Basic Energy Science Advisory Committee Meeting
Computational Infrastructure Panel

December 7, 2022

Data Challenges and Drivers at the Light Sources

Nicholas Schwarz, Advanced Photon Source, Argonne National Laboratory
Stuart Campbell, National Synchrotron Light Source II, Brookhaven National Laboratory
Alexander Hexemer, Advanced Light Source, Lawrence Berkeley National Laboratory
Vivek Thampy, Stanford Synchrotron Radiation Lightsource, SLAC
Jana Thayer, Linac Coherent Light Source, SLAC

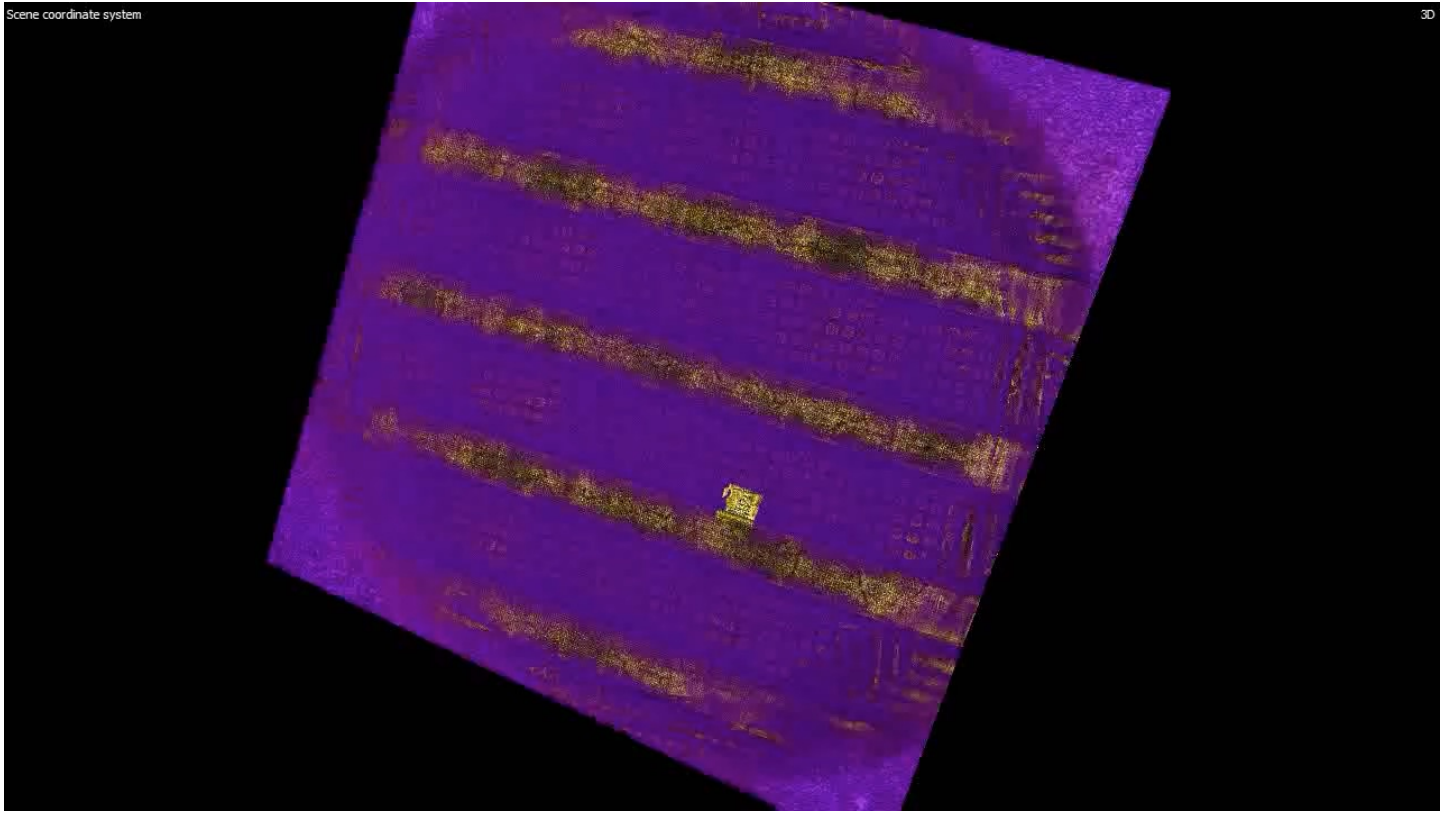
Brookhaven National Laboratory

SLAC NATIONAL ACCELERATOR LABORATORY

BERKELEY LAB

Argonne NATIONAL LABORATORY

Nondestructive Reconstruction of an Integrated Circuit
Weeks of Data Collection & Weeks to Months of Reconstruction Time
Need: Real-Time Analysis and Acquisition



J Klug, Y Jiang, et al (APS), J Hunter, N Weisse-Bernstein, et al (LANL), M Holler, M Guizar Sicairos, et al (PSI).

Achieving the BES Light Source Mission

Advanced computing is crucial to address drastic and rapid increases in data from the facilities

New and more complex experiments

- Multi-modal experiments that combine data from multiple samples, techniques, and facilities
- *In situ* and *in operando* experiments require real-time feedback and autonomous control

Increased source brightness (orders-of-magnitude brighter)

- Due to facility upgrades and accelerator improvements: LCLS-II 1 MHz, APS-U, LCLS-II HE, ALS-U, NSLS-II

Detector advances (orders-of-magnitude faster)

- Increased dynamic range
- Faster readout rates
- Larger pixel arrays

Analyze and reconstruct massive multi-modal data volumes

Identify and classify features and patterns

Merge simulation and experiment data to drive experiments and new results

Execute experiments dynamically using real-time reduction and AI/ML

Scale of the Problem

Over the next decade, the 5 Light Sources are projected to generate ~ 1 exabyte of data/year and will require 10s of petaflop/s to an exaflop/s of peak computing power

1 exabyte/year = 1.5 million movies every day

- We don't need to just watch these movies, we need to look at every frame of every movie, analyze it in near real-time, and make decisions about what to do next
- Complicating this is the fact that there are hundreds of "types" of movies (experiments) that each require their own solutions

1 exaflop/s = 500,000 servers

- This will require up to 1 exaflop/s of peak compute power, fast networks (multiple Tbps), storage, and a robust software infrastructure to support near real-time analysis

New Opportunities in Science

Accelerate science by leveraging computing opportunities over the next decade

Coupling the BES Light Sources to advanced computing capabilities will enable new opportunities over the next decade:

- Leverage the **exabyte (EB) of data** per year across the Light Sources to **increase the scientific knowledge base**
- Utilize **tens of petaflop/s to 1 exaflop/s of on-demand peak computing** to **reduce time to science**
- **Unlock exceptionally challenging experiments** by utilizing billions of core hours per year
- Fully leverage **AI/ML and digital twin** capabilities to **extract information** from complex data, **steer experiments**, **design experiments and facilities**, and use shared data for **ML-driven discovery**

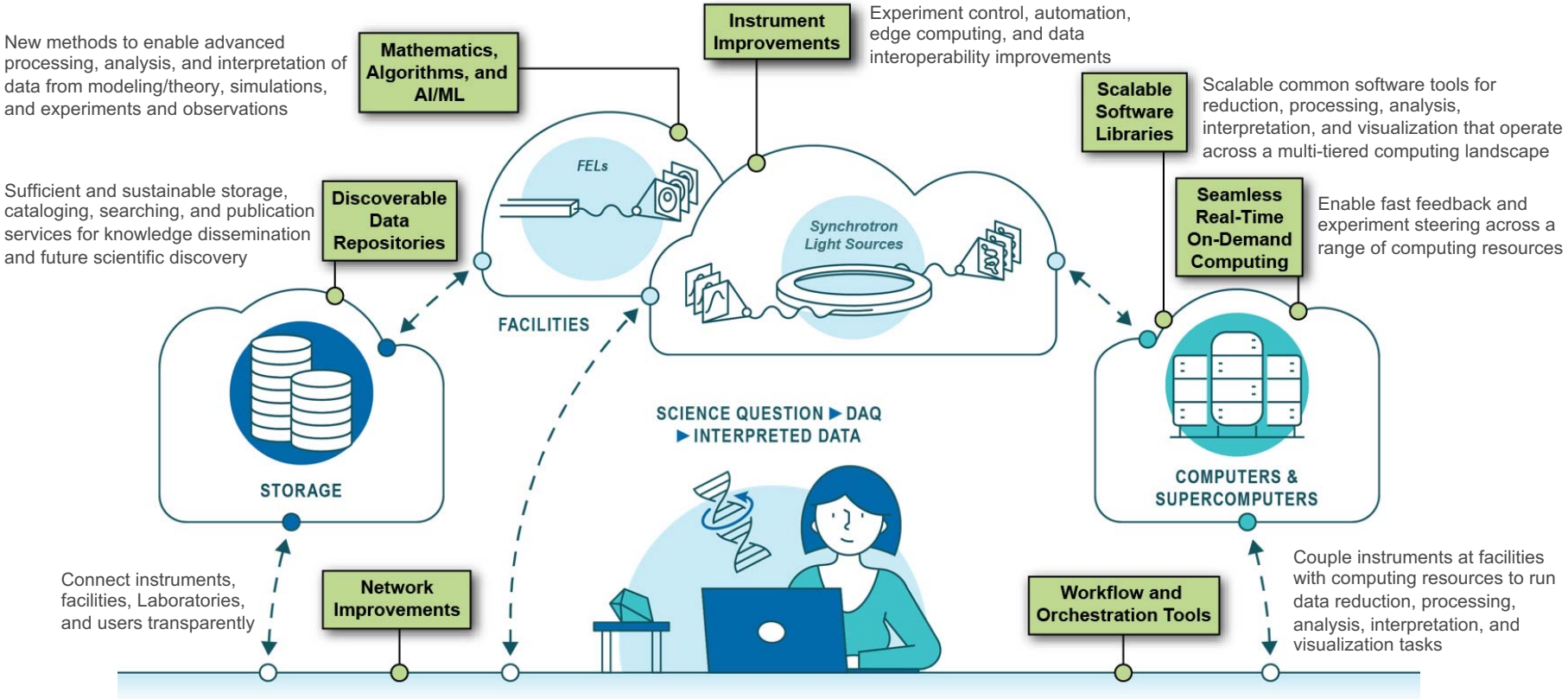
Unified solutions across the facilities, leveraging efficiencies of scale, can provide facility users with the ability to easily and transparently manipulate their data, and will better enable new scientific opportunities

Additional Changes

Advanced computational infrastructure will play a key role in addressing these needs

- Our **user community is diverse** coming from a wide variety of backgrounds and domains
- The **digital divide** amongst the scientific user community is rapidly increasing
- Wide-spread increased interest in **data interoperability** and in **FAIR data** and **open data**
- **Remote access** to experiments and facilities has been shown, particularly during the pandemic, to aid the productivity of facility users; potential to open more avenues to facility utilization by underserved and underrepresented universities and institutions

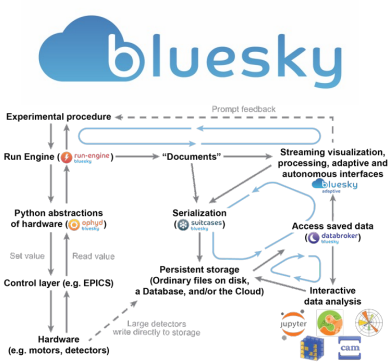
Key Areas That Must be Built to Meet Current and Upcoming Needs



Recent Efforts that Help Point the Path Forward

bluesky from NSLS-II

- New software and framework for experiment control at the Light Sources
- Increasing adoption across the Light Source facilities
- Creates a common path for collecting metadata and creating data interoperability

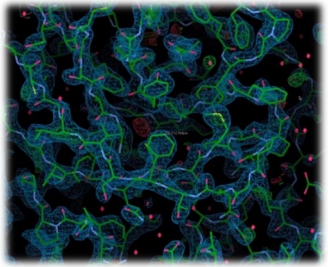


BES Data Solution Task Force Pilot Project

- Develop synergistic and sustainable approach to software solutions across facilities
- Leverage tools and expertise from all BES Light Sources by integrating complementary components

NERSC-LCLS LLANA Data Analytics Project

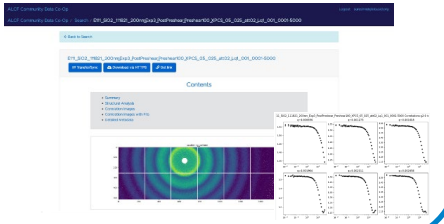
- Implemented an automated SFX pipeline at NERSC
- Demonstrated LCLS/NERSC workflow effort during COVID-19 LCLS experiments



SARS-Cov-2 main protease structures at physiological temperature (image credit: Hasan DeMirci)

ALCF Polaris for On-Demand Workflows

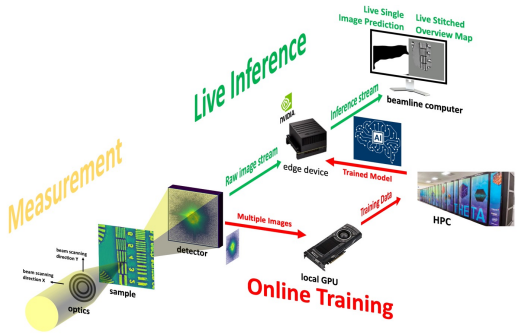
- Utilizing ALCF Polaris system for continuous on-demand data processing of APS and NSLS-II data
- Globus tools provide workflow and web portal services



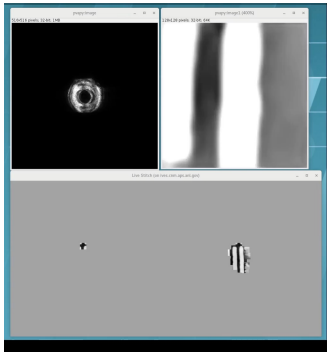
Recent Efforts that Help Point the Path Forward

AI-Enabled Real-Time Ptychography Reconstructions

Live detector images



Live inference

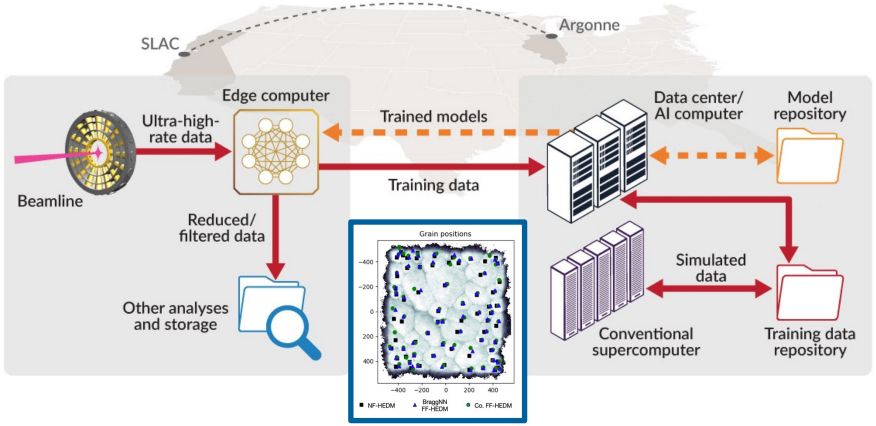


Cumulative image

Computationally intensive phase retrieval methods are replaced with neural network models (PtychoNN) that learn to invert raw coherent imaging data to sample amplitude and phase in a single shot.

Babu, A. V., Zhou, T., Kandel, S., Bicer, T., Liu, Z., Judge, W., Ching, D., Jiang, Y., Veseli, S., Henke, S., Chard, R., Yao, Y., Sirazitdinova, E., Gupta, G., Holt, M. V., Miceli, A., Cherukara, M. J., "Deep learning at the edge enables real-time, streaming ptychography," Paper Forthcoming.

AI-Enabled Bragg Diffraction Data Processing



Workflow depicting data movement, ML training, and inference at the edge between SLAC and Argonne. The pseudo-Voigt profiling approach is replaced with a neural network model (BragNN).

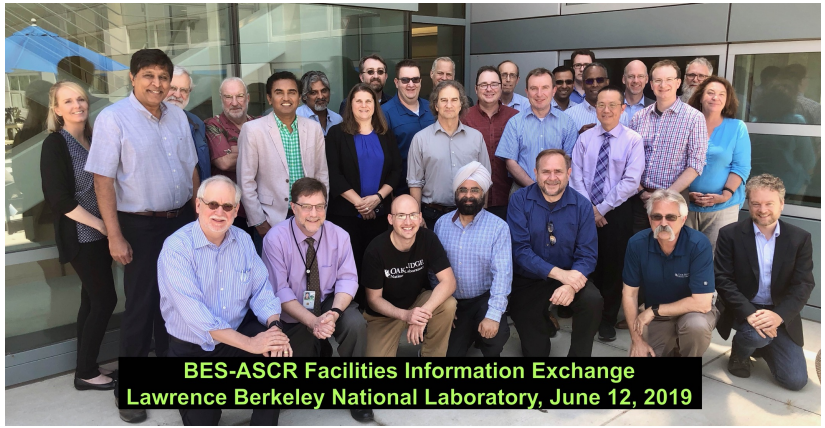
Z. Liu, Ali, A., Kenesei, P., Miceli, A., Sharma, H., Schwarz, N., Trujillo, D., Yoo, H., Coffee, R., Layad, N., Thayer, J., Herbst, R., Yoon, C., Foster, I., *Bridging Data Center AI Systems with Edge Computing for Actionable Information Retrieval*, to appear in the Proceedings of the 3rd Annual Workshop on Extreme-Scale Experiment-in-the-Loop Computing (XLOOP 2021) held in conjunction with SC'21, November 19, 2021.

Funding provided by DOE-SC/BES award: Actionable Information from Sensor to Data Center.

Collaborative Efforts Among Light Source and Computing & Networking Facilities

Light Source and computing & networking facilities have been discussing a strategic path

- Established a working group with membership from the BES Light Sources and the ASCR supercomputing and networking facilities, observers from the US neutron sources and Nano Science Research Centers (NSRCs), and CAMERA
- Report outlining needs and a balance of responsibilities between BES and ASCR facilities and researchers, and white papers describing initial joint ASCR and BES projects and efforts
- Enthusiastic participation in the DOE/SC Integrated Research Infrastructure Architecture Blueprint Activity from BES-funded researchers and facilities



Argonne NATIONAL LABORATORY | Leadership Computing Facility

NERSC National Energy Research Scientific Computing Center

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

ESnet ENERGY SCIENCES NETWORK

CAMERA Applied Math

Summary

Exciting opportunities to accelerate science with advanced computational infrastructure!

- The demand for computing and data resources is increasing by multiple orders-of-magnitude requiring a transformation in the way facilities handle data
- Data and computational systems will now be a key limiting factor for science
- Accelerate science by leveraging computing opportunities
- Potential to open more avenues to facility utilization by underserved and underrepresented universities and institutions
- Identified key areas that must be built to meet upcoming needs
- Recent collaborative efforts help point the path forward
- The full potential of the BES facility mission will be realized by coupling the intrinsic capabilities of the facilities with advanced computational infrastructure

Thank you for your time!