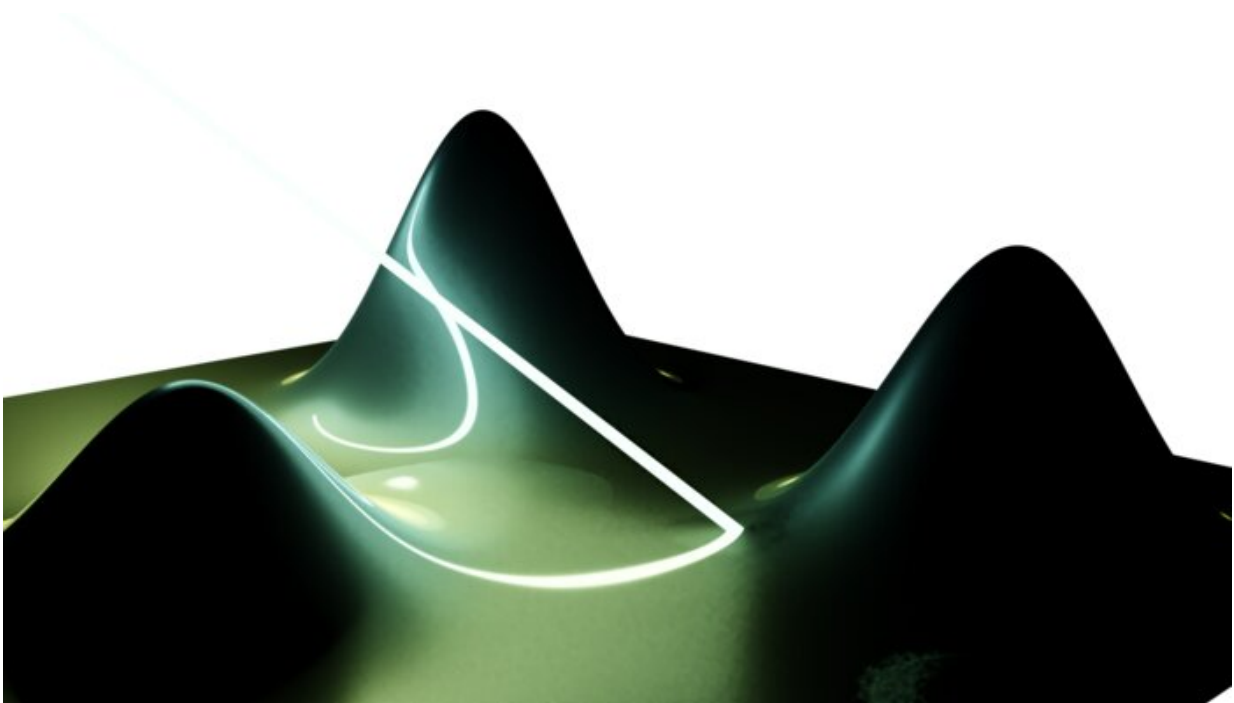# Berkeley Lab's CAMERA leads international effort on autonomous scientific discoveries

July 28 2021

An artistic illustration of a mixture of Gaussian processes and a light or particle beam passing through. The image alludes to the inner workings of the algorithm inside gpCAM, a software tool developed by researchers at Berkeley Lab's CAMERA facility to facilitate autonomous scientific discovery. Credit: Marcus Noack, Berkeley Lab

Experimental facilities around the globe are facing a challenge: their instruments are becoming increasingly powerful, leading to a steady

increase in the volume and complexity of the scientific data they collect. At the same time, these tools demand new, advanced algorithms to take advantage of these capabilities and enable ever-more intricate scientific questions to be asked—and answered. For example, the ALS-U project to upgrade the Advanced Light Source facility at Lawrence Berkeley National Laboratory (Berkeley Lab) will result in 100 times brighter soft X-ray light and feature superfast detectors that will lead to a vast increase in data-collection rates.

To make full use of modern instruments and facilities, researchers need new ways to decrease the amount of data required for [scientific discovery](#) and address data acquisition rates humans can no longer keep pace with. A promising route lies in an emerging field known as autonomous discovery, where algorithms learn from a comparatively little amount of input data and decide themselves on the next steps to take, allowing multi-dimensional parameter spaces to be explored more quickly, efficiently, and with minimal human intervention.

"More and more experimental fields are taking advantage of this new optimal and autonomous data acquisition because, when it comes down to it, it's always about approximating some function, given noisy data," said Marcus Noack, a research scientist in the Center for Advanced Mathematics for Energy Research Applications (CAMERA) at Berkeley Lab and lead author on a new paper on Gaussian processes for autonomous data acquisition published July 28 in *Nature Reviews Physics*. The paper is the culmination of a multi-year, multinational effort led by CAMERA to introduce innovative autonomous discovery techniques across a broad scientific community.

Stochastic Processes Take the Lead

Over the last few years, autonomous discovery methods have become more sophisticated, with stochastic processes (for instance, Gaussian

process regression [GPR]) emerging as the method of choice for steering many classes of experiments. The success of GPR in steering experiments is due to its probabilistic nature, which allows us to make decisions based on the uncertainty of the current model. This is what lies at the heart of gpCAM, a software tool developed by CAMERA.

"In contrast to deep learning, stochastic processes can be used to make decisions based on relatively small datasets, and they provide uncertainty estimates which can optimize the learning process," Noack said.

While CAMERA's initial research efforts have focused primarily on synchrotron beamline experiments, a growing number of scientists in other disciplines are now seeing the advantages of incorporating autonomous discovery techniques into their experimental project workflows. In April, a workshop on autonomous discovery in science and engineering sponsored by CAMERA and chaired by Noack attracted hundreds of scientists from around the world, reflecting the expanding interest in this emerging field.

"We are still in the early days with this, but much progress has been made in the past year," said Martin Böhm, an instrument scientist in the spectroscopy group of Institut Laue-Langevin in Grenoble, France, and a co-author on the Nature Reviews Physics paper. "For spectrometry, for example, it offers a new way of doing experiments and lets the instruments do the work, which results in time savings for users." Other potential application areas include physics, math, chemistry, biology, materials science, environmental studies, drug discovery, computer science, and electrical engineering.

Multiple Uses Emerging

For example, John Thomas, a post-doctoral research fellow in Berkeley Lab's Molecular Foundry, is using photo-coupled scanning probe

microscopy to understand material properties of thin-film semiconducting systems and has been working with gpCAM to enhance these efforts.

"Nanoscale applications that make use of artificial intelligence and machine learning algorithms, specifically for scanning probe systems, have been an interest in the Weber-Bargioni group [at the Foundry] for some time," Thomas said. "We became interested in using Gaussian processes toward autonomous discovery in the summer of 2020."

The group recently completed an application that makes use of gpCAM within a Python-to-LabVIEW interface, where, with some user input for initialization, gpCAM drives an atomically sharp probe across a semiconductive two-dimensional material for hyperspectral data collection. Images obtained represent a convolution of both electronic and topographic information, and point spectroscopy extracts local electronic structure.

"Autonomous driving of scanning probe instruments, without the need for constant human operation, can optimize tool performance for engineers and scientists by continuing experiments during off-business hours or providing routes for simultaneous tasks within a given workflow; that is, the tool can be set up for an autonomous run while the user can efficiently make use of the time allowed," Thomas said. "As a result, we can now use Gaussian processes to map out and identify defective regions in 2D heterostructures with sub-Ångström resolution."

Aaron Michelson, a graduate researcher in the Oleg Gang group at Columbia University working on DNA origami-based self-assembly, is just beginning to apply gpCAM to his research. For one project, it is helping him and his colleagues investigate the thermal annealing history of DNA origami superlattices at the nanoscale; in another, it's being used to mine large datasets from 2D X-ray microscopy experiments.

"DNA nanotechnology in the pursuit of self-assembling functional material often suffers from a limited ability to sample the large parameter space for synthesis," he said. "Either this requires a large volume of data to be collected or a more efficient solution to experimentation. Autonomous discovery can be directly incorporated in both mining large datasets and guiding new experiments. This allows the researcher to steer away from mindlessly making more samples and puts us in the driver's seat to make decisions."

"Noack's work and leadership have brought together a broad, interdisciplinary co-design community. This sort of scientific community building is at the heart of what CAMERA tries to do," said CAMERA Director James Sethian, a co-author on the *Nature Reviews Physics* paper.

More information: Marcus M. Noack et al, Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities, *Nature Reviews Physics* (2021). [DOI: 10.1038/s42254-021-00345-y](#)

Provided by Lawrence Berkeley National Laboratory