INDRA-ASTRA Seamless data processing from DAQ to data analysis

Hindu mythology

INDRA Deity of lightning, thunder, rains and river flows

INDRA-ASTRA Indra’s weapon

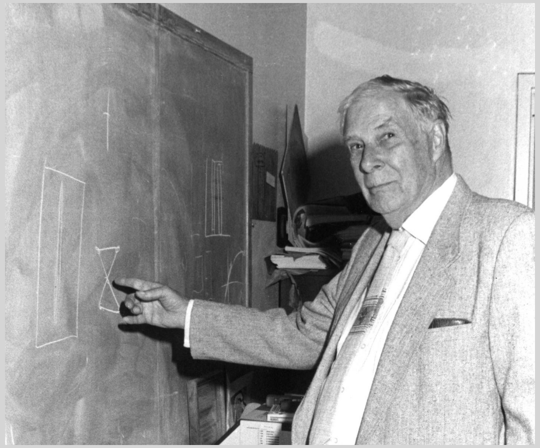
Nuclear Physics

INDRA Facility for **I**nnovations in **N**uclear **D**ata **R**eadout and **A**nalysis

INDRA-ASTRA LDRD on streaming readout

Markus Diefenthaler

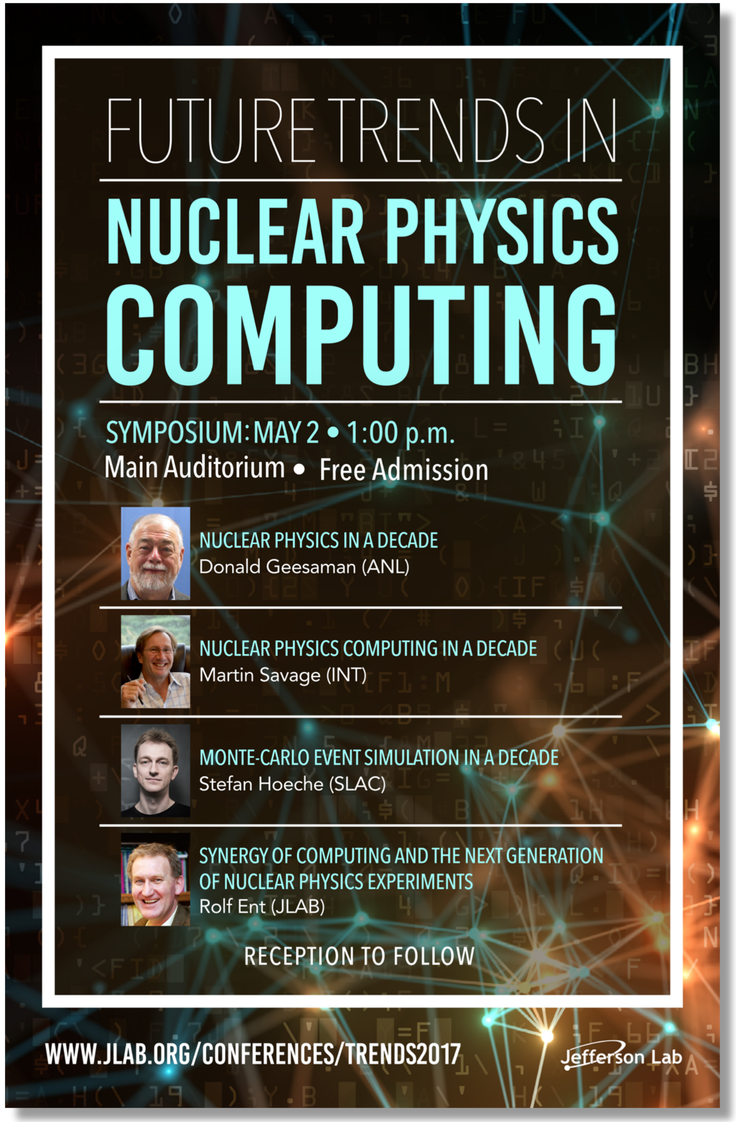
The role of Software & Computing



Richard Hamming (1962) *"The purpose of computing is **insight**, not numbers."*

Martin Savage (INT, 2017) *"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."*

Towards the next-generation Nuclear Physics research model



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

- All scientists of all levels, worldwide, should be enabled to actively participate in the NP data analysis.
- To achieve this goal, we must develop analysis toolkits using modern and advanced technologies while hiding that complexity.
- We must emphasize **data** as much as **analysis**. Experimental data must be open access, **readily accessible** and in a self-describing formats.

Compute-detector integration to deliver **analysis-ready data from the DAQ system**:

- responsive alignment and calibrations in *real time / online*
- *real-time / online* event reconstruction and filtering
- *real time / online* physics analysis

The role of Streaming Readout

Think out of the box

- The way analysis is done has been largely shaped by kinds of computing that has been available, e.g., **trigger systems**.
- This is an unique opportunity for Nuclear Physics to think about new possibilities and paradigms that can and should arise, e.g., **Streaming Readout**.

Challenges in data acquisition

- precision of the science depends on statistics which leads to:
  - development of detectors that can handle high rates
  - improvements in trigger electronics - faster so can trigger at high rates.
- beam time is expensive so data mining or taking generic datasets shared between experiments is becoming popular:
  - loosen triggers to store as much as possible
  - think about sharing datasets from the start
  - science is global – so distribute and share data handling
- some experiments are limited by event-pileup, overlapping signals from different events, hard to untangle in firmware.
  - leads to different readout schemes, often trigger-less

Benefits of Streaming Readout

Move complexity from hardware to software

- enhances flexibility
- reduces complexity
- relaxes hard time constraints
- allows more scientists to contribute (merge of DAQ, online and maybe also offline groups)
- allows cost-effective use of resources
- easier integration of detectors with large data rate (e.g., EIC Vertex detector with 240 GB/s)
- allows to select events based on the information of all detectors

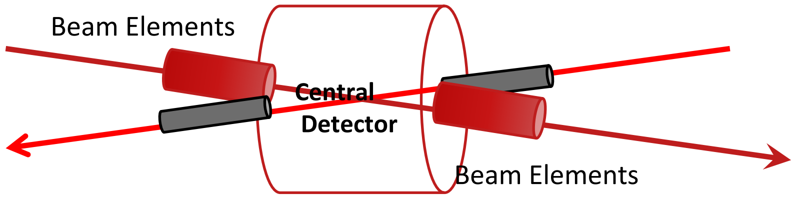
Possible challenges

- real-time integration with accelerator
- how to commission / understand accelerator, detector, and analysis software at the same time

Machine-Detector Interface

Integrated interaction region and detector design to optimize physics reach

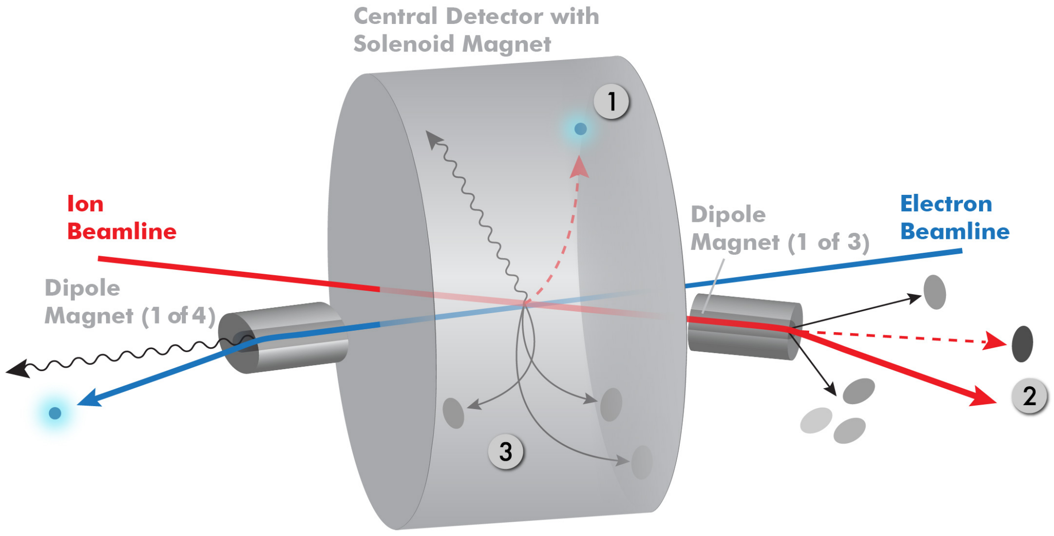
The aim is to get **~100% acceptance** for all final state particles, and measure them with good resolution.



Experimental challenges:

- beam elements limit forward acceptance
- central Solenoid not effective for forward

Possible to get ~100% acceptance for the whole event.



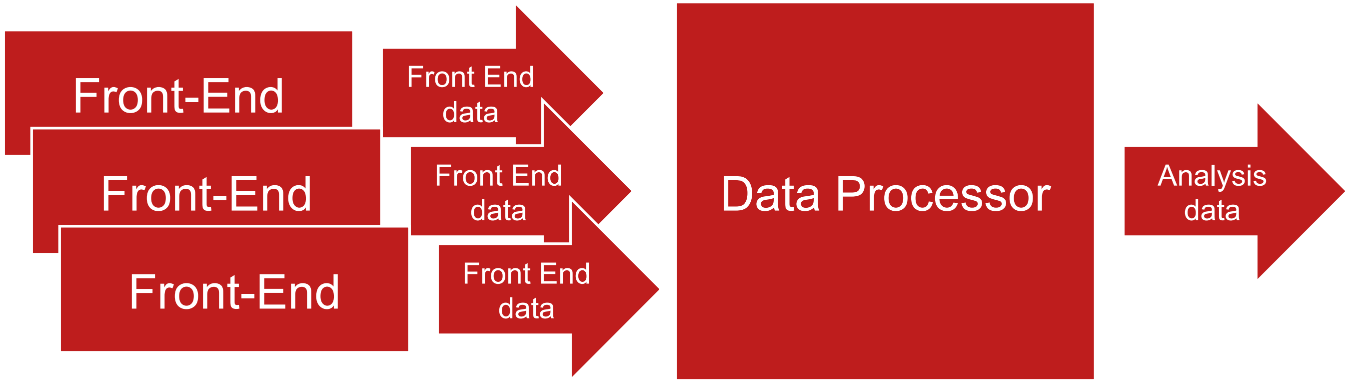
Beyond Machine-Detector Interface

Integration of DAQ, analysis and theory to optimize physics reach

Integration of DAQ, analysis and theory

- research model with seamless data processing from DAQ to data analysis:
  - not about building the best detector
  - but the best detector that fully supports streaming readout, fast alignment and calibration, and reconstruction algorithms for near real-time analysis

Streaming Readout and Real-Time Processing



Data Processor (Software & Computing)

- assembles the data into events
- outputs data suitable for final analysis (**Analysis data**)

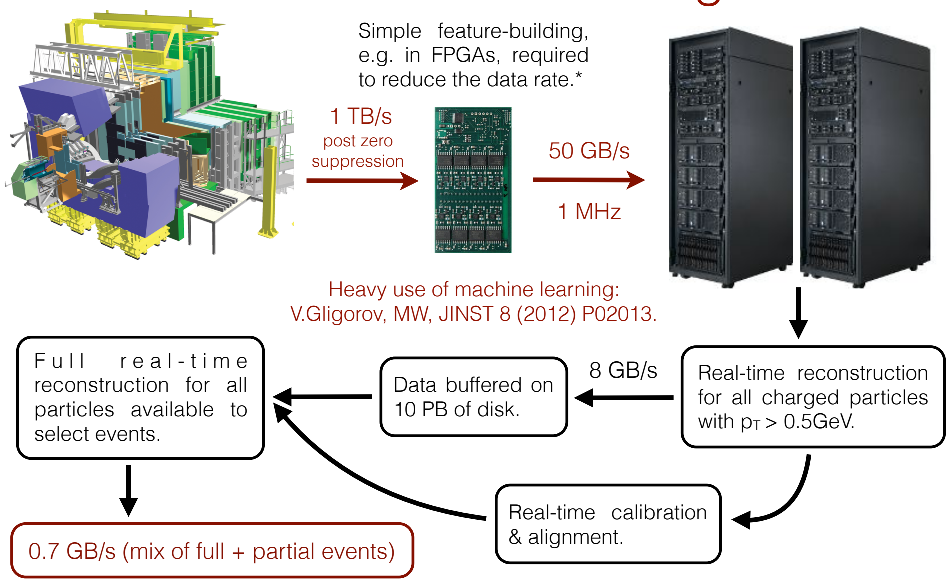
Features

- ideal for AI
- automated anomaly detection
- automated alignment and calibration
- (near) real-time event reconstruction
- event selection and/or labeling into analysis streams
- (near) real-time physics analysis
- responsive detectors (conscious experiment)

LHCb Example

JINST 8 (2013) P04022

Real-Time Processing



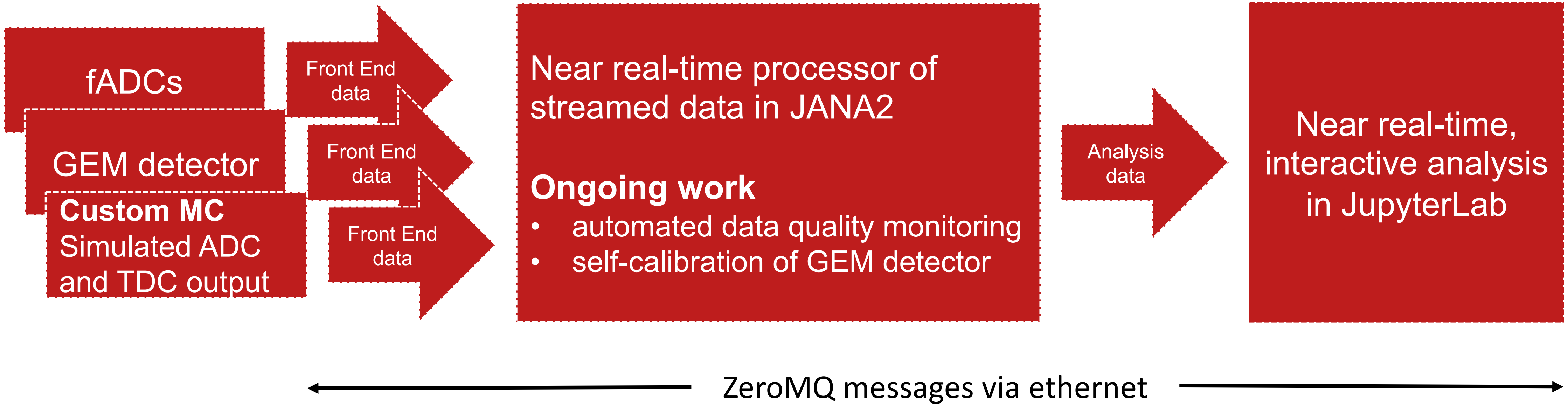
*LHCb will move to a triggerless-readout system for LHC Run 3 (2021-2023), and process 5 TB/s in real time on the CPU farm.

INDRA-ASTRA: Seamless integration of DAQ and analysis using AI

LDRD goal

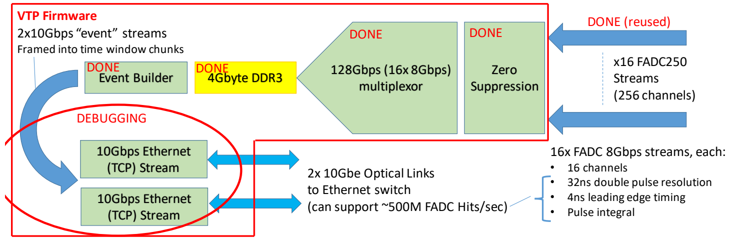
prototype components of streaming readout at NP experiments
→ integrated start to end system from detector read out through analysis
→ comprehensive view: no problems pushed into the interfaces

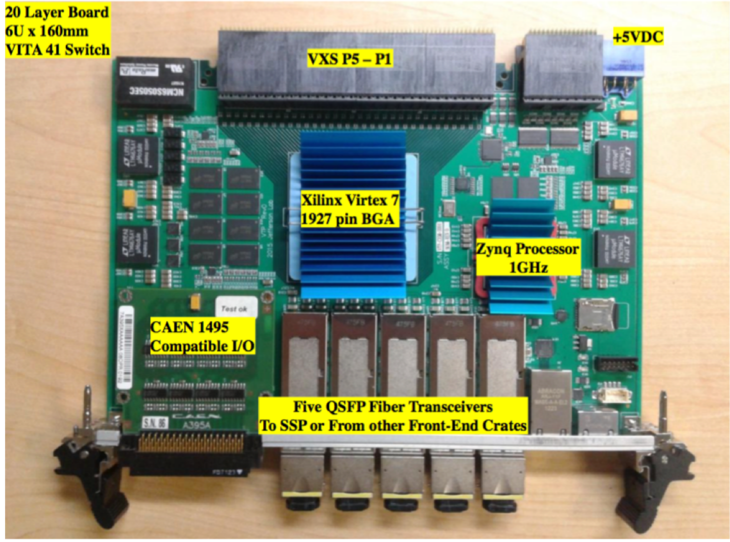
prototype (near) real-time analysis of NP data
→ inform design of new NP experiments



Prototype components of streaming readout

Testing streaming readout
of 250 MHz fADCs
Work with Ben Raydo

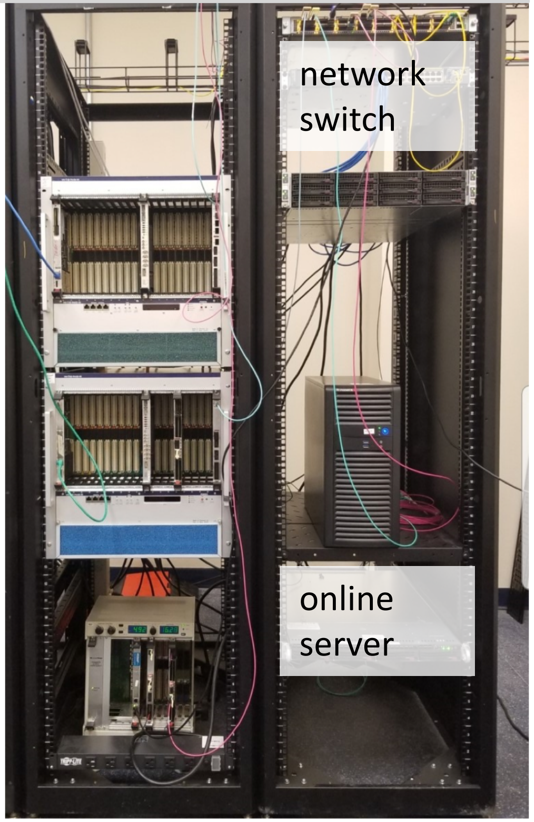




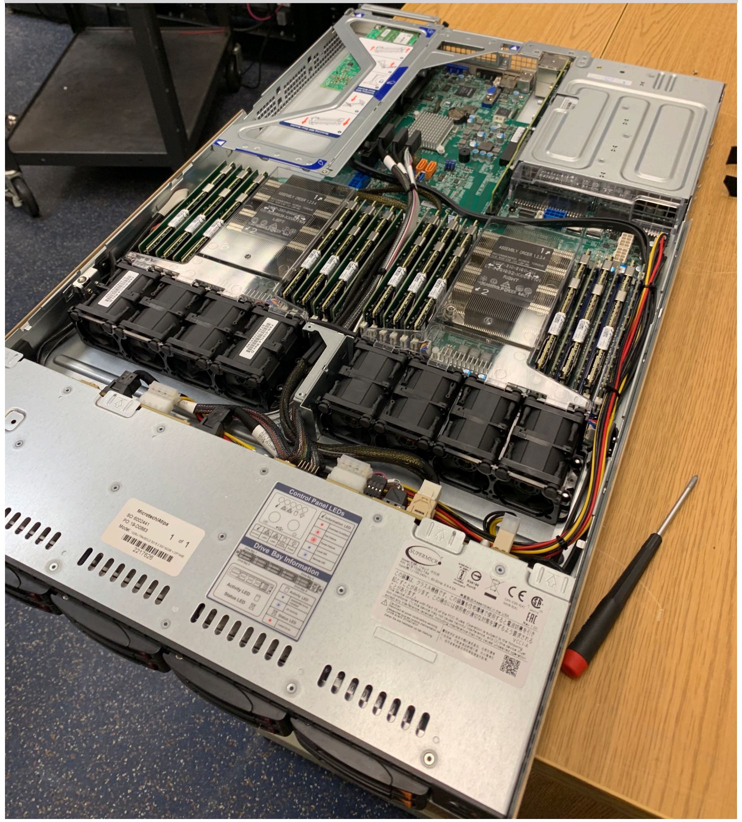
Test stand with front end data source, FPGA preprocessing device, network, and online server

VETROC TDC as front end data source

KCU1500 FPGA development board in PCI



INDRA-ASTRA file server
1.5 TB of memory (Intel® Optane™)
1.0 TB of solid state storage



Prototype detector for streaming readout

- We are able to stream trigger-less GEM data (768 channels) in DAS mode at 45 Gb/s via 5 ALICE FEC's
  - GEM ➛ FEC ➛ TRORC (30 Gb/s) ➛ PC Memory ➛ Disk
- We are able to stream triggered GEM data in DSP mode where a global readout threshold is applied and a programmable window of streamed data is captured by the T-RORC
  - Keeps the data volume to memory (and to disk) at a manageable level
- Will modify the T-RORC firmware to suppress the transmission of unnecessary sync packet data to memory and disk
  - Sync packets keep the serial links from the SAMPAs active when there is no hit data to send
- Then we can acquire data in a truly continuous fashion

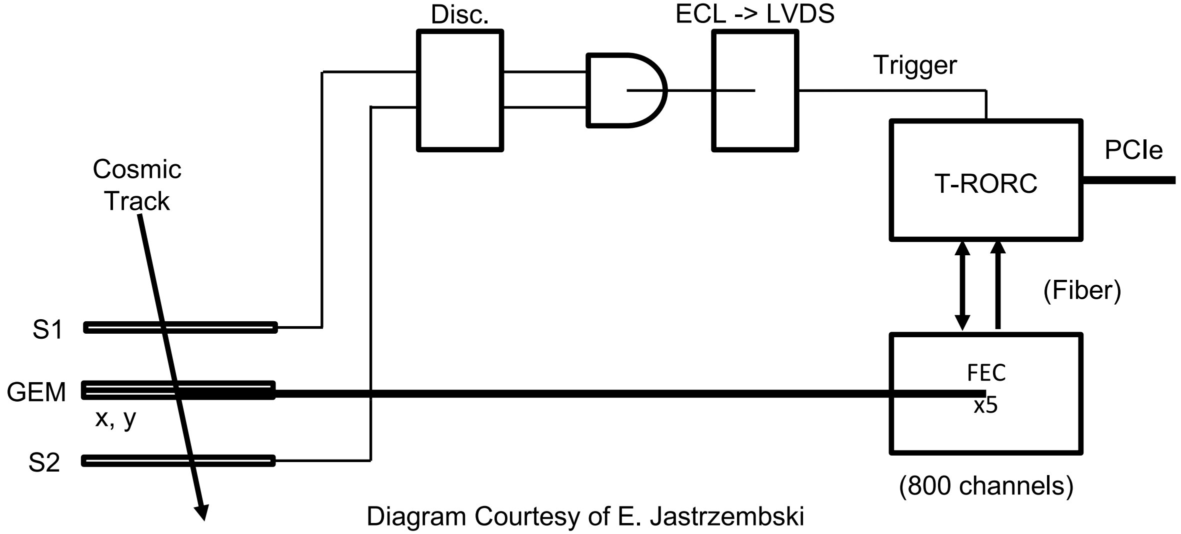
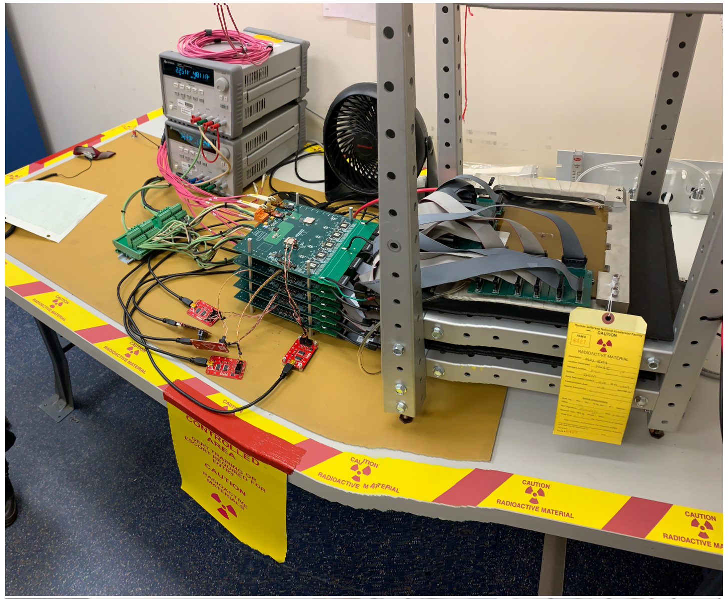
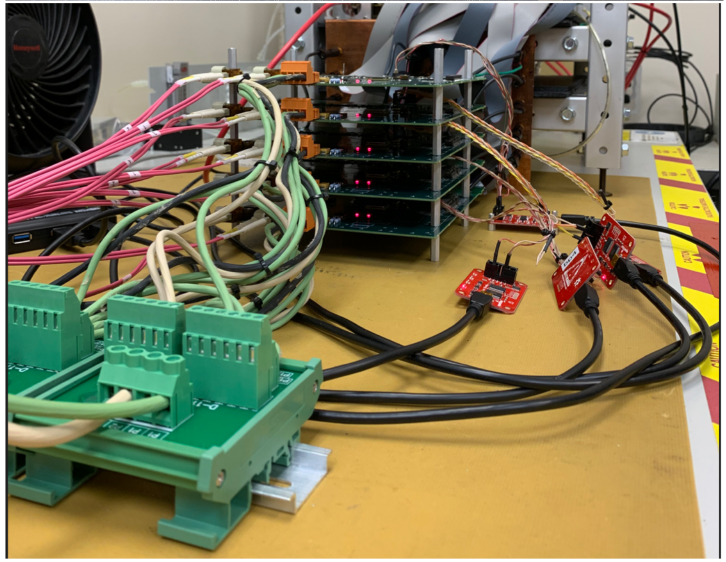


Diagram Courtesy of E. Jastrzembski





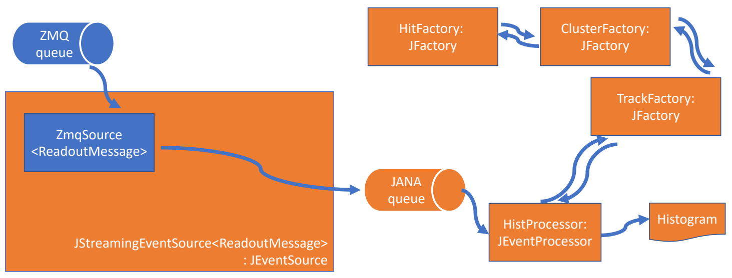
Prototype real-time analysis

Eric Pooser

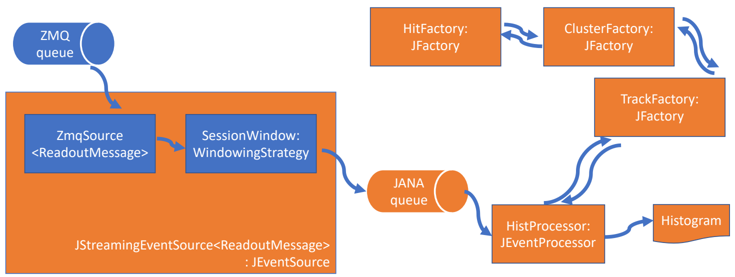
Developed: **Framework for parallel processing of streamed data**

- in collaboration with LDRD project on JANA2 (PI: David Lawrence; Nathan Brei)

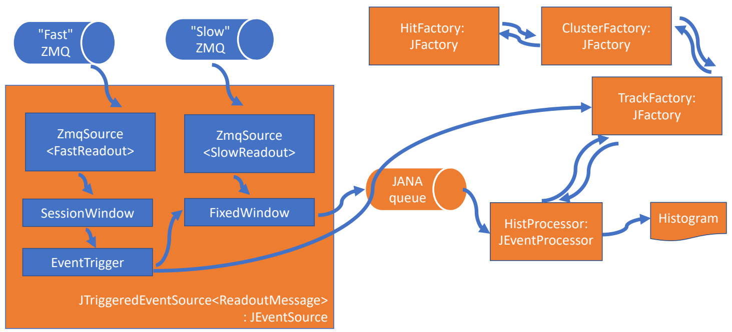
Streaming Data Readout, no event building



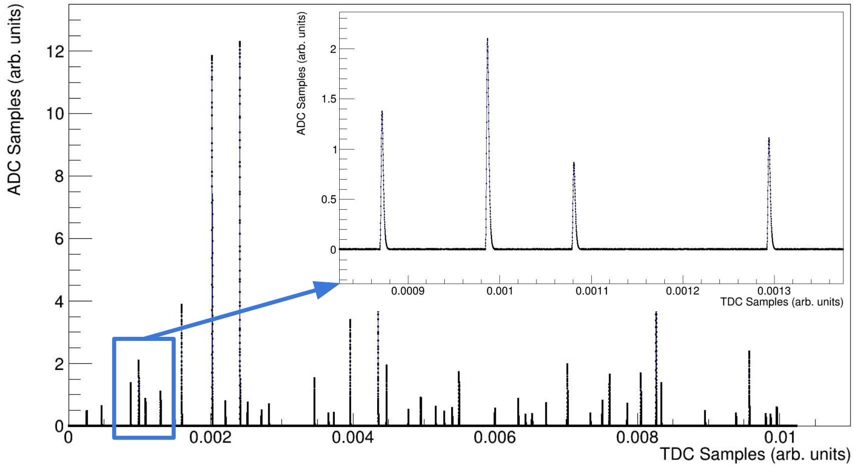
Streaming Data Readout, with event building



Streaming Data Readout, with software trigger



Developed: **Toy Detector Simulations**



Streaming Readout VI, May 14, 2020.

Developed: **Messaging library**

- work with Graham Heyes

- receiving data from streaming data source
- sending it efficiently over the network (TCP)
  - using the ZeroMQ messaging protocol
  - using EIC Streaming Readout protocol
- subscribing to a stream of data

tested at rates up to ~50 Gbit/s

Workflow tool for near real-time analysis of streamed data

JANA2

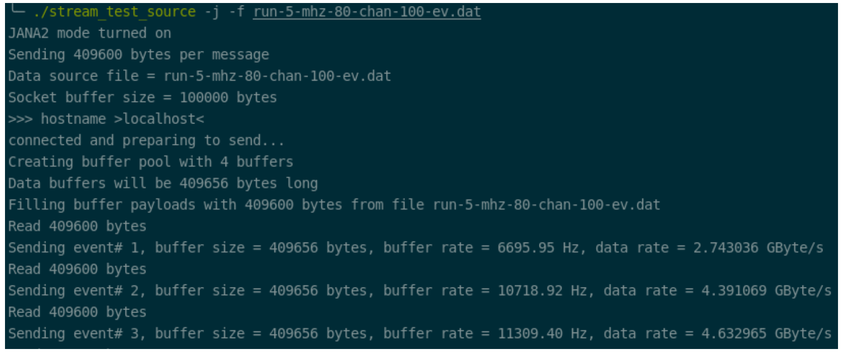
- C++ streamed data processing framework
- thread-level parallelism for streamed data
- supports event building: stream windowing
- stream windowing whenever two streams are merged

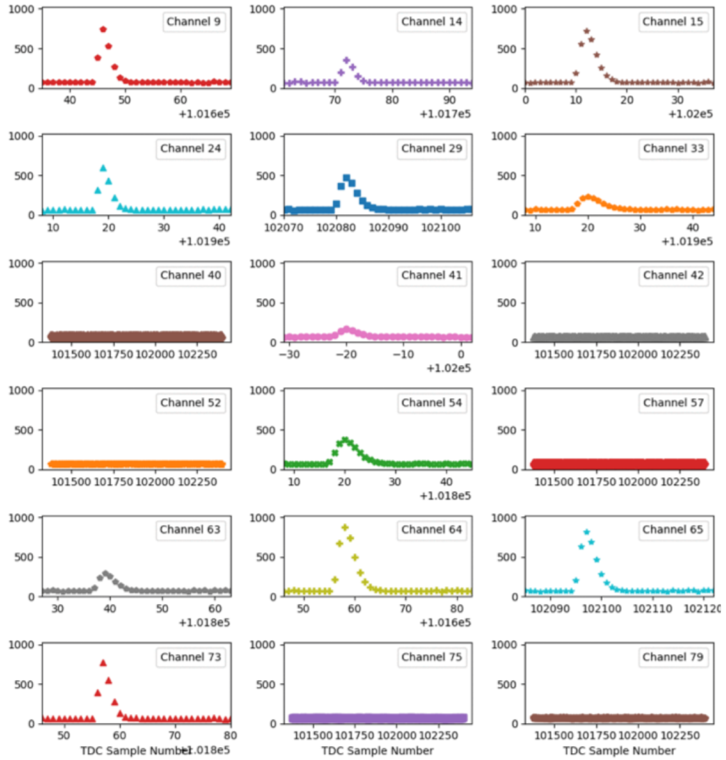
JANA2 design goals

- support streaming as an optional plugin, but use it to inform API improvements
- keep deserialization, transport, and windowing orthogonal to each other
- make these components reusable
- keep JANA responsible for thread-level parallelism; use ZeroMQ for node-level parallelism

JANA2 streaming plugins (Eric Pooser)

- decoding streams of MC and/or detector data ✓
- visualization of streamed data in near real time ✓
  - fully extensible in JupyterLab ✓ (Dmitry Romanov will help)





Automatic detector calibrations

Project with Abdullah Farhat and Yuesheng Xu (Department of Mathematics, Old Dominion University)

Approach

- start with baseline calibration
- monitor detector performance online and automatically identify changes
- re-calibrate when detector performance changes

Take advantage of contemporary ML techniques

- determination of when the underlying probability distribution over streamed data changes
- **unsupervised learning problem for anomaly detection** that can be solved using deep neural networks
- recent **Learn-by-Calibration (LbC) Neural Networks** models produce prediction intervals without any prior assumption on the underlying distribution of data samples, trained by optimizing non-smooth objective and regularity terms. Setting a threshold on the size and shape of these prediction intervals set criteria for detecting anomalous detector behavior
- **Status**
  - toy model being developed
  - work on GEM detector prototype (pedestal analysis)

Outlook

Markus Diefenthaler
mdiefent@jlab.org

- **Vision** Develop Machine-Detector Interfaces into a **Machine-Detector-Analysis Interface** with analysis-ready data from the DAQ system

- **Understand** requirements for DAQ / Electronics and Software & Computing, also in light of a **Machine-Detector-Analysis Interface**

- **INDRA-ASTRA** Rapid prototyping of streaming readout and online / real-time calibration and analysis