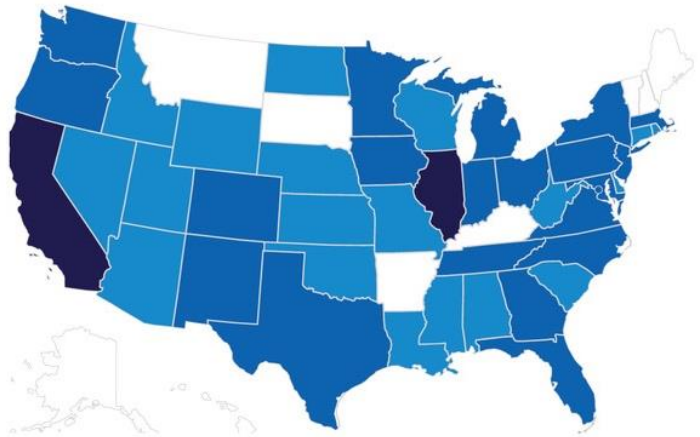
230+ publications

2023 ALCF Users by Affilliation



01	Academia	857
02	Government	620
03	Industry	147

2023 U.S. ALCF Users by State



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

ALCF Systems Evolution

5.7 TF

IBM BG/L
2004

557 TF

Intrepid
IBM BG/P
2007

10 PF

Mira
IBM BG/Q
2012

11.7 PF

Theta
Intel-Cray XC40
2017

15.6 PF

ThetaGPU
NVIDIA
DGX A100
2020

44 PF

Crux
HPE-AMD
+
Polaris
HPE
2021

2 EF

Aurora
Intel-HPE
2023

JLSE (2013)

AI Testbed (2020)

Edge Testbed (2021)

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

Argonne NATIONAL LABORATORY | Argonne Leadership Computing Facility

LEADERSHIP COMPUTING FACILITY UPGRADE – MISSION NEED STATEMENT

- "This falls short for scientific applications in the ***CY 2026-2031*** timeframe that will need a ***5x-10x increase in application performance***."
- "DOE goals for the Energy Earthshots are ambitious [...] Leadership computing will accelerate and optimize efforts across the Earthshots, and ***more advanced capabilities will*** allow optimization within and across the Earthshots to ***maximize impact***."
- "***Industries*** of the future, a bipartisan priority, are ***heavily interdependent with progress in HPC*** – these include Quantum Information Science, Quantum networking, Artificial Intelligence, and microelectronics."
- "The next generation leadership computing ecosystem will be ***designed to interface*** the IRI environment that ***supports automated workflows, data integration, and AI technologies*** and integrates distributed resources and advanced data technologies with edge-to-exascale capabilities that reduce the time from experiment and observation to scientific insight."

PROJECT LEADERSHIP TEAM

Jini Ramprakash
ALCF-4 Project Director

Kevin Harms
ALCF-4 Technical Director

Susan Coghlan
ALCF-4 Project Advisor

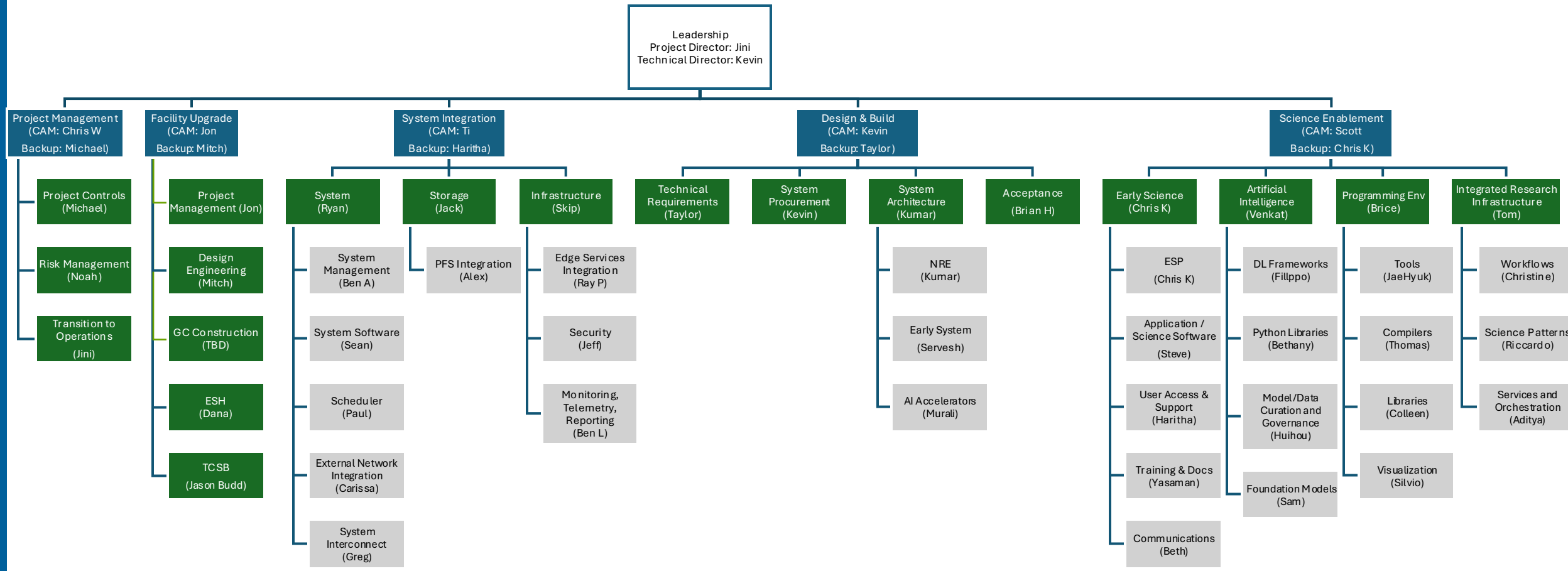
Ti Leggett
ALCF-4 Project Deputy

Taylor Childers
ALCF-4 Technical Deputy

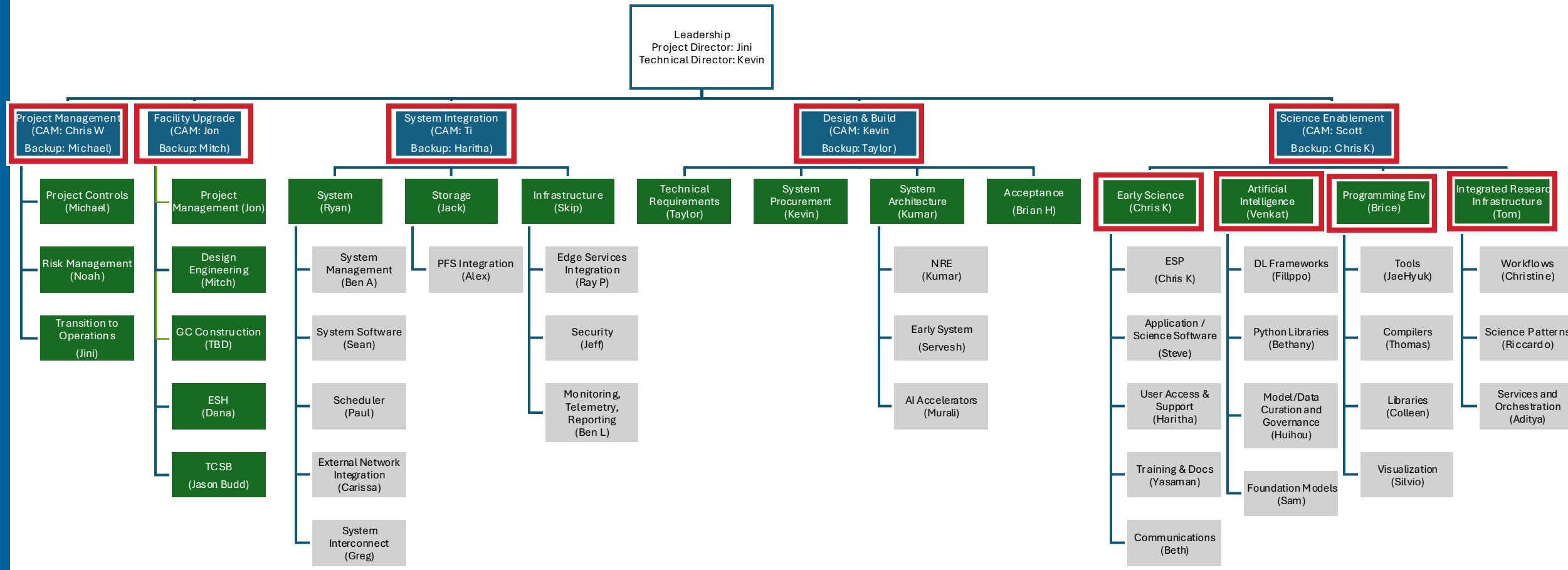
ALCF-4 Project
- Focus on O413.3b requirements
- Schedule, budget, reviews, …

ALCF-4 Technical
- Focus on system build & science requirements
- RFP/SOW, procurement, …

PROJECT ORGANIZATION
(WORK BREAKDOWN STRUCTURE)

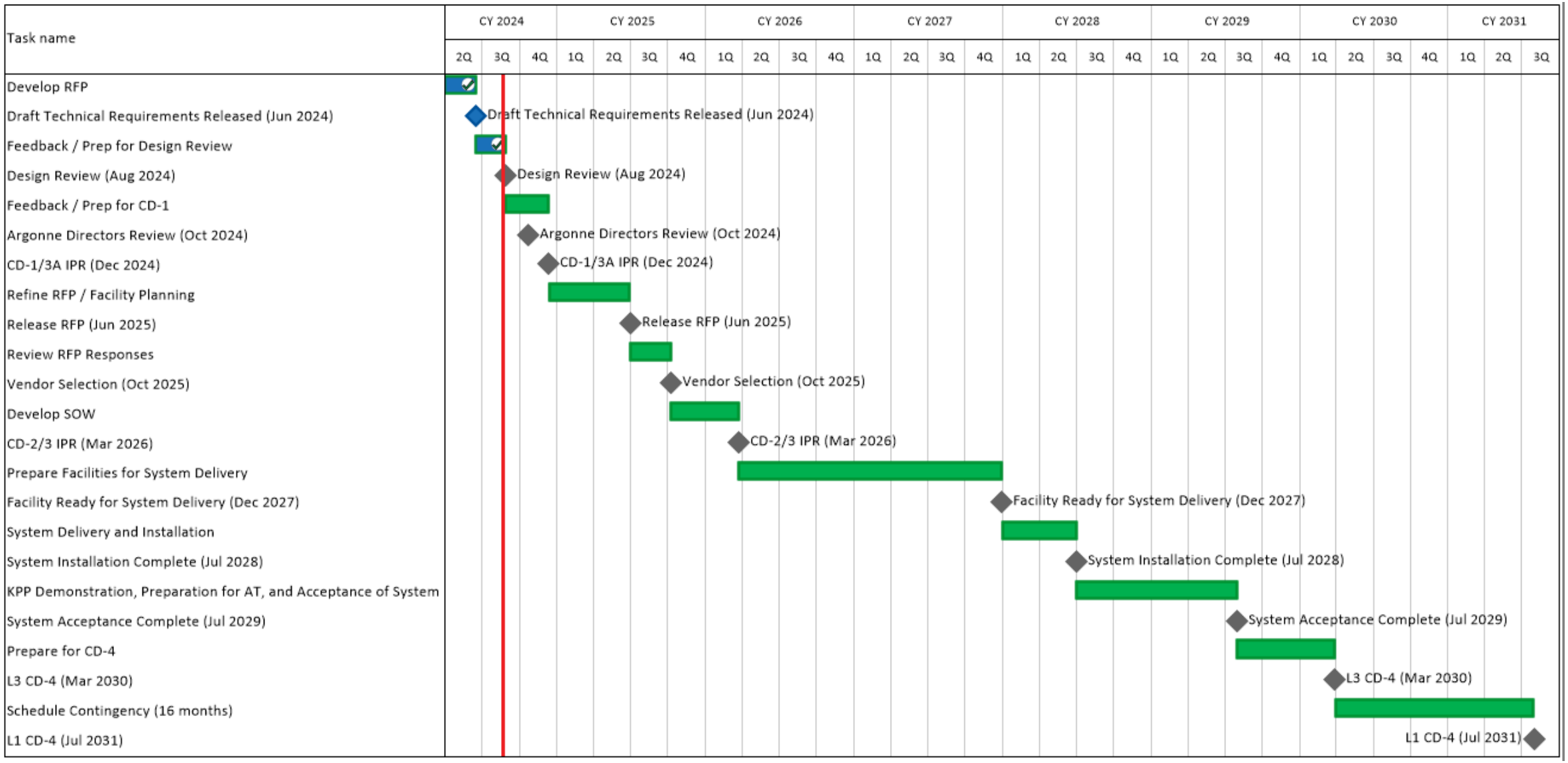


PROJECT ORGANIZATION
(WORK BREAKDOWN STRUCTURE)



* CAMs indicated with red frame

SCHEDULE AT A GLANCE



PROJECT GOALS

- 5x-10x improvement in application performance over Aurora
- Support traditional HPC computation, AI, and data intensive computation
- Investigate potential to provide faster deployment and realization of new technology for future systems beyond ALCF-4

KEY PERFORMANCE PARAMETERS

Scope	Threshold	Objective
Application Performance	3x GeoMean	≥ 5x GeoMean
Early Science Program Applications	8 applications INCITE ready	15 applications INCITE ready
System Power	≤ 40MW	≤ 30MW
Programming Environment	1 compiler, 1 debugger, 1 learning framework	2 compilers, 2 debuggers, 2 learning frameworks
IRI capabilities	Support 3 IRI practice areas*	Support all 6 IRI practice areas*

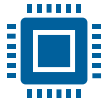
PRACTICE AREAS

Resource Co-Operations

Allocations/provisioning of multiple heterogeneous resources across multiple facilities for large collections of scientific programs must be aligned in time and planned. IRI requires new levels of cooperation, collaboration, co-scheduling, and joint planning across facilities and across DOE programs.

Cybersecurity and Federated Access

Users require a distributed research infrastructure with seamless access and consistent services while the infrastructure must be operated according to cybersecurity requirements and policies set at the federal level. Operators of user facilities also have different missions, and thus different requirements, across the lab complex. Balancing these constraints can also lead to sources of impedance. Novel secure design patterns and architectures will be required to support open science-integrated architecture for seamless scientific collaboration.



User Experience

Understanding evolving users' needs and experiences is critical for technologists to develop effective IRI solutions. This area is central for building an effective IRI. Strategies for enabling users, including requirements gathering, user-centric (co)-design, liaising approaches, and related topics, have been proposed. This topic has implications for all other practice areas.

Workflows, Interfaces and Automation

Users need to systematically and easily assemble system components to support IRI science cases in the form of end-to-end pipelines. Users should be able to manage these overlays and middleware effectively across facilities.

Scientific Data Lifecycle

Users need to manage their data (along with metadata) across facilities from inception to curation, archiving, dissemination, and publication. Technologists need to understand the requirements across different communities to develop solutions appropriate for an IRI and the principles of effective data management to provide a FAIR- based data pipeline with end user-focused interfaces.

Portable / Scalable Solutions

Users and technologists need their applications to move/translate across heterogeneous facilities (be portable) and go from smaller to larger resources (be scalable).

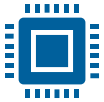
PRACTICE AREAS

Resource Co-Operations

Allocations/provisioning of multiple heterogeneous resources across multiple facilities for large collections of scientific programs must be aligned in time and planned. IRI requires new levels of cooperation, collaboration, co-scheduling, and joint planning across facilities and across DOE programs.

Cybersecurity and Federated Access

Users require a distributed research infrastructure with seamless access and consistent services while the infrastructure must be operated according to cybersecurity requirements and policies set at the federal level. Operators of user facilities also have different missions, and thus different requirements, across the lab complex. Balancing these constraints can also lead to sources of impedance. Novel secure design patterns and architectures will be required to support open science-integrated architecture for seamless scientific collaboration.



User Experience

Understanding evolving users' needs and experiences is critical for technologists to develop effective IRI solutions. This area is central for building an effective IRI. Strategies for enabling users, including requirements gathering, user-centric (co)-design, liaising approaches, and related topics, have been proposed. This topic has implications for all other practice areas.

Workflows, Interfaces and Automation

Users need to systematically and easily assemble system components to support IRI science cases in the form of end-to-end pipelines. Users should be able to manage these overlays and middleware effectively across facilities.

Scientific Data Lifecycle

Users need to manage their data (along with metadata) across facilities from inception to curation, archiving, dissemination, and publication. Technologists need to understand the requirements across different communities to develop solutions appropriate for an IRI and the principles of effective data management to provide a FAIR- based data pipeline with end user-focused interfaces.

Portable / Scalable Solutions

Users and technologists need their applications to move/translate across heterogeneous facilities (be portable) and go from smaller to larger resources (be scalable).

SUMMARY

- LCF Upgrade Mission Need Statement includes
  - Significant increase in leadership computational and data science capabilities
  - Advanced capabilities to maximize impact
  - Interface the IRI environment that supports automated workflows, data integration, and AI technologies
- Project Goals align with LCF Upgrade Mission Need Statement
  - 5x-10x improvement in application performance over Aurora
  - Support traditional HPC computation, AI, and data intensive computation
  - Investigate potential to provide faster deployment and realization of new technology for future systems beyond ALCF-4
- Core team members bring experience and expertise to the ALCF-4 Project

MAPPING CHARGE QUESTIONS TO TALKS

1. Is the technical approach appropriate to support the ALCF-4 Mission Need requirements?
   - Project Overview, Technical Overview and Early Science (Speakers: Jini Ramprakash, Kevin Harms & Chris Knight)
2. Are the RFP technical requirements reasonable, clear, and consistent with the goals and objectives for the ALCF-4 project?
   - Technical Requirements, Benchmarks (Speakers: Taylor Childers, Chris Knight)
3. Does the ALCF facility upgrade plan support the system requirements specified in the RFP for the onsite options?
   - Facilities (Speaker: Jon Cisek)
4. Have the major technical risks and appropriate mitigation strategies been correctly identified for this stage of the project?
   - ALCF-4 Risks Review (Speakers: Noah Legenski & Jini Ramprakash)


Argonne
NATIONAL LABORATORY
Argonne Leadership
Computing Facility