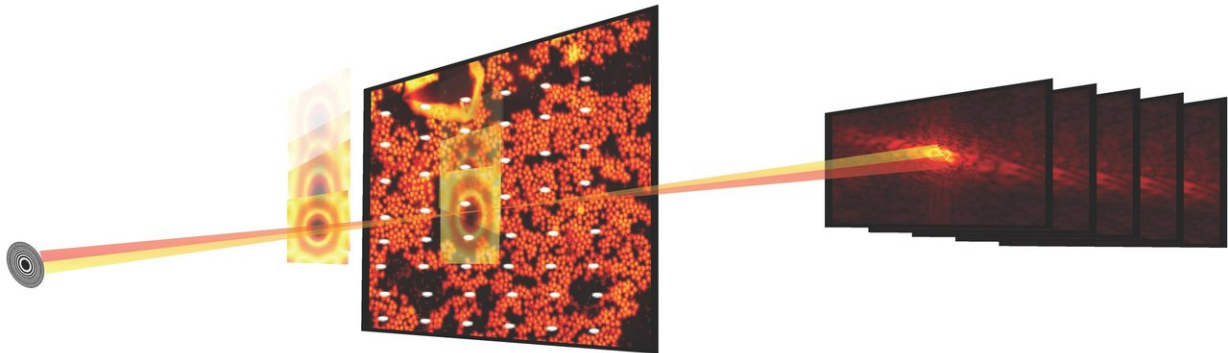


Berkeley Lab paves the way for real-time ptychographic data streaming

June 5 2018, by Kathy Kincade



In scientific experiments, ptychographic imaging combines scanning microscopy with diffraction measurements to characterize the structure and properties of matter and materials. Advances in detectors and x-ray microscopes at light sources have made it possible to now measure a ptychographic dataset in seconds. Credit: Stefano Marchesini, Berkeley Lab

What began nearly a decade ago as a Berkeley Lab Laboratory-Directed

Research and Development (LDRD) proposal is now a reality, and it is already changing the way scientists run experiments at the Advanced Light Source (ALS)—and, eventually, other light sources across the Department of Energy (DOE) complex—by enabling real-time streaming of ptychographic image data in a production environment.

In scientific experiments, ptychographic imaging combines scanning microscopy with diffraction measurements to characterize the structure and properties of matter and materials. While the method has been around for some 50 years, broad utilization has been hampered by the fact that the experimental process was slow and the computational processing of the data to produce a reconstructed image was expensive. But in recent years advances in detectors and X-ray microscopes at light sources such as the ALS have made it possible to measure a ptychographic dataset in seconds.

Reconstructing ptychographic datasets, however, is no trivial matter; the process involves solving a difficult phase-retrieval problem, calibrating optical elements and dealing with experimental outliers and "noise." Enter SHARP (scalable heterogeneous adaptive real-time ptychography), an algorithmic framework and computer software that enables the reconstruction of millions of phases of ptychographic image data per second. Developed at Berkeley Lab through an international collaboration and rolled out in 2016, SHARP has had a demonstrable impact on productivity for scientists working at the ALS and other light sources across the Department of Energy complex.

Now an inter-government agency funded collaboration of scientists from Berkeley Lab's DOE-funded Center for Advanced Mathematics for Energy Research Applications (CAMERA), the ALS and STROBE, the National Science Foundation's Science and Technology Center, has yielded another first-of-its-kind advance for ptychographic imaging: a software/algorithmic pipeline that enables real-time streaming of

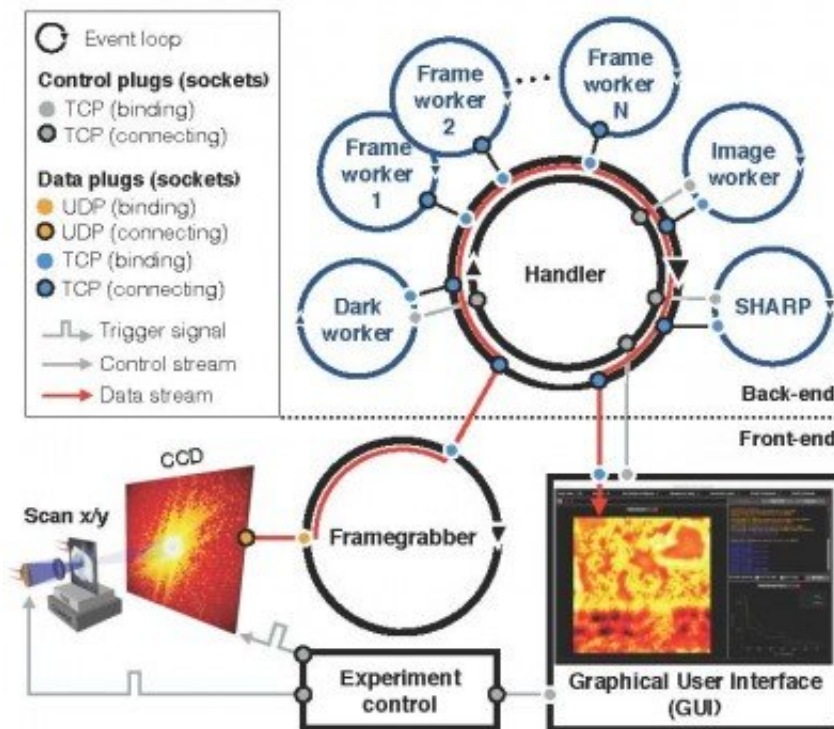
ptychographic image data during a beamline experiment, providing throughput, compression and resolution as well as rapid feedback to the user while the experiment is still running.

Integrated, Modular, Scalable

A successor of the Nanosurveyor concept first introduced in a 2010 LDRD proposal, the modular, scalable Nanosurveyor II system—now up and running at the ALS—employs a two-sided infrastructure that integrates the ptychographic image data acquisition, preprocessing, transmission and visualization processes. The front-end component at the beamline comprises the detector, frame-grabber, a scalable preprocessor and graphical user interface (GUI), while the back-end component at the compute cluster contains SHARP for high-resolution image reconstruction combined with communication channels for data flow, instructions and synchronization. In operation, once a new scan has been triggered by the experiment control, the frame-grabber continuously receives raw data packets from the camera, assembles them to a frame and, in parallel, preprocesses a set of raw frames into a clean and reduced block for the computational back end. Incoming blocks are distributed among the SHARP workers in the computation back end, and reduced data are sent back to the front end and visualized in a GUI. The new streaming software system is being rolled out at the ALS for active use in experiments on the beamlines, such as the COSMIC X-ray microscopy platform that recently came online.

"What we have succeeded in doing is making very complex computation and data workflow as transparent as possible," said David Shapiro, a staff scientist in the Experimental Systems Group at the ALS who has been instrumental in facilitating the implementation and testing of the new system. "Once the software is going, microscope operators have no idea they are generating gigabytes of data and it's going through this pipeline and then coming back to them."

ALS users typically are able to use the beamlines for about two days, during which they may study about a dozen samples, Shapiro noted. This means they only have a few hours per sample, and they need to make a number of decisions about the data within those hours. The pythographic real-time streaming capability gives them real-time feedback and helps them to make more accurate decisions.



The modular, scalable Nanosurveyor II system—now up and running at the ALS—employs a two-sided infrastructure that integrates the pythographic image data acquisition, preprocessing, transmission and visualization processes. Credit: Lawrence Berkeley National Laboratory

"The fact that SHARP can now take data in real time and process it immediately, even if the experiment is not complete, makes it high-performance and operational for science in a production environment,"

said Stefano Marchesini, a CAMERA scientist and the PI on the original LDRD proposal.

Production-ready Environment

Integrating all of the components was one of the biggest challenges in implementing Nanosurveyor II, as was getting it ready to function in a production-ready environment versus a testbed, according to Hari Krishnan, a computer systems engineer at CAMERA who was responsible for developing the pipeline's communication engine. It took a multiyear effort to bring the real-time data-processing capabilities to fruition, with contributions from mathematicians, computer scientists, software engineers, physicists and beamline scientists. In addition to Marchesini and Krishnan, CAMERA scientists Pablo Enfedaque and Huibin Chang were involved in developing the SHARP interface and the communication engine. In addition, Filipe Maia and Benedict Daurer of Uppsala University, which has a collaboration with CAMERA, played an integral in the development of the streaming software. Meanwhile, physicist Bjoern Enders of Berkeley's Physics Department and STROBE was focused on the detector and preprocessor, and Berkeley Physics Ph.D. student Kasra Nowrouzi was involved with pipeline implementation.

"It was important to separate the system's responsibilities and tasks and then find a common interface to communicate over," Enders said of the team effort. "This put part of the responsibility back into the beamline. All of us worked together to cleanly separate the streaming algorithm from the preprocessing part so we can ensure that the beamline actually works."

Looking ahead, Nanosurveyor II's scalability will be critical as the next generation of light sources and detectors come online. The pipeline's data rates mirror that of the detector; the current detector operates at a

rate of up to 400 megabytes per second and can generate a few terabytes of data per day. Next-generation detectors will produce data 100 to 1,000 times faster.

"The streaming solution works as fast as the data acquisition at this point," Enfechaque said. "You can reconstruct the ptychographic image at the speed that the data is being captured and pre-processed."

In addition, the team is looking to enhance the pipeline's capabilities by incorporating machine-learning algorithms that can further automate the data analysis process.

"The goal is to make it seamless," Krishnan said. "Right now what the users look at is a coarse image to make quick decisions and a finer image to make more complex decisions. Eventually we want to get away from the raw frames and data altogether and end up in this realm where all decisions and all points are just a single [reconstructed image](#). All these other intermediate things disappear."

"The data-streaming capability brings analysis into the time scale of human decision making for operators, but it also facilitates the future, which would be machines making decisions and guiding the measurements," Shapiro added.

Provided by Lawrence Berkeley National Laboratory

Citation: Berkeley Lab paves the way for real-time ptychographic data streaming (2018, June 5) retrieved 17 July 2024 from <https://phys.org/news/2018-06-berkeley-lab-paves-real-time-ptychographic.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.