

Smart imaging of materials lets national labs look to solving big energy problems

13 August 2015, by Dawn Levy



The “Big, Deep and Smart Data Analytics in Materials Imaging” workshop included a welcome from ORNL’s Associate Laboratory Director for Physical Sciences Michelle Buchanan (top left), activities (top right) and poster sessions (bottom left) at which researchers shared findings, as well as panel discussions (bottom right, with from left, Oak Ridge’s Hans Christen, Brookhaven’s Eric Stach, Berkeley’s James Schuck, Argonne’s Dean Miller, and Oak Ridge’s Sergei Kalinin.) Credit: Oak Ridge National Laboratory, U.S. Dept. of Energy; photographer Genevieve Martin, photo collage Allison Gray

In the Stone, Bronze and Iron Ages, the state of the art of materials science defined technology’s zenith and accelerated economies. Now, in the Information Age, data is beginning to drive the development of advanced materials, from photovoltaics for solar energy and superconductors for efficient electrical transmission to safer nuclear power plants and stronger, lighter vehicles with better batteries. Recently, the Department of Energy Office of Science’s Nanoscale Science Research Centers (NSRCs) at Argonne, Brookhaven, Lawrence Berkeley, Los Alamos/Sandia and Oak Ridge national laboratories jointly organized a workshop at Oak Ridge National Laboratory to discuss opportunities

and challenges as imaging and data sciences merge. Those efforts will likely aid the [Materials Genome Initiative](#), which aims to speed new materials to the global marketplace.

“Combining materials science with [big data](#) could produce a new field, akin to the merger of biology and engineering that created bioengineering,” said Sergei Kalinin, an organizer of the workshop, director for ORNL’s Institute for Functional Imaging of Materials and a theme leader at Oak Ridge’s NSRC.

Companies like Google and Facebook have long grappled with a volume, variety and velocity of data that characterizes it as “big.” Members of the scientific community, however, have differing degrees of experience with “big data.” Physicists sifting through mountains of data from a collider experiment to find signs of an exotic subatomic particle, for example, have more experience with it than do materials scientists examining images of a failed battery material, who often cherry-pick data related to the failure but leave the rest of the data unexamined.

That unmined data may hold vast riches. To reveal them, big data approaches must get deeper and smarter. Big data approaches, which help tell scientists what has happened or is still happening in a material, are all about uncovering correlations. In contrast, deep data strategies—which help tell scientists why something happens by injecting physics-based information into the data-analysis process—use theory to inform experiment and vice versa. Smart data tactics, on the other hand, integrate data analytics directly into the discovery process via a feedback to do all that better.

With its big-data focus, industry isn’t advancing the deep- or smart-data approaches needed to accelerate advances in materials for energy applications. “Big data means correlation and ignores causation,” Kalinin said. A deeper, smarter

approach that merges imaging data with physical laws may allow scientists to understand the causes of problems in existing materials and predict the behaviors of designed materials. But that strategy depends on directly transferring atomically resolved data from scanning transmission electron microscope (STEM) and X-ray experiments to high-performance computing resources for analysis and visualization.

"Powerful imaging techniques demand increasingly large bursts of computing power to drive their data analysis," said David Skinner, who leads strategic partnerships between the National Energy Research Scientific Computing Center (a DOE Office of Science User Facility at Berkeley) and research communities, instrument/experiment data science teams and the private sector. "Accessing shared high-performance computing through fast networks is an increasingly interesting prospect for these data-driven instruments."

Kalinin said, "Facebook and Google use and re-use information already on the web. Our ground floor is to build an instrumental infrastructure that can stream data to the web. Traditionally, imaging instruments were not developed to provide uninterrupted data to the web, so only a small fraction gets analyzed. We need to develop data pipelines."

That was a major message at the recent "[Big, Deep and Smart Data Analytics in Materials Imaging](#)" workshop, which convened researchers from imaging disciplines—electron, scanning probe and optical microscopies; focused X-ray beams; neutron and chemical imaging; atom probe tomography—with experts in mathematics, statistics and computation.

Entering the nano zone

The meeting, held at ORNL's Spallation Neutron Source (SNS), a DOE Office of Science User Facility, attracted nearly 150 researchers from 16 universities, eight DOE national labs, four companies and three other government agencies. The event featured 47 posters, 33 talks, four panel discussions and several lab tours.

Participants discussed opportunities for integrating advanced [data analytics](#) and theory into imaging science. They shared scientific achievements, discussed approaches for theory and experiment to augment each other, and explored prospects for the development of instruments empowered by data-analytic tools. They got a clear picture of ongoing efforts in data-intensive imaging within the NSRCs—DOE user facilities that are open to researchers in industry, academia and government who need to create, characterize and understand nanomaterials and systems.

ORNL's Hans Christen gave an overview of the NSRCs hosting the meeting—the Center for Nanophase Materials Sciences (CNMS) at Oak Ridge, Center for Functional Nanomaterials at Brookhaven, Center for Integrated Nanotechnologies at Los Alamos and Sandia, Center for Nanoscale Materials at Argonne and Molecular Foundry at Berkeley. Christen, who directs the CNMS, said nanoscience centers provide users worldwide with complementary capabilities for multidisciplinary research, as well as specialized capabilities that leverage ties to their collocated facilities.

NSRC users gain insight from experts in modeling, simulation and theory. They employ premier equipment for structural and functional imaging. And they synthesize "materials by design," living out the dream of atom-by-atom assembly that Richard Feynman described in his famed 1959 talk outlining the vision for nanotechnology.

Promising merger

The workshop's speakers shared information about promising projects that merge data and imaging sciences. Still, challenges remain. While data is understood numerically, imaging is not—yet.

A looming challenge is unifying the language of microscopic data to establish common definitions for the "information content" of images. ORNL microscopist Albina Borisevich said she no longer "takes pictures" of materials but instead collects ever-increasing amounts of quantitative data from them. That data provides information about material properties and structures at atomic resolution with a

precision approaching that of X-ray and neutron characterization tools. Engaging advanced computational approaches brings new capabilities in [data analysis](#), such as allowing the analysis of physics and chemistry reflected in picometer-level details of images. "Cross-pollination of different imaging disciplines with computational flavor is already bearing unexpected fruit," she said. "Implementation of the scanning-probe-like beam control allows us to use electron microscopy to fabricate the smallest 3D structures."

James Sethian, a mathematics professor at the University of California–Berkeley, spoke about CAMERA (Center for Applied Mathematics for Energy Research Applications), a pilot project he directs at Lawrence Berkeley National Laboratory that's supported by the DOE Office of Science's Basic Energy Sciences (BES) and Advanced Scientific Computing Research (ASCR) programs. CAMERA convenes interdisciplinary teams of mathematicians, experimental scientists and software engineers to build mathematical models and algorithms for tools critical to users of DOE facilities. "When these teams work together, they can make sense of the deluge of data, and provide the insight to turn data into information that can accelerate our scientific understanding," he emphasized. He described work on [ptychography](#) (which combines scattering and sub-atomic-resolution imaging), image analysis, chemical informatics, GISAXS (grazing-incidence small-angle X-ray scattering) and fast methods for electronic structure calculations.

ORNL mathematician Rick Archibald provided an overview of the [ACUMEN](#) (Accurate Quantified Mathematical Methods for Neutron Science) project, funded by ASCR and focused on the mathematical challenges of scientists at SNS and CNMS. To bring high-performance computing to the massive data sets generated by scientific experiments at ORNL, ACUMEN's partners develop next-generation algorithms for scalable analytics.

ORNL software engineer Eric Lingerfelt described the Bellerophon Environment for Analysis of Materials (BEAM) software system. It will, for the first time, enable instrument scientists at CNMS to leverage ORNL's powerful computational platform

to perform near real-time analysis of experimental data via parallel processing using a web-deliverable, cross-platform Java application. Developed jointly with CNMS's Stephen Jesse, the BEAM system also offers robust long-term data management services and the ability to transmit data files over ORNL's high-speed network directly to the Compute and Data Environment for Science (CADES), a facility providing advanced algorithms and powerful computing for the analysis of vast amounts of data from neutron and X-ray scattering experiments. "BEAM users can easily manipulate remote directories and data in their private storage area on CADES as if they were browsing their local workstation," Lingerfelt said.

Nigel Browning of Pacific Northwest National Laboratory (PNNL) described methods, statistics and algorithms to extract information from images obtained using aberration-corrected electron microscopy, which enables very high-resolution imaging of increased data quality and quantity. Compressive sensing, for example, pays attention to bits of a sample and uses signal processing to fill in the blanks. Kerstin Kleese van Dam, also of PNNL, spoke about streaming analysis of dynamic imaging experiments that promise to capture evolving processes in materials under operating conditions.

Managing unprecedented data streams is a big challenge. Fortunately, the colocation of NSRCs with other facilities grappling with this elephantine issue gives DOE nanocenters a huge advantage in finding solutions. RHIC, an accelerator at Brookhaven looking at the quark gluon plasma, and ATLAS, a detector at CERN's Large Hadron Collider, are both high-energy physics projects that generate lots of data. The RHIC & ATLAS Computing Facility at Brookhaven manages the data for both. Eric Stach, who leads the Electron Microscopy Group in the Center for Functional Nanomaterials at Brookhaven, noted that the RHIC/ATLAS detector curated 160 petabytes of data in 2013 and will surpass 200 petabytes this year. So materials scientists have learned a lot from nearby physicists—a boon because a single STEM instrument can produce a data flow similar to that of the ATLAS detector, according to Stach. He said, "The introduction of sensitive new detectors

and ultra-bright sources is leading to an explosion of people; it is the defects in them that make them rich materials data—we expect to have more than 20 interesting." Sumpter echoed this sentiment: petabytes generated each year at the user facilities "Defects are not doom if you understand what they at Brookhaven. That's the data equivalent of one- are and do." fifth of every Google search done in 2013."

Big data and mathematical methods can build the bridge needed to link theory to experiment. One problem has been data takes longer to process (e.g., a month on an 8-core computer) than to acquire (say, 10 hours). For ORNL's Borisevich and Andrew Lupini, that problem had a solution. They will be able to acquire ultrafast data from STEM and pipe it directly to a high-performance computing infrastructure for analysis. That infrastructure includes Titan, America's most powerful supercomputer, which is housed at the Oak Ridge Leadership Computing Facility, a DOE Office of Science User Facility. Titan has 299,008 CPU cores to guide simulations while accompanying GPUs handle hundreds of calculations simultaneously.

Experiment and theory work hand in hand to show how the real structure and function of a material compare with the ideal. Experiment helps inform and validate theory and theory-based models. Highly resolved imaging techniques provide information about atoms that is put into the theory. Scientists create a model based on theory and use it to make inferences about a material's properties. Theory can connect pieces given from experiment, such as how properties change upon, for instance, introducing a dopant. Researchers can fill in information and move forward to design better materials.

Neutron scattering gives detailed information about bulk materials, whereas microscopy can examine interfaces and surfaces at high spatial resolution. Combining multimodal experiment and theory advances the advent of materials by design. "This is the first time in history we've matched experiment with theory," ORNL theorist Bobby Sumpter said. "We should have some success."

Success may mean understanding structural deviations called "defects" in atomically ordered materials, such as crystals. As theoretical physicist Sir Charles Frank used to say, "Crystals are like

Combining neutron scattering and microscopy can inform theories and models for predicting properties of designed materials. Kalinin said, "Once we have the infrastructure to stream our data from microscopes and we can measure structures and properties, we can start to build libraries of structure–property relationships on the single-defect level. We can verify libraries against X-ray and neutron scattering methods and know if a library is complete."

At SNS, which provides the world's most intense pulsed neutron beams for scientific research and industrial development, experiments on 20 beam lines measure the structures and dynamics of materials in diverse applications from biology to additive manufacturing. The data has a lifecycle from the time the neutron hits the detector to the identification of scientifically interesting aspects. These data sets are huge, and scientists can't keep up with them.

Thomas Proffen directs the Neutron Data Analysis and Visualization Division in ORNL's Neutron Sciences Directorate and heads the Center for Accelerating Materials Modeling (CAMP), funded by BES for direct integration of simulation and modeling into the analysis loop for data from neutron experiments. Direct integration allows scientists to refine theoretical models against experimental observations, use models to predict where new experimental measurements should be performed, and analyze some data at the user facility before taking it to the home institution for full analysis. "Neutron events are streamed and processed live, allowing a near-real-time view of collected data so a scientist running the experiment can make decisions on the fly," Proffen said. "To visualize data, we play with everything from virtual reality headsets to volume rendering on parallelized servers."

Recommending a path forward

The meeting concluded with a panel discussion,

moderated by CNMS's Christen, of what DOE's nanoscience centers can and should do to help the scientific community strengthen the link between data and imaging sciences.

"We're transitioning from imaging being a qualitative tool to a quantitative tool," Kalinin pointed out. "This requires hardware platforms and, lo and behold, because of the previous investment made by DOE and other funding agencies, these platforms exist. These platforms happen to be extremely data-intensive in terms of data generation and extremely demanding in terms of analytics. And at the same time we have all the capabilities enabled by supercomputing. This feels like a perfect storm."

To use knowledge from imaging to improve predictive theories, scientists need a proper mathematical framework implemented on machines they can use, Kalinin said.

Stach of Brookhaven's Center for Functional Nanomaterials said the lack of staff specifically associated with data management and data methods is a problem. "We don't have a central resource (for users and staff) where there's deep expertise in these areas," he said. Having on-board experts acting as liaisons to centers at the forefront of the big-data tsunami, such as light and neutron sources, could spur opportunities to gain knowledge of best practices and incorporate different approaches.

"We're still at the place where there's a lot of 'artisanal software' being put together," said James Schuck, director of the Imaging and Manipulation of Nanostructures Facility at Berkeley's Molecular Foundry. "This is where we need people to start talking to one another. The data modeling people are starting to talk to the data generators; that's key. However, there's still plenty of jargon, which is a sign they're half talking past each other." The solution is having people who spend their time thinking about this. "This is an opportunity DOE and the NSRCs have. There's a reason the NSRCs are at the [national] labs. It's because they are collocated with other facilities. There's a chance for real interaction to occur."

Dean Miller of Argonne's Center for Nanoscale Materials agreed, emphasizing the need for standardized data formats. "We need teamwork," he said. "Vendors may have a proprietary format, but we can push them to make sure that their software will provide the data in an open-source data format as well."

Miller said his electron microscopy center is collocated with other facilities at the forefront of big-data management. "We're benefitting tremendously from [researchers at the collocated Advanced Photon Source] tackling many of these problems," he said. As the nation's nanocenters were built, community involvement and participation from other facilities were critical. "In terms of now addressing these challenges for data, that same strategy holds: We need community involvement to guide us."

"We absolutely need to be doing continuous education," Kalinin said to an audience and panel enthusiastic about the suggestion. He cited the success of specialized "boot camp" workshops, e.g., in computing topics. "Those of us who got PhDs 10 years ago or more have very limited if any clue about the power of the modern statistical methods—computation and so on. As part of our regular job routine, the chances of doing it are remote," he said. "Without this, we simply won't be able to learn tools in the areas which are vital to us."

A university professor in the audience said, "I push my students, but they don't have the tools to use." She suggested students and teachers as well as scientists could benefit from boot-camp-style online teaching modules. Stach praised the National Science Foundation's [nanoHUB](#) tool to disseminate computer programs helpful in nanoscience and nanotechnology. "The effectiveness of online education is very clear," he said.

Schuck confessed awe at the fact that a colleague can take three-dimensional tomographs, and while the experiment is ongoing, analyze and reconstruct the data in multiple ways and re-ask questions. "I couldn't do that three years ago. We've seen knowledge accelerate," he said. "If a tutorial is made, I'd like to take that class." He pointed out that many attendees of the DOE nanoscience-

centers workshop were also planners of a [data](#)-related symposium at the upcoming meeting of the Materials Research Society and could suggest workshops or tutorials to benefit the broader community.

Provided by Oak Ridge National Laboratory

APA citation: Smart imaging of materials lets national labs look to solving big energy problems (2015, August 13) retrieved 15 November 2019 from <https://phys.org/news/2015-08-smart-imaging-materials-national-labs.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.