As large-scale experimental facilities, such as light sources, particle colliders, and astronomical surveys, continue to advance their capabilities, researchers must analyze increasingly massive datasets in their efforts to make new discoveries.

To address these growing computing demands, US Department of Energy (DOE) Office of Science Leadership Computing facilities are expanding their scope from primarily providing computing cycles for simulations to deeply integrating into science projects, serving as hubs for research communities at any large-scale observatory or experiment, anywhere in the world.

This article describes two recent projects that have harnessed Argonne Leadership Computing Facility (ALCF) supercomputing resources in a novel way to support major fusion science and particle collider experiments.

**On-Demand Analysis Fuels Frontier Science**

Nuclear fusion promises to provide a safe and clean source of virtually limitless energy. Developing efficient devices that fulfill that potential, sustainably, demands a scientific and engineering effort that is a major area of plasma physics research today. One such device, called a tokamak, is an experimental machine designed to harness the energy of fusion using a powerful magnetic field to confine plasma in the shape of a torus (a doughnut-like shape). Achieving a stable plasma equilibrium requires precise control of the magnetic field lines that wind around the torus in a helical shape.

Scientists at the DIII-D National Fusion Facility, a DOE Office of Science User Facility in San Diego, California, are exploring magnetic confinement of fusion energy by conducting fast-paced plasma physics experiments. These experiments involve creating six-second pulses of confined plasma every 15 to 20 minutes. Planning for each
new pulse is informed by data analysis of the previous pulse, and for some experiments, a fusion science analysis code called SURFMN can assist in that planning. However, as the fine-grid analysis by SURFMN itself takes 20 minutes to complete on dedicated local resources, it hasn’t been possible to obtain these results on a between-pulse timescale.

The ALCF solved this timing mismatch problem by automating and shifting the analysis step to its data analysis cluster, which computed the analysis of every single pulse and returned the results to the research team in a fraction of the time required by the computing resources locally available at DIII-D.

Now, a fine-grid SURFMN run takes under three minutes to complete and interacts directly with the DIII-D databases for automated input/output. The additional computing power also allowed SURFMN runs to be conducted with higher complexity, using a finer grid for the Fourier analysis, yielding more precise results than researchers typically obtain on local systems.

The new service leverages code developed for a previous collaboration with CERN’s Large Hadron Collider (LHC) in Switzerland. In 2015, Argonne scientists used an award of 50 million computing hours on ALCF’s IBM Blue Gene/Q Mira to simulate particle collision events generated by LHC experiments. To facilitate the flow of jobs between CERN and Argonne, the Balsam service was developed to seamlessly interface with LHC systems and execute the workflow on ALCF resources. For DIII-D, Balsam was adapted to interact with the DIII-D systems and to run fusion analyses instead.

The DIII-D research program, operated by private contractor General Atomics for the DOE through the Office of Fusion Energy Sciences, is a large international program with nearly 100 participating institutions and a research team of over 500 users, all working to establish the scientific basis for the optimization of the tokamak approach to fusion energy production. The DIII-D program has helped inform the design of ITER, a worldwide fusion effort now under construction in southern France, including the development of the physics basis for key ITER issues and advanced ITER operation.

Supercomputers Help Address LHC’s Growing Computing Needs

In 2017, experiments conducted with the LHC will generate around 50 Phytes of data that must be simulated and analyzed to aid in the facility’s search for new physics discoveries. That already-massive amount of data is expected to grow exponentially when CERN’s upgrade to the High Luminosity LHC (HL-LHC) becomes operational in 2025.

Since 2002, LHC scientists have relied on the Worldwide LHC Computing Grid for all their data processing and simulation needs. The grid is an international distributed computing infrastructure that links 170 computing centers across 42 countries, allowing data to be accessed and analyzed in near real time by an international community of more than 10,000 physicists working on the LHC’s four major experiments: ATLAS (see Figure 1) and CMS, which use general-purpose detectors to investigate the largest range of physics possible; and ALICE and LHCb, which use detectors specialized for specific phenomena.

The computing grid continues to serve the LHC community well, but the researchers are looking into new technologies, including the addition of some of the world’s most powerful supercomputers, to help meet their growing computing demands.

Over the past few years, ATLAS researchers at Argonne have been part of an effort to use Mira to perform calculations for the ATLAS experiment, whose user community of several thousand researchers spans 38 countries. ALCF researchers helped the ATLAS team scale Alpgen, a Monte Carlo-based event generator, to run efficiently on the ALCF system, enabling the simulation of millions of LHC collision events in parallel, while freeing the rest of LHC’s computing grid to run other, less compute-intensive jobs.

The ultimate goal is to create an end-to-end workflow on ALCF computing resources capable of handling the ATLAS experiment’s intensive computing tasks—event generation, detector simulations, reconstruction, and analysis.

The ALCF is working to deploy an Open Science Grid-developed “gateway” software tool, HTCondor-CE, to

![Figure 1. Artist’s representation of the ATLAS detector at CERN’s Large Hadron Collider, showing particles produced in the aftermath of the collision between two high-energy protons (the truck shown in lower left is depicted for scale). ATLAS researchers have simulated billions of LHC collision events on DOE leadership computing resources. Image credit: Taylor Childers, Joseph A. Insley, and Thomas LeCompte, Argonne National Laboratory.](image-url)
authorize remote users and provide a resource provisioning service that will interact with the ALCF’s job scheduler, Cobalt. With HTCondor-CE in place, ATLAS researchers will be able to run simulations on ALCF resources without having to authenticate every job submission with a CRYPTOCard (a security token that provides controlled access to ALCF login systems). To test the HTCondor-CE setup, the Argonne team has successfully carried out end-to-end production jobs on ALCF’s data analysis cluster, paving the way for larger runs on larger systems in the near future.

Once the simulation workflow is in place at the ALCF, ATLAS researchers will benefit greatly from the ability to perform multiple computing steps at one facility. They will no longer have to store intermediate data, which is required when carrying out different simulation steps at different locations. The ALCF workflow will also reduce compute time by bypassing the setup and finalization time that is required for each step when performed independently. The team also has plans to extend this workflow to their ongoing work with DOE computing resources at the Oak Ridge Leadership Computing Facility and the National Energy Research Scientific Computing Center.

Integration with CERN and the DIII-D facility are early examples of how leadership computing facilities like the ALCF are expanding their scope to identify and deploy services that will enable new and existing projects to make better use of their resources. Leadership computing power will accelerate scientific simulations far beyond what is possible today, while also providing big data analytics and machine learning capabilities across a wide variety of science and engineering domains and disciplines, including large-scale experimental and observational efforts.

Acknowledgments
The Argonne Leadership Computing Facility is a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357. A portion of this material is based on work supported by the DOE Office of Science, Office of Fusion Energy Sciences, using the DIII-D National Fusion Facility, a DOE Office of Science User Facility under awards DE-FC02-04ER54698, DE-AC52-07NA27344, DE-FG02-04ER54761, and DE-AC04-94AL85000. The ATLAS work is supported by the DOE Office of Science, Office of High Energy Physics.

Reference