Self-driving Experimentation at Accelerator-based Light Sources

Tekin Bicer, Justin M. Wozniak, Zhengchun Liu and Rajkumar Kettimuthu

Argonne National Laboratory

Overview Technological advancements in accelerator-based light sources will enable experimentation with extremely high spatial and temporal resolutions, complicating management and analysis of measurement data. Even the current data acquisition rates can exceed tens of Gbps rates at synchrotron light sources, resulting in petabyte scale data for long running experiments[1]. Planned upgrades such as the Argonne Advanced Photon Source Upgrade (APS-U), scheduled for June 2022, will provide two to three orders of magnitude higher brightness that will produce overwhelming quantities of experiment data for scientific processing[2] bringing the aforementioned type of experiments to exabyte scale. The next upgrade, which will happen in **20-25 year time period** (if it follows the same timeline as the current upgrade) will be diffraction limited and may provide another order of magnitude or so increase in brightness and data.

High-performance computing (HPC) systems at the exascale and beyond are expected to handle these experimental dataset and their analysis workloads, however today's software systems and user work patterns will not scale to meet the data transfer and computational demands of these. Consider an experiment where a scientist wants to capture a phenomena with microsecond temporal resolution at Tbps data generation rate. Storing experimental data and processing it using local beamline resources will be infeasible. Using remote exascale systems, however, might accommodate computational demands, however utilizing these resources can introduce latency and result in missing the phenomena.

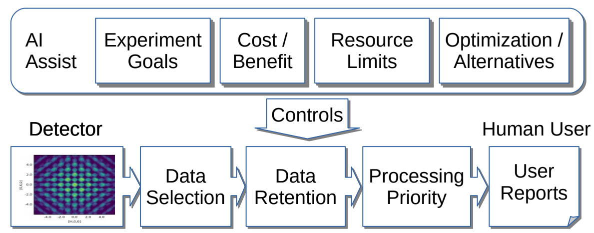

Figure 1: Synopsis of AI-assisted decisions during data acquisition analogous to those faced in vehicle control.

With this in mind, we will discuss the following decision points based on: (i) limitations on the cost and performance of **data movement**, while considering workflows' latency constraints over wide area networks; (ii) limitations in the **response time** capability of the compute resources, especially while utilizing remote resources, with respect to sample changes and resulting analysis; and (iii) the need for assistance when determining the **value** of raw and derived data when not all data can be stored or even propagated for full processing. These cannot be actuated by simple triggers; they involve making intelligent, latency-sensitive decisions, as in the vehicle control case.

We consider possible data acquisition and management challenges in the future, and discuss **self-driving experimentation** with AI-guided data collection and reduction methods. As shown in the figure, AI assistance will be pervasive in the acquisition pipeline, as the data selection, retention, and processing behavior of the pipeline will have to be autonomically controlled in the presence of extremely high data rates and emergent decision points. Multiple high-level objectives will thus have to be presented to the AI assistant, including some depiction of experiment goals, cost/benefit values for data and analysis artifacts, limitations on overall computing and data acquisition resources, and the opportunities for alternative selection and optimization. Although not limited to these, we will initially expand our discussion on the following topics:

Selective data acquisition will consider the region of interest to the experiment: AI/ML can detect the relevant regions on the sample and collect data only on from these parts. Challenges will include finding relevant parts to collect in scientific cases with variable goals that are not easy to specify in advance.

AI assistance in the data reduction pipeline will enable intelligent decision making about data reduction and discard during collection and processing. Both lossy and lossless compression will be available alternatives for both dense and sparse measurement data, with application-specific valuation on data, making for a range of complex decisions regarding what data must be passed on for further processing and/or ultimately stored and reported to users. Challenges include quantifying/predicting the cost of re-running the experiment if critical data is discarded.

Dynamically changing features will be monitored by AI/ML, such as chemical reactions, moving features, and biological sensitivity. The observed rate of change in evolving phenomena will trigger adjustments in data acquisition such that collected data is sufficient to satisfy scientific goal requirements (e.g., meeting a reconstruction quality threshold under constraints). Challenges include possibly extreme latency requirements for fast-moving experiments and integrating simulation and active learning into these decisions.

[1] Rapid Analysis of Various Emerging Nanoelectronics: https://www.iarpa.gov/index.php/research-programs/raven

[2] APS-U: https://www.aps.anl.gov/APS-Upgrade