**INDRA-ASTRA**: Evaluation & Development of Algorithms & Techniques for Streaming Detector Readout

**Hindu mythology**

**INDRA** Deity of lightning, thunder, rains and river flows

**INDRA-ASTRA** Indra’s weapon

**Jefferson Lab**

**INDRA** Facility for Innovations in Nuclear Data Readout and Analysis

**INDRA-ASTRA** LDRD on streaming readout

J. Bernauer (SBU), D. Blyth*(ANL), C. Cuevas*, M. Diefenthaler, A. Farhat**(ODU), W. Gu*, G. Heyes*, E. Pooser (JLAB → GTRI), B. Raydo*, Dmitry Romanov**

* Only FY19, ** Only FY20
Towards the next-generation research model in Nuclear Physics

Science & Industry remarkable advances in electronics, computing, and software over last decade

Evolve & develop Nuclear Physics research model based on these advances

Role of computing Data processing from DAQ to analysis largely shaped by kinds of computing that has been available
Example Trigger-based readout systems

Advances in electronics, computing, and software Unique opportunity to think about new possibilities and paradigms
Example Streaming readout systems
**History** Enabling the Next Generation Research Model for Nuclear Physics (Chris Cuevas, Graham Heyes, Markus Diefenthaler, Rik Yoshida)

- Why think about this?
  - Nuclear Femtography is a new science requiring an understanding of multi-dimensional correlations. The current NP research model is arguably not-adequate.
  - NP research model has not changed for over 30 years. There have been remarkable advances in computing, analytics and microelectronics capabilities. We want to see if these advances can fundamentally improve the research model.

- How to think about this.
  - Rethink the way experiment is compared to theory. Do computing advances enable a different way to do this? (Event Level Analysis, ELA)
  - Rethink the way experimental data is handled. Are there ways to speed up the analysis, in the context of ELA, of the data given modern compute and data handling capabilities. (Additive Event Model, AEM)
  - Rethink the way we read out the data from the Front-End and how they are assembled into events, in view of ELA and AEM. (Streaming Readout)

- Each item ELA, AEM, and Streaming Readout has its own merits and implementing them on their own has immediate advantage.

- However, most efficient ELA implies AEM, and ultimate efficiency of the Research Cycle can be achieved when Streaming Readout feeds AEM → ELA.

- ELA + AEM + Streaming Readout (EASR?) could be a model for accelerating discovery in NP.
**Integration of DAQ, analysis and theory to optimize physics reach**

• seamless data processing from DAQ to analysis using streaming readout

**Definition of streaming readout**

• data is read out in continuous parallel streams that are encoded with information about when and where the data was taken.

**Advantages of streaming readout**

• opportunity to streamline workflows
• take advantage of other emerging technologies, e.g. AI / ML

**Integration of DAQ, analysis and theory to optimize physics reach**

• opportunity for near real-time analysis using AI / ML (alignment, calibration, reconstruction)
• opportunity to accelerate science (significantly faster access to physics results)
Seamless integration of DAQ and analysis using AI/ML

**LDRD goal**

Prototype components of streaming readout at NP experiments
→ integrated start to end system from detector read out through analysis
→ comprehensive view: no problems pushed into the interfaces

Prototype near real-time analysis of NP data
→ inform design of new NP experiments

- fADCs
- GEM detector
- Custom MC Simulated ADC and TDC output
- Near real-time processor of streamed data in JANA2
- Analysis data
- Near real-time, interactive analysis in JupyterLab
- ZeroMQ messages via ethernet
Streaming readout tests

AIEC and INDRA-ASTRA projects discussion, June 9, 2021.

Developed streaming readout simulations

Demonstrated how to integrate any MCEG into streaming readout

Streaming readout of fADC250

TDIS Streaming Readout Prototype

Near real-time processor of streamed data in JANA2

Near real-time, interactive analysis in JupyterLab
Developed messaging library

- receiving data from streaming data source
- sending it efficiently over the network (TCP)
- re-broadcasting the data using the ZeroMQ messaging protocol
- subscribing to a stream of data tested at rates up to ~50 Gbit/s

JANA2 for parallel processing of streamed data (with LD2014)

INDRA-ASTRA file server

- 1.5 TB of memory (Intel® Optane™)
- 1.0 TB of solid state storage
Streaming readout analysis

Near real-time processor of streamed data in JANA2

ZeroMQ messages via ethernet

Streaming plugins (data, MC)
- decoding of streamed data
- visualization of streamed data
- automated data-quality monitoring
- online calibrations
- fully extensible in JupyterLab

AIEC and INDRA-ASTRA projects discussion, June 9, 2021.
Automated data-quality monitoring and calibrations

Automatically identify changes in the underlying probability distribution

Re-calibrate in case of changes

Full re-calibration

Adapted and extended ADaptive WINdowing technique (ADWIN)

- detecting distribution changes, concept drift, or anomalies in data streams
- detection with established guarantees on the rates of false positives and false negatives
- keeps a sliding window $W$ with the most recently read data
- whenever two sufficiently large sub-windows of $W$ have sufficiently different means, then it is likely the corresponding expected values are different, and the older portion of the window is dropped
- prediction intervals without any prior assumption on the underlying distribution of data samples
Address Challenges of Autonomous Control and Experimentation

**Follow-up Project**

Develop a prototype for a fully automated, responsive detector system as a first step towards a fully automated, self-conscious experiment

R&D fully integrated with:
- streaming readout efforts at Jefferson Lab
- AI/ML initiatives at Jefferson Lab and affiliated universities

**Team**

**Jefferson Lab**
- ENP M. Diefenthaler, E. Jastrzembski, H. Szumila-Vance, and CLAS12 involvement
- CST D. Lawrence, V. Gyurjyan, and we for sure will work with N. Brei

**Catholic University of America**
- Physics T. Horn (interested)

**Old Dominion University**
- Applied Numerical Mathematics R. Fang, A. Farhat, Y. Xu
<table>
<thead>
<tr>
<th>Detector Readout</th>
<th>ML</th>
<th>Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup ERSAP prototype</td>
<td>Compare ADWIN2 and MULTISCALE algorithms in their capability • for identifying gradual and sudden changes in streaming data and • for distinguishing these changes from noise • for online capability</td>
<td>Develop a service in ERSAP to report changes and noise levels to the streaming readout system</td>
</tr>
<tr>
<td><strong>GEM detector</strong> Monitor pedestals in real time</td>
<td><strong>GEM detector</strong> Monitor pedestal in real time</td>
<td><strong>GEM detector</strong> Automatically correct for pedestals in real time</td>
</tr>
<tr>
<td><strong>GEM detector</strong> Study cosmics and establish baseline efficiency and tracking using current, non-ML approaches</td>
<td><strong>GEM detector</strong> Identify cosmic over background • learn what is a good efficiency for detector plane</td>
<td><strong>GEM detector</strong> Monitor efficiency in real time (e.g., some region not seeing hits) and automatically report on changes</td>
</tr>
<tr>
<td><strong>Calorimeter</strong> pending on CLAS12</td>
<td><strong>Calorimeter</strong> pending on CLAS12</td>
<td><strong>Calorimeter</strong> pending on CLAS12</td>
</tr>
<tr>
<td>Prepare tests in Halls A/C and B</td>
<td></td>
<td>Identify limitations of automated system, e.g. by rapid changes of threshold or additional noise sources Test integrated system in parasitic beam test</td>
</tr>
</tbody>
</table>
**INDRA-ASTRA**: Evaluation & Development of Algorithms & Techniques for Streaming Detector Readout

Markus Diefenthaler
mdiefent@jlab.org

New possibilities and paradigms for JLab 12 GeV and EIC

- seamless data processing from DAQ to analysis using streaming readout
- opportunity for near real-time analysis (auto-alignment, auto-calibration, near real-time reconstruction)
- opportunity to accelerate science

**LDRD project: INDRA-ASTRA**

- prototyped components of streaming readout
- prototyped near real-time analysis
- **Collaboration with Experimental Physics Software and Computing Infrastructure (EPSCI)**
  - INDRA-ASTRA fulfills need for rapid prototyping tools for DAQ and detector R&D for Nuclear Physics and beyond
  - develop INDRA-ASTRA into plugin for near real-time monitoring and auto calibrations at JLab 12 GeV and EIC

“The purpose of computing is insight, not numbers.”
(Richard Hamming)