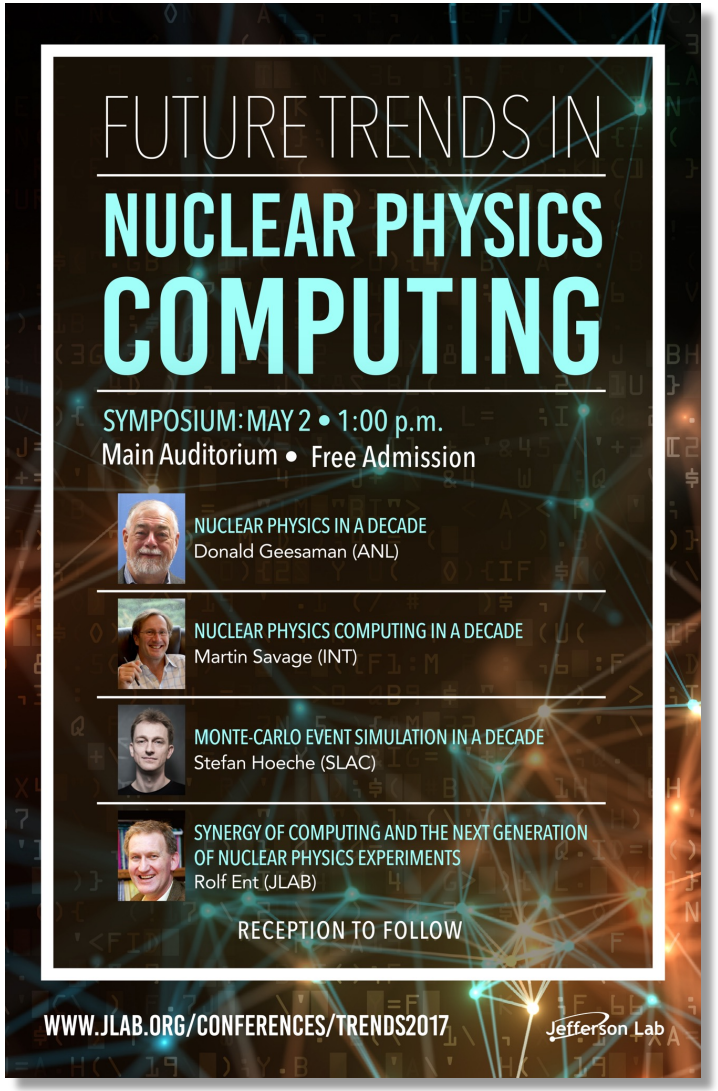
Streaming Workshop X –
Vision and Opportunities for Streaming Readout at EPIC

Markus Diefenthaler

EIC^2 EIC Center at Jefferson Lab

Jefferson Lab

Future Trends



Details on https://www.jlab.org/FTNPC

Donald Geesaman (ANL, former NSAC Chair) *"It will be **joint progress of theory and experiment** that moves us forward, not in one side alone."*

Obvious path

- Sharing data early with theory.

Martin Savage (INT) *"The next decade will be looked back upon as a **truly astonishing period in Nuclear Physics** and in our understanding of fundamental aspects of nature. This will be **made possible by advances in scientific computing** and in how the Nuclear Physics community organizes and collaborates, and how DOE and NSF supports this, to take full advantage of these advances."*

We can make a difference

- AI/ML for autonomous control and experimentation is a tremendous opportunity.
  - We need to bring AI/ML into EPIC (see the slides from the software & computing sessions).
- Streaming readout using AI/ML as a new paradigm for seamless data processing from DAQ to analysis.

Our Vision for Software & Computing at the EIC

Rapid turnaround of data for the physics analysis and to start the work on publications:

- **Problem** Data for physics analyses and the resulting publications available after O(1year) due to complexity of NP experiments (and their organization).
  - Alignment and calibration of detector as well as reconstruction and validation of events time-consuming.
- **Goal** Analysis-ready data from the DAQ system.
- **Solution** Compute-detector integration with AI at the DAQ and analysis level.

More details https://eic.github.io/activities/principles.html

Principle 2:
We will have an unprecedented compute-detector integration:

- We aim for autonomous alignment and calibration.
- We aim for a rapid, near-real-time turnaround of the raw data to online and offline productions.

Streaming Readout: Trigger-less data acquisition

Definition of Streaming Readout

- Data is digitized at a fixed rate with thresholds and zero suppression applied locally.
- Data is read out in continuous parallel streams that are encoded with information about when and where the data was taken.
- Event building, filtering, monitoring, and other processing is deferred until the data is at rest in tiered storage.

Advantages of Streaming Readout

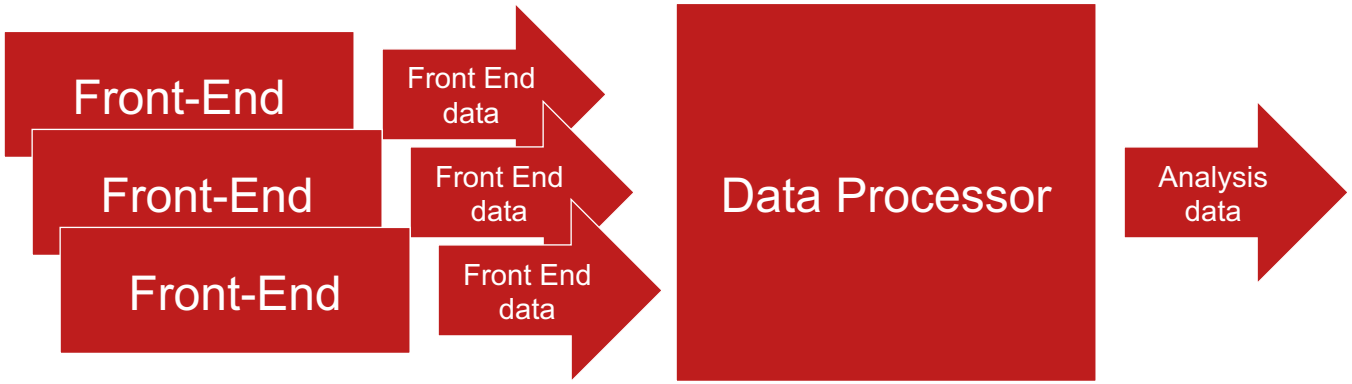
- Simplification of readout (no custom trigger hardware and firmware):
  - Trigger-less readout:
    - Beneficial for experiments that are limited by event-pileup or overlapping signals from different events as well as multi-purpose experiments where one would anyway loosen triggers.
  - Data flow is controlled at source (and at no stage by *back pressure).*
- Opportunity to streamline workflows.
- Take advantage of other emerging technologies.

Integration of DAQ, analysis and theory to optimize physics reach

Research model with seamless data processing from DAQ to data analysis

- Building the best detector that fully supports streaming readout and AI/ML:
  - FastML for alignment, calibration, and reconstruction in near real time.
    - Applications and Techniques for Fast Machine Learning in Science (*Front.Big Data* 5 (2022) 787421)
  - AI for intelligent decisions
- For rapid turnaround of data for the physics analysis and to start the work on publications.

Streaming Readout and (near) real-time processing



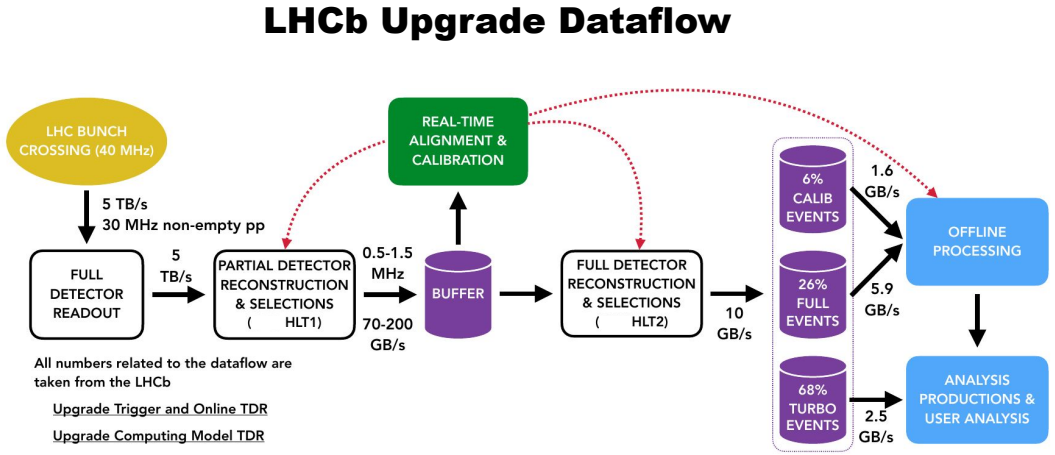
Data Processor

- Assembles data into physics events.
- Outputs data suitable for physics analyses and the resulting publications.

Features

- **FastML**
  - Autonomous alignment
  - Autonomous calibration
  - Reconstruction
  - Event filtering based on full event information
  - Autonomous anomaly detection
- **AI**
  - Responsive detectors
  - Conscious experiment

Allen Example from LHCb

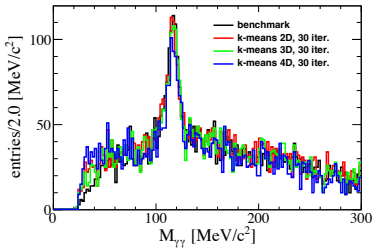


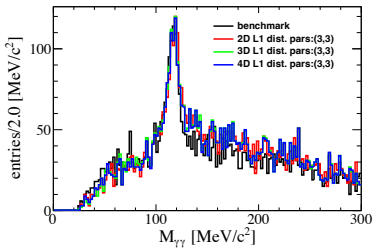
HLT1 challenge: reduce 5 TB/s to 70-200 GB/s in real-time with high physics efficiency

On-Beam Validation of Streaming Readout at Jefferson Lab

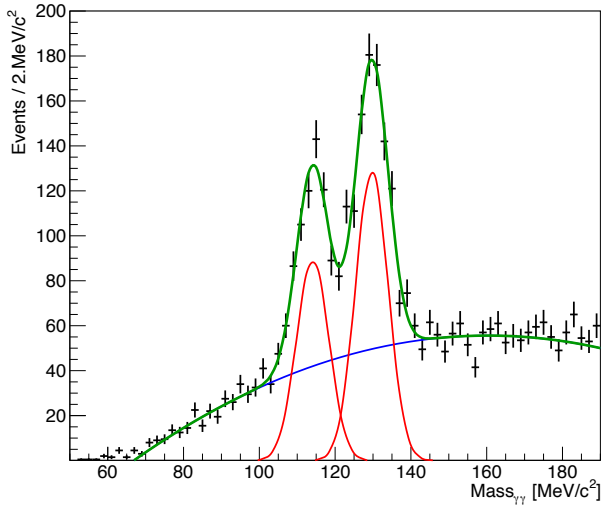
Tests included AI-supported real-time tagging and selection algorithms (*Eur.Phys.J.Plus* 137 (2022) 8, 958)

- Standard operation of **Hall-B CLAS12** with high-intensity electron-beam
- Streaming readout of forward tagger calorimeter and hodoscope
- Measurement of inclusive π^0 hadronproduction





- Prototype of EIC PbWO4 crystal EMCAL in **Hall-D Pair Spectrometer**
- Calorimeter energy resolution of SRQ compatible with triggered DAQ.



Streaming Readout Workshop Series

Organized by EIC Streaming Readout Consortium

2017	**SR I**	(online)		
2018	**SR II**	(MIT)	**SR III**	(CNU)
2019	**SR IV**	(Italy)	**SR V**	(BNL)
2020	**SR VI**	(JLab)	**SR VII**	(BNL)
2021	**SR XIII**	(MIT)	**SR IX**	(ORNL)
2022	**SR X**	(JLab)		

- Community efforts towards streaming readout at the EIC.
- Established streaming readout as default for the E(P)IC detector.
- Forum to discuss many advances in microelectronics and computing (AI/FastML, heterogeneous computing, storage, networking, etc.).

Streaming Readout X, co-organized with EPIC DAQ and electronics WG:

- Review the progress on streaming readout electronics, computing, and software.
- Discuss the future priorities for the EIC Streaming Readout Consortium.
- Mini town hall meeting on streaming readout technologies of the NP community.

Questions from Streaming Readout X

Ongoing work in the EPIC DAQ and Electronics WG

- Is there any need for R&D?
- Can we build a simple test setup? How will we scale it up? Can we use it for test beam?
- **We have to define the clock distribution. How will it be done?**
  - Timing system needs to allow for simultaneous test of the detector components.
- How is the stream aggregation done?
- How are we building events? Do we need to build events online?
- Generalize electrical - optical interface
- What protocols are used for the DAQ?
- Hardware and software (data handling, communication; calibrations, reconstruction, analysis)
- Interface of streaming readout and experimental control, including slow control, and also accelerator control.
  - We need data quality monitoring for each layer of the read out and data processing, including feedback for accelerator control.
- How will we handle firmware and software updates?
- How do we coordinate the purchase of front-end electronics?
- How do we coordinate the purchase of other components, e.g., GPUs?

Questions from Streaming Readout X

Common Discussion of EPIC Computing and Software and DAQ and Electronics WGs on

Seamless Data Processing from DAQ to Analysis Using Streaming Readout and AI/ML:

- **What are the boundaries between DAQ, online and offline data processing? Will this be fully integrated?**
- **What are the computing resources needed for the streaming readout?**
  - What are the available and affordable resources?
- **For each detector component**:
  - How will we handle calibrations?
  - Do we need a triggered system for calibrations?
  - What are the requirements for calibrations?
  - What would be the required turnaround time for calibrations?
- **For the integrated detector:**
  - How do we align the central and for-forward regions?
- **How can we manage background and noise reliably?**
- **What are the biases in the design and implementation of streaming readout and AI/ML and how to prevent them?**

For all these questions: We need **simulations of the entire data stream**. We are ready to implement digitization based on the streaming readout. Then we can start with algorithm development.

Summary

Markus Diefenthaler

mdiefent@jlab.org

We are working to **accelerate science**:

- **Goal** Analysis-ready data from the DAQ system.
- **Solution** Seamless data processing from DAQ to analysis using streaming readout and AI/ML in near real-time.
- We have the advances in scientific computing, we now need to organize and to collaborate to take full advantage of these advances.
- How will the **EPIC Computing and Software** and **DAQ and Electronics WGs work together**?

Many opportunities for **autonomous control and experimentation**.