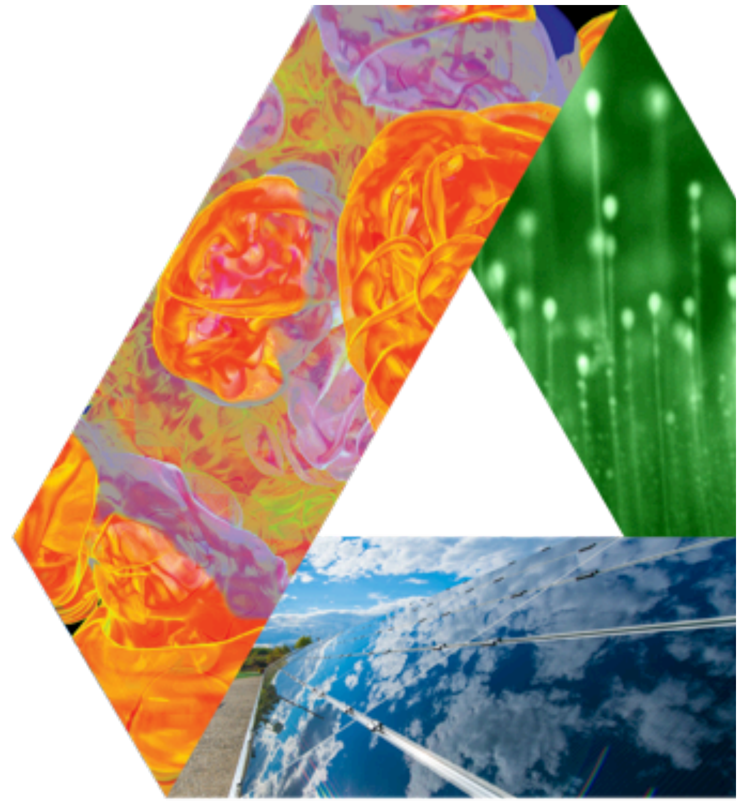
What are Supercomputers Good For?



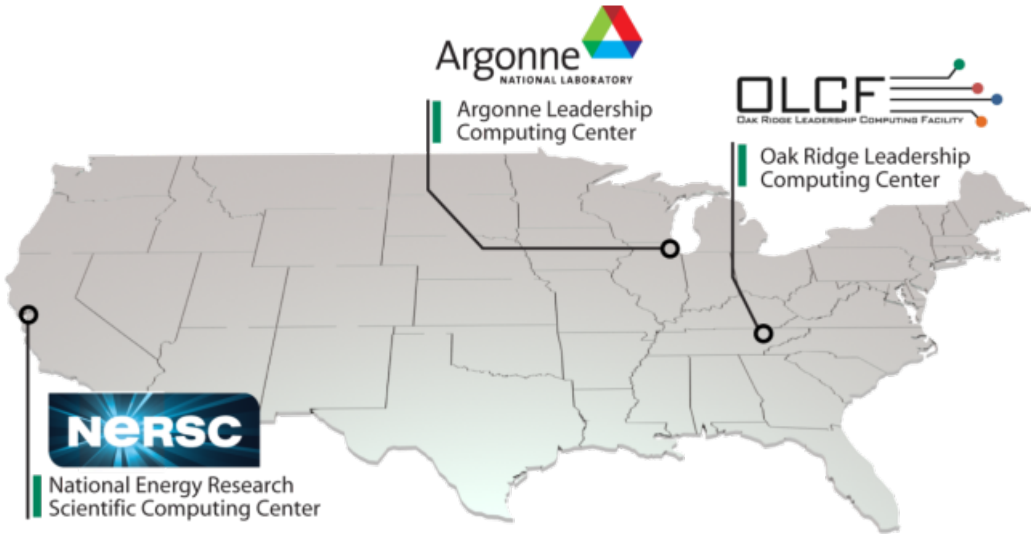
Katherine Riley

ALCF Director of Science

Argonne National Laboratory

10 May 2016

DOE'S OFFICE OF SCIENCE COMPUTATION USER FACILITIES



NERSC
Edison is 2.57 PF

ALCF
Mira is 10 PF

OLCF
Titan is 27 PF

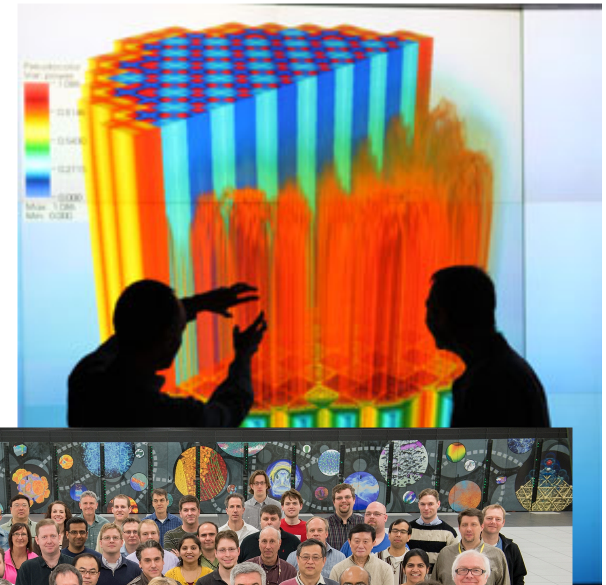
- DOE is leader in open High-Performance Computing
- Provide the world's most powerful computational tools for open science
- Access is free to researchers who publish
- Boost US competitiveness
- Attract the best and brightest researchers

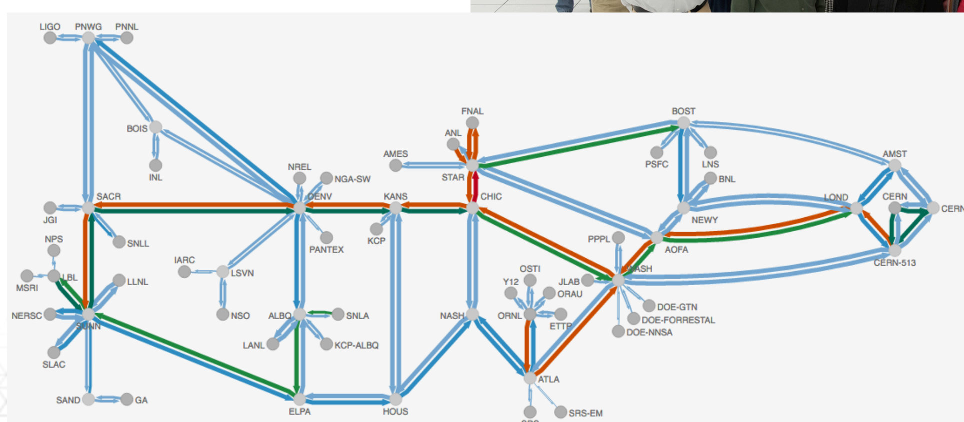
WHAT IS THE LEADERSHIP COMPUTING FACILITY (LCF)?

- 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).

LEADERSHIP IS NOT JUST BIG COMPUTE

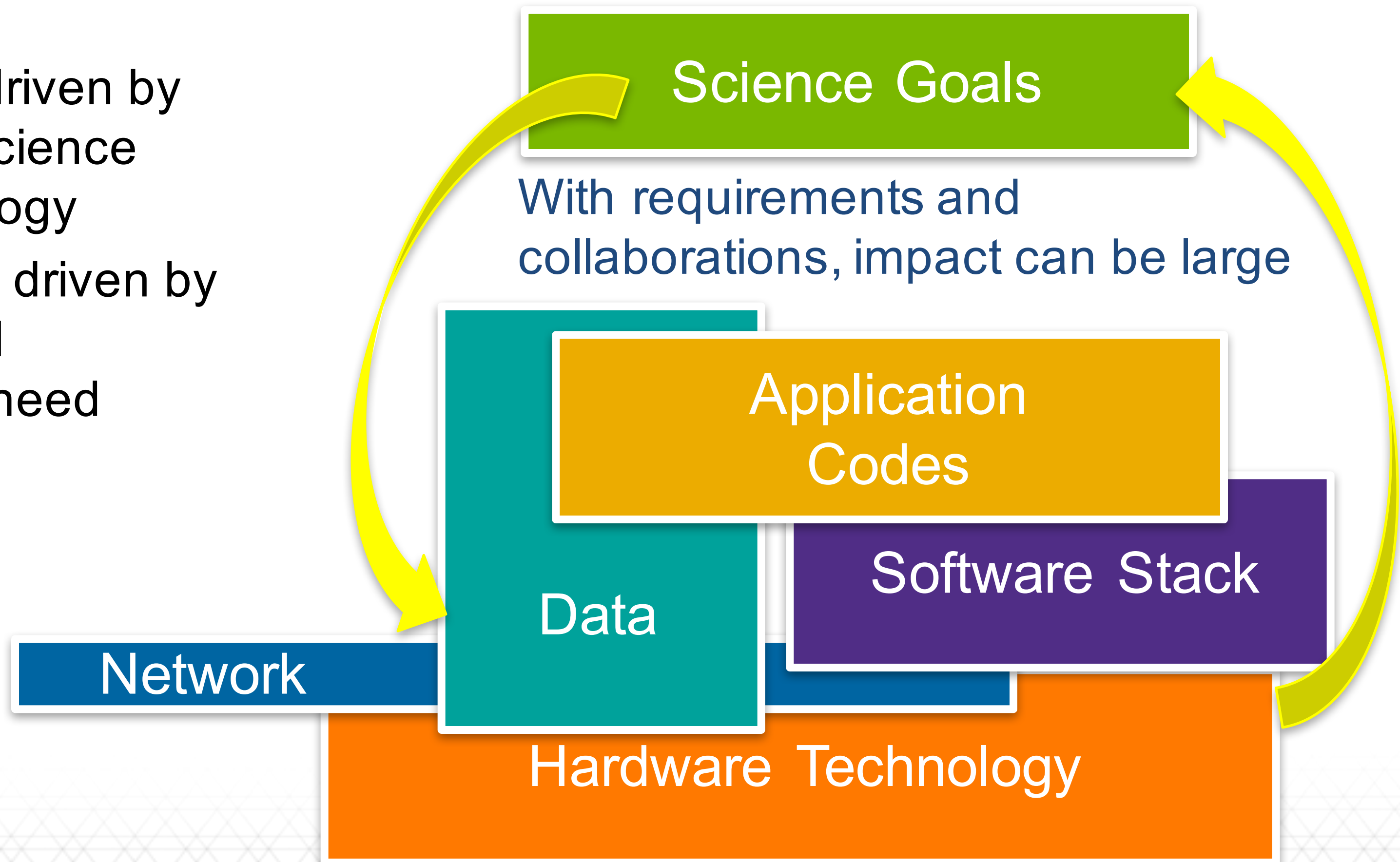
- Software Stack & Ecosystem
- Large-scale data capabilities
  - Storage - 100s of petabytes
  - Data analysis and visualization : post and *in situ*
  - Data discovery, deep learning
- World leading network interconnecting facilities
  - (100 Gb/s ⟹ 1 Tb/s)
- ***People***
  - Computational Scientists
  - Performance Engineers
  - Visualization & Data Analysis
  - Data Science



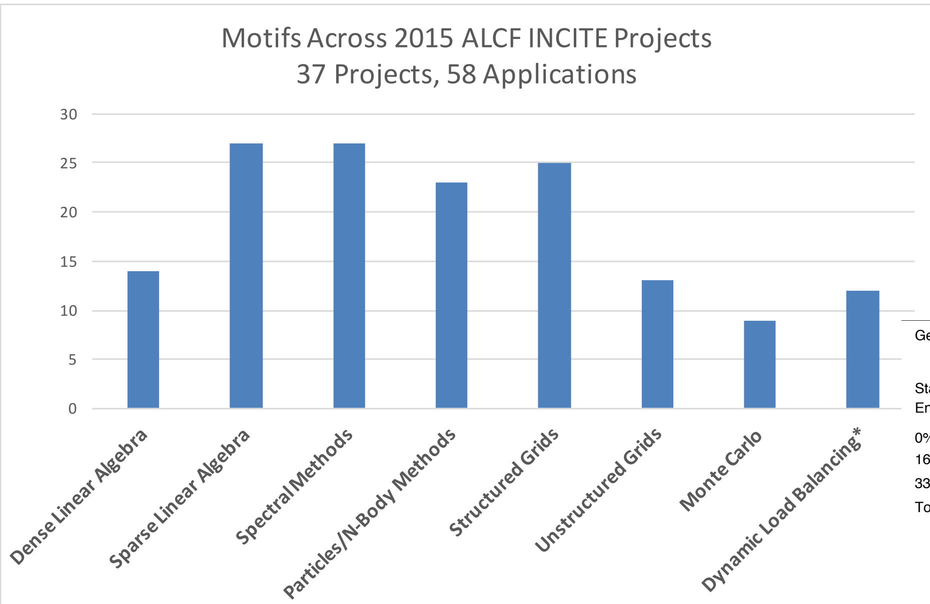


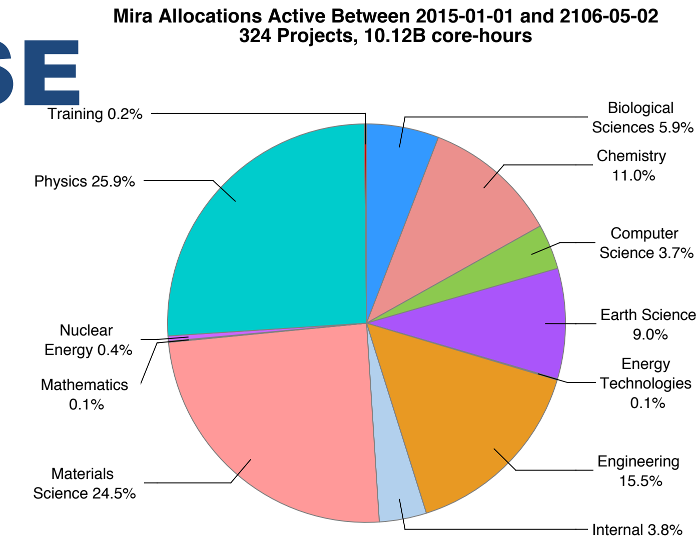
FACILITIES ARE DRIVEN BY SCIENCE COLLABORATION

- **Hardware** driven by balancing science and technology
- **Ecosystem** driven by science and application need



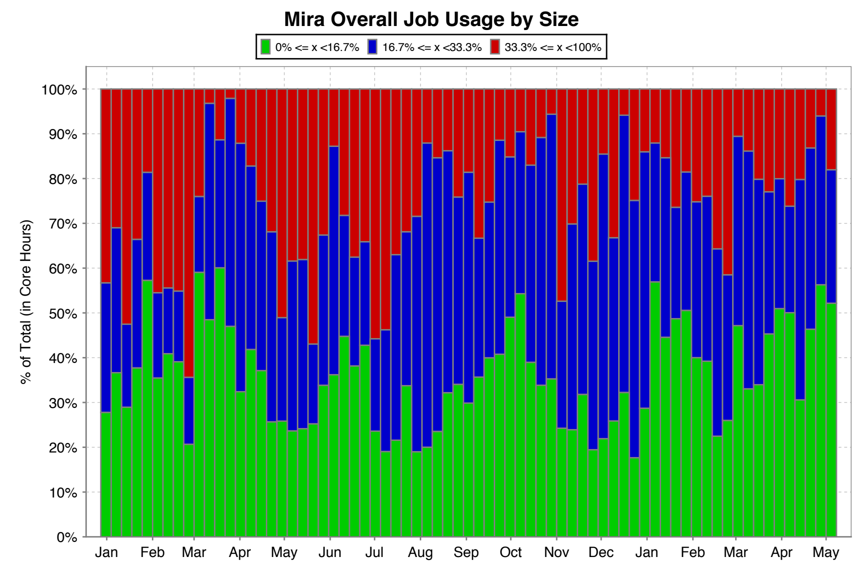
ALCF WORKLOAD: DIVERSE IN METHOD AND DOMAIN





Generated on 2016-05-02

Start	2015-01-01
End	2016-05-01
0% <= X < 16.7%	2.8B
16.7% <= X < 33.3%	3.0B
33.3% <= X <= 100%	2.0B
Total	7.8b



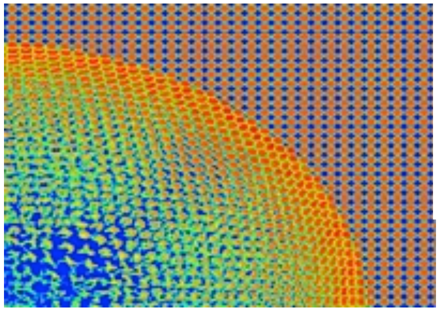
Top Methods in 2015 by INCITE Allocated Time

- Climate
- DFT
- Reactive and non-reactive MD
- Engineering – fluid flow, combustion
- Quantum Monte Carlo
- QCD

USE MODELS EVOLVING

- Traditional:
  - Run a big simulation
  - Generate lots of data
  - (bring it to home institution)
- Traditional still around but quickly transitioning
  - What do you do when you generate 10PB in a short time?
  - Supporting requirements are supercomputing scale
- Complex co-scheduling of analysis, *in situ*
- Complex workflows, Data Science
  - Deep learning, parameters spaces
  - UQ, V&V
- Streaming experimental data
- Analysis of data from community data bases of simulation and experimental data
- Real time
- Containers
- And more

ALCF SCIENCE HIGHLIGHTS



Public Safety

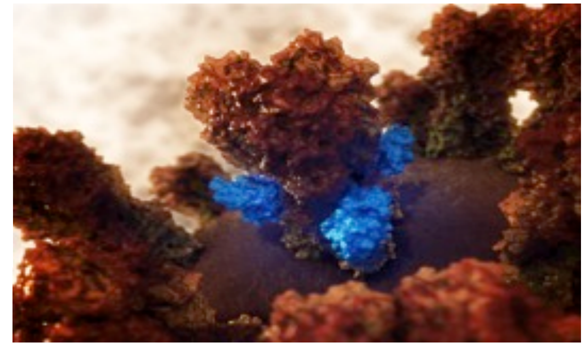
Identifying the critical packing density of conventional explosives for safe transport.

M. Berzins, University of Utah

Designer Proteins

Enabling artificially designed proteins with targeted properties, including disease-fighting drugs.

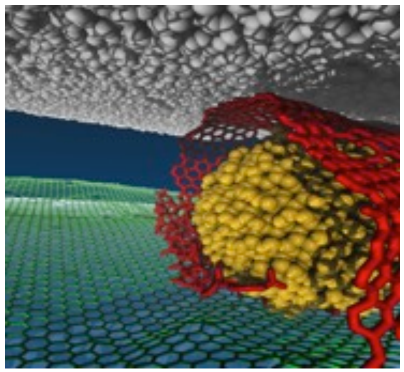
D. Baker, University of Washington



Materials Discovery

Observing a mechanism for eliminating friction at the macroscale with potential to achieve superlubricity for a wide range of mechanical applications.

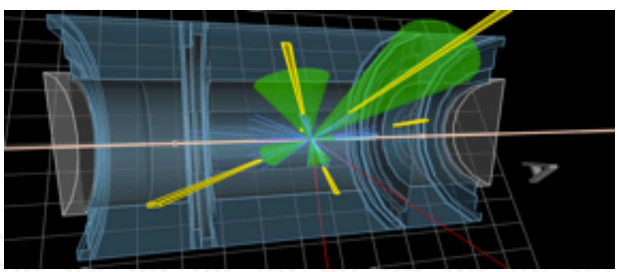
S. Sankaranarayanan, Argonne National Laboratory



Superconductivity

Providing advanced calculations to identify the electronic mechanisms for high-temperature superconductivity in the cuprates.

L. Wagner, University of Illinois at Urbana-Champaign



Supporting LHC

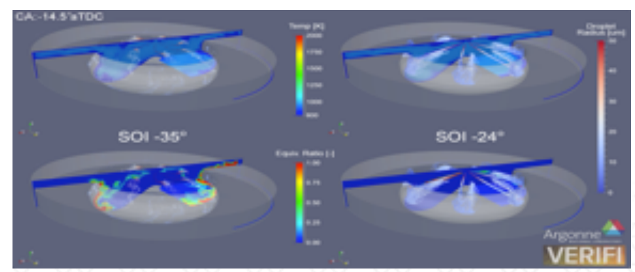
Helping to understand experiments at the LHC by routinely simulating millions of collision events in parallel.

T. LeCompte, Argonne National Laboratory

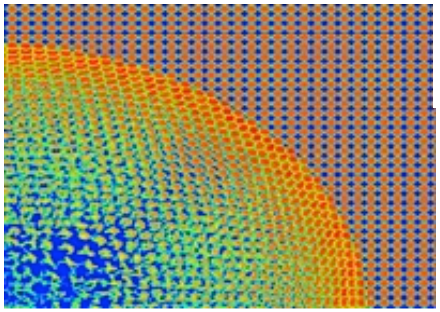
Better Engines

Partnering with industry to develop simulation capabilities that can improve the fuel economy of vehicles.

S. Som, Argonne National Laboratory

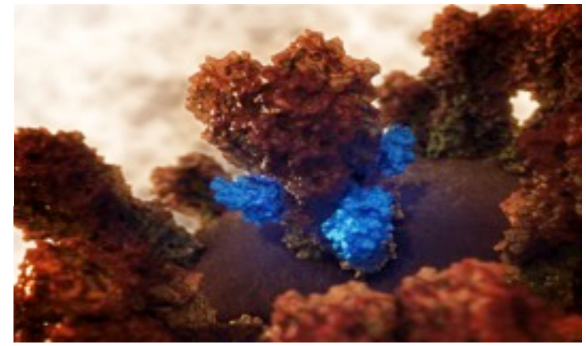


ALCF SCIENCE HIGHLIGHTS

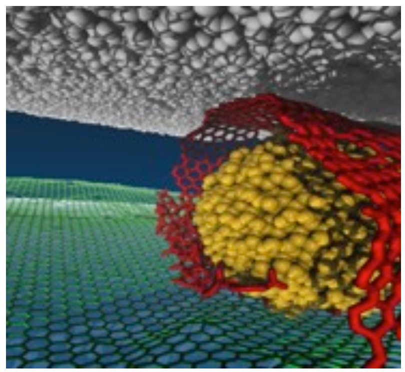


Traditional large simulations
generating large data set

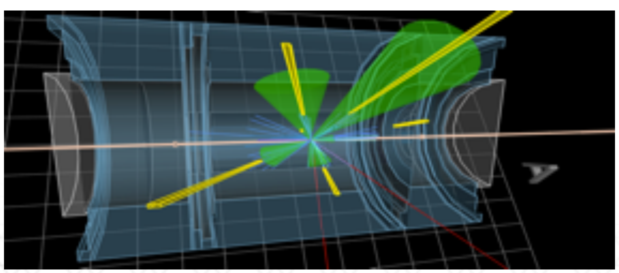
Accelerating
time-to-solution
Enhancing experimental
data.



Experimental Coupling
Augmenting experiment
Pseudo-real-time

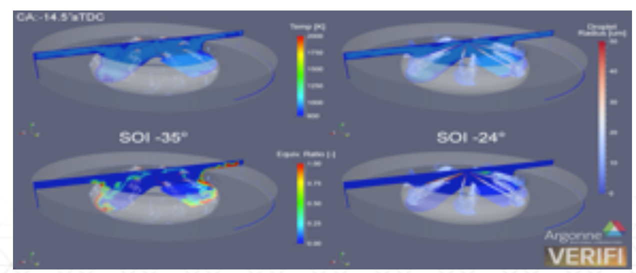


Building and augmenting large
datasets.
Ensembles & workflows



Large scale
experimental data
analysis.
WAN Workflows

Uncertainty
quantification,
parameter studies,
ensembles of large
simulations, workflows



BOOSTING LIGHT SOURCE PRODUCTIVITY WITH SWIFT ALCF DATA ANALYSIS

H Sharma, J Almer (APS); J Wozniak, M Wilde, I Foster (MCS)

Impact

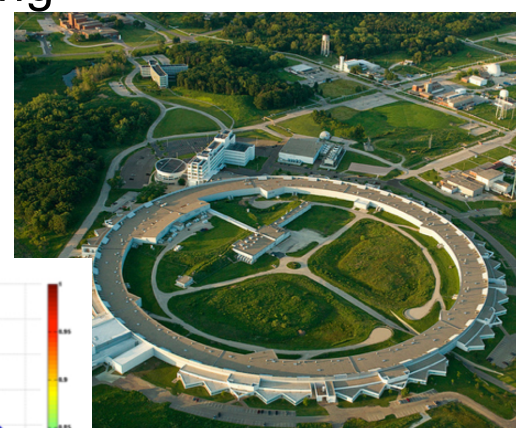
- APS scientists use Mira to process data from live HEDM experiments, providing real-time feedback to correct or improve in-progress experiments

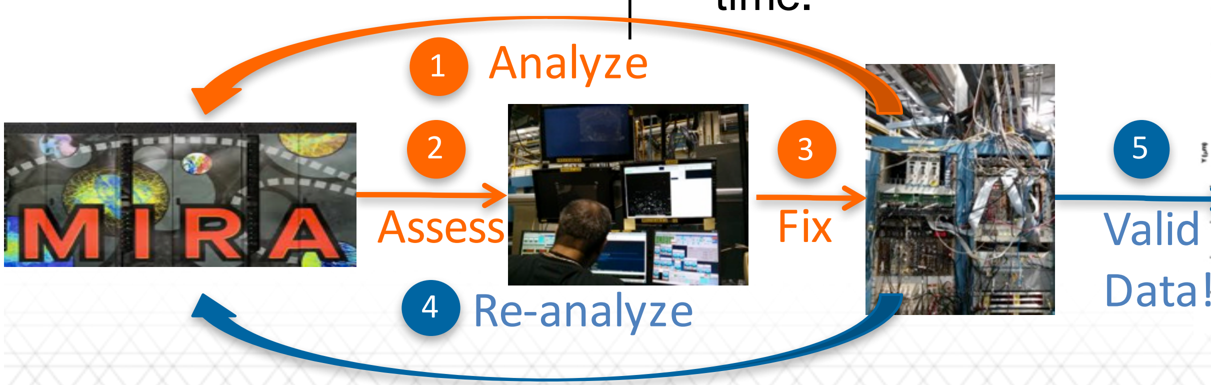
Accomplishments

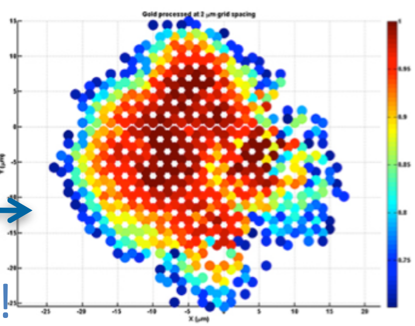
- Real-time analysis of experimental steering
- **Cable flaw was found and fixed at start of experiment,** saving an entire multi-day experiment and valuable user time and APS beam time.

Status

- Workflow is established
- Augmenting real-time scheduling

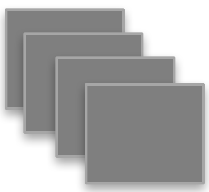






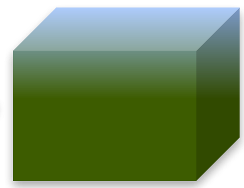
Red indicates higher statistical confidence in data

EMBRACING BIG DATA PROCESSING FOR HPC ANALYTICS

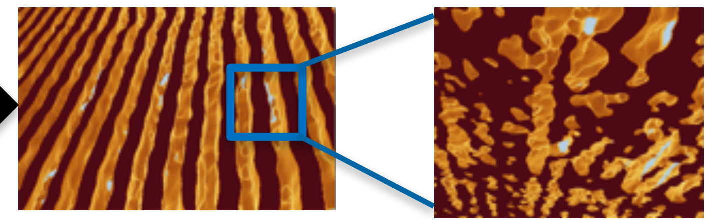


130 images from APS

3D segmentation differencing with SPARK on ALCF Cooley cluster



3D volume output



3D visualization

- Apache SPARK provides for an intuitive data-centric way for scientists to analyze their datasets
- Application scientists can use a large set of analytics packages, including image processing and machine learning, developed in Python, Java, etc., seamlessly for analysis with SPARK

3D visualization using Visus on output of a parallel analysis performed with SPARK on the ALCF Cooley cluster for data imaged at APS. Objective: study the difference between two 3D segmentations of data imaged at APS to understand how the algorithm selected affects the estimated nanoscale structure.

(Credits: Paul Nealey and Tamar Peretz (IME, ANL/UChicago), Cameron Christensen (Utah), Nicola Ferrier, Venkat Vishwanath, Jiaxing Ren (Argonne))

THE EMERGING "SUPERFACILITY" MODEL

Experimental facilities are being transformed by new detectors, advanced mathematics, robotics, automation, programmable networks.

1) Detectors soon capable of generating **terabit data streams.**

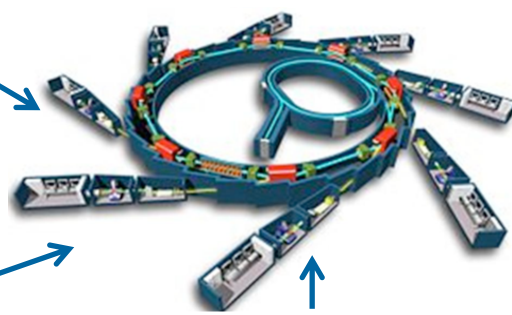
2) Computational tools for analysis, data reduction & feature extraction ***in situ***, using advanced algorithms and special-purpose hardware.

3) Increase scientific throughput from robotics and automation software.

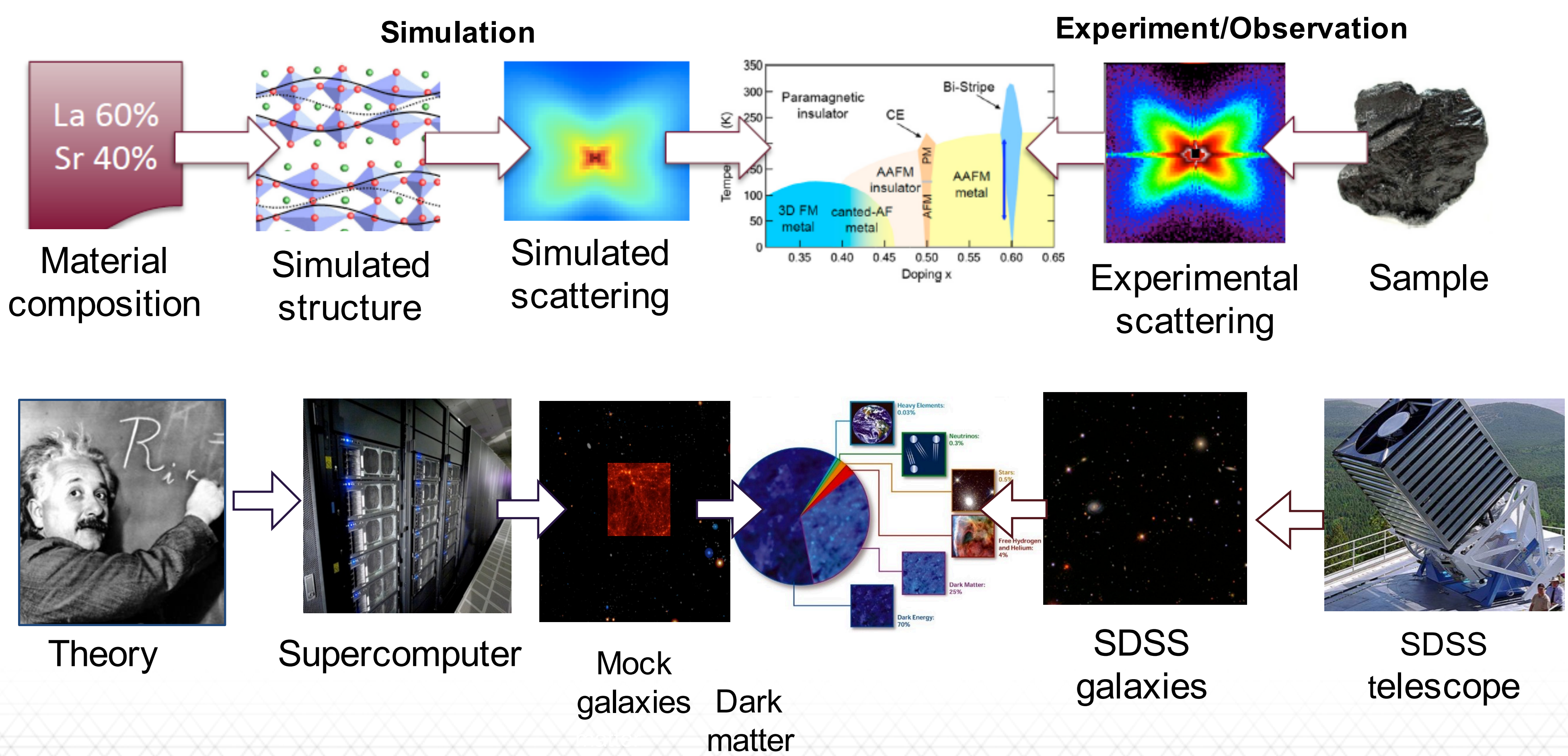
4) **Post-processing:** reconstruction, inter-comparison, simulation, visualization.

5) **Curation:** Data management and sharing, with federated identity management and flexible access control.

6) Integration of experimental and computational facilities in real time, using programmable networks and workloads.



BIG SIMULATION + BIG DATA



QUESTIONS