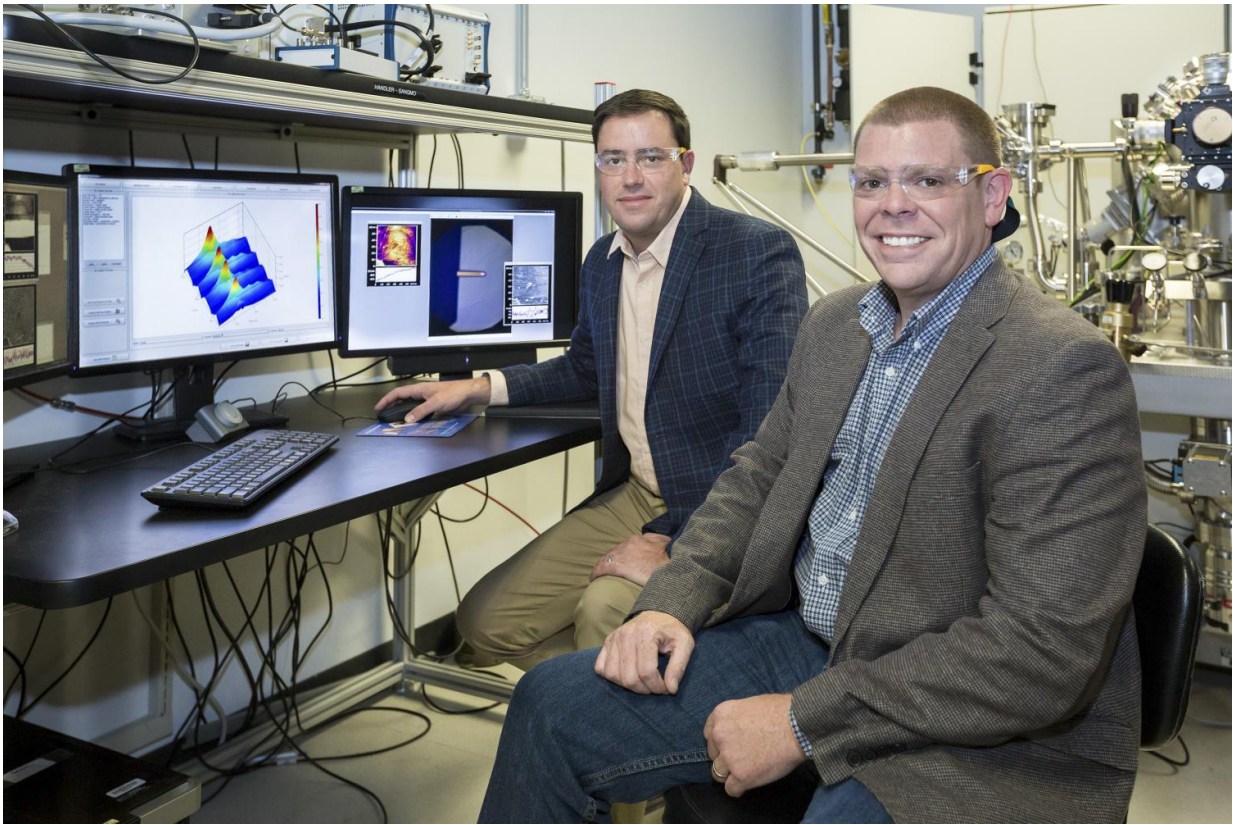


An accelerated pipeline to open materials research

July 21 2016, by Jonathan Hines



This image shows ORNL software engineer Eric Lingerfelt (right) and Stephen Jesse (left) of ORNL's Center for Nanophase Materials Sciences led the development of the Bellerophon Environment for Analysis of Materials (BEAM), an ORNL platform that combines the lab's state-of-the art imaging technologies with advanced data analytics and high-performance computing to accelerate materials science research. Credit: ORNL

Using today's advanced microscopes, scientists are able to capture exponentially more information about the materials they study compared to a decade ago—in greater detail and in less time. While these new capabilities are a boon for researchers, helping to answer key questions that could lead to next-generation technologies, they also present a new problem: How to make effective use of all this data?

At the Department of Energy's Oak Ridge National Laboratory, researchers are engineering a solution by creating a novel infrastructure uniting the lab's state-of-the-art imaging technologies with advanced data analytics and high-performance computing (HPC). Pairing experimental power and computational might holds the promise of accelerating research and enabling new opportunities for discovery and design of advanced [materials](#), knowledge that could lead to better batteries, atom-scale semiconductors, and efficient photovoltaics, to name a few applications. Developing a distributed software system that delivers these advanced capabilities in a seamless manner, however, requires an extra layer of sophistication.

Enter the Bellerophon Environment for Analysis of Materials (BEAM), an ORNL platform that combines scientific instruments with web and data services and HPC resources through a user-friendly interface. Designed to streamline [data analysis](#) and workflow processes from experiments originating at DOE Office of Science User Facilities at ORNL, such as the Center for Nanophase Materials Sciences (CNMS) and Spallation Neutron Source (SNS), BEAM gives materials scientists a direct pipeline to scalable computing, software support, and high-performance cloud storage services provided by ORNL's Compute and Data Environment for Science (CADES). Additionally, BEAM offers users a gateway to world-class supercomputing resources at the Oak Ridge Leadership Computing Facility (OLCF)—another DOE Office of Science User Facility.

The end result for scientists is near-real-time processing, analysis, and visualization of large experimental datasets from the convenience of a local workstation—a drastic improvement over traditional, time-consuming data-analysis practices.

"Processes that once took days now take a matter of minutes," said ORNL software engineer Eric Lingerfelt, BEAM's lead developer.

"Once researchers upload their data into BEAM's online data management system, they can easily and intuitively execute advanced analysis algorithms on HPC resources like CADES's compute clusters or the OLCF's Titan supercomputer and quickly visualize the results. The speedup is incredible, but most importantly the work can be done remotely from anywhere, anytime."

Building BEAM

A team led by Lingerfelt and CNMS's Stephen Jesse began developing BEAM in 2015 as part of the ORNL Institute for Functional Imaging Materials, a lab initiative dedicated to strengthening the ties between imaging technology, HPC, and data analytics.

Many of BEAM's core concepts, such as its layered infrastructure, cloud data management, and real-time analysis capabilities, emerged from a previous DOE project called Bellerophon—a computational workflow environment for HPC core collapse supernova simulations—led by the OLCF's Bronson Messer and developed by Lingerfelt. Initially released in 2010, Bellerophon's database has grown to include more than 100,000 data files and 1.5 million real-time rendered images of more than 40 different core-collapse supernova models.

Applying and expanding Bellerophon's compute and data strategies to the materials realm, however, presented multiple new technical hurdles.

"We spent an entire year creating and integrating the BEAM

infrastructure with instruments at CNMS," Lingerfelt said. "Now scientists are just starting to use it."

Through BEAM, researchers gain access to scalable algorithms—code developed by ORNL mathematicians and computational scientists to shorten the time to discovery. Additionally, BEAM offers users improved data-management capabilities and common data formats that make tagging, searching, and sharing easier. Lowering these barriers for the materials science community not only helps with verification and validation of current findings but also creates future opportunities for scientific discovery. "As we add new features and data-analysis tools to BEAM, users will be able to go back and run those on their data," Lingerfelt said.

A year to hours

One of the first data processing workflows developed for BEAM demonstrates its far-reaching potential for accelerating materials science.

At CNMS, users from around the world make use of the center's powerful imaging instruments to study materials in atomic detail. Conducting analysis of users' data, however, oftentimes slowed scientific progress. One common analysis process required users to format data derived from an imaging technique called band excitation atomic force microscopy. Conducted on a single workstation, the analysis oftentimes took days. "Sometimes people would take their measurement and couldn't analyze it even in the weeks they were here," Jesse said.

By transferring the microscopy data to CADES computing via the BEAM interface, CNMS users gained a 1,000-fold speedup in their analysis, reducing the work to a matter of minutes. A specialized fitting algorithm, which was re-implemented for utilization on HPC resources by ORNL mathematician Eirik Endeve, played a key role in tightening

the feedback loop users relied upon to judge whether adjustments needed to be made to their experiment. "We literally reduced a year of data analysis to 10 hours," Lingerfelt said.

BEAM is also proving its worth at SNS—the most intense pulsed neutron [beam](#) system in the world—by tightening the interplay between theory and experiment. Working with Jose Borreguero from the Center for Accelerating and Modeling Materials at SNS, the BEAM team created a workflow that allows near-real-time comparison of simulation and neutron scattering data leveraging CADES computing. The feedback helps neutron scientists fine-tune their simulations and guides subsequent experiments. In the future, machine-learning algorithms could fully automate the process, freeing up scientists to focus on other parts of their work. "Humans, however, will still be at the center of the scientific process," Lingerfelt said.

"We're not here to replace every single step in the workflow of a scientific experiment, but we want to develop tools that complement things that scientists are already doing," he said.

Adding to the toolbox

Now that BEAM's infrastructure is in place, Lingerfelt's team is collaborating with advanced mathematics, [data](#), and visualization experts at ORNL to regularly augment the software's toolbox.

"Once we've created a fully functioning suite, we want to open BEAM up to other material scientists who may have their own analysis codes but don't have the expertise to run them on HPC," Lingerfelt said. "Down the line we would like to have an open science materials-analysis library where people can validate analysis results publicly."

Currently Lingerfelt's team is developing a suite of algorithms to

conduct multivariate analysis, a highly complex, multidimensional analytic process that sifts through vast amounts of information taken from multiple instruments on the same material sample.

"You need HPC for this type of [analysis](#) to even be possible," Jesse said. "We're gaining the ability to analyze high-dimension datasets that weren't analyzable before, and we expect to see properties in materials that weren't visible before."

More information: E.J. Lingerfelt et al, BEAM: A Computational Workflow System for Managing and Modeling Material Characterization Data in HPC Environments, *Procedia Computer Science* (2016). [DOI: 10.1016/j.procs.2016.05.410](#)

Provided by Oak Ridge National Laboratory

Citation: An accelerated pipeline to open materials research (2016, July 21) retrieved 30 June 2024 from <https://phys.org/news/2016-07-pipeline-materials.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.