AI in Streaming Data Processing at the EIC

Markus Diefenthaler (Jefferson Lab)

Overview

In my presentation, I will **focus on ePIC**, the primary experiment at the EIC. I will provide an **overview of the compute-integration of the ePIC experiment** and **highlight streaming readout and AI applications at ePIC**, showcasing their potential to redefine experimental capabilities and precision.

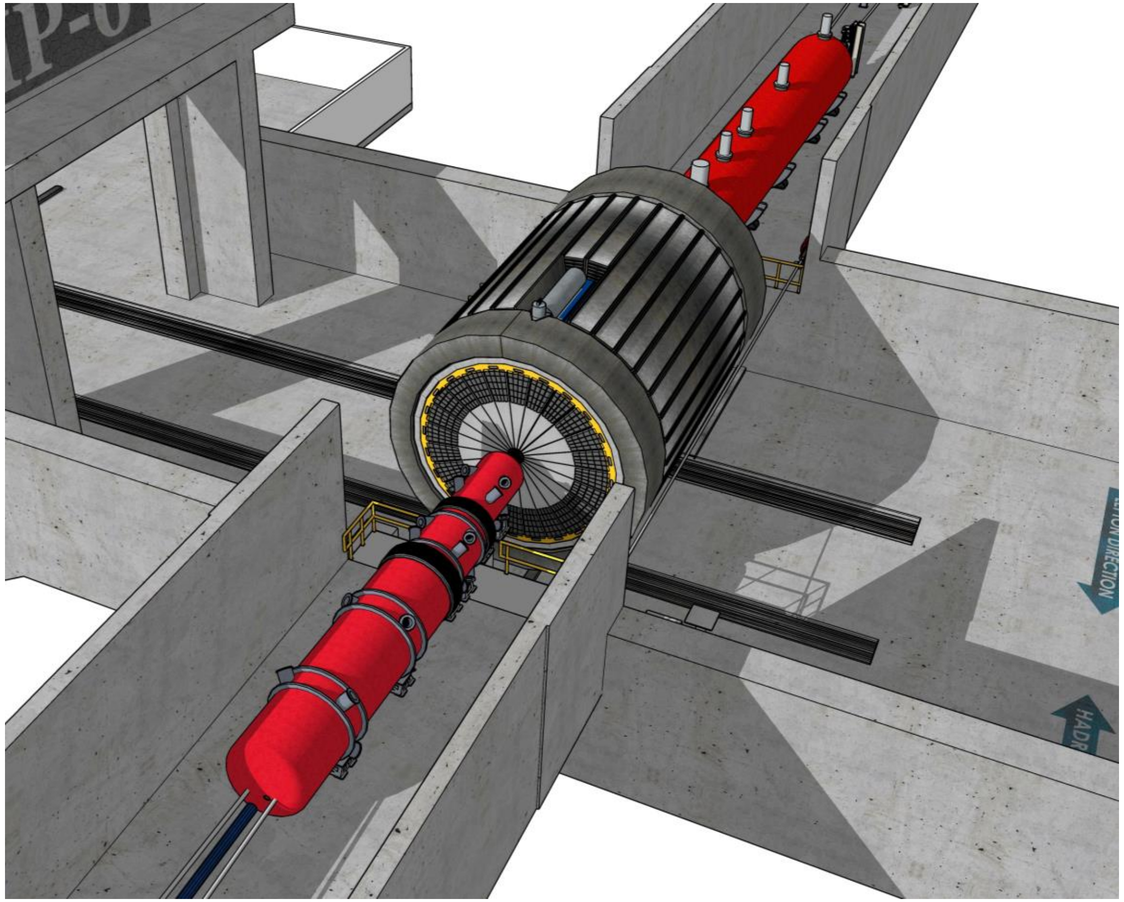
ePIC in a Nutshell

- ePIC is a **highly integrated**, **multi-purpose** experiment.
- The **ePIC Experiment** is being developed by the **ePIC Collaboration** in partnership with the EIC Project.
- The **ePIC Collaboration** was established in 2022 and is international in scope, comprising over 900 collaborators from 178 institutions across 25 countries and 5 world regions.

The Highly-Integrated ePIC Experiment

Integrated Interaction and Detector Region (90 m)

Get close to full acceptance for all final state particles, and measure them with good resolution. All particles count!



Compute-Detector Integration

Seamless data processing from detector readout to analysis using streaming readout and streaming computing.

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 data processing is deferred to streaming computing.

Advantages of Streaming Readout

- Simplification of readout (no trigger hardware, firmware, decisions) and increased flexibility.
- Event building from holistic detector information.
- Continuous data flow provides detailed knowledge of backgrounds and enhances control over systematics.

Compute-Detector Integration to Maximize Science

Broad ePIC Science Program:
- **Plethora of observables**, with less distinct topologies where every event is significant.
- **High-precision measurements**: Control of systematic uncertainties of paramount importance.

Streaming Readout Capability Due to Moderate Signal Rate:
- **Capture every collision signal**, including background.
- Event selection using all available detector data for **holistic reconstruction**:
  - **Eliminate trigger bias** and provide accurate estimation of uncertainties during event selection.
- Streaming background estimates ideal to **reduce background** and related systematic uncertainties.

	EIC	RHIC	LHC → HL-LHC
Collision species	$\vec{e}+\vec{p}, \vec{e}+A$	$\vec{p}+\vec{p}/A, A+A$	$p+p/A, A+A$
Top x-N C.M. energy	140 GeV	510 GeV	13 TeV
Peak x-N luminosity	10^{34} cm^{-2} s^{-1}	10^{32} cm^{-2} s^{-1}	$10^{34} \rightarrow 10^{35}$ cm^{-2} s^{-1}
x-N cross section	**50 µb**	**40 mb**	**80 mb**
Top collision rate	500 kHz	10 MHz	1-6 GHz
$dN_{ch}/d\eta$	0.1-Few	~3	~6
Charged particle rate	**4M N_{ch}/s**	**60M N_{ch}/s**	**30G+ N_{ch}/s**

Compute-Detector Integration to Accelerate Science

- **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** **Rapid turnaround of 2-3 weeks for data for physics analyses**.
  - Timeline driven by alignment and calibrations.
  - Discussed alignment and calibration procedures and requirements with detector experts. Preliminary information from Detector Subsystem Collaborations indicates that 2-3 weeks are realistic.
- **Solution** Compute-detector integration using:

Streaming readout for continuous data flow of the full detector information.

AI for autonomous alignment and calibration as well as autonomous validation for rapid processing.

Heterogeneous computing for acceleration (CPU, GPU).

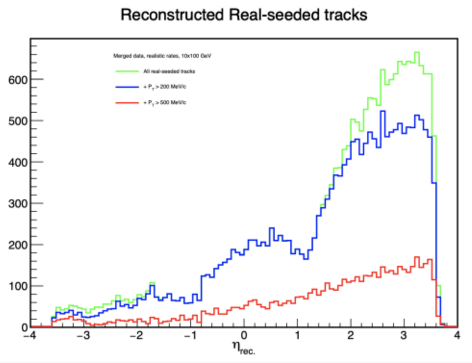
Prototype of Event Reconstruction from Streaming Data

Scope of the first prototype: Track reconstruction only. Demonstrated that we can correlate hits in a realistic time frame to the various events in the time window of the MAPS of 2μs.

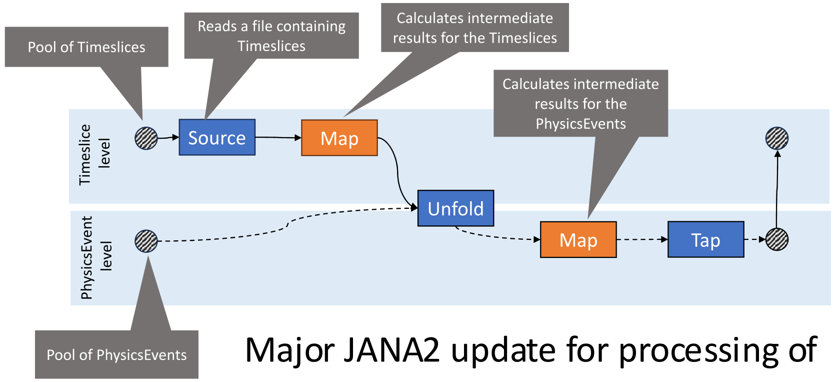
Streaming DAQ → **Timeframes** → **Streaming Computing (JANA2)** → **Reconstructed Events**

Continuous stream of data

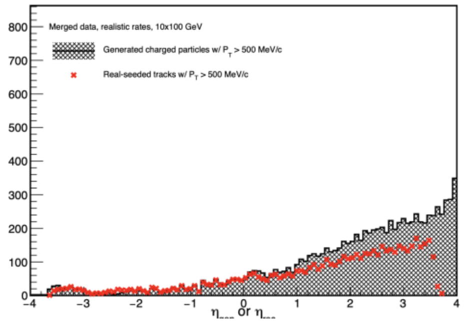
0.6ms of data



- Data transferred in collections called *timeframes* (or timeframes aggregated into super-timeframes).
- Each timeframe includes:
  - Data read from detectors over a time window of 2^{16} cycles of the beam RF, equivalent to 0.6 ms.
  - Channel information and corresponding timing data

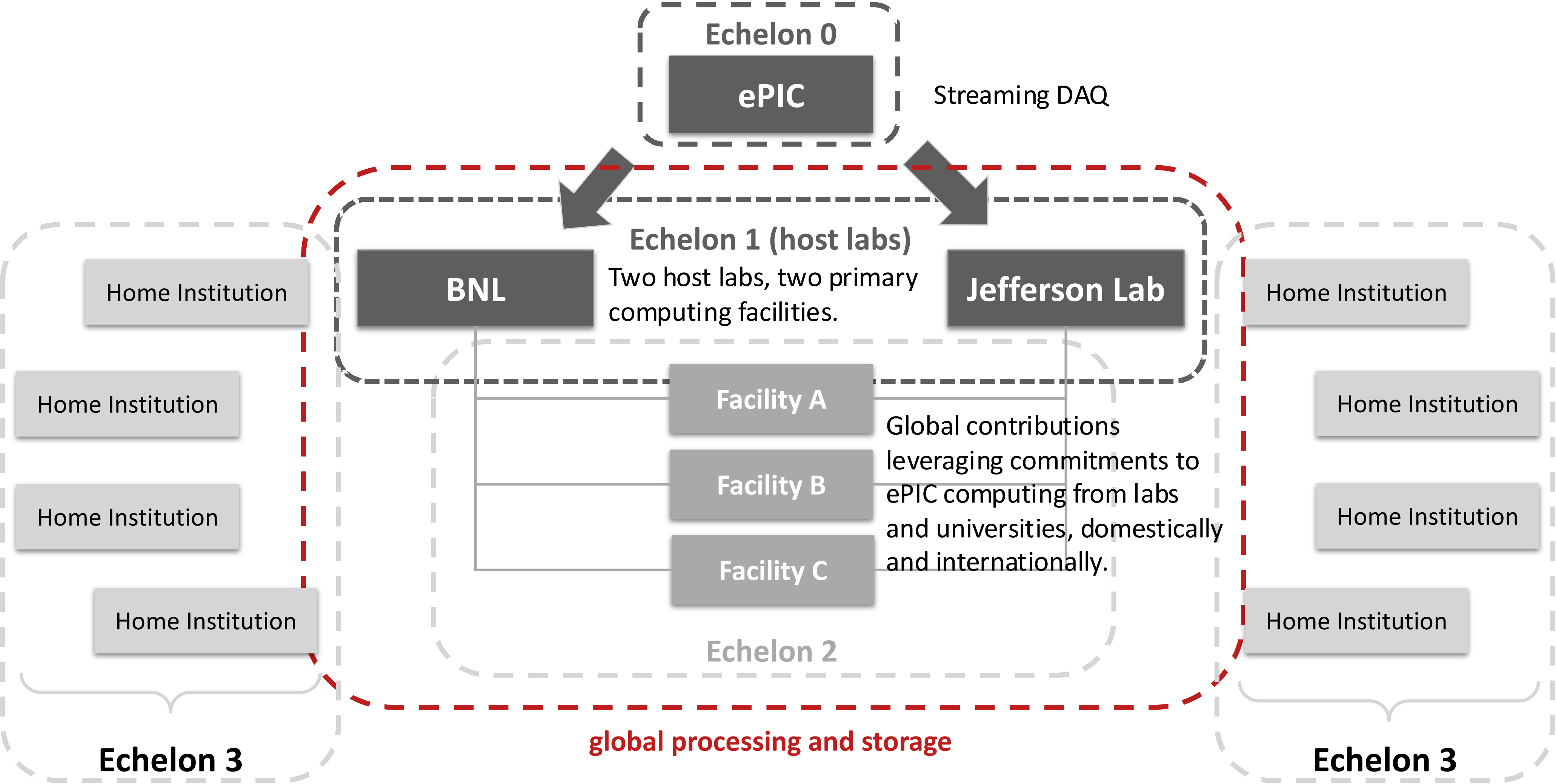


Major JANA2 update for processing of timeframes, events, and subevents.



Jefferson Lab

The ePIC Streaming Computing Model



Supporting the analysis community *where they are* at their home institutes, primarily via services hosted at Echelon 1 and 2.

Computing Resource Needs and Their Implications

For EIC Phase I

Processing by Use Case [cores]	Echelon 1	Echelon 2
Streaming Data Storage and Monitoring	-	-
Alignment and Calibration	6,004	6,004
Prompt Reconstruction	60,037	-
First Full Reconstruction	72,045	48,030
Reprocessing	144,089	216,134
Simulation	123,326	369,979
Total estimate processing	**405,501**	**640,147**

Storage Estimates by Use Case [PB]	Echelon 1	Echelon 2
Streaming Data Storage and Monitoring	71	35
Alignment and Calibration	1.8	1.8
Prompt Reconstruction	4.4	-
First Full Reconstruction	8.9	3.0
Reprocessing	9	9
Simulation	107	107
Total estimate storage	**201**	**156**

O(1M) core-years to process a year of data:

- Optimistic scaling of constant-dollar performance gains would reduce the numbers about 5x:
  - Based on current LHC measure of 15% per year.
  - But the trend is towards lower gains per year.
- Whatever the gains over time, processing scale is substantial!
- Motivates attention to leveraging distributed and opportunistic resources from the beginning.

~350 PB to store data of one year.

Computing resource needs comparable to LHC experiments ATLAS and CMS at their scales today.

ePIC is compute intensive experiment; must ensure ePIC is not compute-limited in its science.

The Role of AI

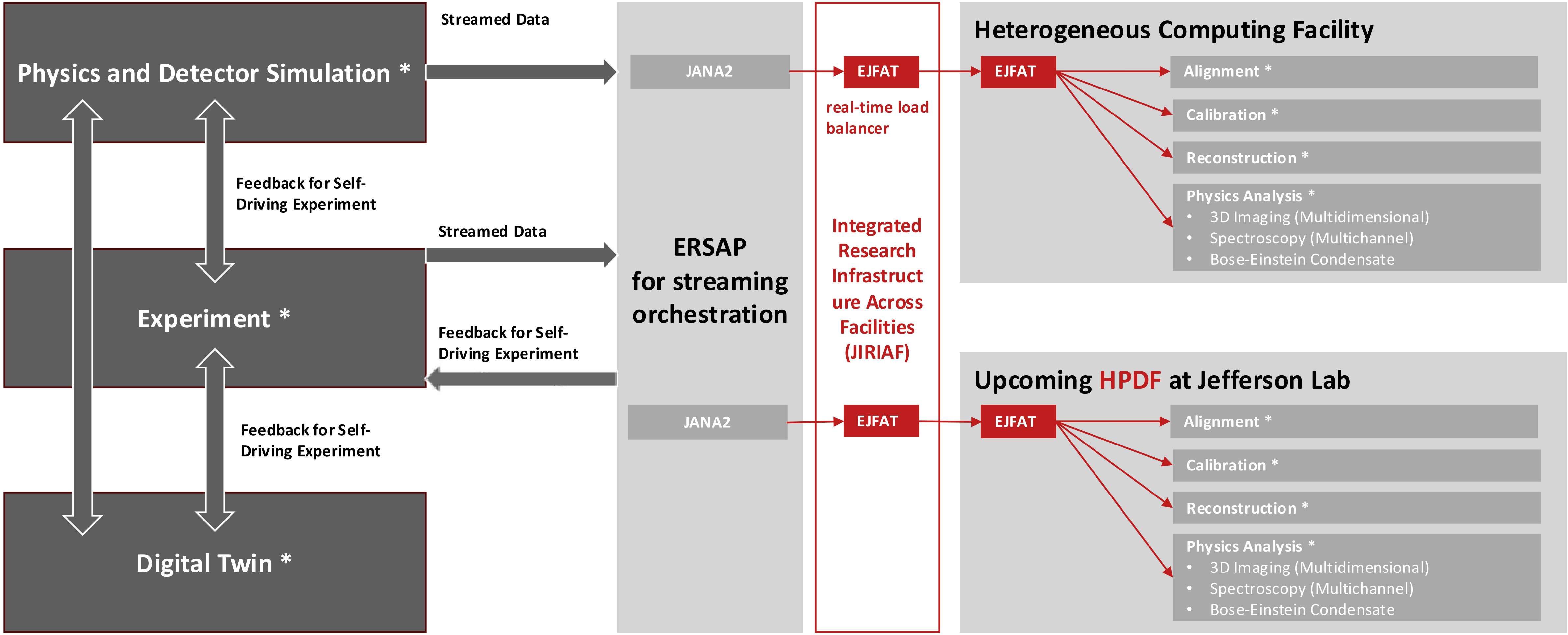
- **Compute-detector integration** using:

Streaming readout for continuous data flow of the full detector information.	**AI** for autonomous alignment and calibration as well as reconstruction and validation for rapid processing.	**Heterogeneous computing** for acceleration.

- AI will **empower the data processing** at the EIC.
  - Rapid turnaround of data relies on autonomous alignment and calibration as all as autonomous validation.

- AI will also **empower autonomous experimentation and control** beyond data processing:
  - Vision for a responsive, cognizant detector system, .e.g., adjusting thresholds according to background rates.
  - Enabled by access to full detector information via streaming readout.

Example: Streaming Computing Developments at Jefferson Lab

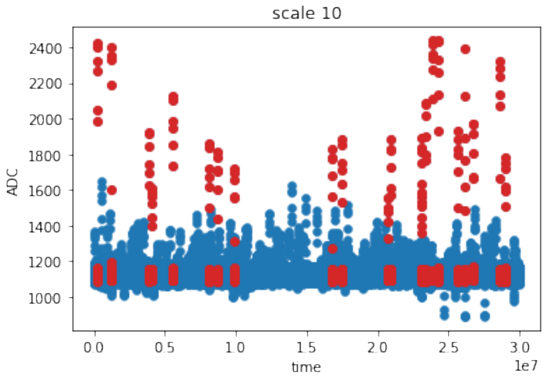


*** AI/ML**

Example: Change Detection

Soon to be published, featuring examples from TIDIS.

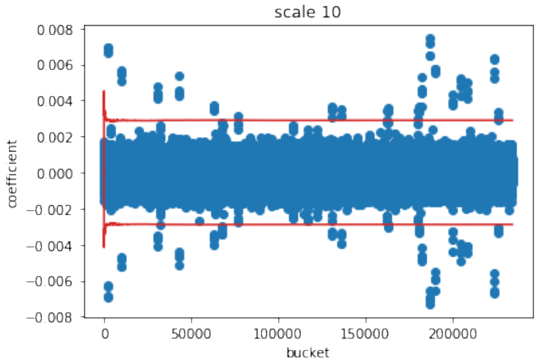
- **Multiscale method** for detection of sudden changes and gradual changes
  - Various test functions for various changes
  - **Fast algorithm for signal analysis**
    - Represent data in multiscale basis:



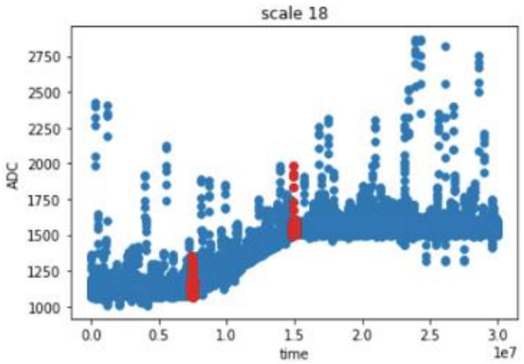
Transform to **coefficient space**:

- Outliers in the distribution → Change.
- **Change detection: Detect outliers**
  - e.g., using IQR

Reference Z. Chen, C.A. Micchelli, C., Y. Xu, *Multiscale Methods for Fredholm Integral Equations*, Cambridge Monographs on Applied and Computational Mathematics (2015)



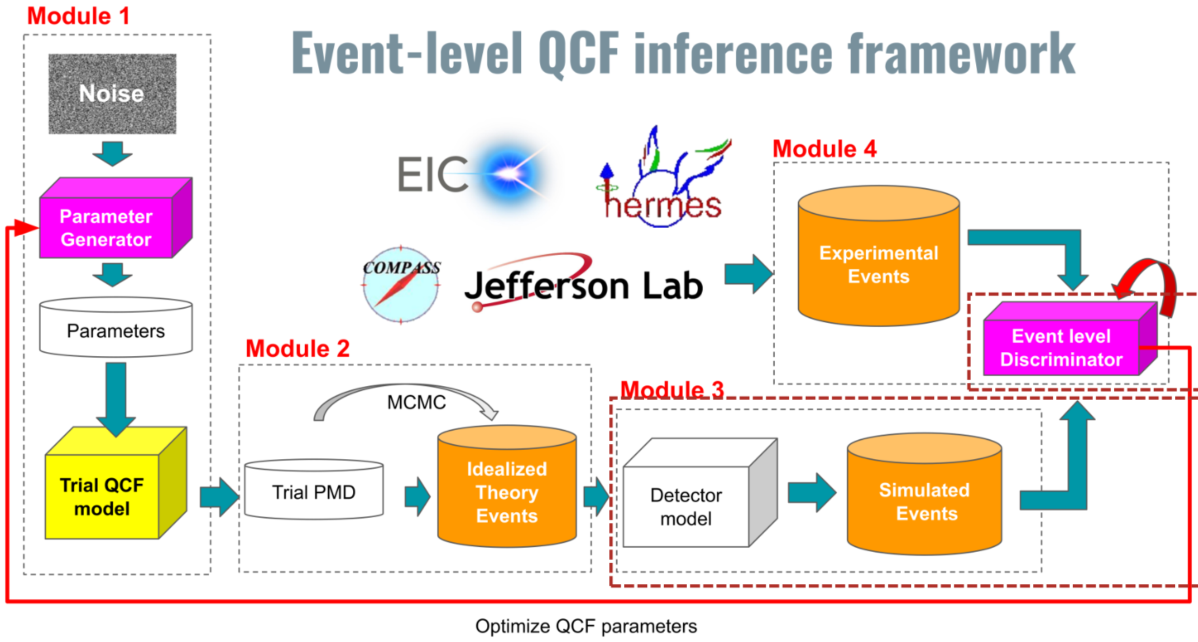
- **Application**:
  - Analyze results from **multiscale method** (in red).
  - Decide on response:
    - Issue alarm,
    - Restart calibration,
    - Start user-defined process.



Example: Streaming Analysis

Example from QuantOm (SciDAC Project).

Example workflow for 3D imaging of quarks and gluons,
extracting quantum correlation functions (QCFs) using **events streamed from the detector**.



Compare experiment and theory at the detector level.

Enabled by folding that means idealized theory events are folding with detector effects to resemble experimental events.

AI used for **modeling of experimental effects**.

Advantages of Folding Approach

Joint theoretical-Experimental Workflow: Match theoretical assumptions and experimental cuts in an unprecedented manner, reducing mismatches between experiment and theory.

Robust approach: Variants in theory can be rigorously studied, and additional experimental data can be incorporated seamlessly.

Summary: AI in Streaming Data Processing at the EIC

ePIC Experiment: A highly integrated, multi-purpose experiment at the EIC, leveraging advanced compute-detector integration to enhance precision measurements:

- **Streaming Readout to Maximize Science**: Continuous detector data acquisition without hardware triggers, enabling holistic event selection, improved background control, and precise uncertainty estimation during event selection.
- **AI-Powered Processing to Accelerate Science**: Autonomous alignment, calibration, and validation for rapid data turnaround, accelerating scientific discoveries.
- **Distributed Computing**: Globally distributed resources to manage compute resource needs efficiently.

Streaming + AI + Distributed Computing = Data Processing at the EIC

AI Examples:

- **Change Detection**: Multiscale analysis for identifying sudden and gradual changes in detector data, enabling responsive system adjustments.
- **Streaming Analysis**: Extracting QCFs in the QuantOm workflow using events streamed from the detector. AI-powered modeling of experimental effects for the folding approach in QuantOm.

Community efforts on AI and data applications for streaming and distributed computing are essential for the data processing at the EIC.