# Streaming Computing Context

Markus Diefenthaler (Jefferson Lab)

# Compute-Detector Integration to Accelerate Science

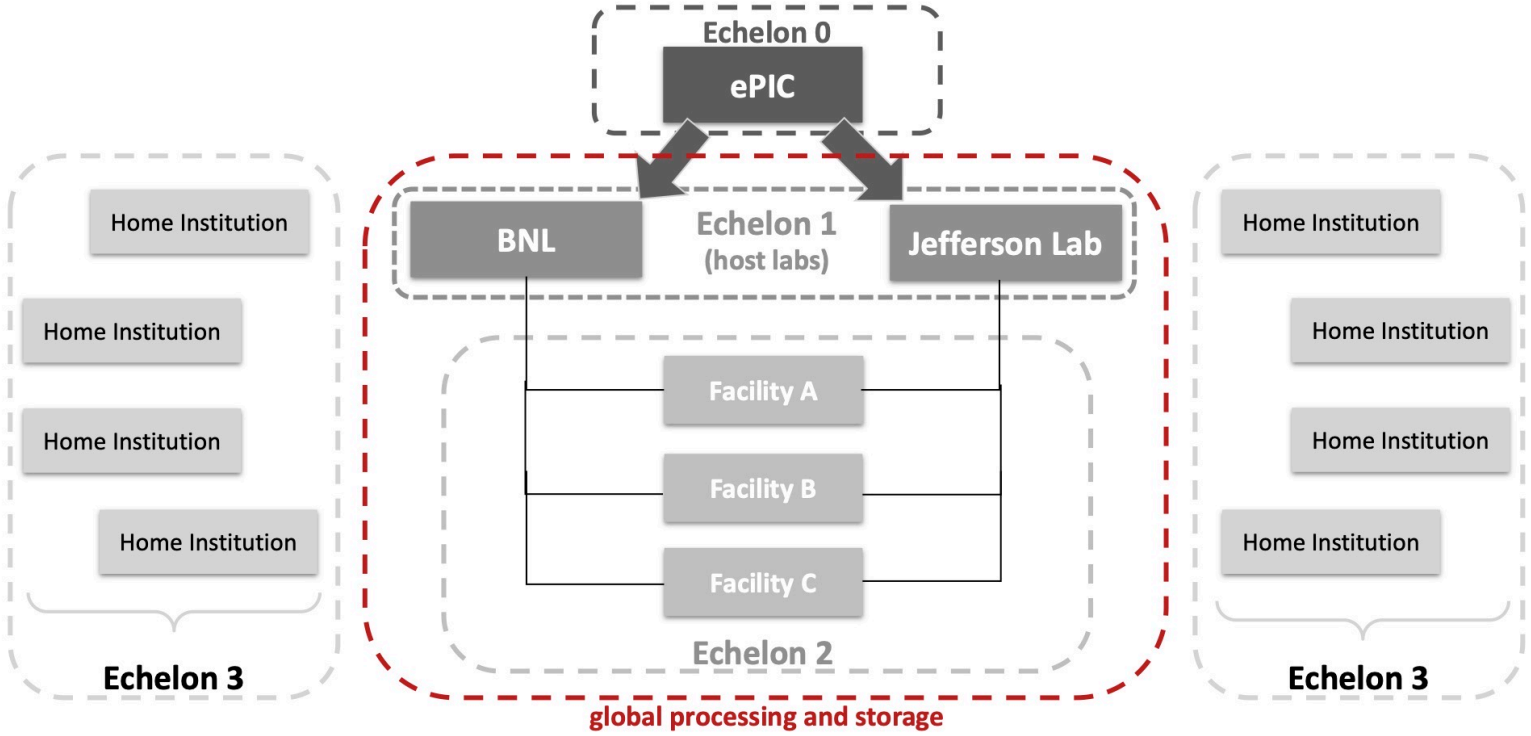
- **Problem** Data for physics analyses and the resulting publications available after O(1year) due to complexity of NP experiments (and their organization).
  - Alignment and calibration of detector as well as reconstruction and validation of events time-consuming.
- **Goal** **Rapid turnaround of 2-3 weeks for data for physics analyses**.
  - Timeline driven by alignment and calibrations. Overview of alignment and calibration workflows.
- **Solution** Compute-detector integration using:

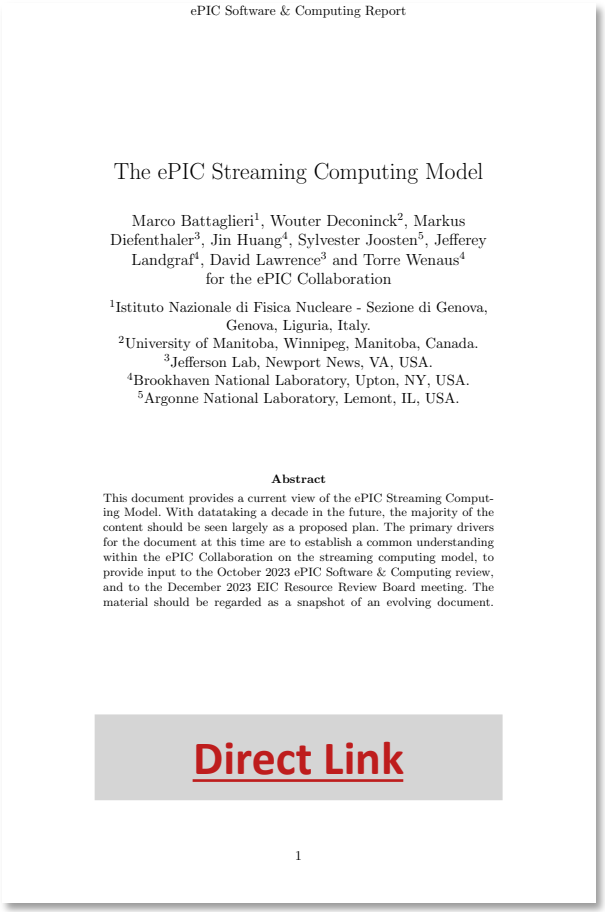
**Streaming readout** for continuous data flow of the full detector information.

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

**Heterogeneous computing** for acceleration.

# Current View of the Computing Model



ePIC Software & Computing Report

The ePIC Streaming Computing Model

Marco Battaglieri[1], Wouter Deconinck[2], Markus Diefenthaler[3], Jin Huang[4], Sylvester Joosten[5], Jefferey Landgraf[4], David Lawrence[3] and Torre Wenaus[4]
for the ePIC Collaboration

[1]Istituto Nazionale di Fisica Nucleare - Sezione di Genova, Genova, Liguria, Italy.
[2]University of Manitoba, Winnipeg, Manitoba, Canada.
[3]Jefferson Lab, Newport News, VA, USA.
[4]Brookhaven National Laboratory, Upton, NY, USA.
[5]Argonne National Laboratory, Lemont, IL, USA.

**Abstract**

This document provides a current view of the ePIC Streaming Computing Model. With datataking a decade in the future, the majority of the content should be seen largely as a proposed plan. The primary drivers for the document at this time are to establish a common understanding within the ePIC Collaboration on the streaming computing model, to provide input to the October 2023 ePIC Software & Computing review, and to the December 2023 EIC Resource Review Board meeting. The material should be regarded as a snapshot of an evolving document.

Direct Link

1

Initial version of a plan set to develop over the next decade.

# Priority Target for the TDR

## Prototype of event reconstruction from realistic frames:

**Purpose**: Demonstrate that we can reconstruct events from Streaming DAQ.

**Purpose**: Estimation of streaming reconstruction time for compute resource planning.

# Prototype of Event Reconstruction from Realistic Frames

**Key Tasks**: We limit the scope of the first study to the track reconstruction only. The key is to demonstrate we can correlate hits in a realistic time frame to the various events in the time window of the MAPS.

- **Electronics and DAQ WG**
  - Define the *realistic* time frames. We need a consensus on the meaning of *realistic*. **Today's Meeting.**

- **Simulation WG**:
  - Prepare simulation productions, using detailed information on FEEs for tracking detectors, utilizing the full, wide MAPS integration window for tracking purposes.
  - Implement and utilize the frame-building infrastructure post-Geant4 and post-digitization.

- **Reconstruction WG**:
  - Integrate Jana2's built-in workflow for supporting frames in and events out in EICrecon.
  - Adapt the reconstruction process to work with frames, making it frames-aware.
  - Demonstrate tracking from realistic frames.

# Current Estimate for Compute Resources

**Streaming DAQ** sends data in **1ms time frames**.

Each time frame corresponds to 10MB of data.

Based on our current detector readout design and when running at peak luminosity and in standard operating conditions.

40% of data bunch crossing related, 60% background.

In a year, we will record 15.5 billion frames.

Assuming a 50% up-time for 6 months.

**Number of expected events** (assuming a 50% up-time for 6 months):

- The event rate at peak luminosity is 500kHz, which gives roughly 4 x $10^{12}$ events.
  - Lower at start of operations, where the luminosity will be lower (but relatively speaking background rate is expected to be higher).
- The expected number of physics events of interest for one year of running at peak luminosity is ~ $10^{10}$.
  - The actual physics events is only a very small fraction of the total physics bunch crossings.

**Number of simulation events**:

- We expect to simulate 10x events for each event of interest, yielding O($10^{11}$) simulated events.

  While considerable ( ~ 60k core years on today's hardware), this should be a realistic target in a decade.

**Core-seconds for simulation and reconstruction (on a typical modern machine):**

- Our current simulations including background take ~17s for simulation and ~ 2-3s for reconstruction, per event.
  - Simulation and reconstruction on event level only.
- **Unknown**: How much this will change once changing to streaming data processing?
- **Priority target for TDR**: Prototype of event reconstruction from realistic frames.