Supporting Experiment and Observation at ALCF

TOM URAM

Scientific Application Developer

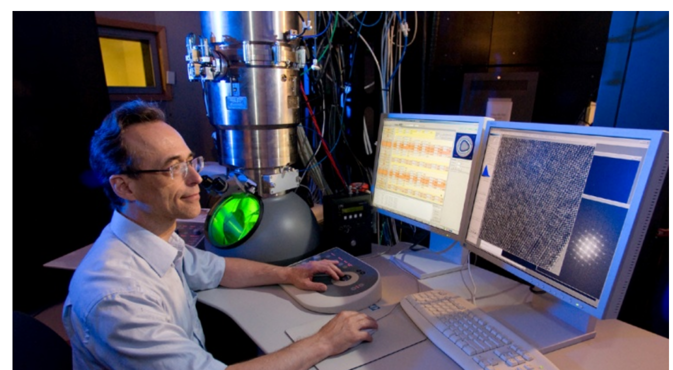
MAGIC Meeting
7 October 2015

Major Scientific User Facilities at Argonne National Laboratory

Advanced Photon Source

Argonne Tandem Linear Accelerator System

Center for Nanoscale Materials



Electron Microscopy Center

This provides us with a rich testbed to work on ideas of how to support experimental sciences.

Argonne Leadership Computing Facility

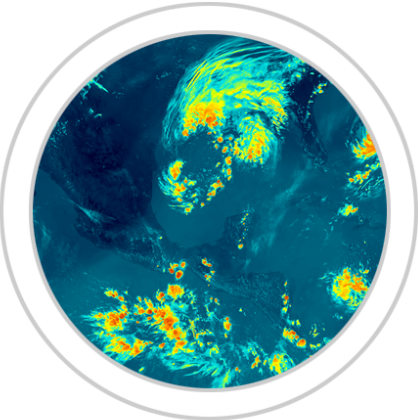
- Established in 2004
- Funded by DOE's ***Advanced Scientific Computing Research*** program.
- Operates as two centers, at Argonne and at Oak Ridge National Laboratory, and are **fully dedicated** to open science.
- Operates two petascale architectures that are **10-100 times more powerful** than systems typically available for open scientific research.

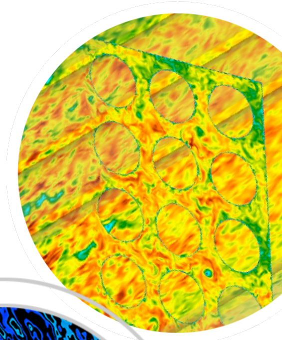
OAK RIDGE National Laboratory

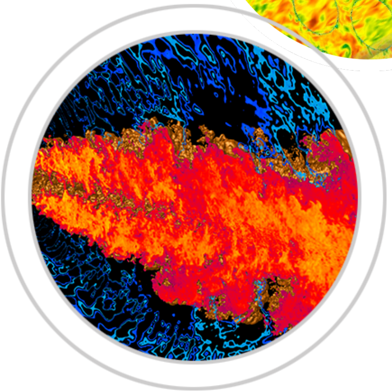
Argonne NATIONAL LABORATORY

What is Leadership Computing?

- A gateway for scientific discovery and a tool for understanding the world around us.
- Scientific breakthroughs lead to advancements that help solve the great scientific, energy, environment, and security challenges of our time.
- The nation that leads the world in HPC will have an enormous competitive advantage in every sector, from energy and environment to manufacturing.







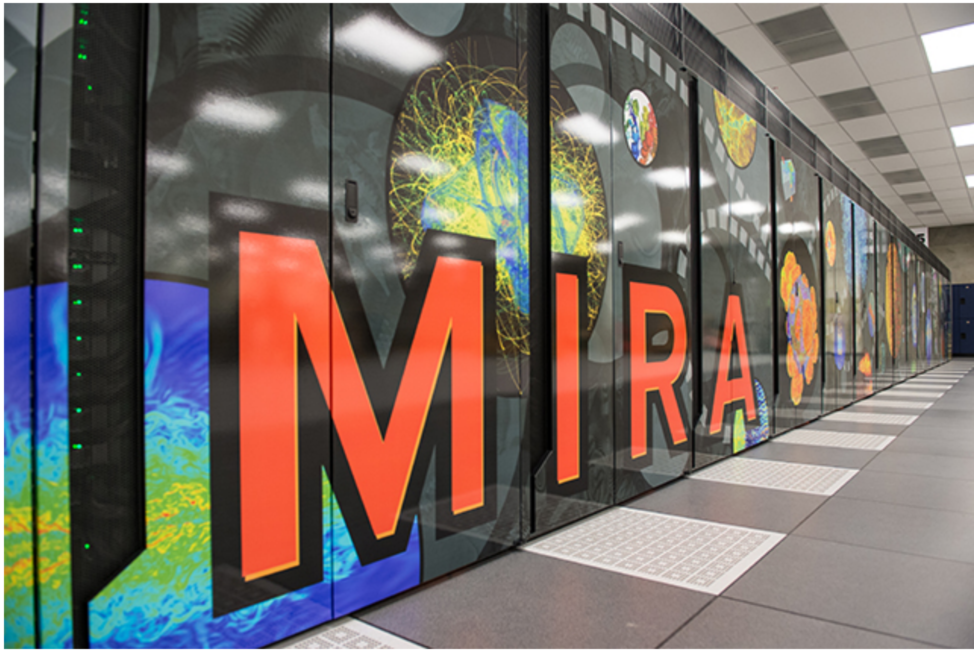
ALCF-2 Systems

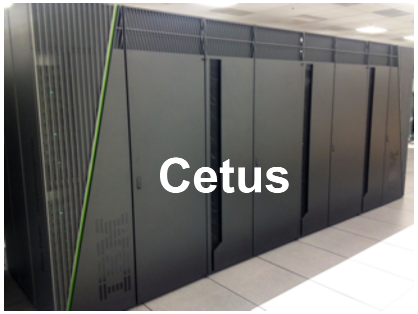
Mira – *IBM BG/Q* *[Production]*
49,152 nodes / 786,432 cores
786 TB RAM
Peak flop rate: 10 PF

Cetus – *IBM BG/Q* *[Production]*
4,096 nodes / 65,536 cores
64 TB RAM
Peak flop rate: 836 TF

Vesta – *IBM BG/Q* *[Testing & Development]*
2,048 nodes / 32,768 cores
32 TB RAM
Peak flop rate: 419 TF

Cooley – *Cray/NVIDIA* *[Production]*
126 nodes / 1512 Intel Haswell CPU cores
126 NVIDIA Tesla K80 GPUs
48 TB RAM / 3 TB GPU memory
Peak flop rate: 223 TF

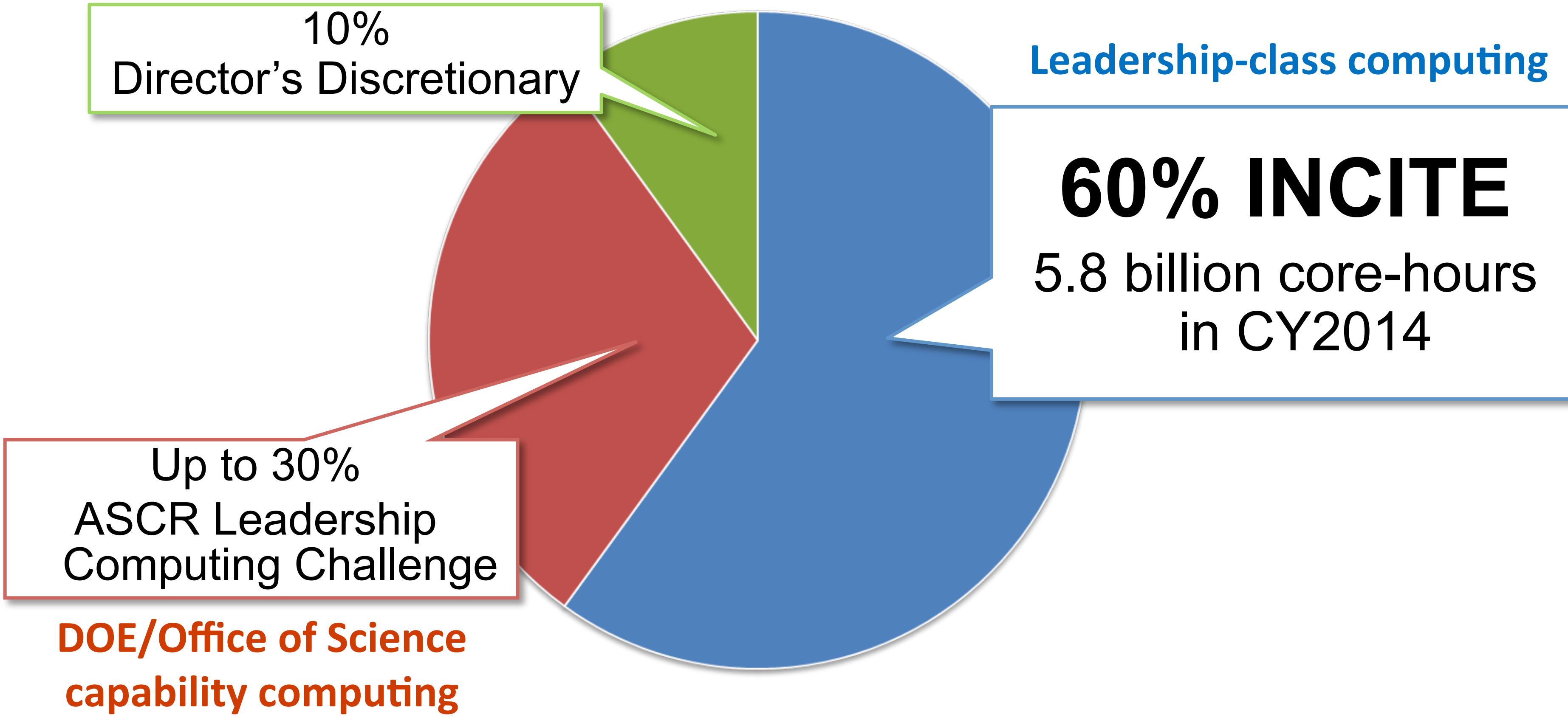




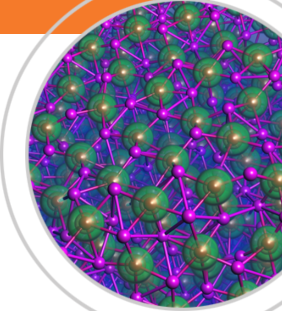
Storage
Home: 1.44 PB raw capacity
Scratch:
- fs0 - 26.88 PB raw, 19 PB usable; 240 GB/s sustained
- fs1 - 10 PB raw, 7 PB usable; 90 GB/s sustained
- fs2 (ESS) - 14 PB raw, 7.6 PB usable; 400 GB/s sustained (not in production yet)

Tape: 21.25 PB of raw archival storage [17 PB in use]

Three primary ways for access to LCF
Distribution of allocable hours



How time on DOE Leadership Computing systems is awarded

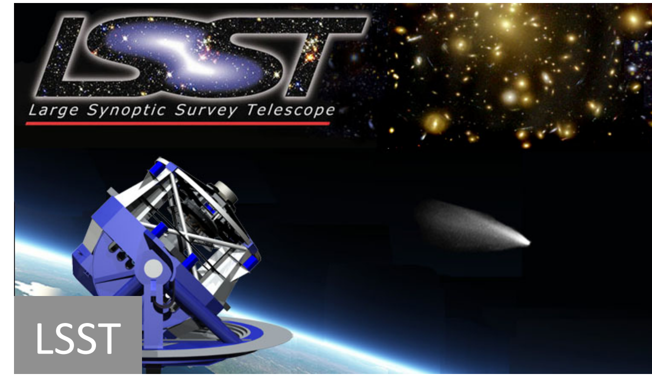


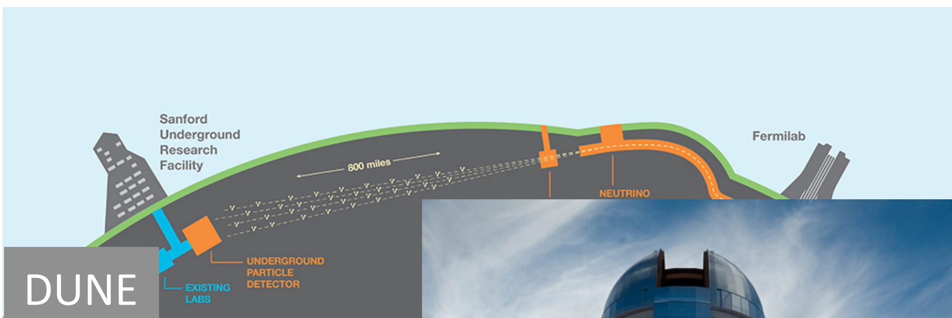
INCITE
- Peer-reviewed program open to any researcher in the world
- 5.8 B core-hours awarded for 2015 on ALCF's IBM BG/Q Mira and OLCF's Cray XK7 Titan.
- Approximately **60 percent** of ALCF resources are allocated through INCITE.

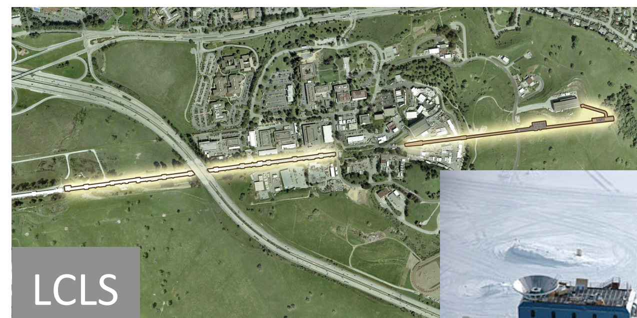
ASCR Leadership Computing Challenge
- Peer-reviewed program
- Allocates up to **30 percent** of the computational resources at ALCF, NERSC, and OLCF.
- Emphasis on high-risk, high-reward simulations in areas directly related to the DOE's energy mission, national emergencies, or for broadening the community of researchers capable of using leadership class resources

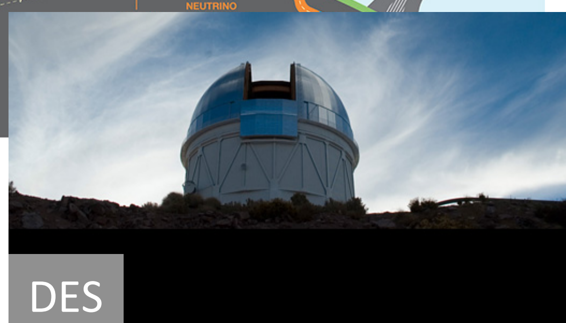
Director's Discretionary
- Peer-reviewed program open to researchers in academia and industry.
- Primarily a "first step" for projects working toward an INCITE or ALCC award.
- Allocates up to **10 percent** of the computational resources at ALCF.

Experimental and Observational Data

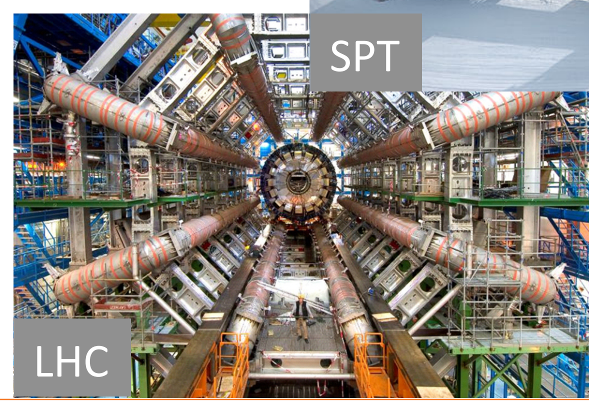






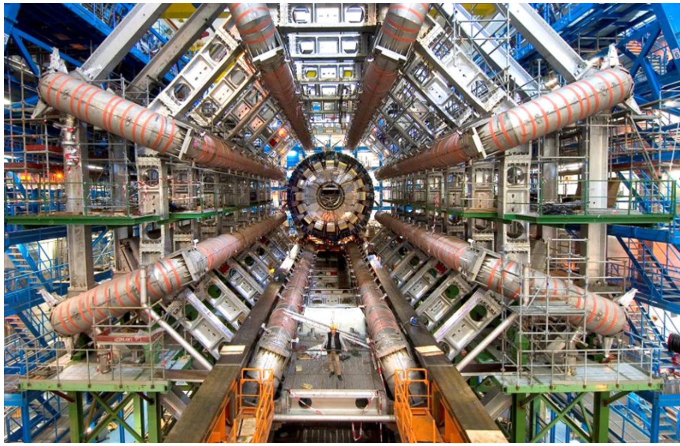


ARM
CLIMATE RESEARCH FACILITY



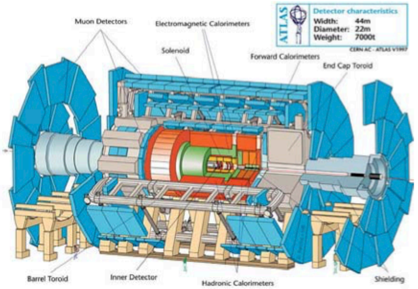
JCESR

LHC Simulation and Experiment
Argonne ATLAS group



How It Works

Real Data Chain



Reconstruction

Simulated Data Chain

Event Generation — Simulates the physics process of interest: produces lists of particles and their momenta

Simulation — Simulates the interaction of these particles with the detector

Reconstruction — Infers particles that must have been present based on the detector response

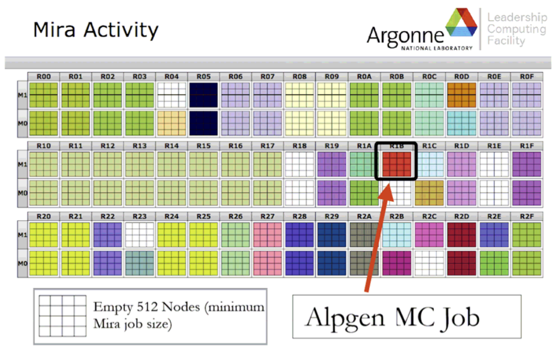
Analysis: comparing the two

Slides from Tom LeCompte (ANL)

LHC Simulation and Experiment
Argonne ATLAS group

Where we were 1 year ago

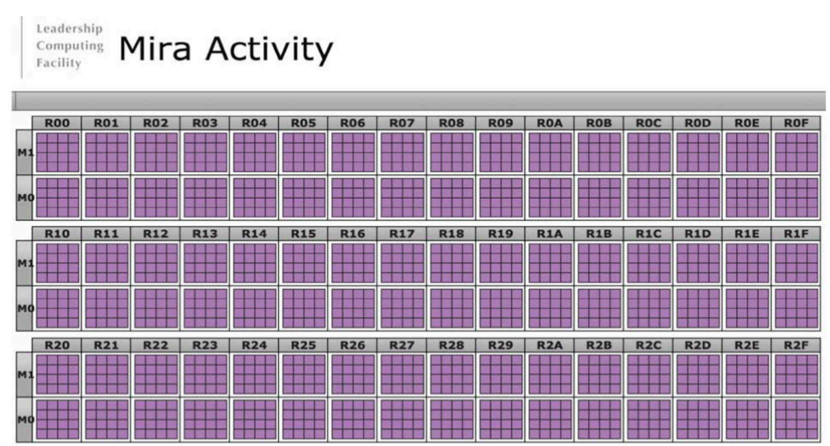
- We could run in the minimum Mira partition (and only in the minimum efficiently)
- Event generation rate was 1/15 of a Grid node
  - ALCF suggests a nominal of 1/10



At this point, we are limited by I/O.

reasons, we normally limit ourselves to 1/3 of the machine: a million parallel processes

- Based on this success, the experiment asked us to do all the Alpgen gheneration for the next two years

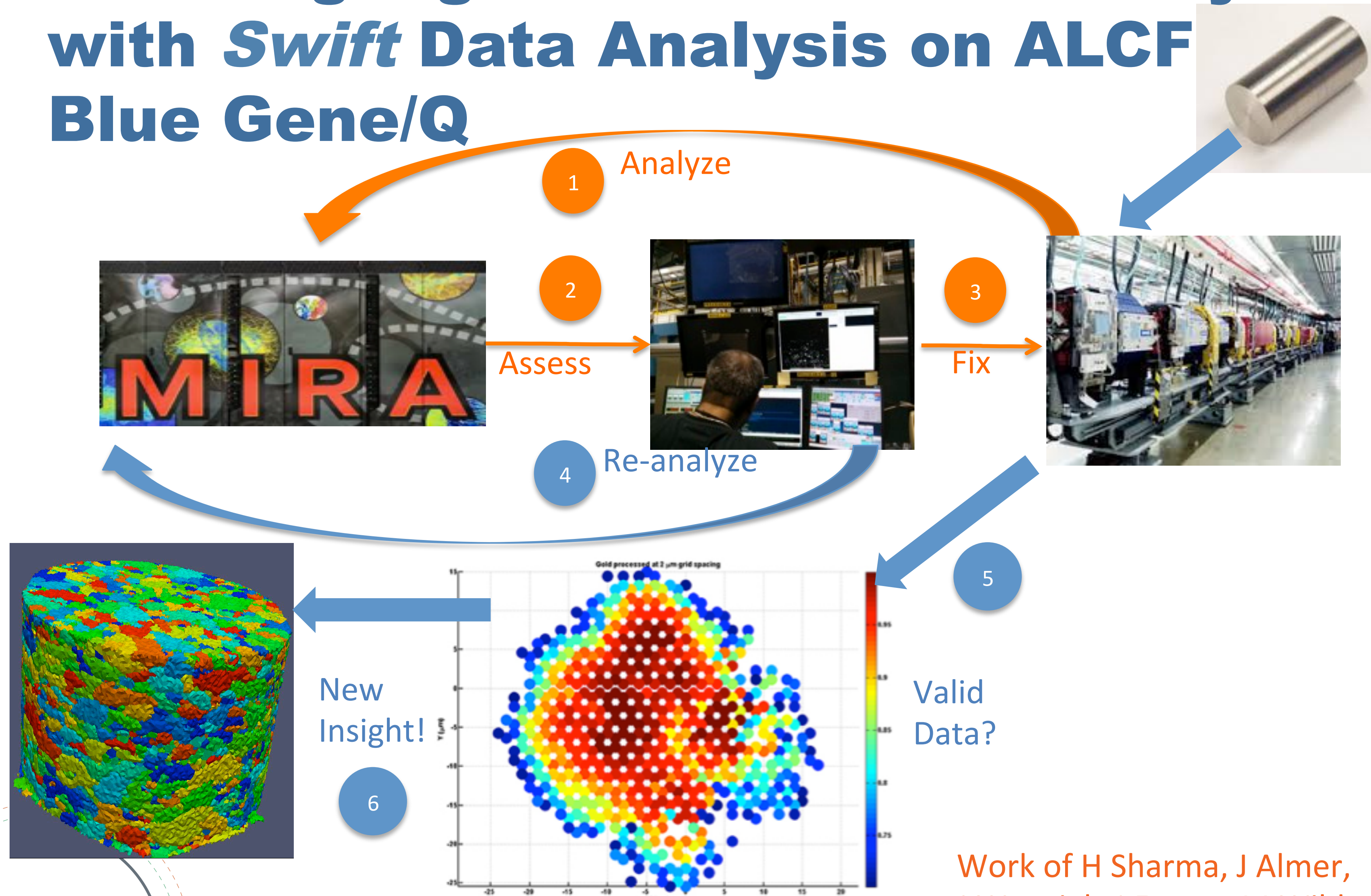


While this job was running, Mira was producing the equivalent computing as 5 or 6 ATLAS Grids.

On our best days, we provide the equivalent computing capacity of the whole ATLAS Grid.

Slides from Tom LeCompte (ANL)

Boosting Light Source Productivity with *Swift* Data Analysis on ALCF Blue Gene/Q



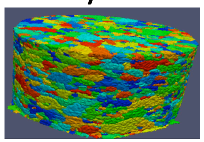
Work of H Sharma, J Almer, J Wozniak, I Foster, M Wilde

Boosting Light Source Productivity with *Swift* ALCF Data Analysis

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

Impact and Approach

- HEDM imaging and analysis shows granular material structure, non-destructively



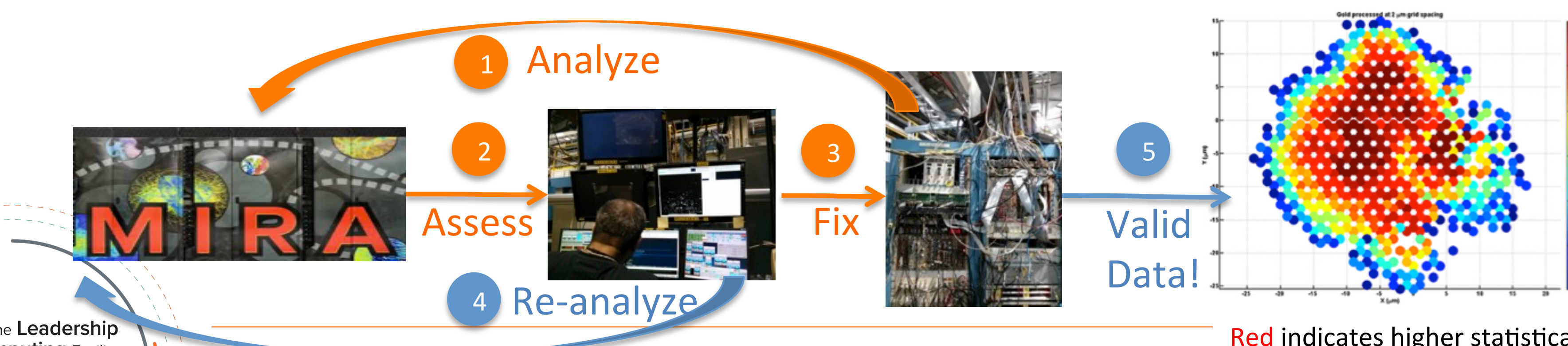
- APS Sector 1 scientists use Mira to process data from live HEDM experiments, providing real-time feedback to correct or improve in-progress experiments
- Scientists working with *Discovery Engines* LDRD developed new *Swift* analysis workflows to process APS data from Sectors 1, 6, and 11

Accomplishments

- Mira analyzes experiment in 10 mins vs. 5.2 hours on APS cluster: > 30X improvement
- Scaling up to ~ 128K cores (driven by data features)
- **Cable flaw was found and fixed at start of experiment,** saving an entire multi-day experiment and valuable user time and APS beam time.
- **In press:** *High-Energy Synchrotron X-ray Techniques for Studying Irradiated Materials*, J-S Park et al, J. Mat. Res.
- *Big data staging with MPI-IO for interactive X-ray science*, J Wozniak et al, Big Data Conference, Dec 2014

ALCF Contributions

- Design, develop, support, and trial user engagement to make *Swift* workflow solution on ALCF systems a reliable, secure and supported production service
- Creation and support of the Petrel data server
- Reserved resources on Mira for APS HEDM experiment at Sector 1-ID beamline (8/10/2014 and future sessions in APS 2015 Run 1)

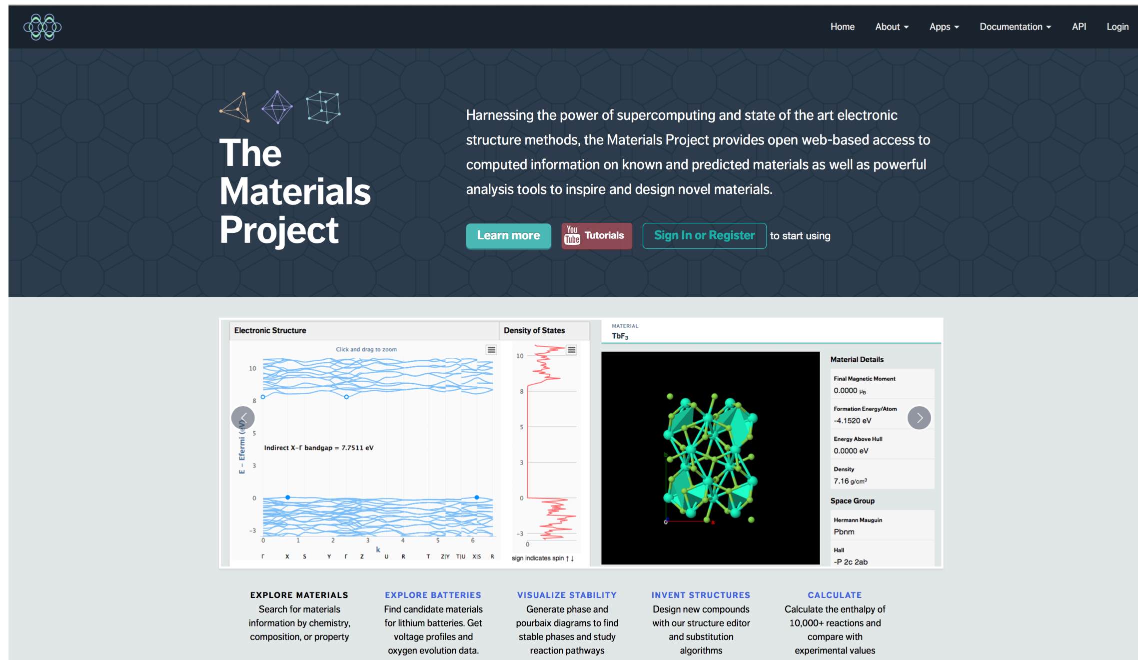


Red indicates higher statistical confidence in data

FireWorks

ALCF Users and Contributions

- Joint Center for Energy Storage Research (JCESR) to drive the Electrolyte Genome efforts within the umbrella of the Materials Project http://materialsproject.org
- Also utilized by the discretionary project NMGC-Mira-2013
- ALCF has developed queue adapters for Fireworks, a variant of the Cobalt scheduler for running Cobalt within Cobalt, and worked to adapt the existing Materials Project infrastructure for use with ALCF resources
- All code, including the ANL Cobalt scheduler are fully open source and available to the wider community.



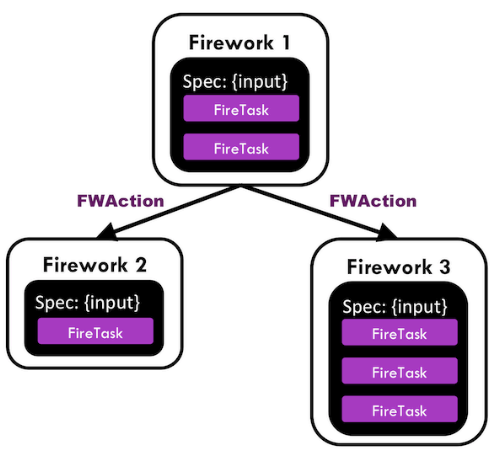
FireWorks

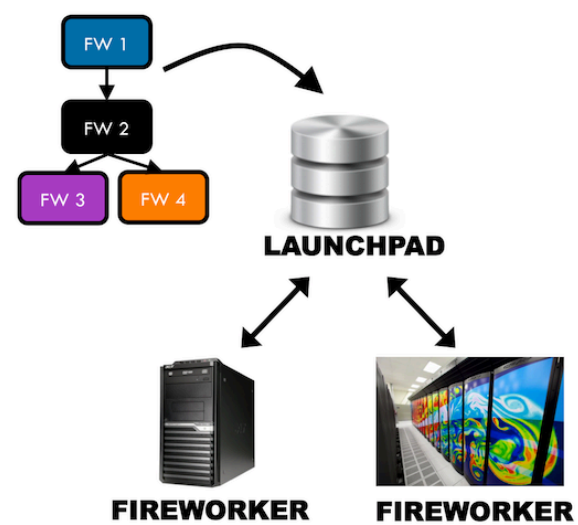
Software Description

- FireWorks is free, open-source software for defining, managing and executing scientific workflows.
- Written in Python and backed by a scalable NoSQL MongoDB database
- FireWorks was developed primarily by Anubhav Jain at Lawrence Berkeley National Lab, using research funding from Kristin Persson for the Materials Project, supported by the U.S. Department of Energy, Batteries for Advanced Transportation Technologies (BATT) and a LDRD grant from LBNL.
- Visit http://pythonhosted.org/FireWorks/ for more information

Key Terms

- **FireTask**: an atomic computing job. It can call a single shell script or execute a single Python function definable through FireWorks or an external program
- **FireWork**: specification for a job composed of one or more FireTasks and their input parameters in JSON.
- **Workflow**: A set of FireWorks which may have dependencies between them
- **LaunchPad** the frontend that acts as a server and manages workflows
- **FireWorker** a compute resource that requests workflows from the LaunchPad, execute them, and send back information.





Challenges

- LCF mission is to support large long running jobs versus many small jobs
  - Addressing via multiple approaches (Cobalt scheduler development, Swift, Fireworks)
- Diverse computational workloads
- Batch versus real-time
  - APS users require fast turnaround of data analysis to ensure the correct data is being captured
- Data movement, archiving, curation