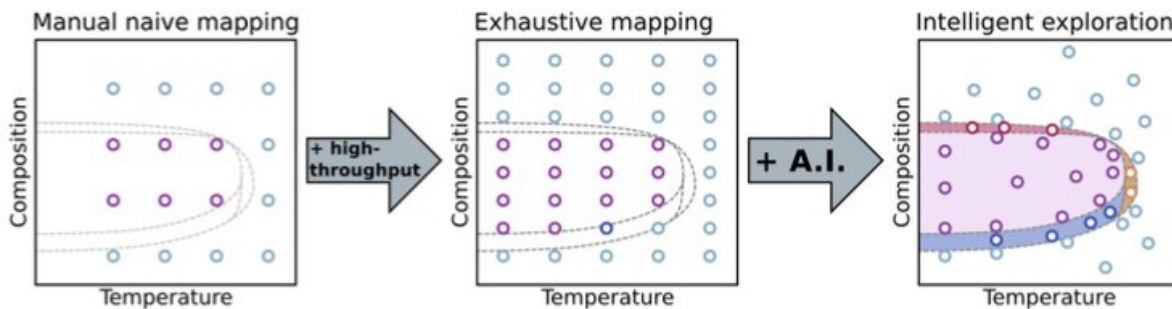


Automatically steering experiments toward scientific discovery

July 28 2021



Moving from a manual to automated experimentation approach enables scientists to more thoroughly explore parameter spaces. With artificial intelligence (AI) decision-making methods, scientists can home in on key parts of the parameter space (here, composition and temperature) for accelerated material discovery. Credit: Brookhaven National Laboratory

In the popular view of traditional science, scientists are in the lab hovering over their experiments, micromanaging every little detail. For example, they may iteratively test a wide variety of material compositions, synthesis and processing protocols, and environmental conditions to see how these parameters influence material properties. In each iteration, they analyze the collected data, looking for patterns and relying on their scientific knowledge and intuition to select useful follow-on measurements.

This manual approach consumes limited instrument time and the attention of human experts who could otherwise focus on the bigger picture. Manual experiments may also be inefficient, especially when there is a large set of parameters to explore, and are subject to human bias—for instance, in deciding when one has collected enough data and can stop an experiment. The conventional way of doing science cannot scale to handle the enormous complexity of future scientific challenges. Advances in scientific instruments and data analysis capabilities at experimental facilities continue to enable more rapid measurements. While these advances can help scientists tackle complex experimental problems, they also exacerbate the human bottleneck; no human can keep up with modern experimental tools!

Envisioning automation

One such facility managing these types of challenges is the National Synchrotron Light Source II (NSLS-II) at the U.S. Department of Energy's (DOE) Brookhaven National Laboratory. By directing light beams, ranging from infrared to hard X-rays, toward samples at experimental stations (beamlines), NSLS-II can reveal the electronic, chemical, and atomic structures of materials. When scientists were designing these beamlines a decade ago, they had the foresight to incorporate automation enabled by machine learning (ML) and artificial intelligence (AI)—now an exploding field—as part of their vision.

"We thought, wouldn't it be great if scientists could not only do measurements faster but also do intelligent exploration—that is, explore scientific problems in smarter, more efficient ways by leveraging modern computer science methods," said Kevin Yager, leader of the Electronic Nanomaterials Group of the Center for Functional Nanomaterials (CFN) at Brookhaven Lab. "In fact, at the CFN, we've defined one of our research themes to be accelerated nanomaterial discovery."

This idea for a highly automated beamline that could intelligently explore scientific problems ended up becoming a long-term goal of the Complex Materials Scattering (CMS) beamline, developed and operated by a team led by Masafumi Fukuto.

"We started by building high-throughput capabilities for fast measurements, like a sample-exchanging robot and lots of in-situ tools to explore different parameters such as temperature, vapor pressure, and humidity," said Fukuto. "At the same time, we began thinking about automating not just the beamline hardware for data collection but also real-time data analysis and experimental decision making. The ability to take measurements very quickly is useful and necessary but not sufficient for revolutionary materials discovery because material parameter spaces are very large and multidimensional."

For example, one experiment may have a parameter space with five dimensions and more than 25,000 distinct points within that space to explore. Both the data acquisition and analysis software to deal with these large, high-dimensional parameter spaces were built in house at Brookhaven. For data collection, they built on top of [Bluesky](#) software, which NSLS-II developed. To analyze the data, Yager wrote code for an image-analysis software called [SciAnalysis](#).

Closing the loop

In 2017, Fukuto and Yager began collaborating with Marcus Noack, then a postdoc and now a research scientist in the Center for Advanced Mathematics for Energy Research Applications (CAMERA) at DOE's Lawrence Berkeley National Laboratory. During his time as a postdoc, Noack was tasked with collaborating with the Brookhaven team on their autonomous beamline concept. Specifically, they worked together to develop the last piece to create a fully automated experimental setup: a decision-making algorithm. The Brookhaven team defined their needs,

while Noack provided his applied mathematics expertise and wrote the software to meet these needs.

By leveraging AI and ML, this algorithm determines the best next measurements to make while an experiment is ongoing. (AI refers to a machine simulating human behavior, while ML is a subfield of AI in which a machine automatically learns from past data.) For the algorithm to start modeling a system, it's as simple as a user defining the inputs and outputs: what are the variables I can control in the experiment, and what am I going to measure? But the more information humans provide ahead of time—such as the expected response of the system or known constraints based on the particular problem being studied—the more robust the modeling will be. Behind the scenes, a Gaussian process is at work modeling the system's behavior.

"A Gaussian process is a mathematically rigorous way to estimate uncertainty," explained Yager. "That's another way of saying knowledge in my mind. And that's another way of saying science. Because in science, that's what we're most interested in: What do I know, and how well do I know it?"

"That's the ML part of it," added Fukuto. "The algorithm goes one step beyond that. It automatically makes decisions based on this knowledge and human inputs to select which point would make sense to measure next."

In a simplistic case, this next measurement would be the location in the parameter space where information gain can be maximized (or uncertainty reduced). The team first demonstrated this [proof of concept](#) in 2019 at the NSLS-II CMS beamline, imaging a nanomaterial film made specifically for this demonstration.

Since this initial success, the team has been making the algorithm more

sophisticated, applying it to study a wide range of real (instead of contrived) scientific problems from various groups, and extending it to more experimental techniques and facilities.

While the default version of the algorithm aims to minimize uncertainty or maximize knowledge gain in an iterative fashion, there are other ways to think about where to focus experimental attention to obtain the most value. For example, for some scientists, the cost of the experiment—whether its duration or amount of materials used—is important. In other words, it's not just *where* you take the data but how *expensive* it is to take those data. Others may find value in homing in on specific features, such as boundaries within a parameter space or grain size of a crystal. The more sophisticated, flexible version of the algorithm that Noack developed can be programmed to have increased sensitivity to these features.

"You can tune what your goals are in the experiment," explained Yager. "So, it can be knowledge gain, or knowledge gain regulated by experimental cost or associated with specific features."

Other improvements include the algorithm's ability to handle the complexity of real systems, such as the fact that materials are inhomogeneous, meaning they are not the same at every point across a sample. One part of a sample may have a uniform composition, while another may have a variable composition. Moreover, the algorithm now takes into account anisotropy, or how individual parameters can be very different from each other in terms of how they affect a system. For example, "x" and "y" are equivalent parameters (they are both positional coordinates) but temperature and pressure are not.

"Gaussian processes use kernels—functions that describe how data points depend on each other across space—for interpolation," said Noack. "Kernels have all kinds of interesting mathematical properties.

For instance, they can encode varying degrees of inhomogeneity for a sample."

Increasing the sophistication of the algorithm is only part of the challenge. Then, Fukuto and Yager have to integrate the updated algorithm into the closed-loop automated experimental workflow and test it on different experiments—not only those done in-house but also those performed by users.

Deploying the method to the larger scientific community

Recently, Fukuto, Yager, Noack, and colleagues have deployed the autonomous method to several real experiments at various NSLS-II beamlines, including CMS and Soft Matter Interfaces (SMI). Noack and collaborators have also deployed the method at LBNL's Advanced Light Source (ALS) and the Institut Laue-Langevin (ILL), a neutron scattering facility in France. The team released their decision-making software, [gpCAM](#), to the wider scientific community so anyone could set up their own autonomous experiments.

In one experiment, in collaboration with the U.S. Air Force Research Laboratory (AFRL), they used the method in an autonomous synchrotron X-ray scattering experiment at the CMS beamline. In X-ray scattering, the X-rays bounce off a sample in different directions depending on the sample's structure. The first goal of the experiment was to explore how the ordered structure of nanorod-polymer composite films depends on two fabrication parameters: the speed of film coating and the substrate's chemical coating. The second goal was to use this knowledge to locate and home in on the regions of the films with the highest degrees of order.

"These materials are of interest for optical coatings and sensors," explained CMS beamline scientist [Ruipeng Li](#). "We used a particular fabrication method that mimics industrial roll-to-roll processes to find out the best way to form these ordered films using industrially scalable processes."

In another X-ray scattering experiment, at the SMI beamline, the algorithm successfully identified regions of unexpected ordering in a parameter space relevant to the self-assembly of block copolymer films. Block copolymers are polymers made up of two or more chemically distinct "blocks" linked together. By identifying these features, the autonomous experiment illuminated a problem with the fabrication method.

"It wasn't hypothetical—we've been working on this project for many years," said CFN materials scientist Gregory Doerk. "We had been iterating in the old way, doing some experiments, taking images at locations we arbitrarily picked, looking at the images, and being puzzled as to what's going on. With the autonomous approach, in one day of experiments at the beamline, we were able to find the defects and then immediately fix them in the next round. That's a dramatic acceleration of the normal cycle of research where you do a study, find out it didn't work, and go back to the drawing board."

Noack and his collaborators also applied the method to a different kind of X-ray technique called autonomous synchrotron infrared mapping, which can provide chemical information about a sample. And they demonstrated how the method could be applied to a spectroscopy technique to autonomously discover phases where electrons behave in a strongly correlated manner and to neutron scattering to autonomously measure magnetic correlations.

Shaping the future of autonomous experimentation

According to Yager, their method can be applied to any technique for which the data collection and data analysis are already automated. One of the advantages of the approach is that it's "physics agnostic," meaning it's not tied to any particular kind of material, physical problem, or technique. The physically meaningful quantities for the decision making are extracted through the analysis of the raw data.

"We wanted to make our approach very general so that it could be applied to anything and then down the road tailored to specific problems," said Yager. "As a user facility, we want to empower the largest number of people to do interesting science."

In the future, the team will add functionality for users to incorporate physics awareness, or knowledge about the materials or phenomena they're studying, if they desire. But the team will do so in a way that doesn't destroy the general-purpose flexibility of the approach; users will be able to turn this extra knowledge on or off.

Another aspect of future work is applying the method to control real-time processes—in other words, controlling a system that's dynamically evolving in time as an experiment proceeds.

"Up until this point, we've been concentrating on making decisions on how to measure or characterize prepared material systems," said Fukuto. "We also want to make decisions on how to change materials or what kinds of materials we want to make. Understanding the fundamental science behind material changes is important to improving manufacturing processes."

Realizing this capability to intelligently explore materials evolving in real time will require overcoming algorithmic and instrumentation challenges.

"The decision making has to be very fast, and you have to build sample environments to do materials synthesis in real time while you're taking measurements with an X-ray beam," explained Yager.

Despite these challenges, the team is excited about what the future of autonomous experimentation holds.

"We started this effort at a very small scale, but it grew into something much larger," said Fukuto. "A lot of people are interested in it, not just us. The user community has been expanding, and with users studying different kinds of problems, this approach could have a great impact on accelerating a host of scientific discoveries."

"It represents a really big shift in thinking to go from the old way of micromanaging experiments to this new vision of automated systems running experiments with humans orchestrating them at a very high level because they understand what needs to be done and what the science means," said Yager. "That's a very exciting vision for the future of science. We're going to be able to tackle problems in the future that 10 years ago people would have said are impossible."

Provided by Brookhaven National Laboratory

Citation: Automatically steering experiments toward scientific discovery (2021, July 28) retrieved 11 May 2024 from <https://phys.org/news/2021-07-automatically-scientific-discovery.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.