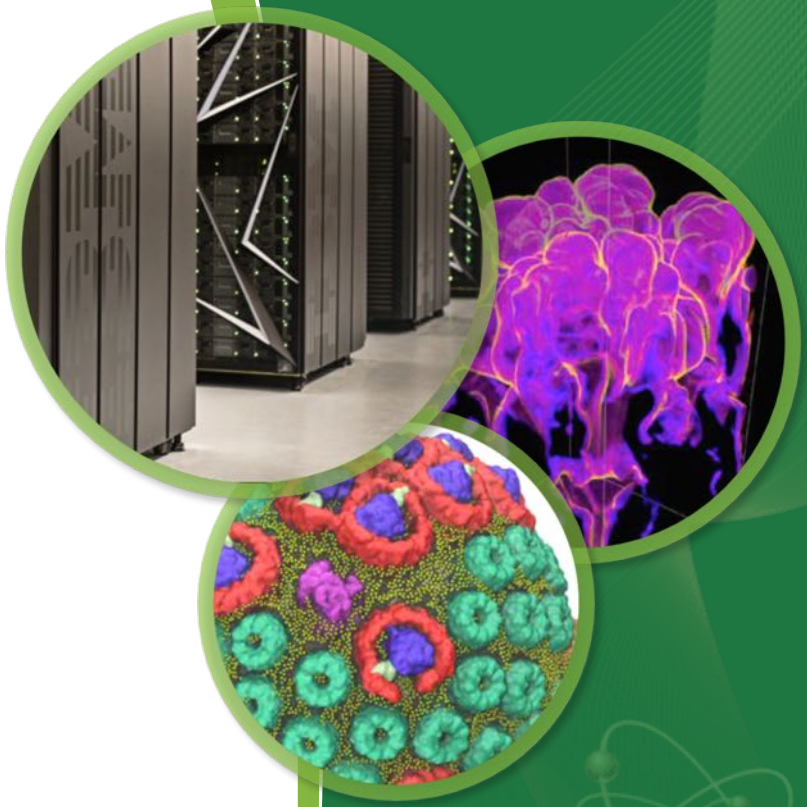
Overview of High Performance Computing Resources at the Oak Ridge Leadership Computing Facility (OLCF)



Tom Papatheodore
Oak Ridge Leadership Computing Facility
Oak Ridge National Laboratory

June 26, 2018

DOE's Office of Science Computation User Facilities

NERSC
Edison is 2.57 PF

ALCF
Mira is 10 PF

OLCF
Titan is 27 PF

- DOE is leader in open High-Performance Computing
- Provide the world's most powerful computational tools for open science
- Access is free to researchers who publish
- Boost US competitiveness
- Attract the best and brightest researchers

What is a Leadership Computing Facility (LCF)?

- Collaborative DOE Office of Science user-facility program at ORNL and ANL
- Mission: Provide the computational and data resources required to solve the most challenging problems.
- 2-centers/2-architectures to address diverse and growing computational needs of the scientific community

- Highly competitive user allocation programs (INCITE, ALCC).
- Projects receive 10x to 100x more resource than at other generally available centers.
- LCF centers partner with users to enable science & engineering breakthroughs (Liaisons, Catalysts).

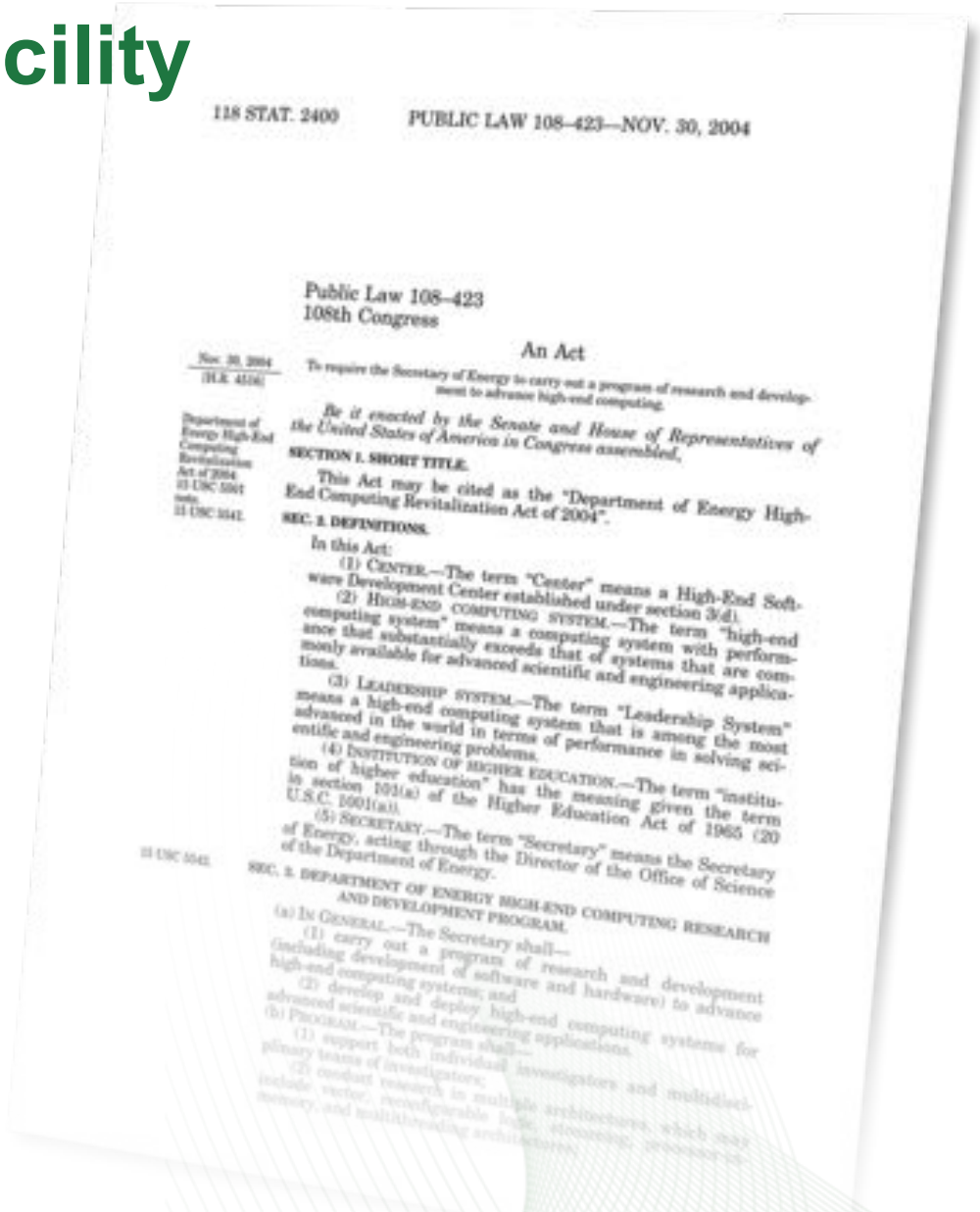
OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

Origin of Leadership Computing Facility

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

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;

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

ORNL has systematically delivered a series of leadership-class systems

On scope • On budget • Within schedule

OLCF-1

OLCF-2

OLCF-3

18.5 TF
2004
Cray X1E
Phoenix

25 TF
2005
Cray XT3
Jaguar

54 TF
2006
Cray XT3
Jaguar

62 TF
2007
Cray XT4
Jaguar

263 TF
2008
Cray XT4
Jaguar

1 PF
2008
Cray XT5
Jaguar

2.5 PF
2009
Cray XT5
Jaguar

1000-fold improvement in 8 years

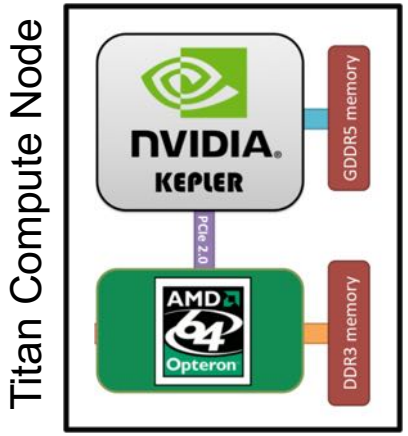
27 PF
2012
Cray XK7
Titan

Titan, five years old in October 2017, continues to deliver world-class science research in support of our user community. We will operate Titan through 2019 when it will be decommissioned.

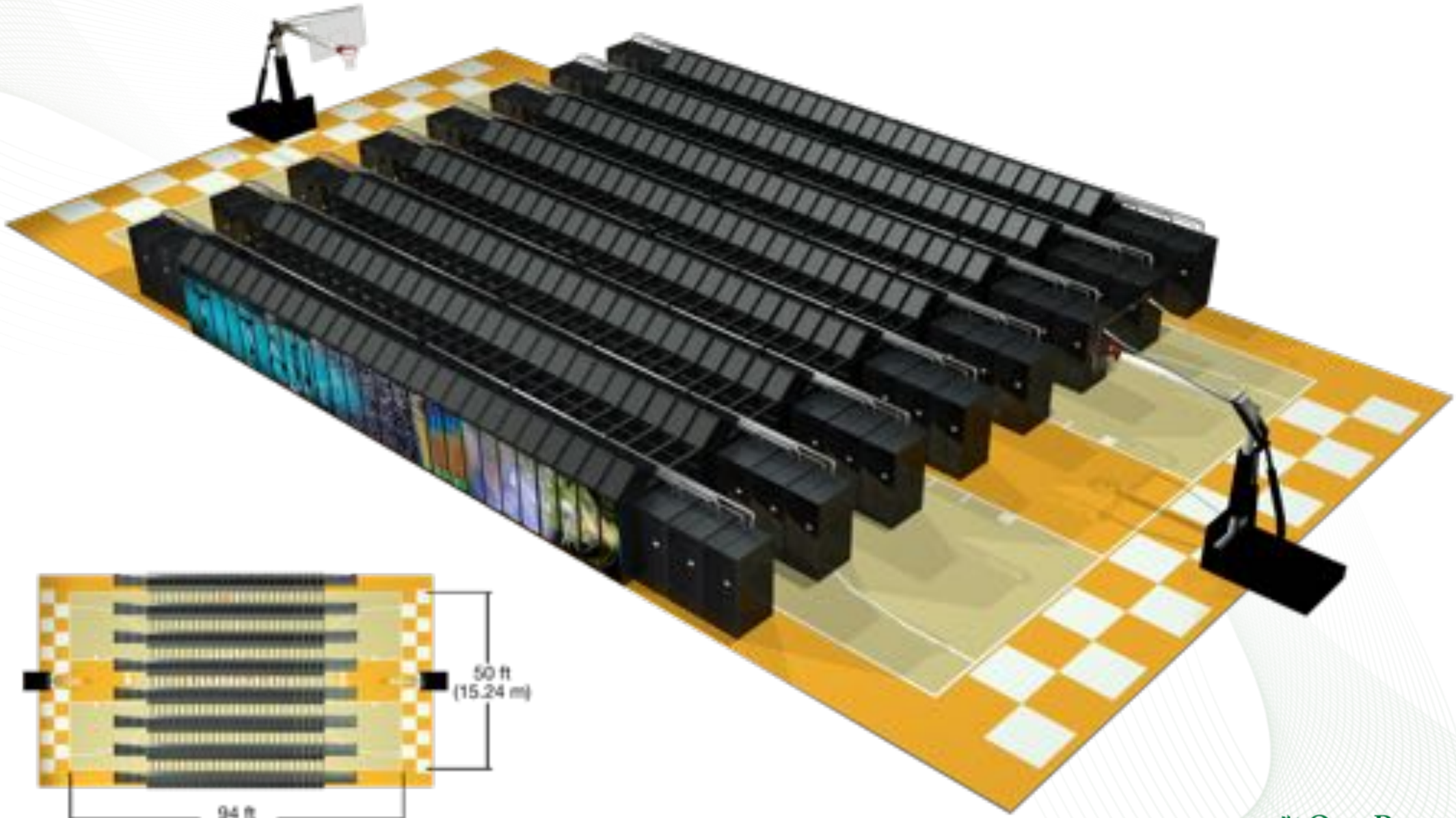
We are building on this record of success to enable exascale in 2021

Titan Overview

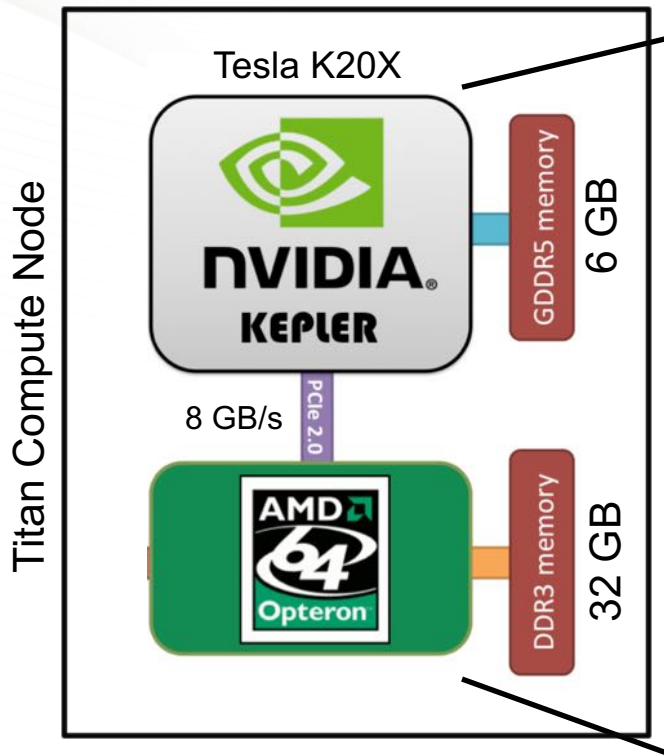
- Cray XK7 – Hybrid Architecture (CPUs + GPUs)
  - CPUs: several cores optimized for serial work
  - GPUs: thousands of core optimized for parallel work
  - Offload compute-intensive regions of code to run on GPUs
- 27 PF (peak performance)
  - FLOPS: floating point operations per second (e.g. 1.7*2.4 = 4.08)
  - 27 quadrillion floating point operations each second
  - How does Titan achieve this peak performance?
    - It has MANY processors!
- 18,688 compute nodes

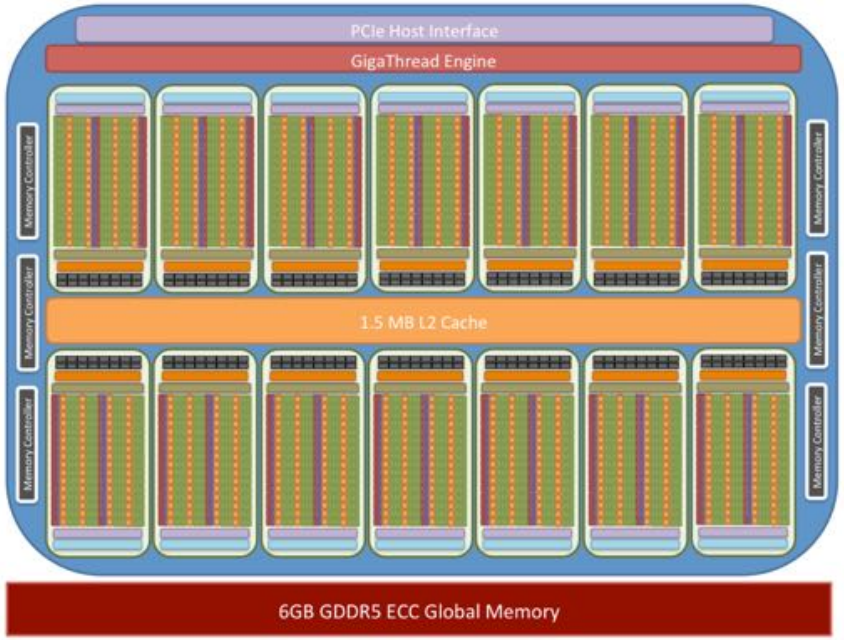


Smallest granularity of the system you can request.

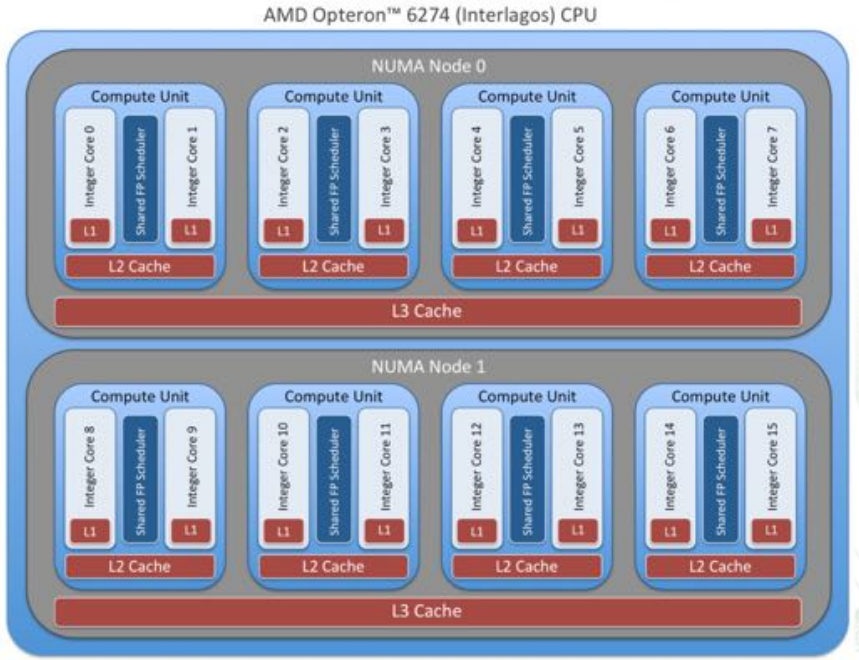

50 ft
(15.24 m)
94 ft
(28.65 m)

Titan Compute Nodes





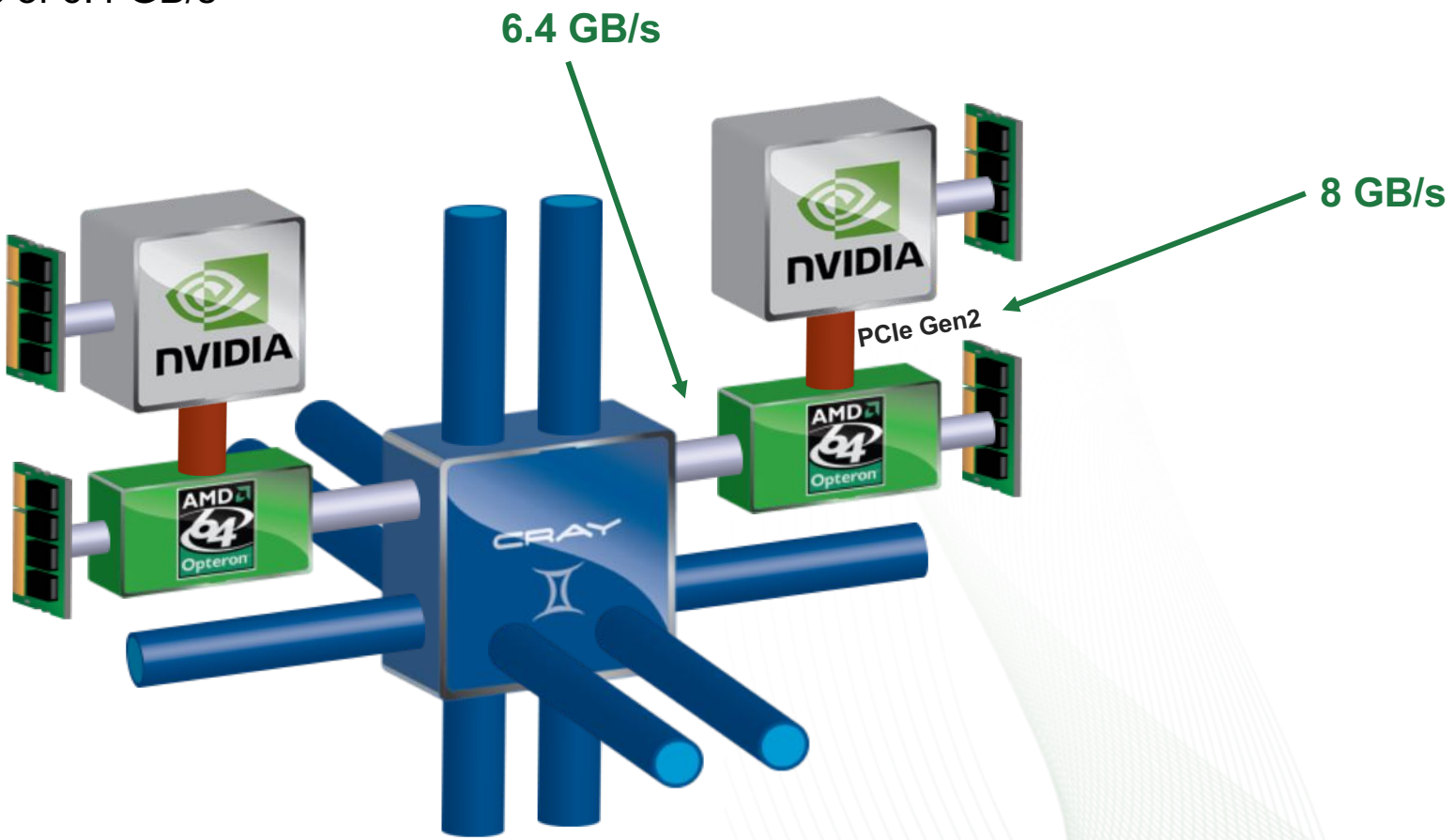
2688 cores



16 cores

Titan Compute Nodes

- Cray Gemini Interconnect
  - Data transfer speeds between nodes of 6.4 GB/s



Titan Overview

- 3 foot raised floor
  - Power cables, interconnect, A/C duct
- A/C units around the room blow air under floor
  - Directed with perforated tiles where needed
- Titan itself needs additional cooling
  - Air blows up through each cabinet
  - Transfers heat by boiling refrigerant
  - Air enters and exits cabinet at same temperature as room (~60 F)

Lustre File System (Atlas)

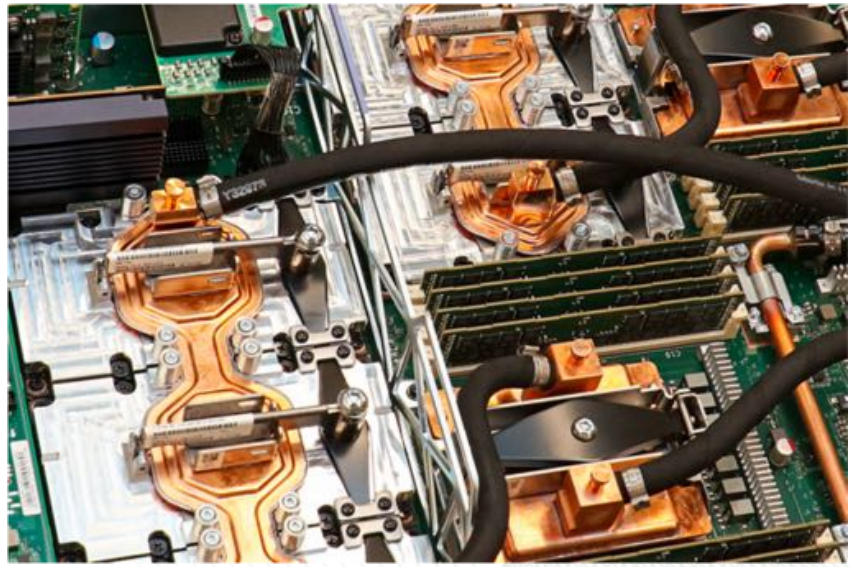
- Temporary storage for production I/O
- 32 PB Capacity
- 1 TB/s read/write (aggregate)
- Mounted on most OLCF systems

High Performance Storage System (HPSS)

- Long-term archival storage resource
- Robotic tape and disk storage

Coming in 2018: Summit will replace Titan as the OLCF's leadership supercomputer

Summit, slated to be more powerful than any other existing supercomputer, is the Department of Energy's Oak Ridge National Laboratory's newest supercomputer for open science.



Summit Overview

OpenPOWER™

Compute System

256 compute racks
4,608 compute nodes
Mellanox EDR IB fabric
200 PFLOPS
10.2 PB Total Memory
~13 MW

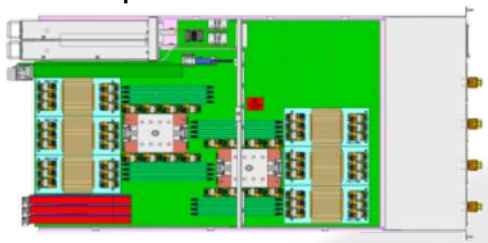
Compute Rack

18 Compute Servers
Warm water (70°F direct-cooled components)
RDHX for air-cooled components

39.7 TB Memory/rack
55 KW max power/rack

Compute Node

2 x POWER9
6 x NVIDIA GV100
NVMe-compatible PCIe 1600 GB SSD



25 GB/s EDR IB- (2 ports)
512 GB DRAM- (DDR4)
96 GB HBM- (3D Stacked)
Coherent Shared Memory

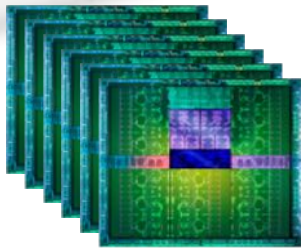
Components

IBM POWER9

- 22 Cores
- 4 Threads/core
- NVLink

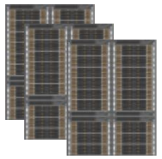
NVIDIA GV100

- 7 TF
- 16 GB @ 0.9 TB/s
- NVLink

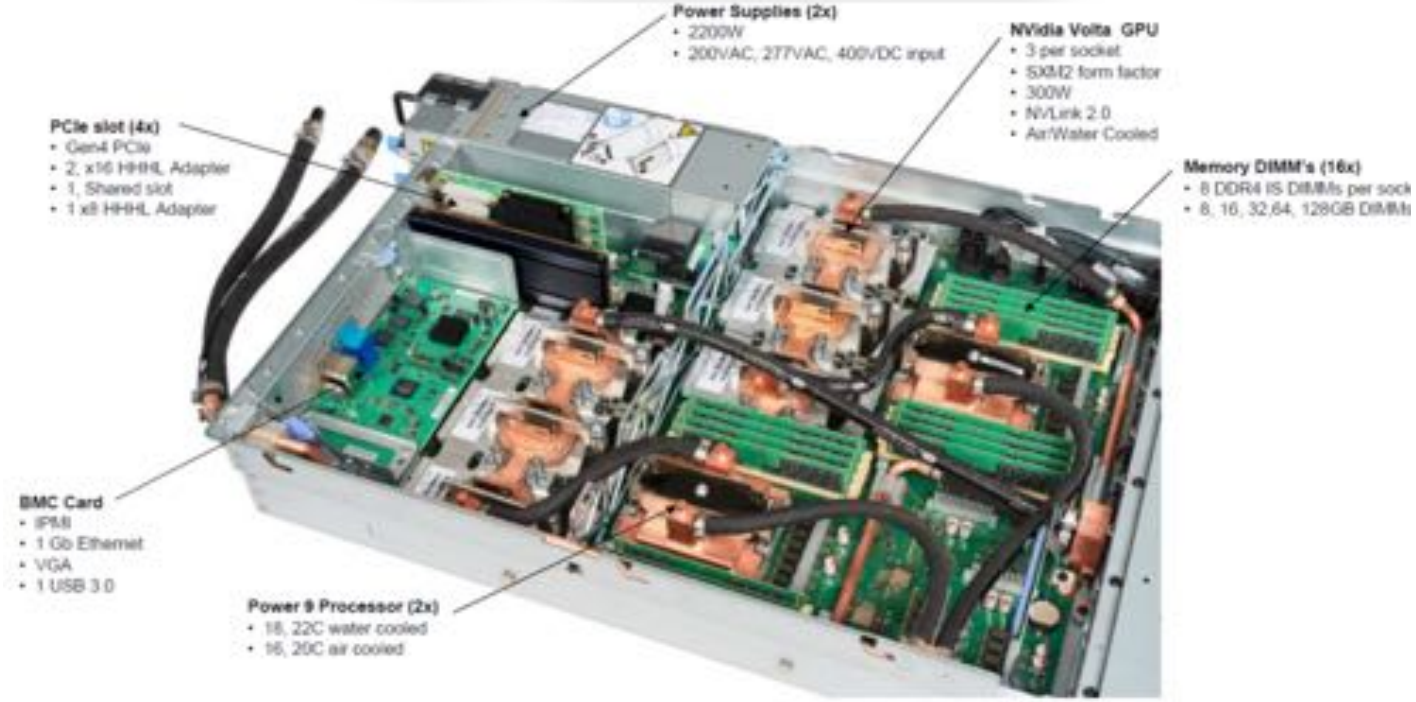


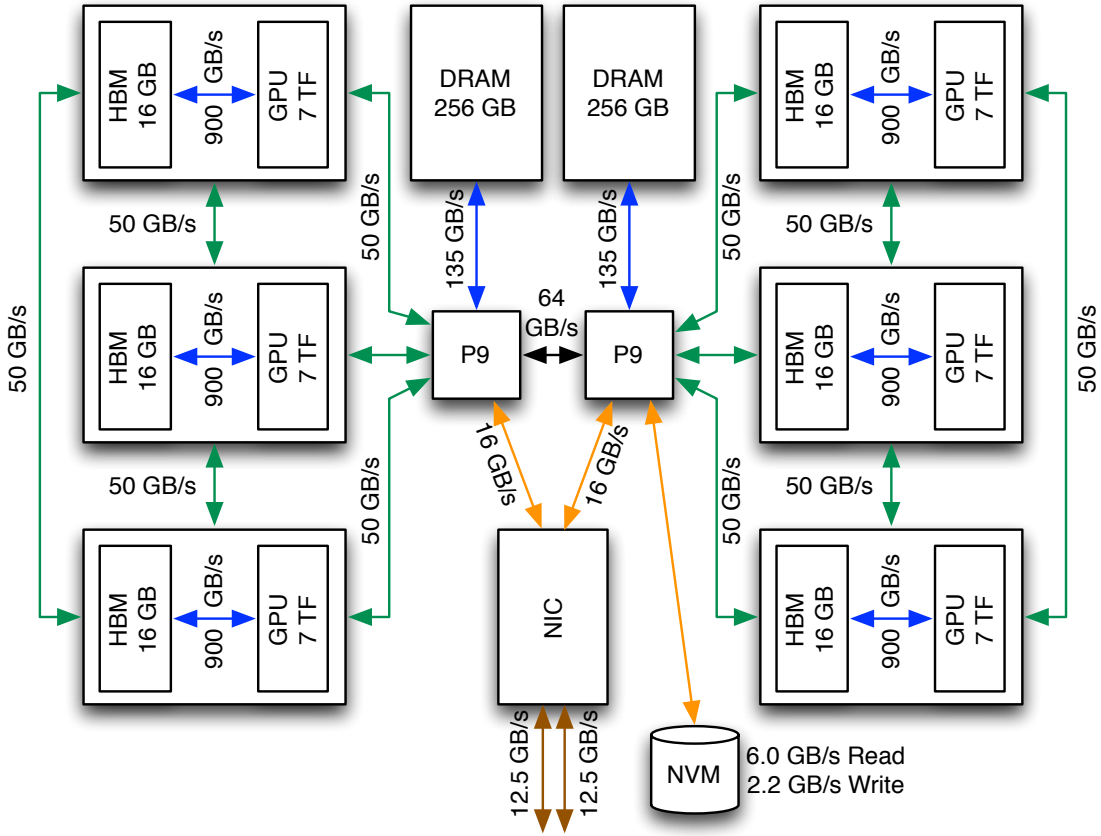
GPFS File System

250 PB storage
2.5 TB/s read, 2.5 TB/s write
(**2.5 TB/s sequential and 2.2 TB/s random I/O)



Summit Node Overview





TF	42 TF (6x7 TF)
HBM	96 GB (6x16 GB)
DRAM	512 GB (2x16x16 GB)
NET	25 GB/s (2x12.5 GB/s)
MMsg/s	83

- HBM/DRAM Bus (aggregate B/W)
- NVLINK
- X-Bus (SMP)
- PCIe Gen4
- EDR IB

HBM & DRAM speeds are aggregate (Read+Write).
All other speeds (X-Bus, NVLink, PCIe, IB) are bi-directional.

Coming in 2018: Summit will replace Titan as the OLCF's leadership supercomputer

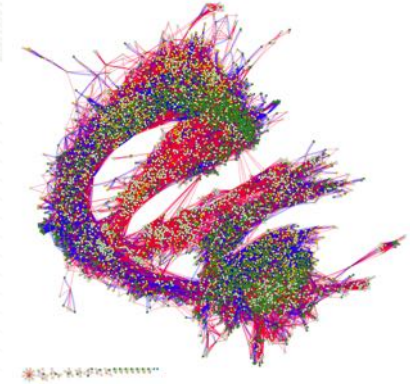
- Many fewer nodes
- Much more powerful nodes
- Much more memory per node and total system memory
- Faster interconnect
- Much higher bandwidth between CPUs and GPUs
- Much larger and faster file system

Feature	Titan	Summit
Application Performance	Baseline	5-10x Titan
Number of Nodes	18,688	4,608
Node performance	1.4 TF	42 TF
Memory per Node	32 GB DDR3 + 6 GB GDDR5	512 GB DDR4 + 96 GB HBM2
NV memory per Node	0	1600 GB
Total System Memory	710 TB	>10 PB DDR4 + HBM2 + Non-volatile
System Interconnect	Gemini (6.4 GB/s)	Dual Rail EDR-IB (25 GB/s)
Interconnect Topology	3D Torus	Non-blocking Fat Tree
Bi-Section Bandwidth	112 TB/s	115.2 TB/s
Processors	1 AMD Opteron™ 1 NVIDIA Kepler™	2 IBM POWER9™ 6 NVIDIA Volta™
File System	32 PB, 1 TB/s, Lustre®	250 PB, 2.5 TB/s, GPFS™
Power Consumption	9 MW	13 MW

Installation Nearing Completion

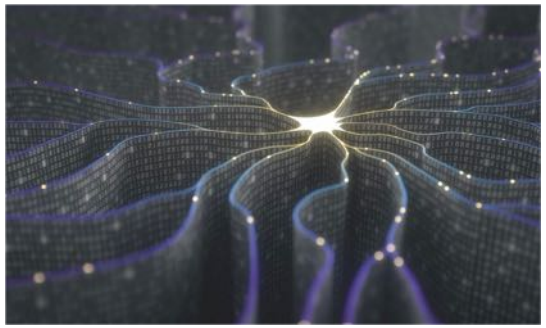
- Hardware installation completed in March
- Continuing to stabilize nodes, disks, and network
- In December, accepted 1,080 of 4,608 nodes to port codes
- OLCF is working with IBM, NVIDIA, Red Hat, and Mellanox to stabilize and debug system software

Science at OLCF

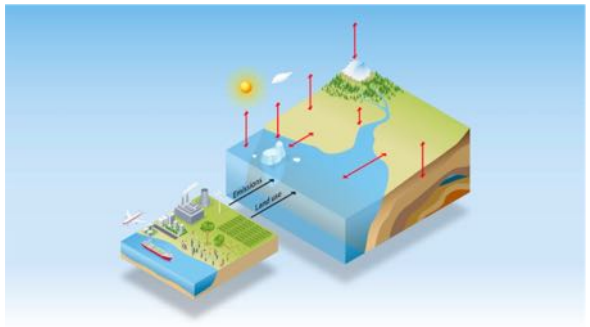


Biology

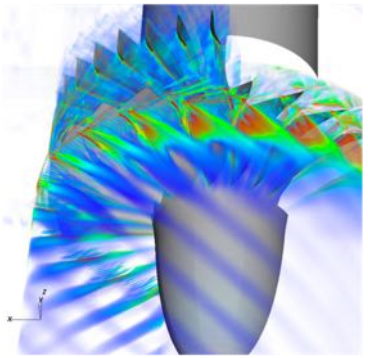
Chemistry



Computer Science

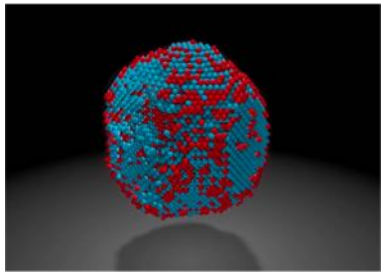


Earth Science

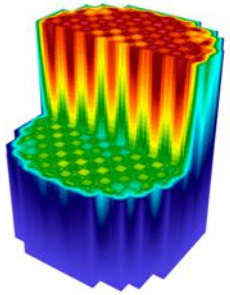


Engineering

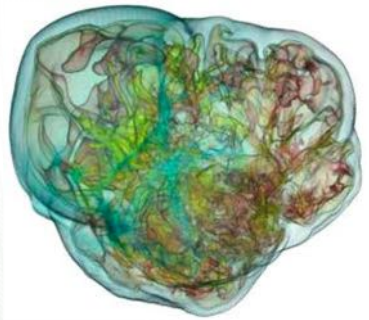
Fusion



Materials Science



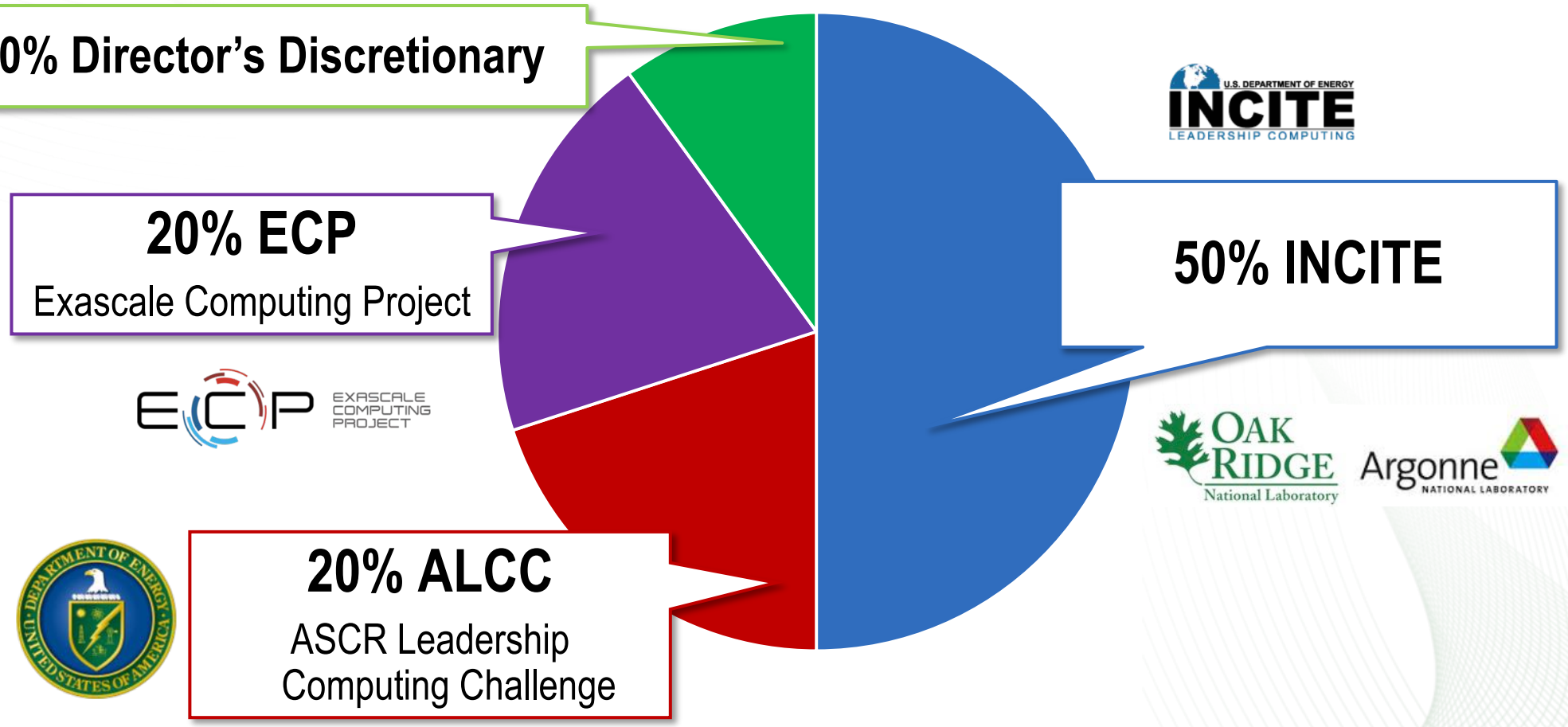
Nuclear Energy



Physics

Four primary user programs for access to LCF

Distribution of allocable hours

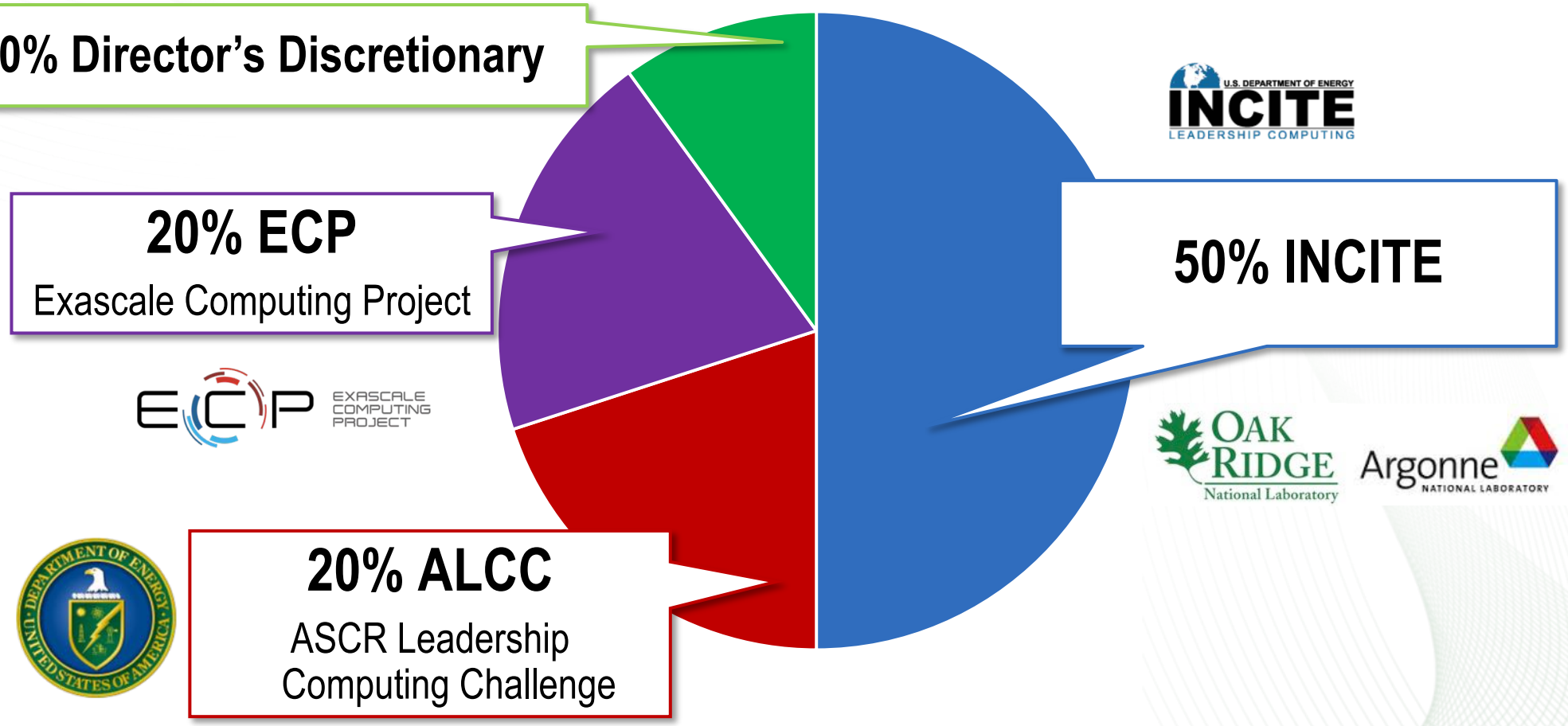


Innovative and Novel Computational Impact on Theory and Experiment (INCITE) Program

- Access to the most capable, most productive, fastest open science supercomputers in the nation
- Applicable to a broad array of science, engineering, and computer science domains
- Proposals must be:
  - High-impact, computationally and/or data intensive campaigns
  - Must take advantage of unique HPC architectures
  - Research that cannot be performed anywhere else.
- For more information visit http://www.doeleadershipcomputing.org/

Four primary user programs for access to LCF

Distribution of allocable hours



Questions?

Tom Papatheodore
papatheodore@ornl.gov