FY 2024 LDRD Concept Paper

Program: DRD

Initiative: Streaming Grand Challenge

Proposal Title: **Streaming Readout Real-Time Development and Testing Platform**

Principal Investigator, Division: David Lawrence , CST

Co-Investigator, Division: Markus Diefenthaler, ENP

Contributors, Division:

Advisor(s), Division: Marco Battaglieri

Summary

- **Rationale:**

In the last five years, Jefferson Lab has made tremendous progress in the rapid data processing and analysis through the integration of streaming readout, AI/ML, and heterogeneous computing. These efforts address the intense data challenges at upcoming NP experiments, such as SoLID or ePIC at EIC, but will also advance the workflows at running experiments.

However, there is currently a gap in the software that will be used to connect the continuous readout of a streaming readout (SRO) data acquisition system to the software being used for the reconstruction and analysis at a compute or data centers such as the foreseen High Performance Data Facility (HPDF). Major pieces of a system capable of performing near real-time DAQ, calibration, and reconstruction exist, but we lack a complete system where all of these pieces can be integrated, and tested at the scale of current and future experiments. Here, systems refers not only to frameworks implementing streaming data models but also to optimized hardware architecture for SRO. Which algorithms to run at each stage of the seamless data processing from the SRO to the analysis and which hardware to use for each of the stages, strongly depends on data rate and complexity of each experiment and needs to be decided for each experiment individually.

This project will fill this gap by developing a design and validation platform for computing models using streaming readout, AI/ML, and heterogeneous computing. The design and validation platform will connect algorithms for the seamless data processing from the SRO to the analysis and run them in various configurations on compute and data centers. This will allow testing and validating streaming compute models for NP experiments before implementing them in full detail. Processing streaming data is a “hot topic” in industry, computer science research, as well as current and future trends in nuclear and high energy physics. As with all “hot topics” the right balance must be found between innovation and implementation. Our LDRD project will guide finding the right balance.

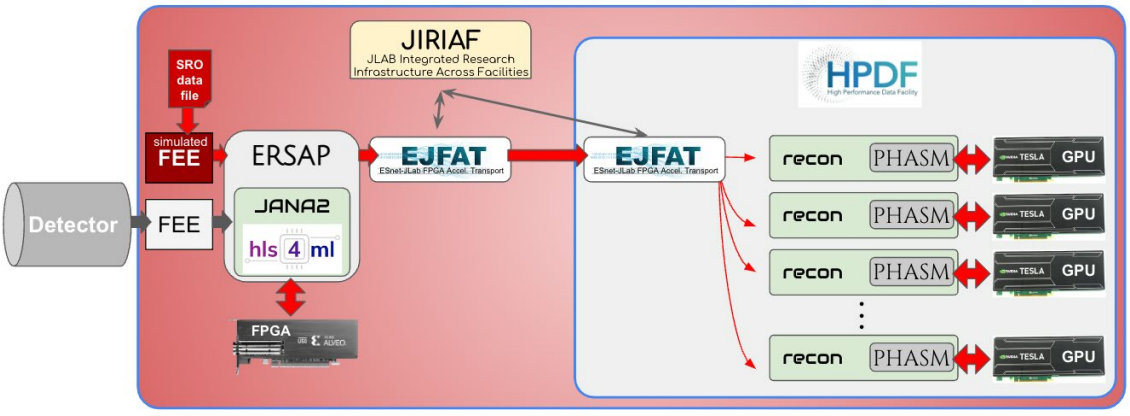


Diagram illustrating an example SRO system where data is streamed from the detector to

numerous nodes at an HTC facility such as HPDF. While the individual pieces exist, a platform that ties them all together coherently does not. Red items and arrows indicate pieces that will be developed as part of the SRO design and validation platform.

- **Description:**
  We will develop a full-scale testing platform for SRO systems utilizing existing tools to access different components of a complete streaming system downstream of the Front End Electronics (FEE). At the end of the project, we will have a platform that implements a fully functioning SRO reconstruction system that takes data from different sources (FEE in beam-on real time tests, simulations or archival beam data) to produce multiple streams at scale, mimicking a large-scale detector. The platform will utilize components from the following projects: ERSAP, JANA, CLARA, PHASM, EJFAT/IRIAD, JIRIAF. It will also include both GPU and FPGA components at various stream stages. The framework will be tested on pseudo- or real data collected from existing experiments and test facilities running at the lab (CLAS12, INDRA, …) or their GEANT-generated proxies. Data will be streamed out on the network from the experimental halls as well as from existing SRO test facilities in the INDRA lab to the Computer Center and processed in real-time by the SciComp farm in CEBAF Center. The full off-line data reconstruction pipeline will be used. When complete, this will provide a full-scale design and validation platform for SRO experiments with near real-time data processing.
- **Mission Relevance:**
  A design and validation platform like this is needed to fully realize SRO systems for experiments including CLAS12 and future experiments like SoLID, and ePIC. SRO systems will include data reduction, consolidation, and filtering algorithms close to the detectors and calibration and reconstruction algorithms further downstream. Final design decisions for such experiments will include which algorithms will be run on which streams, where consolidation occurs, and which type of hardware is best suited for each algorithm.

  JLab is the appropriate place to develop such a platform since copious data are available from current running experiments and several software and hardware tools developed at JLab will be connected by this effort. Moreover, considering the current physics program and the number of scheduled experiments, JLab represents the ideal test bed for the final system in real conditions.

- **Approach:**
  We will utilize existing facilities (e.g INDRA-lab, experimental Halls, …) as well as existing data from previous SRO beam tests and actual experiments. A postdoc will work with local domain experts on the relevant software packages to connect all of the pieces so they may be tested at scale or as close to scale as possible that the existing equipment allows.

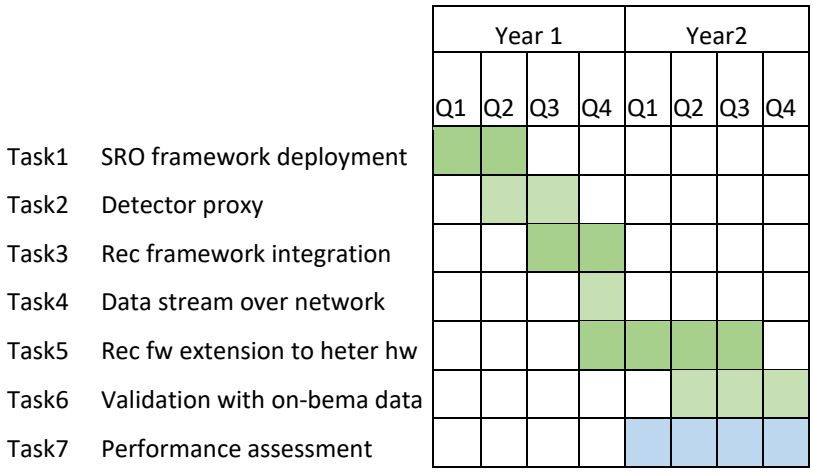
Research Plan

- **Goal:**
  Fully developed software platform that is capable of processing SRO data from CLAS12 detector in real-time or near real-time.

  - Tools for fully simulating a real-time SRO data processing network from Front End Electronics to large compute resource.

- **Objectives:**
  - Deployment of a distributed (quasi) real-time SRO data processing model includes data calibration and full traditional off-line reconstruction.
  - Framework optimization using GEANT-generated and archived beam-on data
  - Optimized framework validation with beam-on tests
  - Assessment of needed network and computing resources
  - Assessment of the performance for different hardware platforms
  - Identify potential issues relevant to a future HPDF in receiving and processing SRO data

- **Tasks:**
  - **Task1: identify or capture SRO formatted data from CLAS12 and INDRA test systems with data tag/filtering capability (output data ready for further offline processing)**
  - **Task2: deployment of a detector proxy that uses archived on-beam and GEANT-generated data as SRO data source**
  - **Task3: integration of the off-line data analysis framework into SRO DAQ framework**
  - **Task4: integration of network tools to stream data from detector (counting house) to remote data center (JLab CC CPU farm)**
  - **Task5: extend the SRO distributed data acquisition and reconstruction framework to heterogenous hardware (including FPGA and GPUs)**
  - **Task6: validation of the optimized pipeline with data streamed in on-beam tests**
  - **Task7: Performance assessment in terms of achieved results in data calibration and reconstruction and quantification of needed computing resources for scaling to future detectors/systems**



- **Measure of Success:**
  - CLAS12 sub-detectors data will be streamed from counting house to Computer Center
  - GPUs and FPGAs will both be demonstrated to process stream data within the analysis framework in conjunction with algorithms run on the CPU
  - AI-supported (quasi) real-time data reconstruction algorithms will be tested on heterogenous hardware
  - Calibrations, based on high level physics quantities (such as pi0 mass, interaction vertex location, straight track finding, …) will be performed in (quasi-) real-time;
  - Bandwidth will be shown to meet or exceed expected bandwidth from a CLAS12 SRO experiment
  - Bottlenecks or limitations for processing SRO data with rates similar to ones expected in operations at high luminosity (CLAS-12 HI-LUMI, SoLID, ...) will be identified.
- **Future Funding:**
  The project is not anticipated to require an additional funding stream. The resulting product will be absorbed into the collection of SRO tools that include CODA, ERSAP, CLARA, JANA, etc.

Proposed Fiscal Year and Total Budgets:
Only staffing and some travel are needed for this. Existing hardware and network facilities at JLab will be used for the project. The following is used for estimating the budget (includes fringe).

Staffing:

PI (SCS): 10%

SRO DAQ programmer (SSIII): 25%

GPU Programmer/Specialist (CSII): 25%

Postdoc: 100%

Travel: 1 domestic + 1 international/year

Resource Needed	FY24 Projected Budget ($K)	FY 25 Projected Budget ($K)
Staff Effort	$326K	$342K
Material, equipment, etc	$10K	$11K
Total Budget Request	$336K	$353K

References:

Heyes, G. (2019, February 12). Streaming Grand Challenge [Slides]. Retrieved from **https://indico.jlab.org/event/307/contributions/4681/attachments/3852/4660/20190212_Grand_Challenge_Overview.pdf**

Ameli, F., Battaglieri, M., Berdnikov, V.V. *et al.* Streaming readout for next generation electron scattering experiments. *Eur. Phys. J. Plus* **137**, 958 (2022). https://doi.org/10.1140/epjp/s13360-022-03146-z

Diefenthaler, M., Farhat, A., Verbytskyi, A. *et al.* Deeply learning deep inelastic scattering kinematics. *Eur. Phys. J. C* **82**, 1064 (2022). https://doi.org/10.1140/epjc/s10052-022-10964-z

J.C. Bernauer, C.T. Dean, C. Fanelli *et al.* Scientific computing plan for the ECCE detector at the Electron Ion Collider NIMA 1047, 167859 (2023). https://www.sciencedirect.com/science/article/pii/S0168900222011512?via%3Dihub