# NERSC Overview
# Perlmutter & Doudna

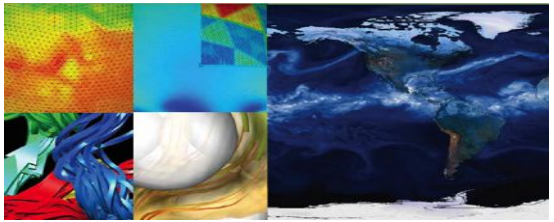
**Jack Deslippe**
N10 Workflow Readiness Lead

July 2025

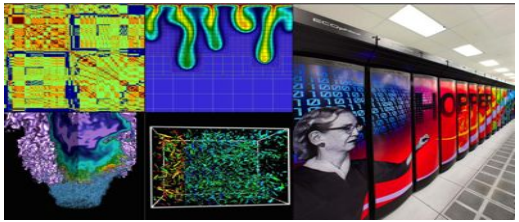
# NERSC: Mission HPC for DOE Office of Science Research

U.S. DEPARTMENT OF ENERGY | Office of Science

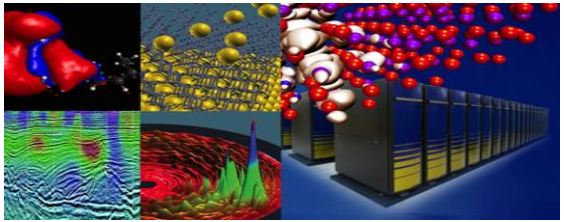
Largest funder of physical science research in the U.S.



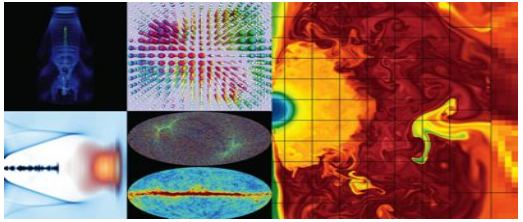
Biological and Environmental Research



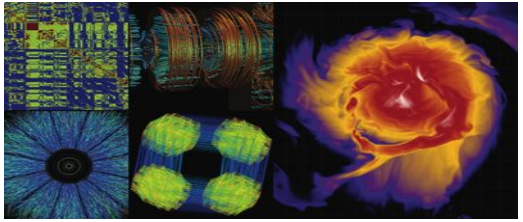
Computing



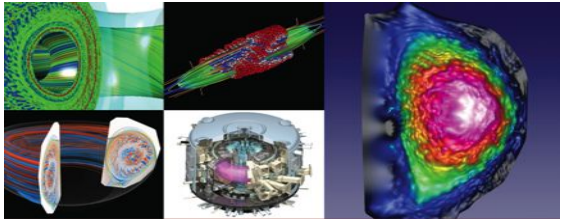
Basic Energy Sciences



High Energy Physics

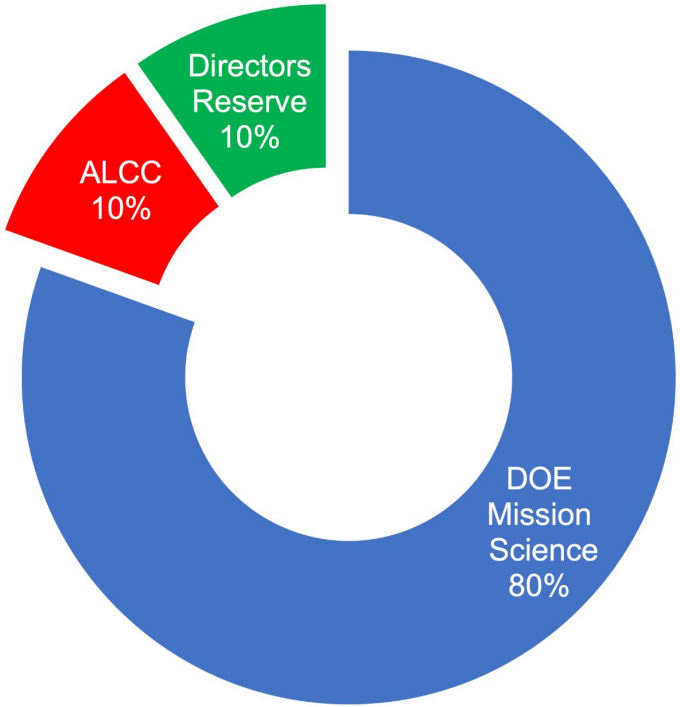


Nuclear Physics



Fusion Energy, Plasma Physics

# Allocation Programs




**Distributed by DOE Office of Science program managers**

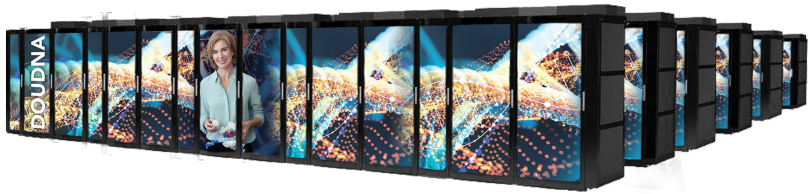
**Competitive awards run by DOE Advanced Scientific Computing Research Office**

**Strategic awards from NERSC**

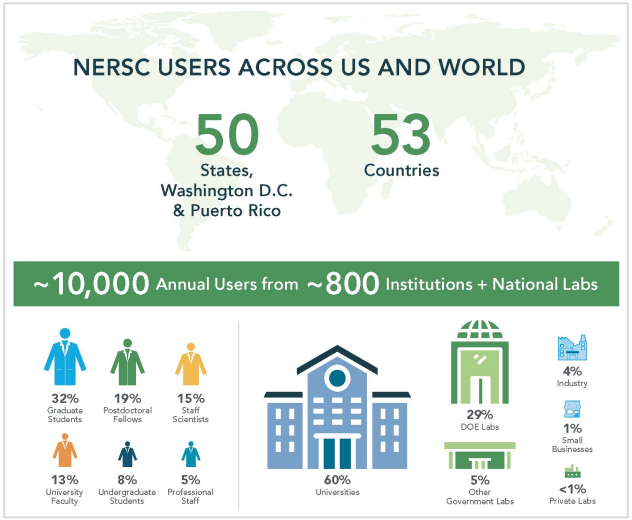
Anyone conducting Office of Science relevant research can apply to use NERSC, but you must demonstrate a need to use the unique capabilities offered by the center.

# NERSC has a dual mission to advance science and the state-of-the-art in supercomputing

- We collaborate with computer companies years before a system's delivery to deploy advanced systems with new capabilities at large-scale
- We provide a highly customized software and programming environment for science applications
- We are tightly-coupled with the workflows of DOE's experimental and observational facilities – ingesting tens of terabytes of data each day
- Our staff provide advanced application and system performance expertise to users

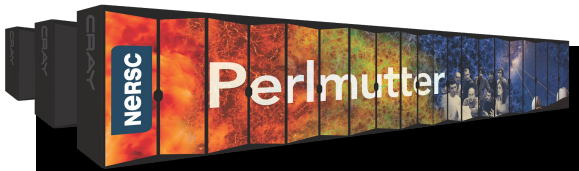



# Success Is Depth and Breadth of Scientific Impact





**NERSC has been acknowledged in 11,173 refereed scientific publications & high profile journals since 2020 ( > 2000 annually )**




## 2024 Nobel Prize in Chemistry

David Baker and his research group used AI and the 1.5 Million GPUs hours on Perlmutter to build "entirely new kinds of proteins."

*"Over the past year, enabled by the extensive resources provided by NERSC, we have made significant progress with both new methods development and application of protein structure prediction, design, and interaction prediction to reveal new biological insights."* ~ David Baker


8 NERSC acknowledgements

Nature (4)
Science (3)
Nature Structural & Molecular Biology (1)

# Nobel-Prize Winning Users

**2013 Chemistry**

*for the development of multiscale models for complex chemical systems*

Martin Karplus

**2011 Physics**

*for the discovery of the accelerating expansion of the Universe through observations of distant supernovae*

Saul Perlmutter

**2006 Physics**

*for the discovery of the blackbody form and anisotropy of the cosmic microwave background radiation*

George Smoot

**2007 Peace**

*for their efforts to build up and disseminate greater knowledge about man-made climate change*

Warren Washington

**2017 Chemistry**

*for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*

Joachim Frank

**2015 Physics**

*for the discovery of neutrino oscillations, which shows that neutrinos have mass*

SNO Collaboration

# Nobel-Prize Winning Users

**2013 Chemistry**

*for the development of multiscale models for complex chemical systems*

Martin Karplus

**2011 Physics**

*for the discovery of the accelerating expansion of the Universe through observations of distant supernovae*

Saul Perlmutter

**2006 Physics**

*for the discovery of the blackbody form and anisotropy of the cosmic microwave background radiation*

George Smoot

**2007 Peace**

*for their efforts to build up and disseminate greater knowledge about man-made climate change*

Warren Washington

**2017 Chemistry**

*for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*

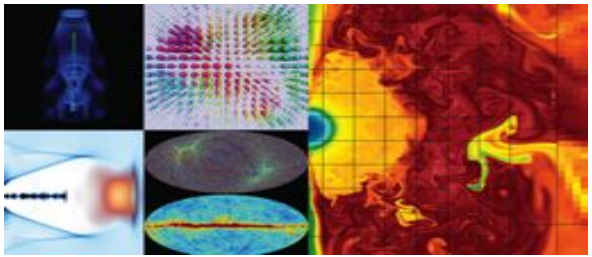
Joachim Frank

**2015 Physics**

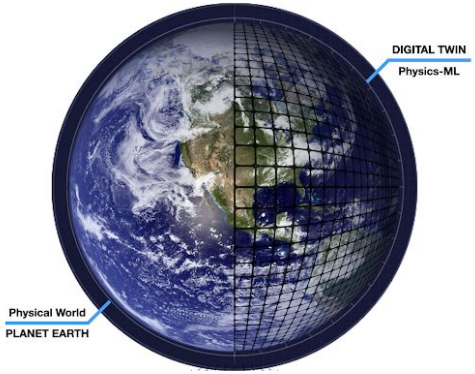
*for the discovery of neutrino oscillations, which shows that neutrinos have mass*

SNO Collaboration

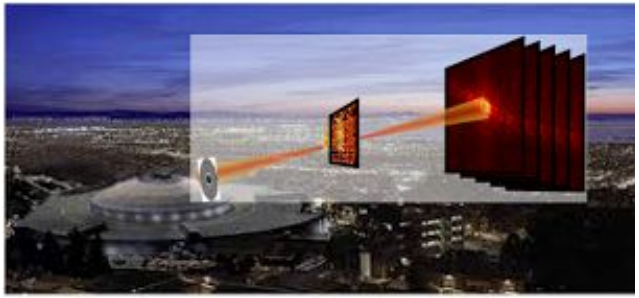
# We Accelerate Scientific Discovery for Thousands of Office of Science Users with 3 Advanced Capability Thrusts



Large-scale applications for simulation, modeling and data analysis



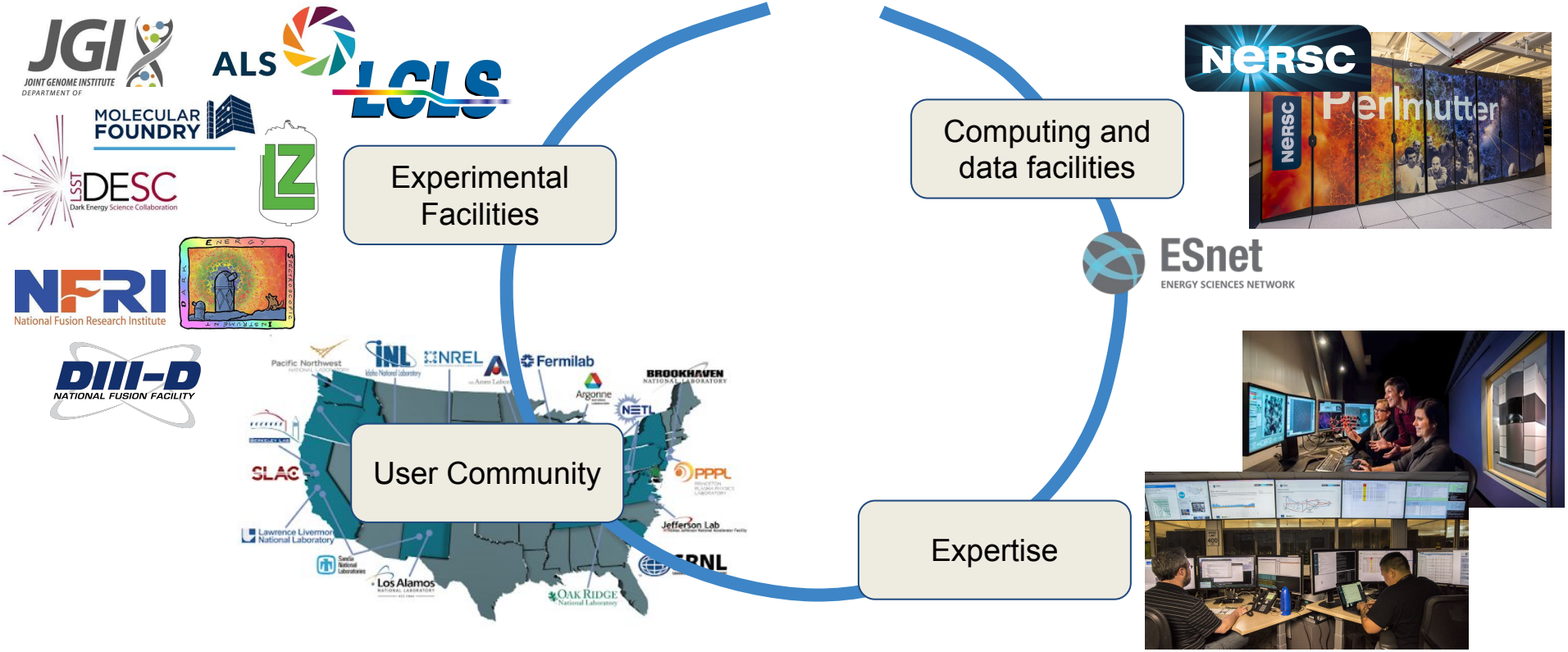

Complex experimental and AI driven workflows



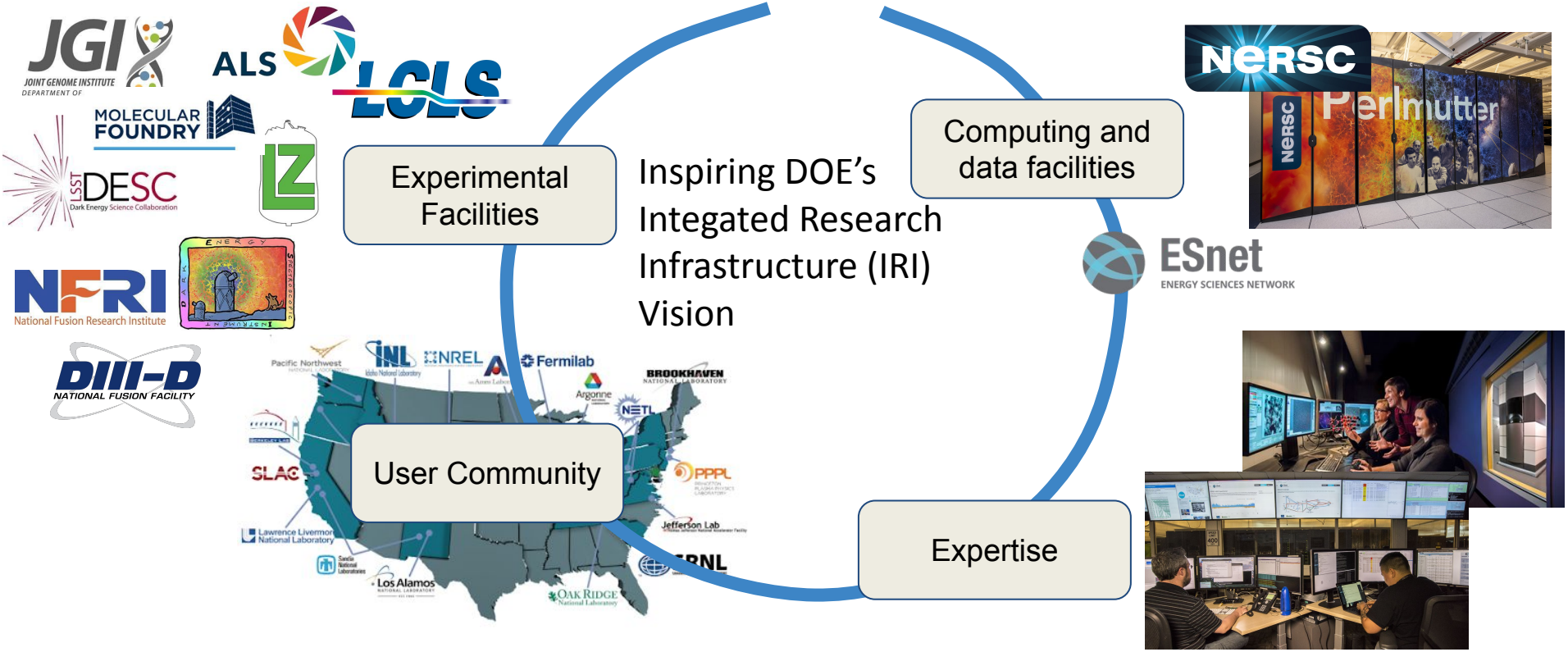
Time-sensitive and interactive computing

**The NERSC workload is diverse with growing emphasis on integrated research workflows**

# “Superfacility” connects experiment and compute facilities with the expertise and community they need for success



# “Superfacility” connects experiment and compute facilities with the expertise and community they need for success



# DOE's Integrated Research Infrastructure (IRI) Vision:

*To empower researchers to meld DOE's world-class research tools, infrastructure, and user facilities seamlessly and securely in novel ways to radically accelerate discovery and innovation*


**3 IRI Science Patterns**

- Time-Sensitive
- Data-Integration
- Long Campaign

**Experimental and Observational User Facilities**


**Advanced Networking**

ESnet
ENERGY SCIENCES NETWORK

**Edge Sensors**

**Local Campus Computing**

**Computing Testbeds**

**Cloud Computing**

**Advanced Computing**


**Researchers**

**Data Management**

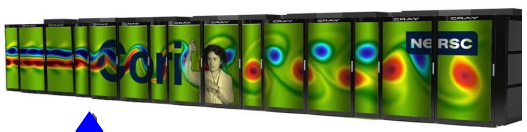
- High Performance Data Facility
- Data Repositories PuRE Data Assets

**Software**

- Software and Applications
- AI Tools Digital Twins

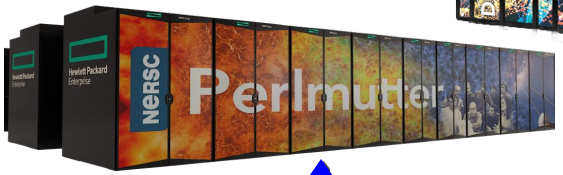
Slide adapted from Ben Brown, ASCR

# NERSC Systems Roadmap



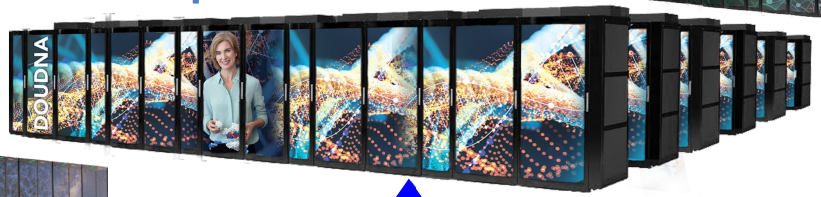

**NERSC-8: Cori**
Manycore CPU architectures

**2016**




**NERSC-9: Perlmutter**
CPU and GPU nodes
Expanded Simulation, Learning & Data

**2020**

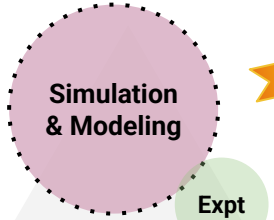


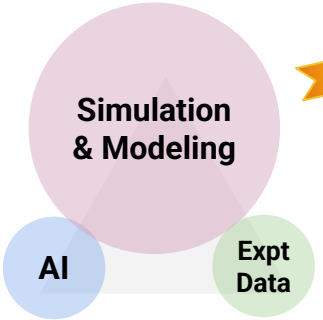

**NERSC-10: Doudna**
Accelerating end-to-end workflows

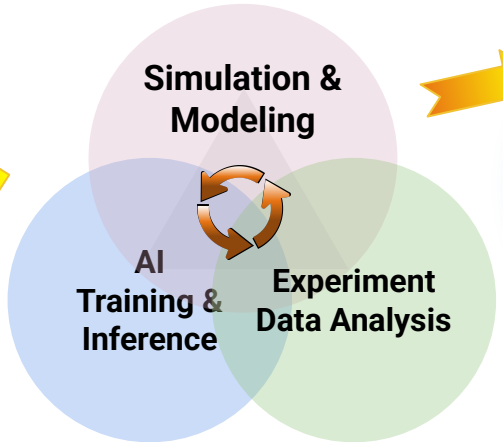
**2026**

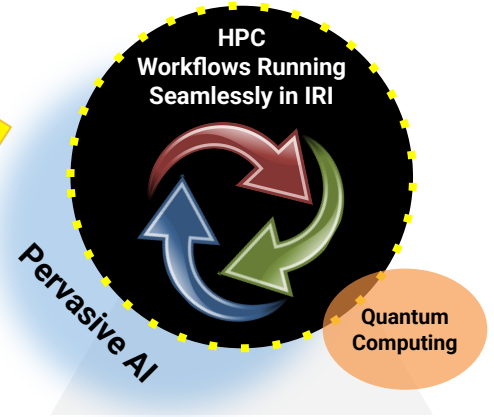
**NERSC-11:**
Beyond Moore

**2030+**



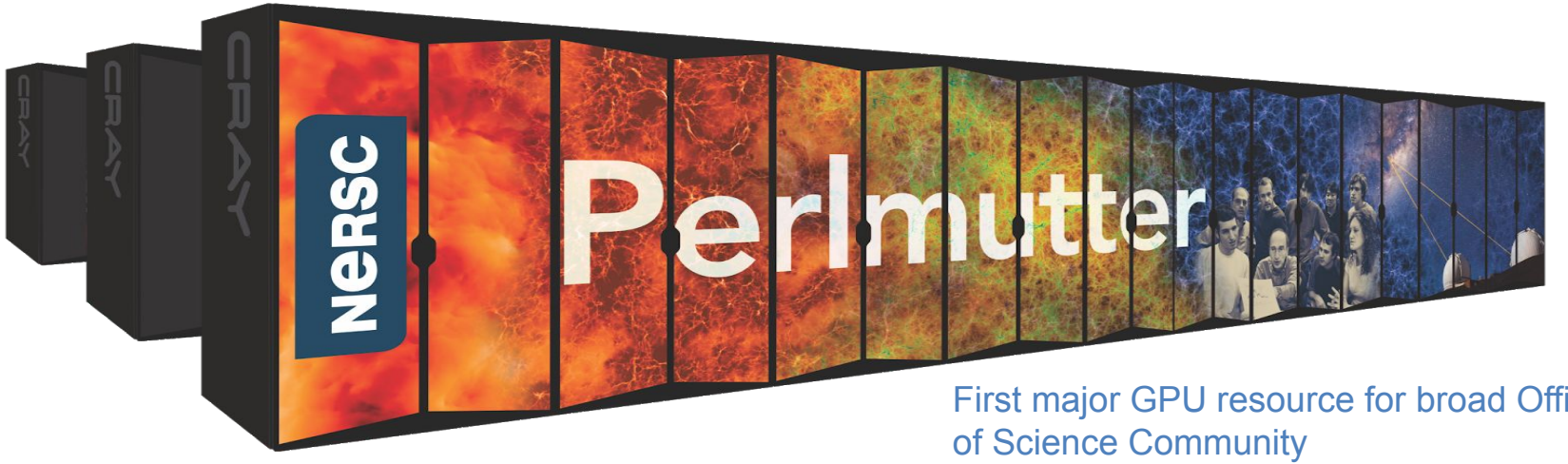



# Perlmutter

Debuted as #5 most powerful supercomputer in the world in 2021; #19 as of November 2024




First major GPU resource for broad Office of Science Community

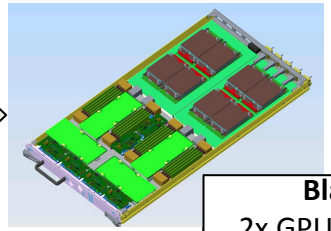
- 1,792 GPU Accelerated Nodes
  - 4 NVIDIA A100 per node
  - 7,168 GPUs
  - 1 AMD “Milan” per node
- 35 PB FLASH scratch filesystem

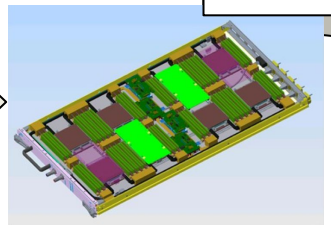
- 3,072 CPU only nodes
  - 2 AMD “Milan” per node
  - 512 GB memory per node
  - 393K cores
- HPE Cray “Slingshot 11” network

# Perlmutter system configuration

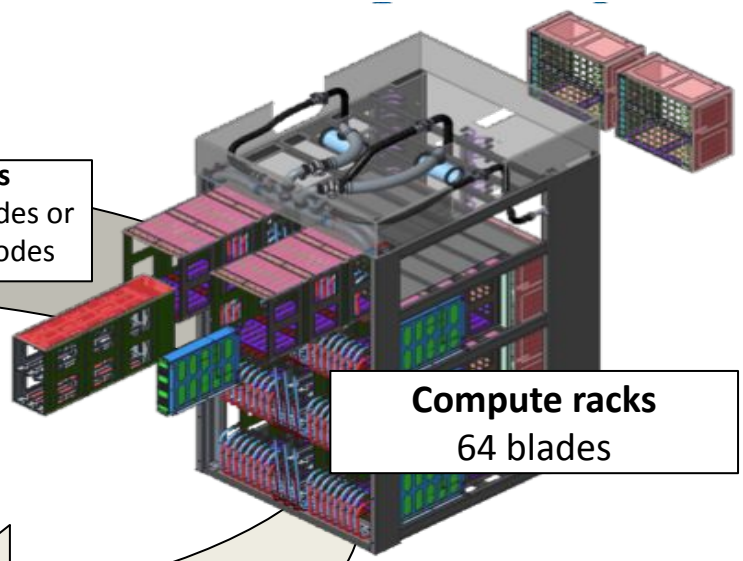
**NVIDIA "Ampere" GPU Nodes**
4x GPU + 1x CPU (>75 TF)
160 GiB HBM + DDR
4x 200G "Slingshot" NICs

**AMD "Milan" CPU Node**
2x CPUs
> 256 GiB DDR4
1x 200G "Slingshot" NIC





**Blades**
2x GPU nodes or
4x CPU nodes



**Compute racks**
64 blades

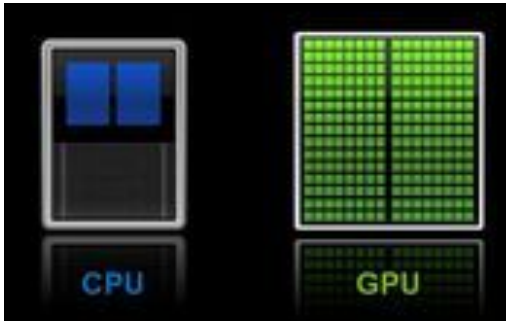
**Centers of Excellence**
*Network*
*Storage*
*App. Readiness*
*System SW*

**Perlmutter system**
GPU racks
CPU racks
~6 MW

# Optimization: CPUs vs GPUs

CPU (Haswell)

- 64 cores
- 2 threads each
- 2x256-bit vectors
- double precision
  - ~2000 way parallelism (64*2*2*8)




GPU (A100)

- 108 SM
- Up to 64 warps per SM (2 active at a time)
- 32 SIMT per warp
- double precision
  - 200,000+ way parallelism (108*64*32)

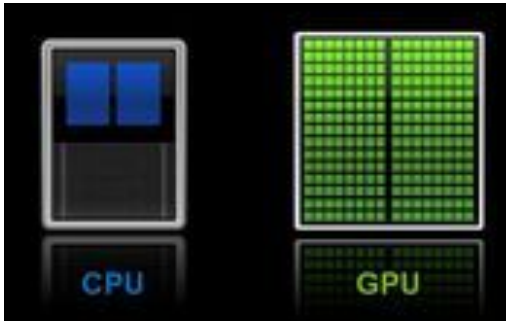
CPU - Speed

GPU - Throughput

# Optimization: CPUs vs GPUs

CPU (Haswell)

- 64 cores
- 2 threads each
- 2x256-bit vectors
- double precision
  - ~2000 way parallelism (64*4*8)




GPU (A100)

- 108 SM
- Up to 64 warps per SM (2 active at a time)
- 32 SIMT per warp
- double precision
  - 200,000+ way parallelism (108*64*32)

CPU - Speed

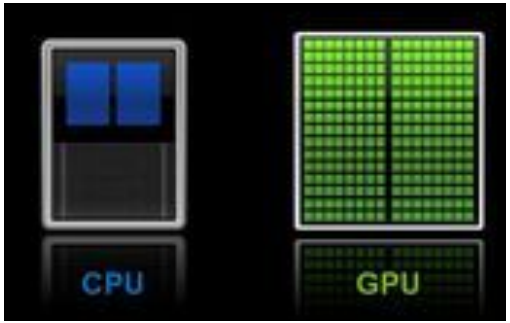
GPU - Throughput

Oversubscribing GPUs (w/ Warps and Streams) helps hide latency, too!

# GPUs vs. CPUs Memory Bandwidth

CPU (Haswell)

- 128GB DDR
- ~120 GB/Sec Memory Bandwidth




GPU (A100)

- 40GB HBM
- 1,500 GB/Sec Memory Bandwidth

PCIe ~ 32 GB/Sec

CPU - Speed

GPU - Throughput

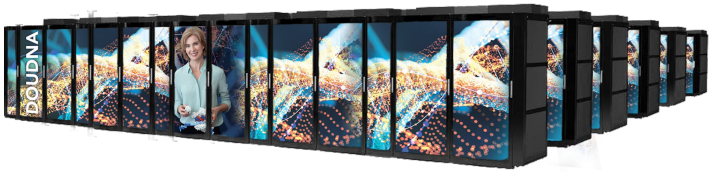
Try to avoid moving data back and forth frequently

# Doudna: A Supercomputer for Complex, Integrated Workflows

***Users require*** an ***integrated ecosystem*** that supports new paradigms for ***data analysis*** *with* ***real-time interactive feedback*** *between experiments and simulations. Users need the ability to* ***search, analyze, reuse,*** *and* ***combine data*** *from* ***different sources*** *into* ***large scale simulations and AI models.***

**NERSC-10 Mission Need Statement**: The NERSC-10 system will accelerate end-to-end DOE SC workflows and enable new modes of scientific discovery through the integration of experiment, data analysis, and simulation.

# *Doudna* System Overview




Designed to provide > 10x performance over *Perlmutter* and support diverse and complex DOE SC workflows.

**NVIDIA User Software**

## Vera-Rubin CPU-GPU

NVIDIA **Vera-Rubin Compute Node**

- Able to support GPU and CPU portions of complex workflows
- Integrated AI and Compute Optimized Partitions

## High Speed Network

NVIDIA **Infiniband XDR (800 Gbps)**

- Quantum-3 Infiniband switches
- ConnectX-8 Infiniband NICs
- Unified Fabric Manager

### External Connectivity

- Skyway-Next IB-to-Eth Gateway

## 2 Types of Storage

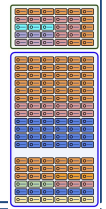
IBM **Platform Storage System (PSS)**
IBM Storage Scale System

VAST **Quality of Service Storage System (QSS)** on Dell PowerEdge Servers

## Workflow Environment

**Workflow Environment Nodes**
Heterogeneous node-types
CPU-only, Vera-Rubin
Air-cooled, Water-cooled

Reconfigurable to support complex IRI workflows (e.g. batch, compile, Jupyter, cloud-native, data transfer, etc.)

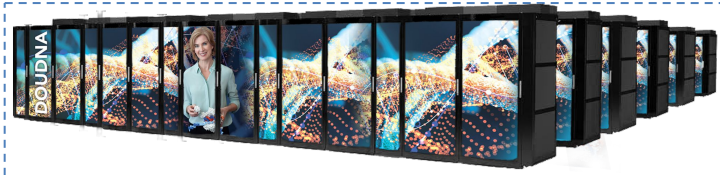


DELL **Dell System Management Tools (OME, iDRAC, Omnia)**

**Dell ORv3 Direct liquid-cooled server technology & integrated rack scalable systems**

# *Doudna* Drives the Integrated Data Center Ecosystem

***Scientific breakthroughs in AI and at Experimental Facilities enabled through high-speed access to DOE SC Community Data***




**> 10x performance over *Perlmutter* and supports diverse and complex DOE SC workflows**

### Workload Optimized Compute Capabilities

- Able to support GPU and CPU portions of complex workflows
- Integrated AI and Compute Optimized Partitions

### 2 Types of Storage Capabilities

**Platform Storage System (PSS) For large-scale simulation & modeling**

**Quality of Service Storage System (QSS) for data intensive AI and Experiment**

### Workflow Environment

**Workflow Environment Nodes**

Reconfigurable to support complex IRI workflows (e.g. batch, compile, Jupyter, cloud-native, data transfer, etc.) with integrated Spin capabilities.

*200 GB/s*

**HPSS Data Archive (375 PB)**
Demand is growing by 30% each year
Data can be reused & leveraged for AI training

*> 800 GB/s*

**Community File System (150 PB)**
Allows sharing data between users, systems and the 'outside' - medium performance

*3.25 TB/s (26 Tbps)*

**NERSC Data Center Network**
**Ethernet LAN**

*> 10 GB/s*

**/home (400 TB)**
Permanent storage and backup for important files

**Data Transfer Nodes**
Transfers data between NERSC storage resources and externally to resources at other sites

ESnet ENERGY SCIENCES NETWORK
2 x 400 Gb/s
2 x 100 Gb/s

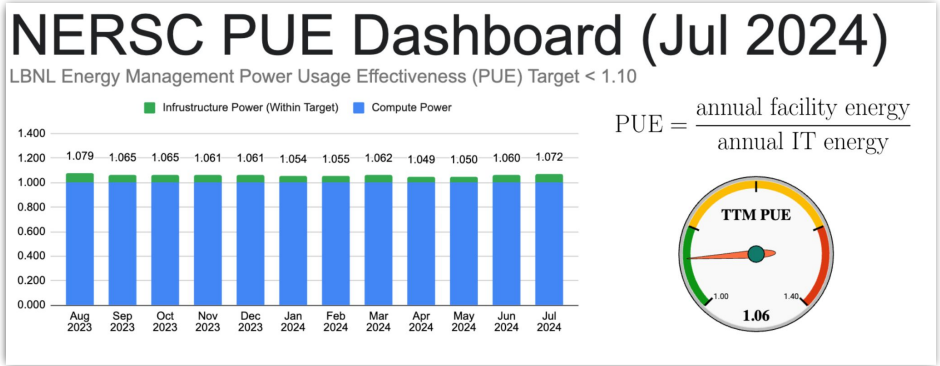
**Experimental Facility** | **ASCR Facility** | **Home Institution** | **Cloud** | **Edge**

# Shyh Wang Hall - a Highly Efficient Data Center

- **Power & Cooling Capabilities (chiller-free)**
  - 25.0 MW peak facility power currently (~20 MW HPC power)
  - 35.0 MW peak facility power after N10 upgrade (~30 MW HPC power)
- **Energy Efficiency**
  - Comprehensive annualized Power Usage Efficiency (PUE) of 1.06 (All compute systems included)
- **Awards and Recognition**
  - USGBC LEED Gold Facility
  - DOE Office of Management Sustainable Performance Division awards:
    - 2020 LBNL's NERSC Efficiency Optimization Team
    - 2021 LBNL's ISO 50001 Implementation
  - Clean Energy Ministerial – 2021 Energy Management Insight Award





# NESAP - NERSC Science Acceleration Program - is Our Vehicle for Preparing Users for Future Systems

Collaboration between **NERSC, Strategic Science Teams**, and **Technology Vendors** to prepare science workflows for systems at NERSC

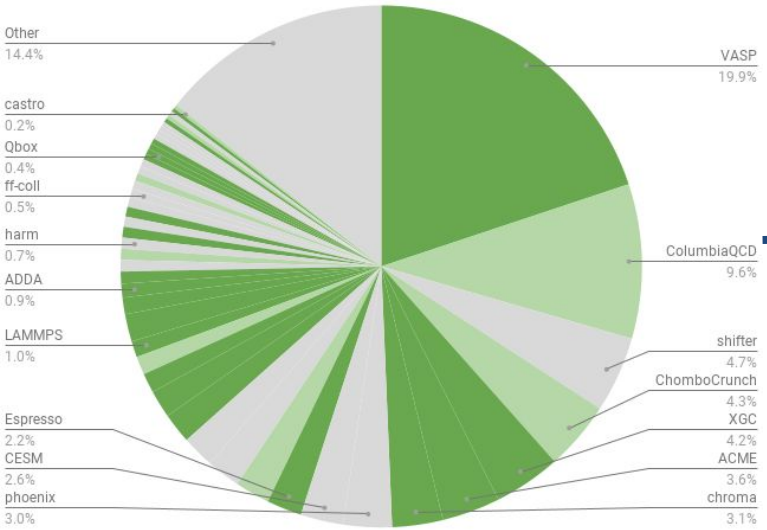
The partnership include:

- Science Application + Technology **CoDesign**
- **Early Access** to Advanced Hardware
- Collaborative **Code Development, Analysis and Optimization**
- Mutual **Staff Embedding, Training**
- Creation of Lessons Learned and Best Practices to be shared with community

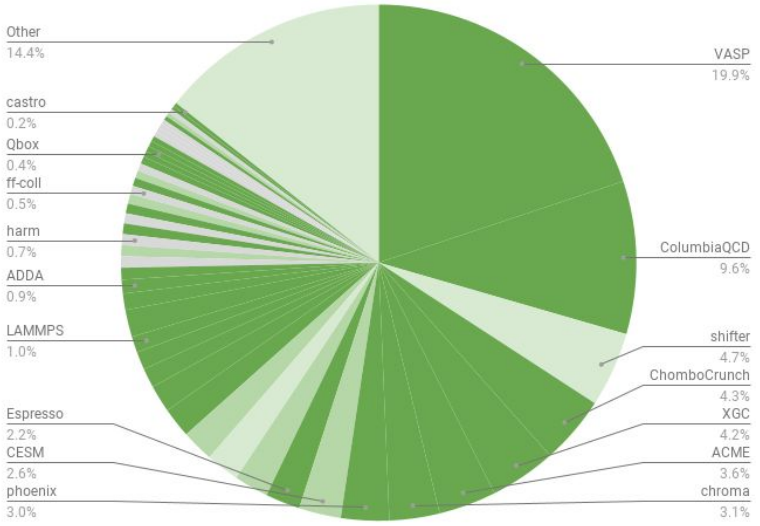
Proven Successful on Cori and Perlmutter

| Application | NESAP Category | Perlmutter Speedup |
|---|---|---|
| FlowGAN | Learning | 89.7 |
| BerkeleyGW | Electronic Structure | 78.3 |
| ExaFEL | Data | 186.9 |
| WarpX | Particles & Grids | 27.2 |
| LAMMPs | Molecular Dynamics | 79.1 |
| Chroma | LQCD | 56.0 |
| **Mean** | | **73.5** |

# NESAP is NERSC's Vehicle for Preparing Users for Current and Future Systems




GPU Readiness of NERSC Workload Before NESAP



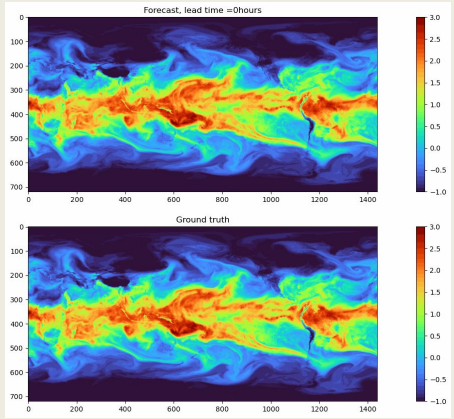

GPU Readiness of NERSC Workload After NESAP

# NESAP and Perlmutter Enabled Adoption of Large-scale and Groundbreaking AI

## FourCastNet

Pathak et al. 2022 **arXiv:2202.11214**
Collab with Nvidia, Caltech, …

- Forecasts global weather at high-resolution.
- Prediction skill of numerical model; 10000s times faster




Jaideep Pathak
former NERSC
Postdoc now NVIDIA

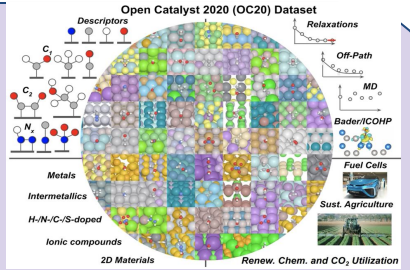
Shashank
Subramanian
Former NERSC
Postdoc now Staff

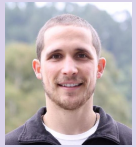
Jared Willard
NERSC Postdoc

## CatalysisDL

Chanussot et al. 2021
Collab with CMU, MetaAI, …
arXiv:2010.09990

- NeurIPS 2021-23 Competitions
- Pre-trained models now used with DFT - e.g. FineTuna; AdsorbML



Brandon Wood
former NERSC
Postdoc now Meta AI

Wenbin Xu
NERSC postdoc

# NERSC is Creating an HPC Workforce

## Current NESAP Postdocs

| Name | Workflow |
|---|---|
| Soham Ghosh | Materials Science |
| Vinicius Mikuni | HEP-ML |
| Andrew Naylor | U.S. CMS, HEP Mini-apps |
| Mukul Dave | WarpX |
| Wenbin Xu | Catalysis, BonDNet |
| Jared Willard | iNAIADS, FourCastNeXT |
| Sam Welborn | Superfacility, NCEM |
| Madan Timalsina | Superfacility, U.S. CMS |

## Recent NESAP PostDocs

| Name | Current Position |
|---|---|
| Taylor Barnes | NSF MOLSSI |
| Andrey Ovsyannikov | Intel |
| Tuomas Koskela | University of Helsinki |
| Mathieu Lobet | CEA |
| Tareq Malas | Meta |
| Zahra Ronaghi | NVIDIA |
| Bill Arndt | LBNL / NERSC |
| Jonathan Madsen | AMD |
| Michael Rowan | AMD |
| Kevin Gott | LBNL / NERSC |
| Muaaz Awan | LBNL / NERSC |
| Laurie Stephey | CalEPA |
| Oisin Creaner | Dublin Institute of Advanced Tech. |
| Ozgur Cekmer | CSIRO |
| Yunsong Wang | NVIDIA |
| Yan Zhang | Velodyne Lidar |
| Hugo Brunie | CEA (France) |
| Dossay Oryspayev | Brookhaven Lab |
| Brandon Wood | AI at Meta |
| Dhruva Kulkarni | LBNL / NERSC |
| Amanda Dufek | LBNL / NERSC |
| Raphael Prat | CEA (France) |
| Jaideep Pathak | NVIDIA |
| Daniel Margala | LBNL / NERSC |
| Muhammad Haseeb | NVIDIA |
| Lipi Gupta | LBNL / NERSC |
| Nestor Demeure | LBNL / NERSC |

# Hackathons - Apply Now - openhackathons.org

## Hackathons Are a Central Component for Preparing for New System

Public GPU Hackathons (https://www.openhackathons.org) - NVIDIA and NERSC (along with ORNL and ANL) provided more team mentors than any other institutions to worldwide events in past years.

**Future hackathons focus on a broader set of themes:**

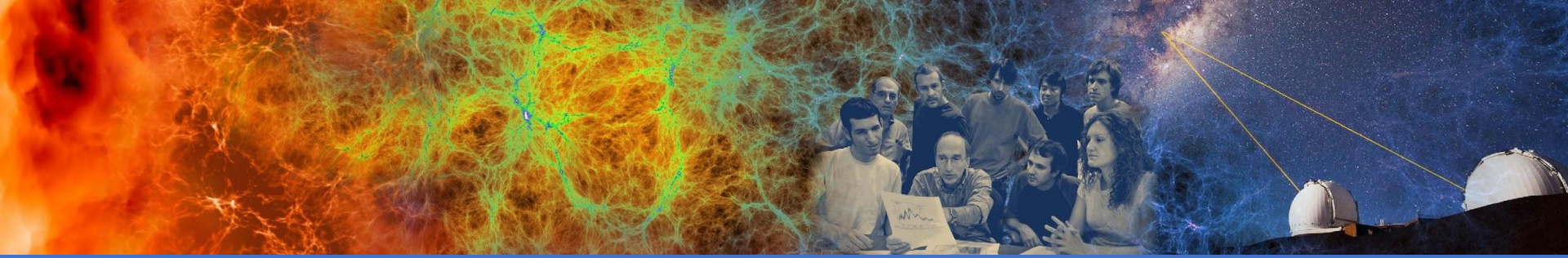
- Distributed AI Training
- AI + ModSim Coupling
- Co-Scheduling Complex Workflows
- Reduced Precision
- Performance Portability
- Container Performance
- Library Development/Optimization
- I/O Optimization
- Communication Optimization

Community GPU Hackathons

Hackathons are impactful events for both technical and sociological reasons. They provide an energetic environment where progress is made quickly.

# Final Thoughts

- NERSC has deployed numerous innovative supercomputers
- *Doudna* will be a first-of-a-kind supercomputer in 2027
- Result has been many scientific breakthroughs and publications
- The next 10-20 years will see great change
  - Expansion of the mission space (complex coupled workflows, Superfacility, IRI, AI)
  - End of Moore's Law
  - An opportunity to use AI to revolutionize science and operations



# Questions?