Argonne NATIONAL LABORATORY

(Re)Introducing Aurora

The Road to Exascale and Beyond

Ti Leggett
Deputy Director of Operations & Deputy Project Director ALCF-3
Argonne Leadership Computing Facility

Argonne National Laboratory

For seven decades, the U.S. Department of Energy's Argonne National Laboratory has excelled in integrating world-class science, engineering, and user facilities to deliver innovative research and technologies and new knowledge that addresses the scientific and societal needs of our nation.

About Argonne

Argonne is a multidisciplinary science and engineering research center located outside Chicago.

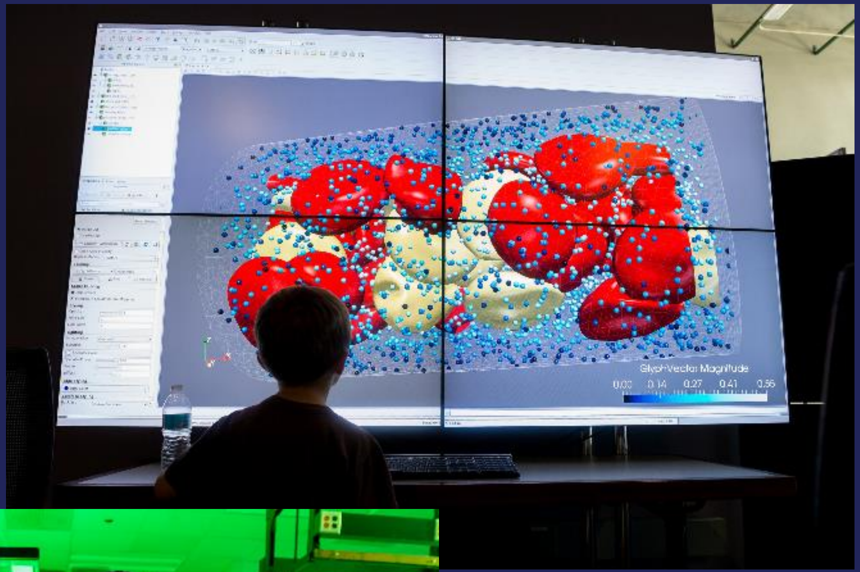
- **Born out of the University of Chicago's work on the Manhattan Project in the 1940s.**
- **Managed by UChicago Argonne, LLC, for the U.S. Department of Energy's Office of Science**
- **Works with universities, industry, and other national labs on questions and experiments too large for any one institution to do by itself.**

Lab at a Glance

Fiscal Year 2017
Budget: $750 million / Procurement: $270 million

Workforce
3,200 total employees
270 postdoctoral scholars
569 graduate and undergrad students
274 joint faculty
8,300 facility users
1,107 visiting scientists

Research
16 research divisions
5 national scientific user facilities
Many centers, joint institutes, program offices
Hundreds of research partners



Argonne NATIONAL LABORATORY

Office of Science User Facilities at Argonne National Laboratory

DOE Office of Science user facilities provide the research community with the most advanced tools for modern science.

- Advanced Photon Source
- Argonne Leadership Computing Facility
- Argonne Tandem Linear Accelerator System
- ARM Southern Great Plains
- Center for Nanoscale Materials

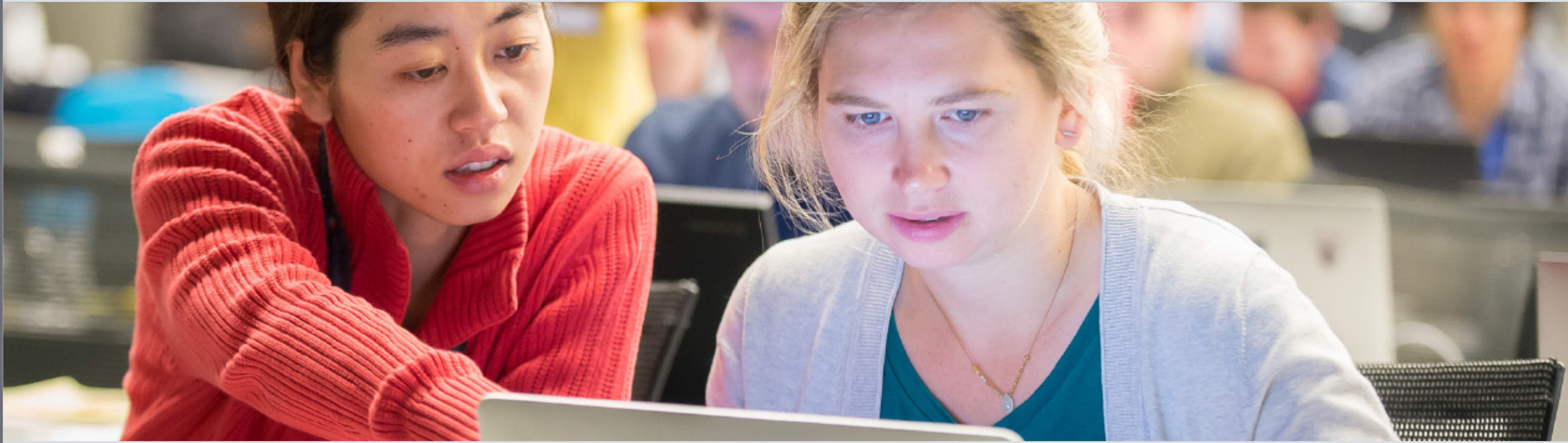
What is the Argonne Leadership Computing Facility?

The Argonne Leadership Computing Facility provides world-class computing resources to the scientific community.

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

A Community of Users

Our community is made up of researchers from academia, industry, and government labs working in a wide range of disciplines.



In-House Expertise

ALCF teams play a critical role in supporting the facility's supercomputing environments, the user community, and their efforts to accelerate scientific discoveries.

ALCF researchers lead and participate in several strategic activities that aim to push the boundaries of what's possible in computational science and engineering.

ALCF Growth and Impact

Hours requested vs. allocated:

~2X per year

~3X per year

Hours allocated	4.9 M	6.5 M	18.2 M	95 M	268 M	889 M	1.6 B	1.7 B	1.7 B	4.7 B	5.8 B	5.8 B	6.2 B	8.4 B
Projects	3	3	15	45	55	66	69	57	60	61	59	56	60	91

Researchers solved the 2D Hubbard model and presented evidence that it predicts HTSC behavior. ***Phys. Rev. Lett*** (2005)

Largest simulation of a galaxy's worth of dark matter, showed for the first time the fractal-like appearance of dark matter substructures. ***Nature*** (2008), ***Science*** (2009)

Unprecedented simulation of magnitude-8 earthquake over 125-square miles. **Proc. SC10**

Calculation of the number of bound nuclei in nature. ***Nature*** (2012)

NIST proposes new standard reference materials from LCF concrete simulations.

Recovery from slow inactivation in potassium channels controlled by H_2O. ***Nature*** (2013)

Carbon-based tribofilms from lubricating oils. ***Nature*** (2016)

Quantitative 3D evolution of colloidal nanoparticle oxidation in solution. ***Science*** (2017)

2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017

Modeling of molecular basis of Parkinson's disease named #1 computational accomplishment. ***Breakthroughs*** (2008)

World's first continuous simulation of 21,000 years of Earth's climate history. ***Science*** (2009)

Largest-ever LES of a full-sized commercial combustion chamber used in an existing helicopter turbine. **Compte Rendus Mecanique** (2009)

OMEN breaks the petascale barrier using more than 220,000 cores. **Proc. SC10**

New method to rapidly determine protein structure, with limited experimental data. ***Science*** (2010), ***Nature*** (2011)

Macroscale superlubricity enabled by graphene nanoscroll formation. ***Science*** (2015)

Ultra-selective high-flux membranes from directly synthesized zeolite nanosheets. ***Nature*** (2017)

What is a Supercomputer?

What is a computer?

- Input
- Output
- Storage
- Process Information

Supercomputing Resources

Our supercomputers are 10 to 100 times more powerful than systems typically used for scientific research.

What's the Difference?

- **A typical laptop can do about 89,000,000,000 math calculations every second**
  - Or 89 gigaFLOPS

- **The fastest supercomputers today can do over 100,000,000,000,000,000 math calculations every second**
  - Or 100 petaFLOPS
  - Over 1 million times as fast as a laptop

ALCF Computing Resources

Mira IBM BG/Q
49,152 nodes
786,432 cores
768 TiB RAM
Peak flop rate: 10 PF

Theta Cray XC40
4,392 nodes
281,088 cores
892 TiB RAM
Peak flop rate: 11.69 PF

Cetus IBM BG/Q
4,096 nodes
65,536 cores
64 TiB RAM
Peak flop rate: 836 TF

Cooley Cray/NVIDIA
126 nodes
1512 Intel Haswell CPU cores
126 NVIDIA Tesla K80 GPUs
48 TB RAM / 3 TB GPU

Iota Intel/Cray XC40
44 nodes
2,816 cores
8.9 TiB RAM
Peak flop rate: 117 TF

Firestone IBM Power8
2 nodes + K80 GPU
20 cores
128 GB RAM
Hybrid CPU/GPU

Storage Capability
Disk
- Mira: ~27 PB of GPFS file system capacity with performance of 240 GB/s on the largest file system (19PB).
- Theta: ~18 PB of GPFS/Lustre file system capacity; 9PB is GPFS and 9.2PB is Lustre.

Tape
- The ALCF has three 10,000-slot libraries using LTO 6 tape technology. The LTO tape drives have built-in hardware compression for an effective capacity of 36-60 PB.

Theta – ALCF's newest production system

Features Intel processors and interconnect technology, a new memory architecture, and a Lustre-based parallel filesystem – all integrated by Cray's HPC software stack

Argonne NATIONAL LABORATORY

What is a Supercomputer Used For?

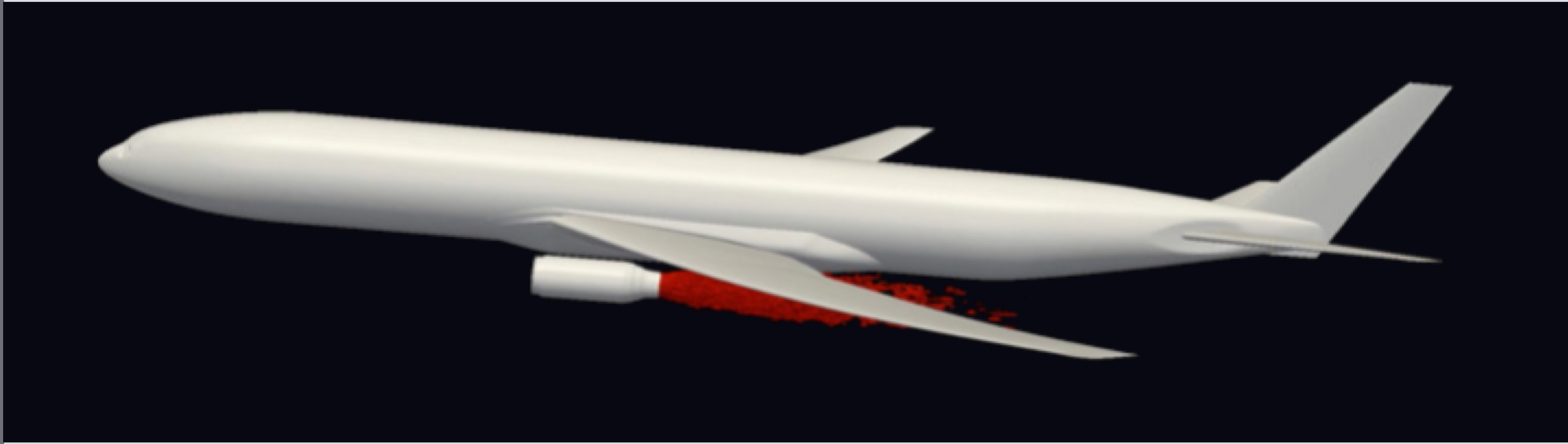
The scientific output of our nation is increasingly supported by computational science – the so-called third pillar of scientific inquiry, along with theory and experiment.

To prepare for future exascale systems, the ALCF is driving a new paradigm for scientific computing.

— Modeling & Simulation
— Data Science
— Machine Learning

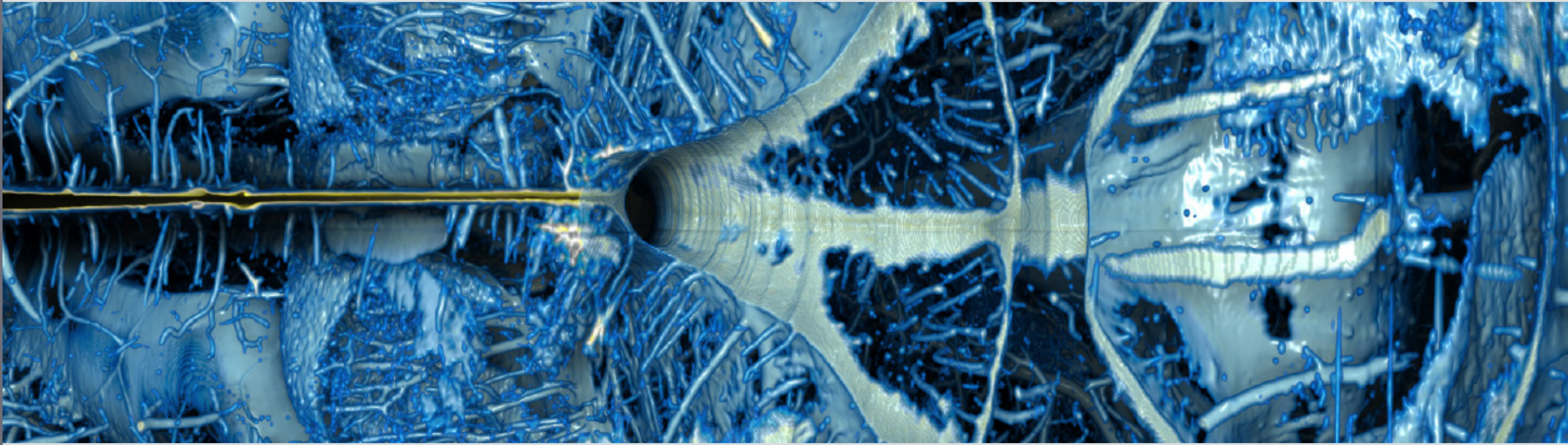
Modeling & Simulation

Simulation can be used to study things that are too big, too small, or too dangerous to study in a laboratory setting.



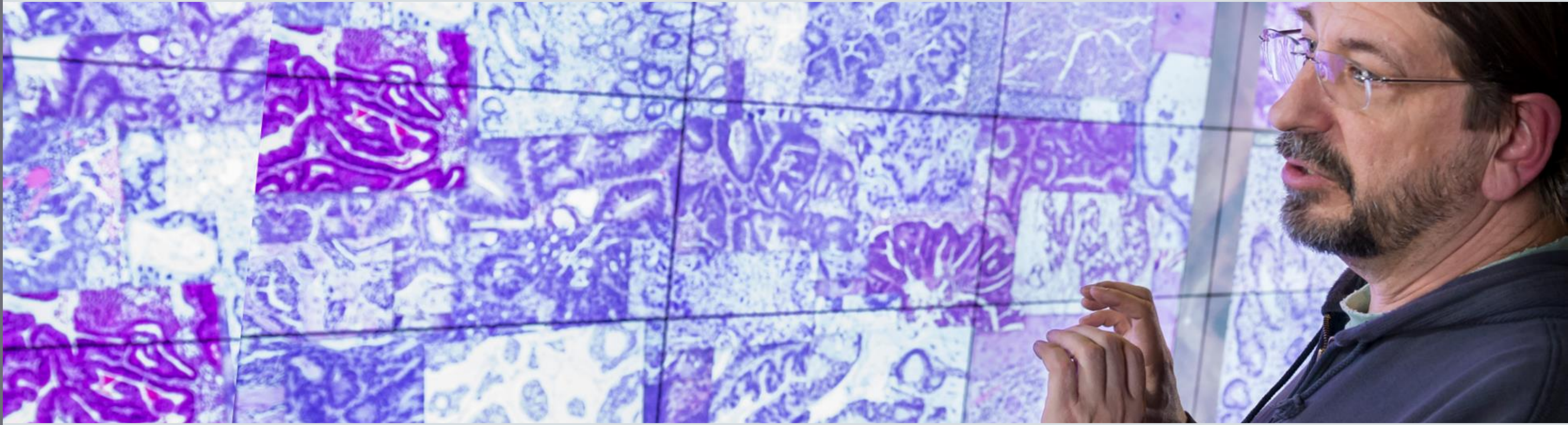
Data Science

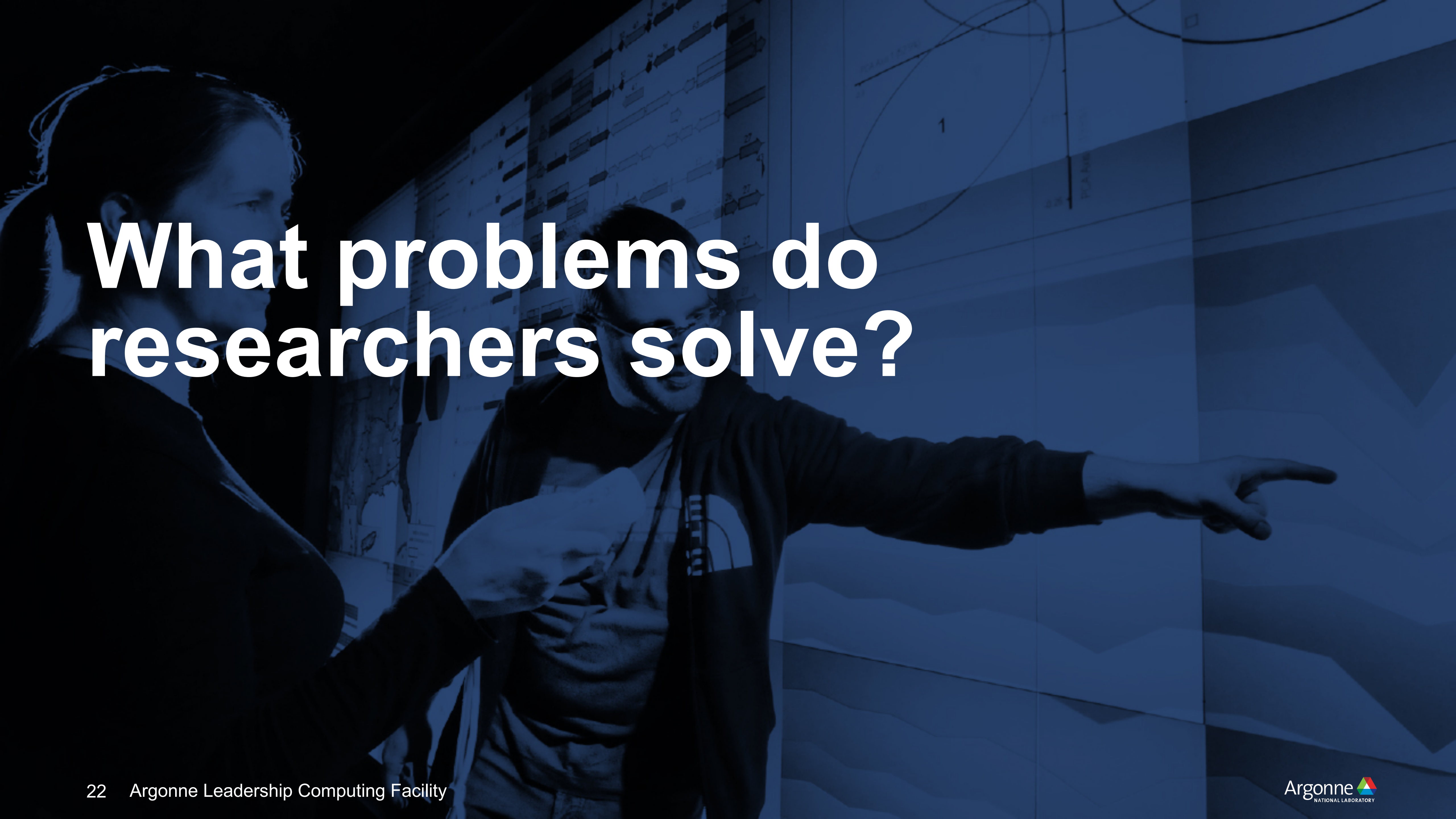
Researchers can glean insights from very large datasets produced by experimental, simulation, or observational methods.



Machine Learning

Machine learning is a type of artificial intelligence that trains computers to discover hidden patterns in data to make novel predictions without being explicitly programmed.





What problems do researchers solve?

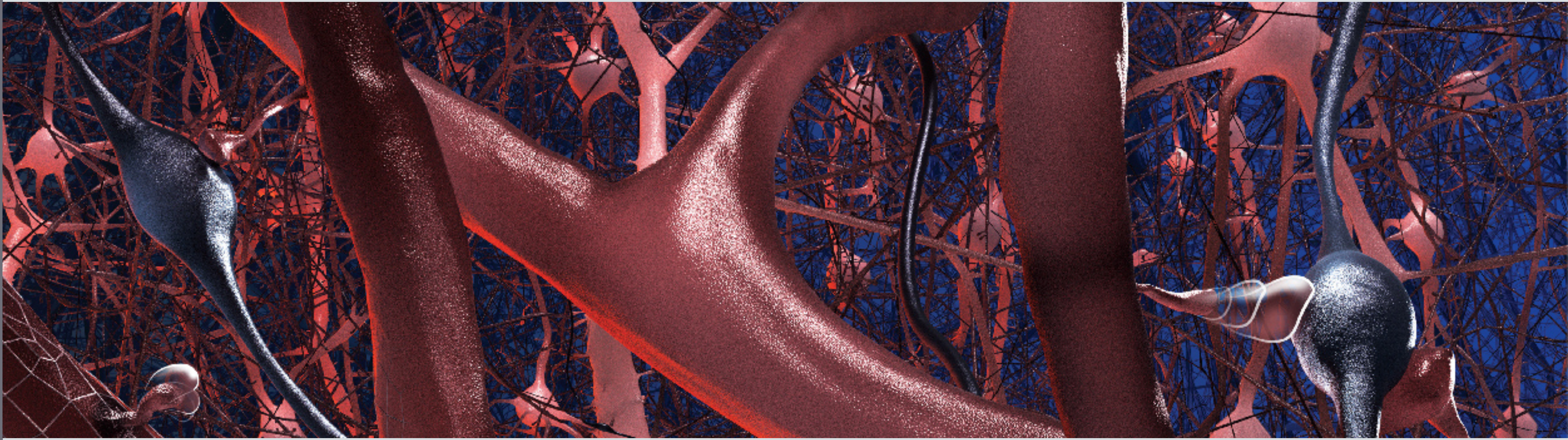
Research teams come to ALCF to work on some of the biggest challenges in the world.

— How does life work?
— What is our universe made of?
— How can we meet our energy needs?
— What technologies are on the horizon?

Biological Systems

From designing new drug therapies to understanding how our brain works, our supercomputers are essential for analyzing biological phenomena in precise molecular terms.

Researchers use simulation and modeling to study the complex behaviors and interactions a wide range of biological systems of increasing complexity—from macromolecular interactions to entire ecosystems.



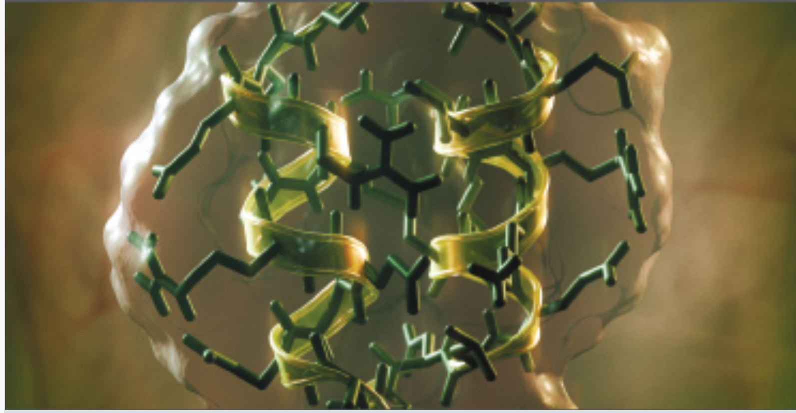
Toward Breakthroughs in Protein Structure Calculation And Design

PI: **David Baker, University of Washington**

Challenge: To design non-natural peptides with protein-like folds and activity.

Impact: Small synthetic peptides could potentially combine the advantages of small-molecule drugs and large protein therapeutics. This work is a major advancement toward designing therapeutic peptides that perfectly complement target molecules for diseases such as Ebola, HIV, antibiotic-resistant bacterial infections, and Alzheimer's.

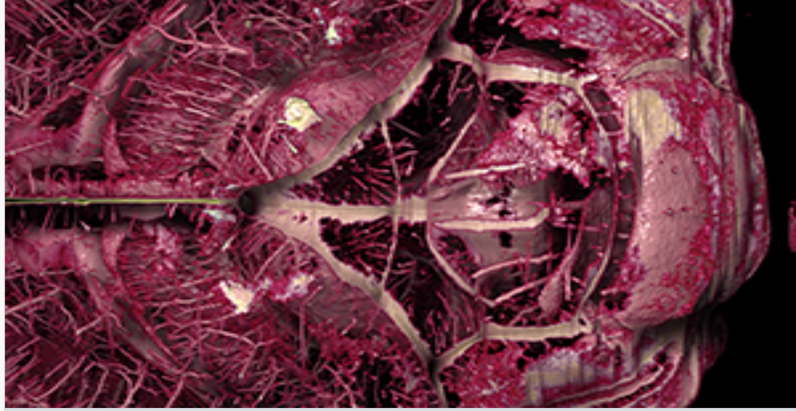
Approach: The Baker team uses Mira to design and verify stable versions of synthetic peptides. The computational design methods and stable scaffolds generated provide a promising starting point for the development of a new generation of peptide-based drugs.



Synthetic, or non-natural, peptides represent a new class of drugs that have potential for greater efficacy and fewer side effects.

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

Large-Scale Computing on the Connectomes of the Brain



PI: **Doga Gursoy, Argonne National Laboratory**

Challenge: To develop extreme-scale, data-centric pipelines for brain science that integrates exascale computational approaches.

Approach: Initial studies will focus on the reconstruction of mice brains utilizing novel imaging and analytical tools to image at the level of individual cells and blood vessels.

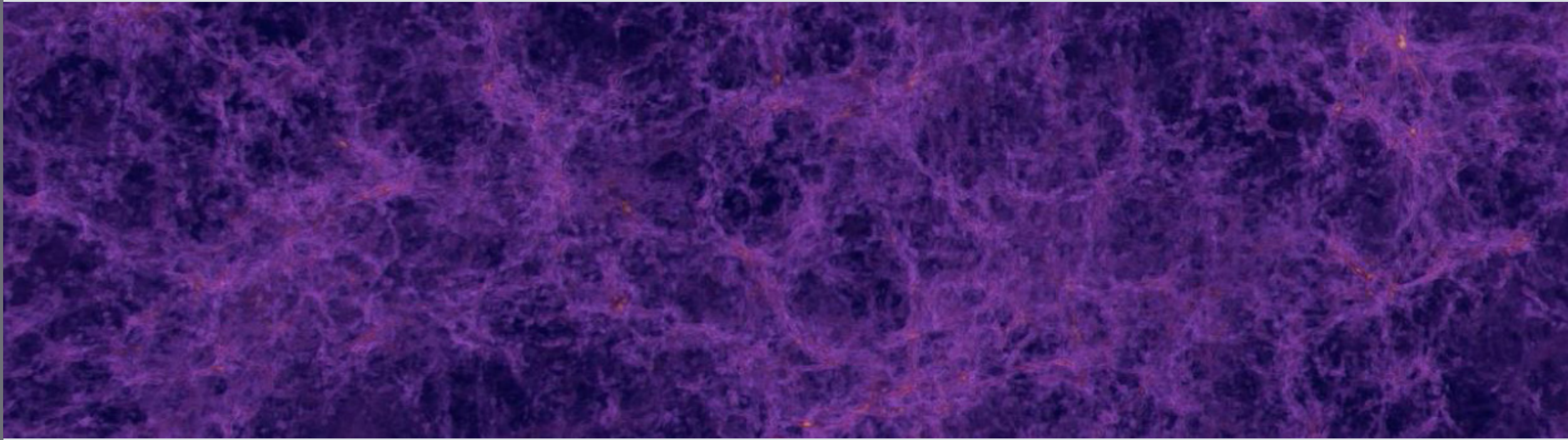
Impact: The workflows, focused on analysis and visualization of experimental data, will help researchers gain invaluable knowledge about disease models, such a Alzheimer's, autism spectrum disorder, and many others. Additionally, the insights gleaned will enable transformative advances in neuromorphic computing.

The combined techniques will, for the first time, allow researchers to compare potential organizational patterns across brains to distinguish which are genetic and which are unique.

The Evolution of Our Universe

Scientists are using our systems to simulate the formation of our universe, from the Big Bang to today.

Our supercomputers simulate the interaction of small bits of matter, represented by particles, with the various laws of physics over time to model galaxies, galaxy clusters, and superclusters.



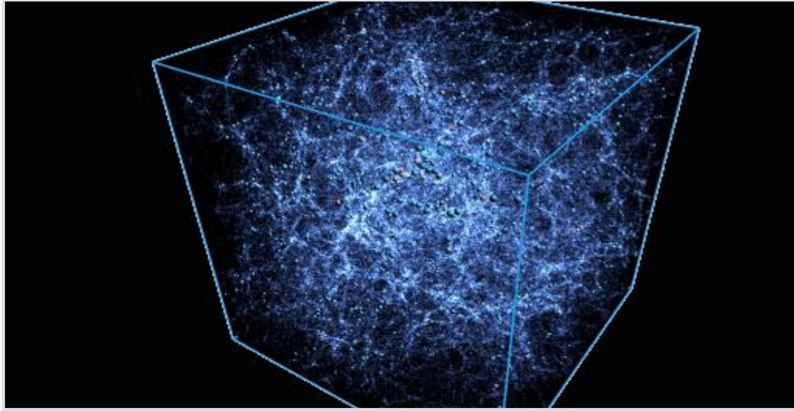
Cosmological Simulations for Large-Scale Sky Surveys

PI: **Salman Habib, Argonne National Laboratory**

Challenge: To generate realistic synthetic observations and sky catalogs to help constrain a host of systematic uncertainties.

Impact: Mira enables cosmology runs with greater resolution and accuracy on much larger simulation volumes, giving researchers the ability to confront theory with observational data from wide-area cosmological surveys.

Approach: This project focuses on generating precision prediction tools for different cosmological observables spanning a large range of parameters, and constructing sophisticated synthetic sky maps from very large high-resolution cosmological simulations.



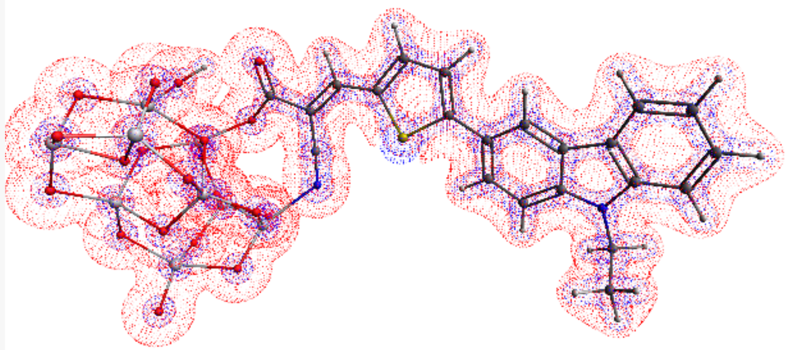
Large-scale sky surveys are key drivers of advances in modern cosmology. This project uses advanced computational tools to build accurate emulators to help resolve the mysteries of dark energy and dark matter.

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

Renewable Energy

Researchers use our supercomputers to explore a variety of processes and technologies aimed at expanding the nation's renewable energy portfolio to help meet growing energy demands. From developing more efficient wind turbines to identifying new materials for solar energy cells, supercomputers are helping accelerate the development of technologies that will ensure a cleaner and more secure energy future.

Molecular Design of Dye-Sensitized Solar Cells



PI: Jaqueline M. Cole, University of Cambridge and Argonne

Challenge: Dye-sensitized solar cells (DSSCs) are a next-generation photovoltaic technology whose transparent and low-cost nature make them a particularly strong contender for “smart windows” — windows that generate electricity from sunlight.

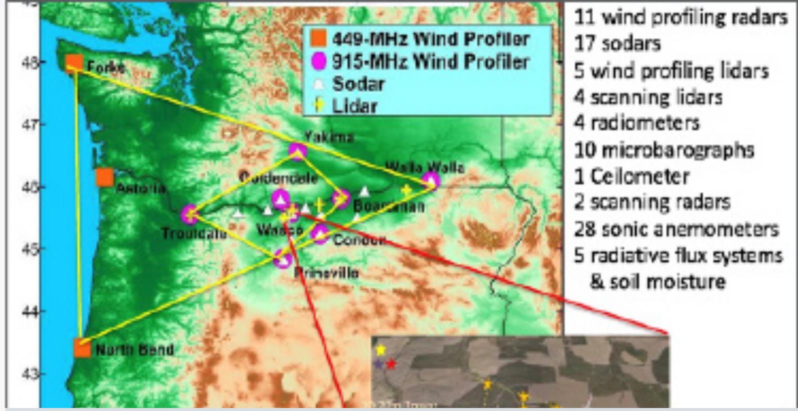
Impact: “Smart windows” that generate electricity from sunlight hold exciting prospects for meeting entire cities’ building energy demands in a fully sustainable fashion.

Approach: To use data science techniques to search through a representative set of all possible chemical molecules and use artificial intelligence to target the chemicals whose molecules have optical properties that would yield optimum device function in DSSCs.

This project marries the latest technical capabilities in natural language processing, machine learning, and quantum-chemical calculations to the world-leading supercomputing resources available at Argonne.

ALCF DATA SCIENCE PROGRAM

Validation of Forecast Models in Complex Terrains



PI: Joe Olson, NOAA

Challenge: To develop a 750-meter resolution forecast model for more accurate wind predictions in complex terrains, such as forests, mountains, and coastlines.

Impact: This model will help optimize how wind power is used on the electric grid and potentially introduce wind energy to new regions.

Approach: The team uses Mira to evaluate an experimental version of NOAA's High-Resolution Rapid Refresh weather model with complex, terrain-specific enhancements. The project simulates an entire year of weather forecasts in the Columbia River Gorge region to evaluate how forecasts have improved at the new resolution with improved physical parameterizations.

Utility operators rely on forecast models to predict how to balance wind energy on the grid with conventional power like coal and nuclear. While forecast models are effective at predicting wind on flat terrain, essential complex terrain data is missing.

ALCC

DIII-D Computing

PI: **David Schissel, General Atomics**

Challenge: Scientists at DIII-D National Fusion Facility run plasma physics experiments involving six-second pulses of confined plasma every 15-20 minutes. Planning for each new pulse is informed by data analysis of the previous pulse, however, the fine-grid analysis takes 20 minutes to complete on local resources.

Approach: General Atomics scientists and ALCF staff established a pipeline to compute the analysis of every single plasma pulse on ALCF resources and return the results to the DIII-D team in time to calibrate the next one. ALCF's advanced capabilities also enabled the between-pulse simulation to run at 4x the resolution previously used at GA.

Impact: ALCF is expanding its services to include near-real-time capabilities that can help large experimental and observational efforts make better use of their resources.

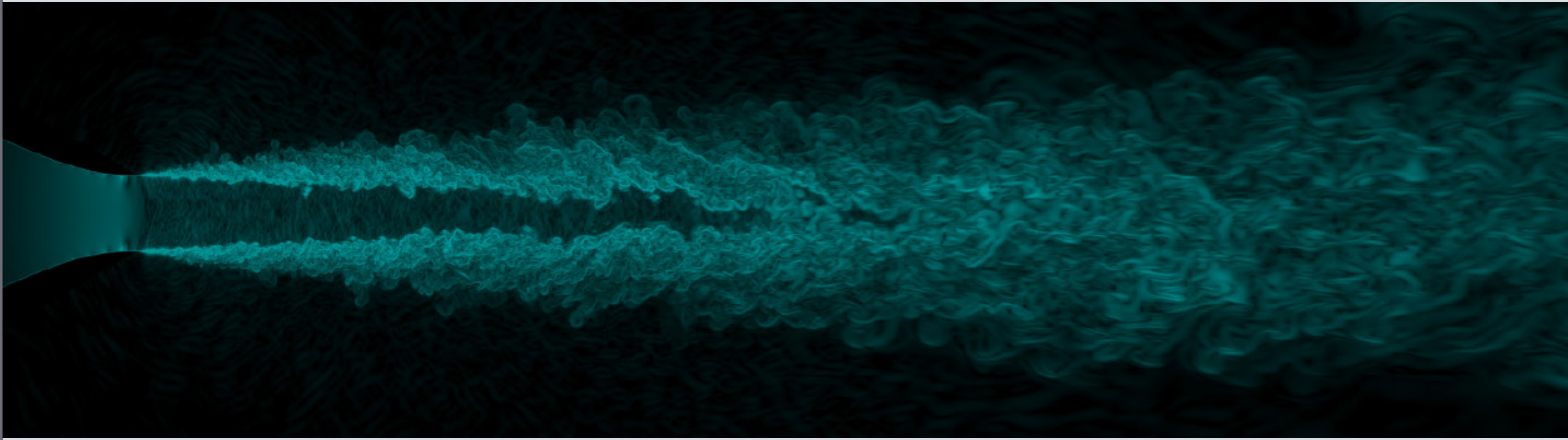
This work is the first instance of an automatically triggered, between-shot fusion science analysis code running on-demand at a remotely located HPC resource.

DIRECTOR'S DISCRETIONARY

Engineering R&D

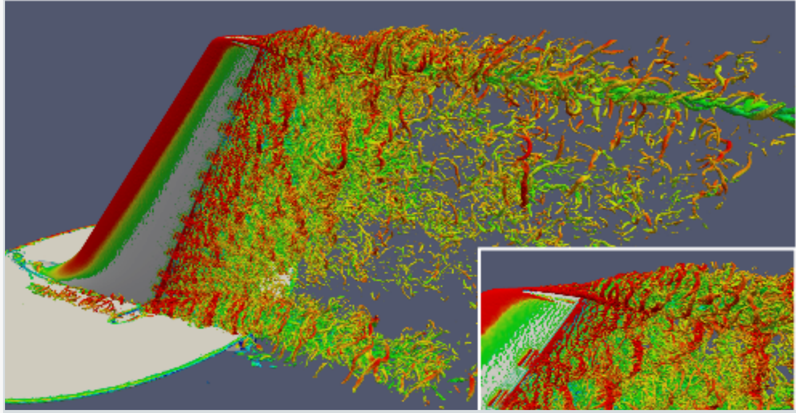
Some of the most advanced codes used at the ALCF are driving industry-leading aerospace engineering R&D—from wing design to engines to tail rudder assemblies.

Our supercomputers provide fundamental insight into the interaction between physical components and physical phenomena on a realistic geometry, without the expense of building and testing multiple physical models.



Argonne NATIONAL LABORATORY

Adaptive DES of a High Lift Wing with Active Flow Control



PI: Kenneth Jansen, University of Colorado Boulder

Challenge: To simulate synthetic jet active flow control on a multicomponent, realistic high lift wing configuration.

Impact: These simulations will provide insights into the interaction between synthetic jets and the main flow on a realistic geometry in aeronautics. Achieving better high lift wing designs could result in significant fuel savings that both reduce operating costs and engine emissions.

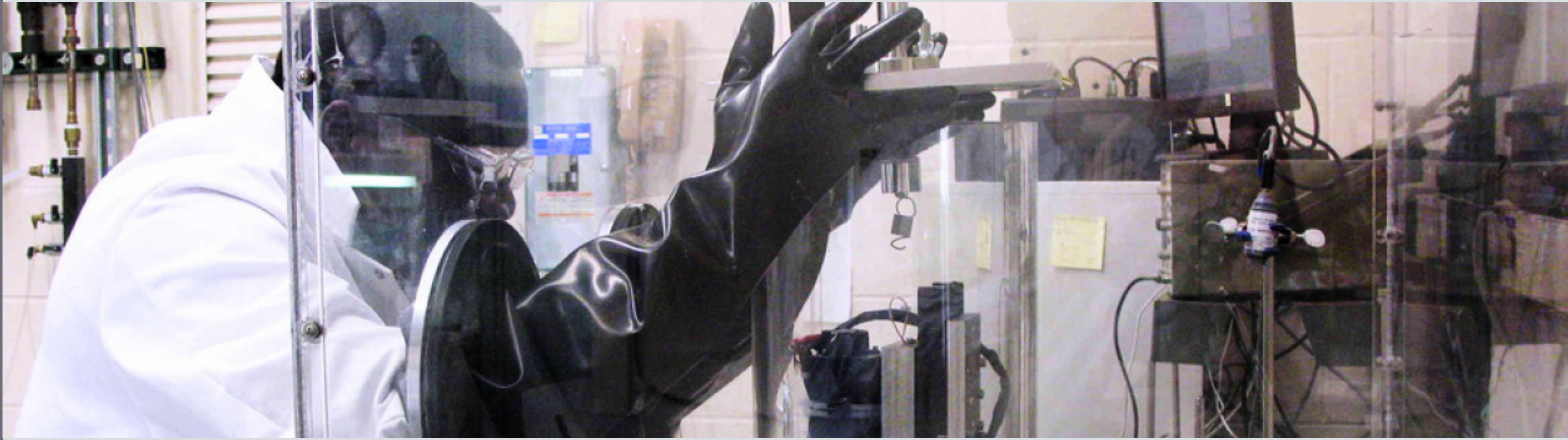
Approach: This team is employing parallel adaptive meshing and parallel solver technology to yield fundamental insights into the complicated physics of flow control on real aircraft configurations.

Aerospace engineers use supercomputers to simulate airplane wing configurations that can provide high lift during takeoff and landing, yet still be able to cruise at altitude such that air is moving smoothly across the wing.

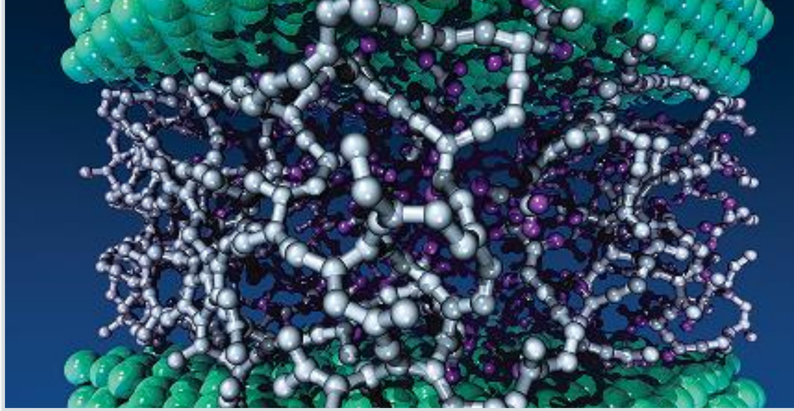
U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

Discovering New Materials

Our supercomputers can be used to simulate the operating conditions that impact energy technologies, study new materials, and test new battery chemistries. Simulations can both identify promising candidates for further R&D and validate experimental results in a matter of days versus years, or even decades.



Reactive Mesoscale Simulations of Tribological Interfaces



PI: **S. Sankaranarayanan, Argonne National Laboratory**

Challenge: To simulate reactive processes, including bond breakage and formation, at electrochemical interfaces on the order of millions of atoms.

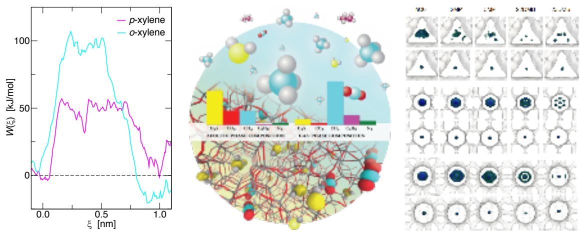
Impact: Simulations can reveal the complex processes that make oils, coatings, electrodes, and other electrochemical interfaces effective. Using Mira, this team discovered a self-healing, anti-wear coating that drastically reduces friction. Their findings are being used to virtually test other potential self-regenerating catalysts.

Approach: The researchers modeled as many as 2M atoms per simulation. Millions of time steps per simulation enabled the team to identify the initial catalytic processes that occur within nanoseconds of machine operation.

When experiments showed that a new film was being regenerated at the interface of an engine coating and base oil, researchers turned to Mira to model the underlying reactive processes. The results led to an amazing discovery.

U.S. DEPARTMENT OF ENERGY INCITE LEADERSHIP COMPUTING

Predictive Modeling of Functional Nanoporous Materials



PI: J. Ilja Siepmann, University of Minnesota

Challenge: To develop improved predictive modeling capabilities to accelerate the discovery and design of nanoporous materials for complex chemical separation and transformation applications.

Approach: Mira was used to interpret experiments and make membrane predictions using a newly developed synthesis method (patent pending).

Impact: The ability to identify optimal zeolites and metal-organic frameworks for specific energy applications has the potential to improve the production of biofuel and petroleum products, and to advance the development of gas storage and carbon capture devices.

This project uses hierarchical screening workflows that involve machine learning, evolutional algorithms, molecular simulations, and high-level electronic structure calculations.

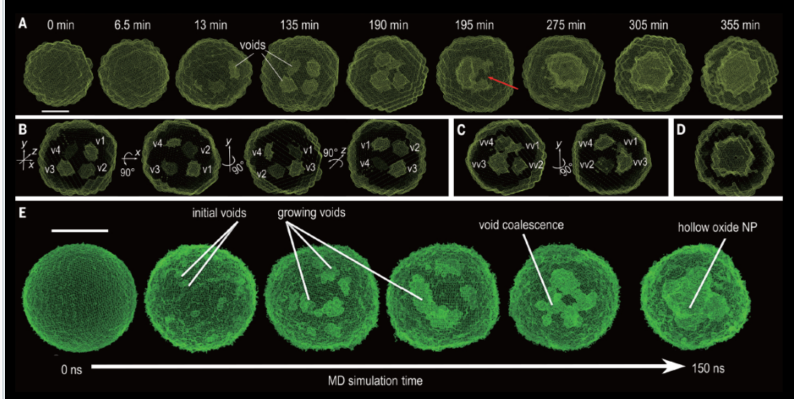
Atomistic Simulations of Nanoscale Oxides and Oxide Interfaces

PI: S. Sankaranarayanan, Argonne National Laboratory

Challenge: To study metal oxidation, a chemical reaction that transforms iron nanoparticles into nanoshells and affects matter at length scales of nanometers to kilometers.

Approach: Mira was used to simulate the iron nanoparticle oxidation process using force-field-based molecular dynamics. An atomistic 'movie' of the process revealed the role played by voids in this chemical reaction (Kirkkendall diffusion). X-ray experiments at the Advanced Photon Source corroborated the MD simulations.

Impact: The results of this work highlights the complex interplay between defect chemistry and defect dynamics in determining nanoparticle transformation and formation.



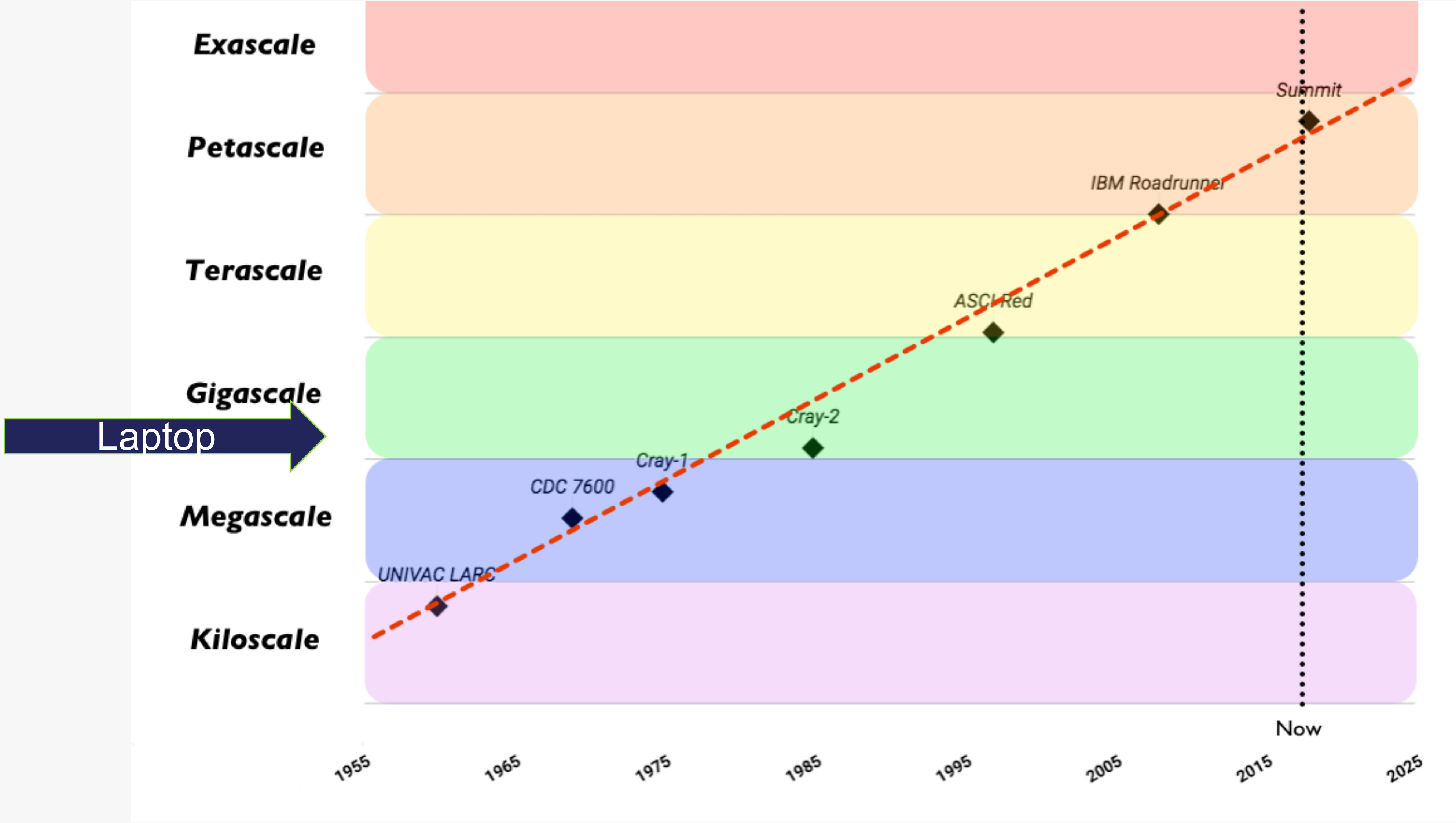
Real-time tracking of the 3D evolution of colloidal nanoparticles in solution is essential for understanding complex mechanisms involved in nanoparticle growth and transformation.

ALCC

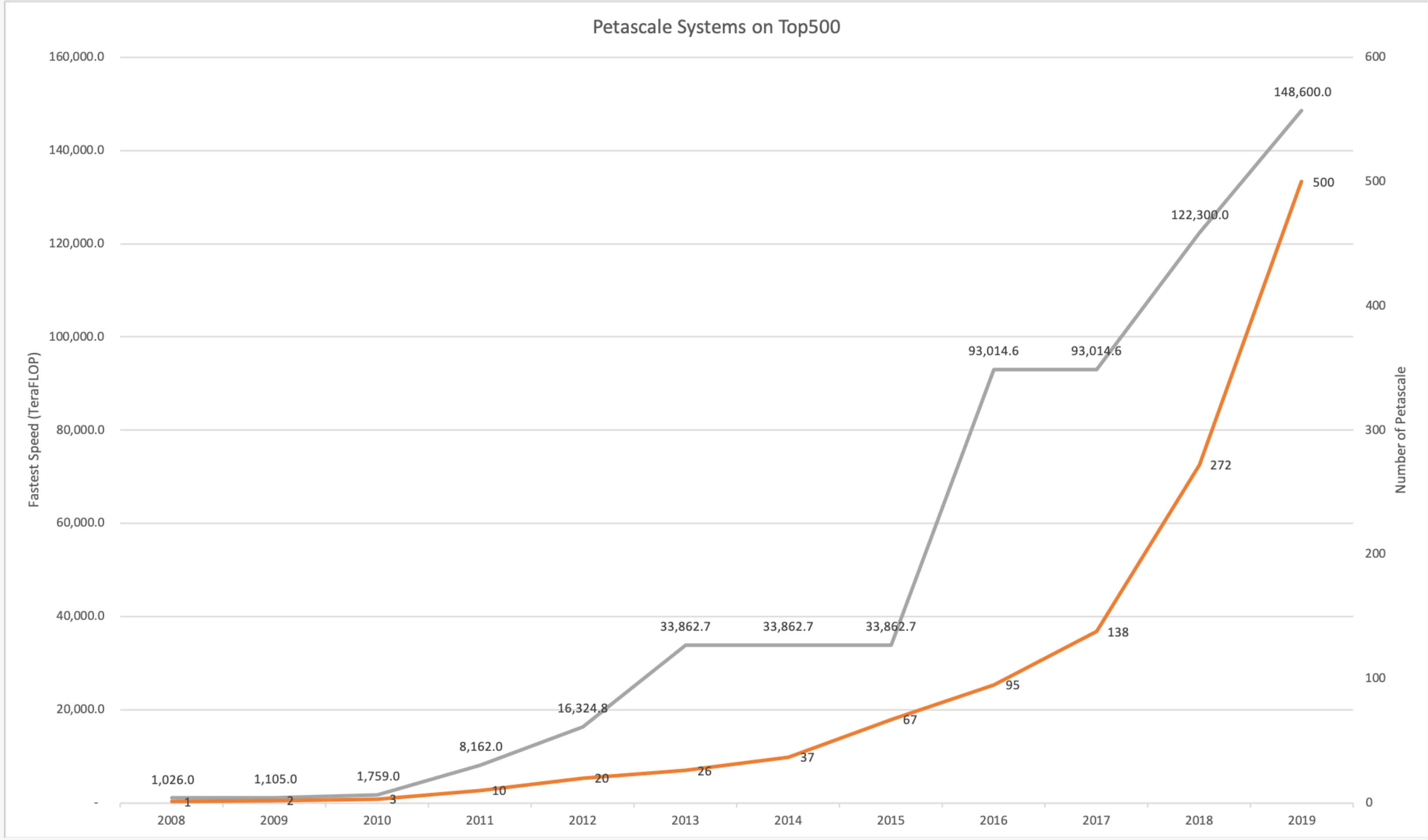
Argonne NATIONAL LABORATORY

The Road to Exascale

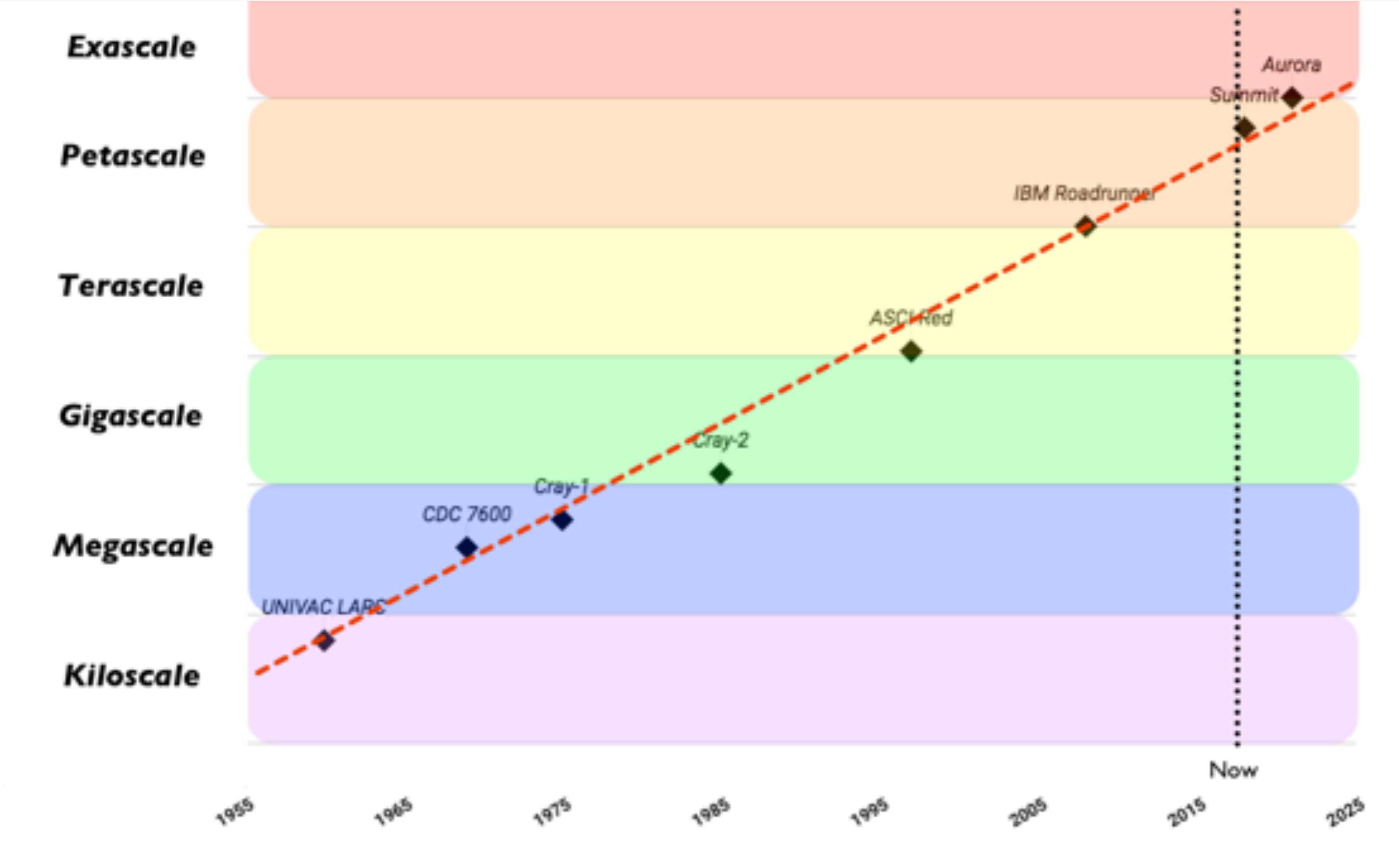
How did we get here?



How did we get here?



To Exascale



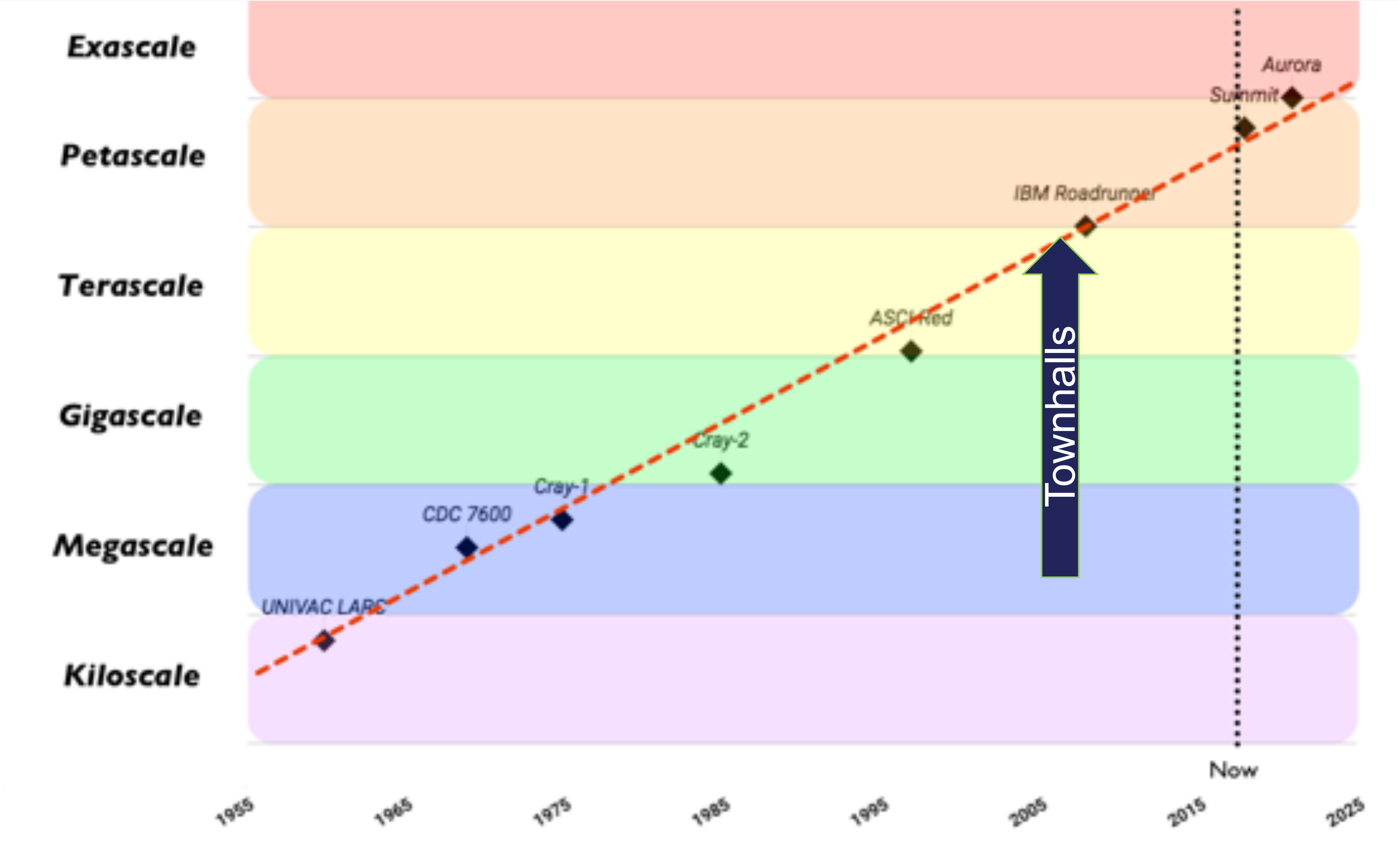
Exascale
Petascale
Terascale
Gigascale
Megascale
Kiloscale
Aurora
Summit
IBM Roadrunner
ASCI Red
Cray-2
Cray-1
CDC 7600
UNIVAC LARC
Now
1955
1965
1975
1985
1995
2005
2015
2025

Why Exascale?*

- Deliver the next era of high-end computational systems
- Build-in cyber defenses including information protection
- Provide technologies to ensure supply-chain integrity
- Attack previously unsolvable problems critical to the nation's national security, science and innovation engine
- Deliver tremendous improvements in electronic component energy efficiency, reducing operations costs of HPC centers and carbon footprints
- Stimulate US competitiveness and intellectual property through aggressive technology development and adoption

*https://www.energy.gov/sites/prod/files/2013/09/f2/20130913-SEAB-DOE-Exascale-Initiative.pdf

To Exascale



Argonne NATIONAL LABORATORY

Timelines

- **2007 – A series of townhalls were conducted to ask the scientific community: *If there were exascale computers, what would you do with them?***
- **2009 – An Exascale Initiative Steering Committee was formed to think about how to achieve exascale[1]**
- **2010 – DOE ASCR lays out a plan to exascale[2]**
- **2013 – Exascale Computing Initiative (ECI) is defined[3]**
- **2015 – President Obama signs *National Strategic Computing Initiative* executive order authorizing funding the ECI**
- **2016 – The Exascale Computing Project (ECP) is started**
- **2018 – The ALCF-3 project is rebaselined from a 180PF machine in 2018 to a >1EF machine in 2021**
- **2021 – The ALCF exascale system, Aurora, will be delivered**

[1] https://computing.ornl.gov/workshops/FallCreek10/presentations/nichols.pdf

[2] https://www.exascale.org/mediawiki/images/6/63/IESPv2-DOE-Helland.pdf

[3] https://www.energy.gov/sites/prod/files/2013/09/f2/20130913-SEAB-DOE-Exascale-Initiative.pdf

Aurora

System Spec	Aurora
Delivery	CY2021
Sustained Performance	≥1EF DP
Compute Node	Intel Xeon scalable processors X^e arch based GP-GPUs
GPU Architecture	X^e arch based GPU Tile based, chiplets, HBM stack, Foveros 3D integration
CPU-GPU interconnect	PCIe
Aggregate System Memory	>10 PB
System Interconnect	Cray Slingshot Dragonfly topology with adaptive routing
Network Switch	25.6 Tb/s per switch, from 64 - 200 Gbs ports (25GB/s per direction)
High-Performance Storage	≥230 PB, ≥25 TB/s (DAOS)
Programming Models	Intel OneAPI, OpenMP, DPC++/SYCL
Software stack	Cray Shasta software stack + Intel enhancements + Data and Learning
Platform	Cray Shasta
# Cabinets	>100

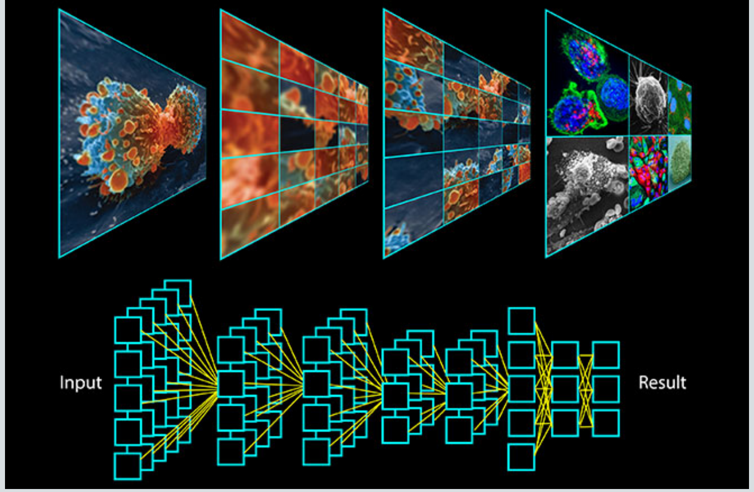
Why Exascale?

- **Advance the understanding of human body and brain and medical treatments**
  - **CANDLE, BRAIN Initiative, Human Brain Project, Blue Brain**

CANDLE: Exascale Deep Learning Enabled Precision Medicine for Cancer*

PI: **R. Stevens, Argonne National Laboratory**

Challenge: This project aims to build and apply a scalable deep neural network environment—the CANcer Distributed Learning Environment (CANDLE)—to address three top challenges of the National Cancer Institute: understanding the molecular basis of key protein interactions; developing predictive models for drug response, and automating the analysis; and the extraction of information from millions of cancer patient records to determine optimal cancer treatment strategies.

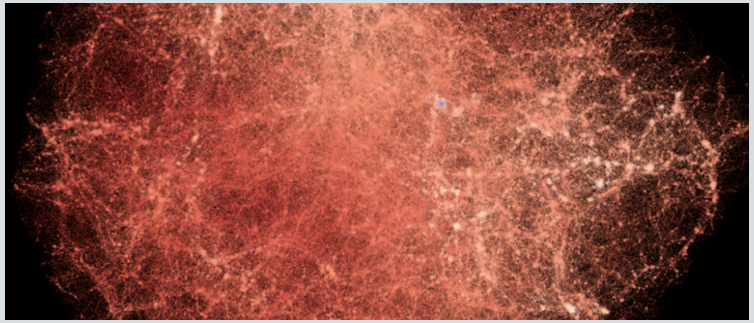


*https://www.exascaleproject.org/project/candle-exascale-deep-learning-enabled-precision-medicine-cancer/

Why Exascale?

- Advance the understanding of human body and brain and medical treatments
  - CANDLE, BRAIN Initiative, Human Brain Project, Blue Brain
- Enable finer and higher resolution imaging and models of the cosmos
  - ExaSky, Square Kilometer Array

ExaSky: Computing the Sky at Extreme Scales*



PI: **S. Habib, Argonne National Laboratory**

Challenge: This project aims to elucidate cosmological structure formation by uncovering how smooth and featureless initial conditions evolve under gravity in an expanding universe to eventually form our complex cosmic web. Modern cosmological observations have led to a remarkably successful model for the dynamics of the Universe. Three key ingredients—dark energy, dark matter, and inflation—are signposts to further breakthroughs, as all reach beyond the known boundaries of the particle physics Standard Model.

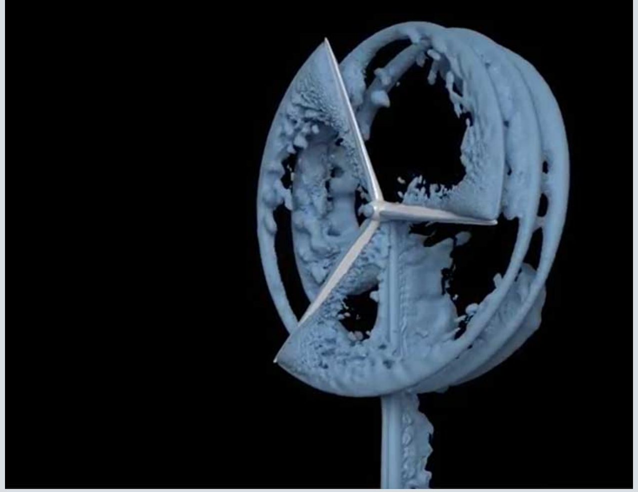
*https://www.exascaleproject.org/project/exasky-computing-sky-extreme-scales/

Why Exascale?

- Advance the understanding of human body and brain and medical treatments
  - CANDLE, BRAIN Initiative, Human Brain Project, Blue Brain
- Enable finer and higher resolution imaging and models of the cosmos
  - ExaSky, Square Kilometer Array
- Advance efficient energy production
  - ExaWind, Fusion Reactors

EXAWIND: Exascale Predictive Wind Plant Flow Physics Modeling*

PI: **M. Sprague, National Renewable Energy Laboratory**



Challenge: Significant plant-level energy losses by turbine-turbine interactions in complex terrain hamper the wide-scale deployment of wind energy on the power grid. This project is focused on predicting the flow physics that govern whole wind plant performance: wake formation, complex terrain impacts, and the effects of turbine-turbine interaction.

*https://www.exascaleproject.org/project/exawind-exascale-predictive-wind-plant-flow-physics-modeling/

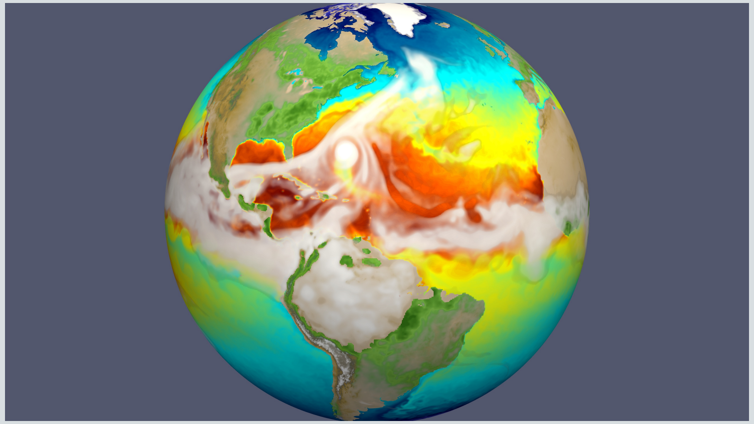
Why Exascale?

- Advance the understanding of human body and brain and medical treatments
  - CANDLE, BRAIN Initiative, Human Brain Project, Blue Brain
- Enable finer and higher resolution imaging and models of the cosmos
  - ExaSky, Square Kilometer Array
- Advance efficient energy production
  - ExaWind, Fusion Reactors
- Model and understand more complex climate interactions
  - Energy Exascale Earth System Model (E3SM)

E3SM-MMF: Cloud-Resolving Climate Modeling of the Earth's Water Cycle*

PI: **M. Taylor, Sandia National Laboratories**

Challenge: This project employs cloud-resolving earth systems that model with throughput necessary for multi-decade, coupled high-resolution climate simulations with substantial reduction of major systematic errors in precipitation via a realistic convective storm treatment. This will improve the ability to assess regional water cycles that directly affect multiple sectors of the US economy (agriculture and energy production).

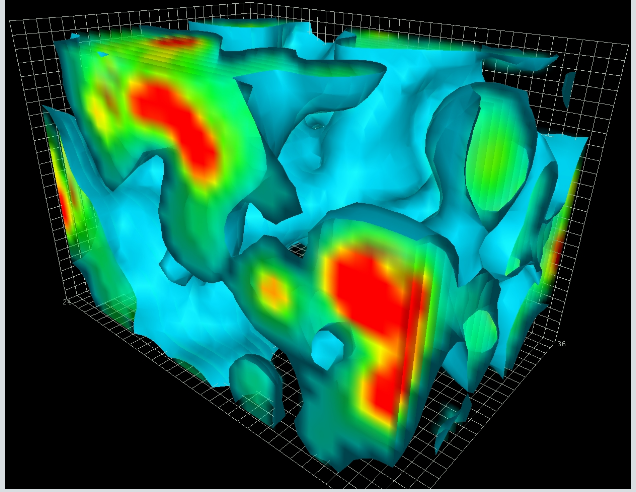


*https://www.exascaleproject.org/project/e3sm-mmf-cloud-resolving-climate-modeling-earths-water-cycle/

Why Exascale?

- Advance the understanding of human body and brain and medical treatments
  - CANDLE, BRAIN Initiative, Human Brain Project, Blue Brain
- Enable finer and higher resolution imaging and models of the cosmos
  - ExaSky, Square Kilometer Array
- Advance efficient energy production
  - ExaWind, Fusion Reactors
- Model and understand more complex climate interactions
  - Energy Exascale Earth System Model (E3SM)
- Materials design and biological/chemical interactions
  - LatticeQCD, EXAALT, ExaAM

LatticeQCD: Lattice Quantum Chromodynamics for Exascale*



PI: **A. Kronfeld, Fermilab**

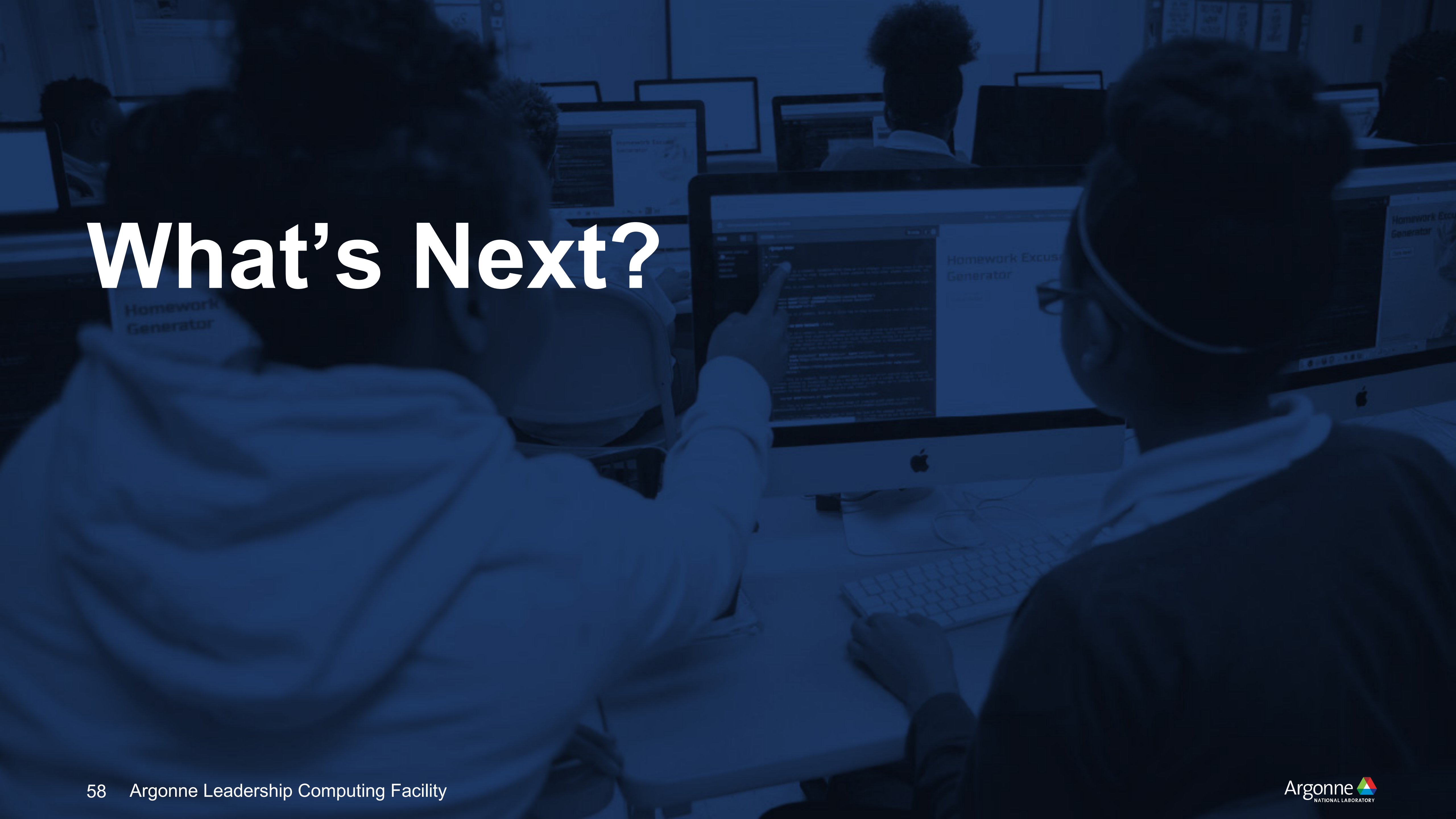
Challenge: Lattice quantum chromodynamics (QCD) calculations are the scientific instrument to connect observed properties of hadrons (particles containing quarks) to fundamental laws of quarks and gluons and critically important to decadal particle and nuclear physics experiments. To elucidate tiny effects of yet-to-be-discovered physics beyond the standard model, particle physics needs QCD simulations accurate to ~0.10% and nuclear physics needs QCD-computed properties and interactions of hadrons and light nuclei on much larger volumes than possible today.

*https://www.exascaleproject.org/project/latticeqcd-lattice-quantum-chromodynamics-exascale/

Argonne NATIONAL LABORATORY

What Are the Challenges?

- Power and Cooling
  - Today’s supercomputers can use almost 20 MW
  - A city of 80,000 needs about 45 MW
  - Breakdown of Dennard Scaling
- Space
  - Today’s supercomputers can be made of almost 50,000 nodes
- Programming
  - Updating codes to efficiently use an exascale system
- Von Neumann bottleneck



What's Next?

Moore’s Law

- **The number of transistors in an integrated circuit doubles every two years**
  - **This has driven the exponential performance increases seen over the last 50 years**
- **Driving factors include:**
  - **Efficiencies in manufacturing processes**
  - **Shrinking the size of the transistor**
    - In 1971, transistors were 10 µm or 10,000 nm
    - Today, they’re about 10-14 nm, with 7 nm on the near horizon
  - **CPU clock rate increases**
    - In 1971, the Intel 4004 speed was 740 kHz
    - Today, they’re as fast as 4.7 GHz or 4,700,000 kHz

Moore’s Law is slowing

- Despite having held up for 50 years, physics is catching up
- After 5 nm, laws of physics become a hurdle: quantum tunneling
- Increasing clock frequencies produces diminishing returns
- CPUs can only be so big before communication breaks down
- Von Neumann bottleneck is becoming more restrictive

Computers Continue Evolving

- First there were business computers (1950s)
  - Large, expensive, exclusive
- Then, personal computers (1970s)
  - Small, “affordable”, ”accessible”
- Now, edge computing (2000s)
  - Smaller, cheaper, ubiquitous
- Future, accelerators and specialized processors (2020s)
- Future, quantum computing (2040s+)

Alternative Architectures

- Accelerators
  - GPUs, FPGAs, etc.
  - Neuromorphic
    - Design processors to mimic the brain
- Quantum Computing
  - Use quantum physics to achieve more compute power with less
  - Qubits (Quantum Bits)

Artificial Intelligence

- President Trump signed *Maintaining American Leadership in Artificial Intelligence* executive order earlier this year
  - Invest in AI Research and Development
  - Unleash AI Resources
  - Set AI Governance Standards
  - Build the AI Workforce
  - International Engagement and Protecting Our AI Advantage
- Like the exascale townhalls in 2007, just completed AI townhalls
  - What would scientists do with AI machines?
  - What opportunities for breakthrough science are there with AI?
  - How to leverage AI in simulation, data, and the facility?
  - How to develop Uncertainty Quantification (UQ)?
  - Are the AI architectures being developed what is needed for science AI?

Thank You!

Learn more at: alcf.anl.gov