Echelon 1 Role and Timeline

Markus Diefenthaler (Jefferson Lab)

ePIC Software and Computing Organization

Spokesperson's Office

Detector | **Software & Computing** | Physics

Software and Computing Coordinator
Markus Diefenthaler (Jefferson Lab)

Cross-cutting Working Group:

- *Data and Analysis Preservation*

Deputy Coordinator (Operations)
Wouter Deconinck (U. Manitoba)

Operation Working Groups:

- Production
- User Learning
- Validation

Deputy Coordinator (Development)
Dmitry Kalinkin (UKY)

Development Working Groups:

- Physics and Detector Simulation
- Reconstruction Framework and Algorithms
- *Analysis Tools*

Deputy Coordinator (Infrastructure)
Torre Wenaus (BNL)

Infrastructure Working Groups:

- Streaming Computing Model
- *Multi-Architecture Computing*
- *Distributed Computing*

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} → 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

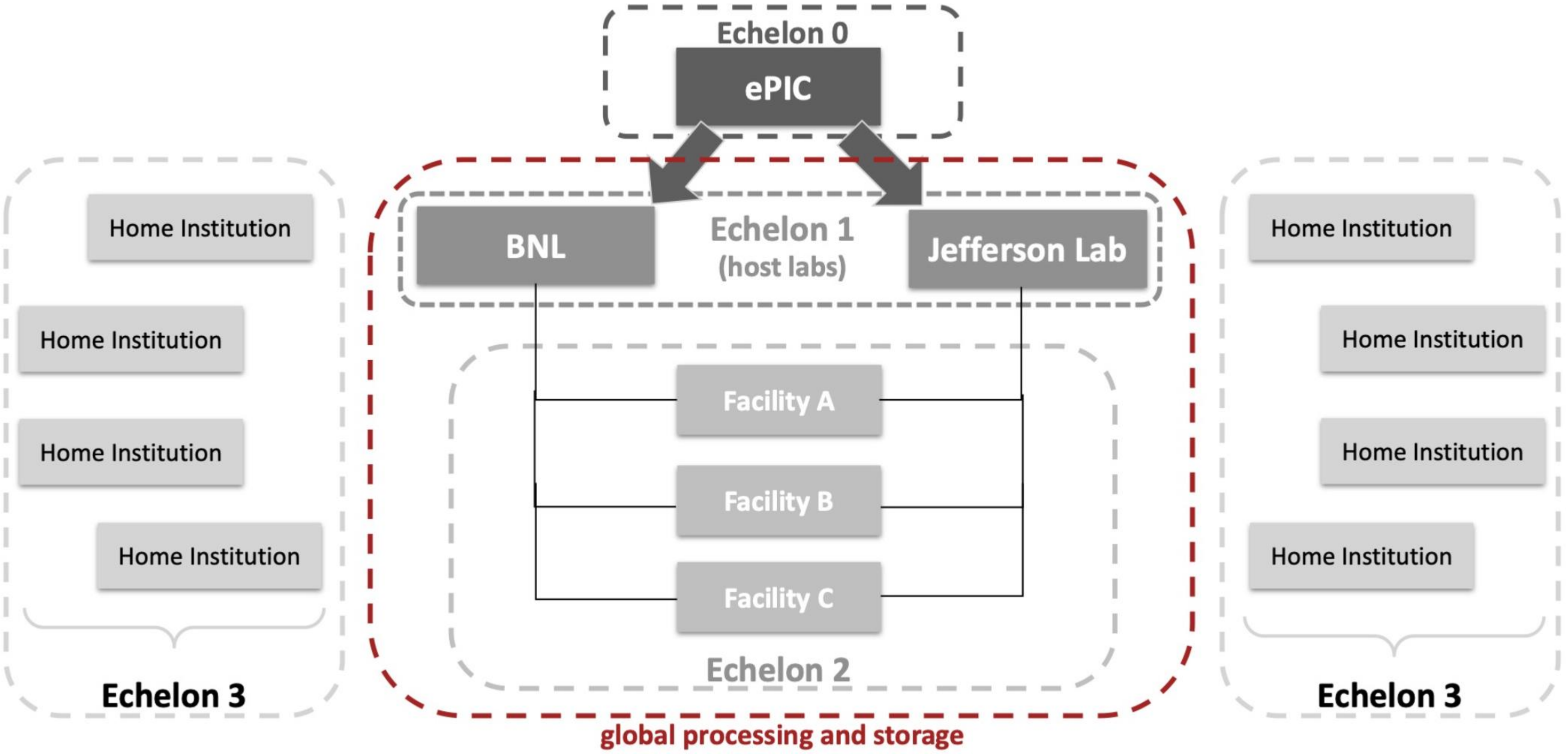
- **Problem** Data for physics analyses and the resulting publications available after O(1year) due to complexity of NP experiments (and their organization).
  - Alignment and calibration of detector as well as reconstruction and validation of events time-consuming.
- **Goal** **Rapid turnaround of 2-3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

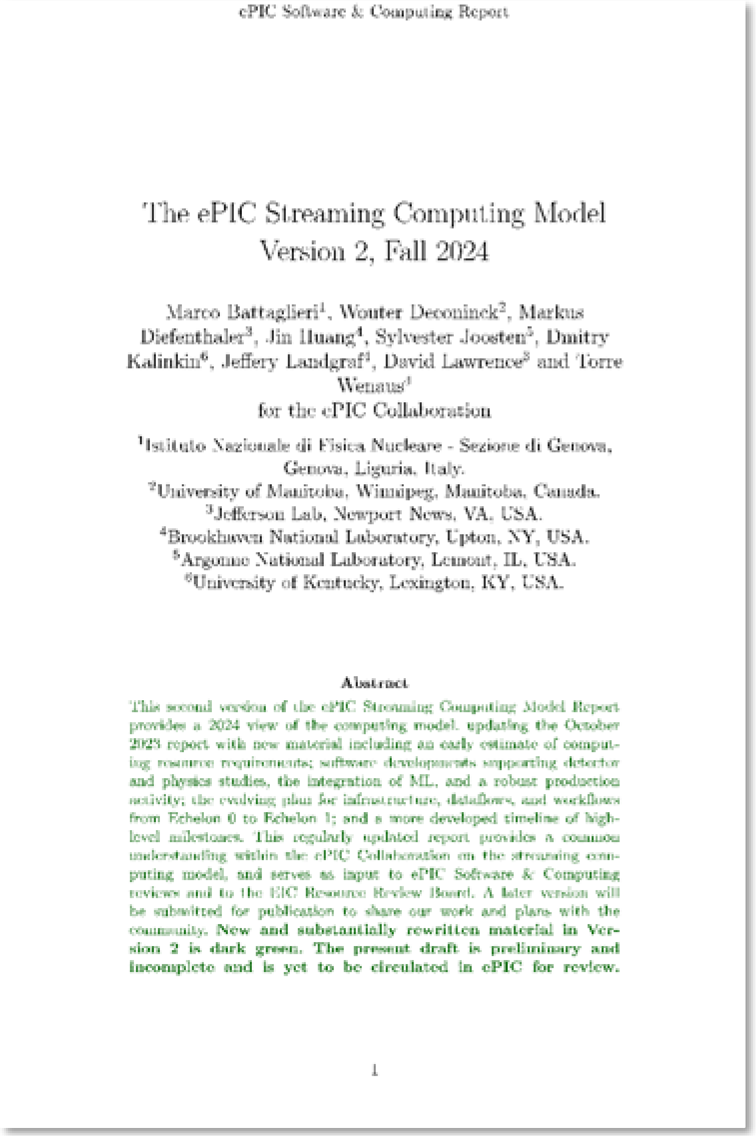
Heterogeneous computing for acceleration (CPU, GPU).

ePIC Streaming Computing Model



Echelon 0: ePIC Streaming DAQ.
Echelon 1: Two host labs, two primary ePIC computing facilities.
Echelon 2: Global contributions leveraging commitments to ePIC computing from labs and universities, domestically and internationally.
Echelon 3: Supporting the analysis community *where they are* at their home institutes, primarily via services hosted at Echelon 1 and 2.

ePIC Software & Computing Publication


ePIC Software & Computing Report

The ePIC Streaming Computing Model
Version 2, Fall 2024

Marco Battaglieri[1], Wouter Deconinck[2], Markus Diefenthaler[3], Jin Huang[4], Sylvester Joosten[5], Dmitry Kalinkin[6], Jeffery 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.
[6]University of Kentucky, Lexington, KY, USA.

Abstract

This second version of the ePIC Streaming Computing Model Report provides a 2024 view of the computing model, updating the October 2023 report with new material including an early estimate of computing resource requirements; software developments supporting detector and physics studies, the integration of ML, and a robust production activity; the evolving plan for infrastructure, dataflows, and workflows from Echelon 0 to Echelon 1; and a more developed timeline of high-level milestones. This regularly updated report provides a common understanding within the ePIC Collaboration on the streaming computing model, and serves as input to ePIC Software & Computing reviews and to the EIC Resource Review Board. A later version will be submitted for publication to share our work and plans with the community. **New and substantially rewritten material in Version 2 is dark green. The present draft is preliminary and incomplete and is yet to be circulated in ePIC for review.**

1

Status and Plans

- We already have a draft of the paper (https://doi.org/10.5281/zenodo.14675920).
- As the next step, the draft is currently under review by the EIC Computing and Software Advisory Committee (ECSAC) for technical feedback. We expect to receive their comments by April 15 and will meet on May 12 to discuss their feedback.
- Following that discussion, we will circulate the revised draft within the collaboration for a final round of input before submission.

Use Cases and a Word on Funding

Use Case	Echelon 0	Echelon 1	Echelon 2	Echelon 3
Streaming Data Storage and Monitoring	✓	✓		
Alignment and Calibration		✓	✓	
Prompt Reconstruction		✓		
First Full Reconstruction		✓	✓	
Reprocessing		✓	✓	
Simulation		✓	✓	
Physics Analysis		✓	✓	✓
AI Modeling and Digital Twin		✓	✓	

Echelon 0	Included in EIC Project Detector funding.
Echelon 1	Part of future EIC Operations funding.
Echelon 2	Extraordinary resources to accelerate progress on EIC Science.
Echelon 3	Extraordinary resources to accelerate EIC analysis.

Echelon 1 and 2 Computing

Use Case	Echelon 1	Echelon 2
Streaming Data Storage and Monitoring	✓	
Alignment and Calibration	✓	✓
Prompt Reconstruction	✓	
First Full Reconstruction	✓	✓
Reprocessing	✓	✓
Simulation	✓	✓

Assumed Fraction of Use Case Done Outside Echelon 1		
Alignment and Calibration	50%	Driven mainly by where the experts are.
First Full Reconstruction	40%	Assuming Echelon 2 ready by start of operations.
Reprocessing	60%	Can scale to whatever our Echelon 2 resources are. Allows Echelon1 to focus on prompt and first full reconstruction.
Simulation	75%	Good candidate to do outside Echelon 1.

Storage Resource Estimates

Actual needs in 2034.

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**

Echelon 1 sites arrive data, two copies
One copy (can and may be more) across Echelon 2 sites for alignment, calibration, and reconstruction use cases.

Computing Resource Estimates

Actual needs in 2034.

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**

See prompt reconstruction.

Roughly 10% of data stream.

Must keep up with data taking; assume 2x headroom.

Reprocessing includes simulation as well as data.

Simply adding together the core counts is an overestimate. Reconstruction core hours used only part time.

Networking Estimates

Echelon 0: The raw data from the ePIC Streaming DAQ (Echelon 0) will be replicated across the host labs (Echelon 1). At the highest luminosity of 1e34, the data stream from the ePIC Streaming DAQ is estimated at 100 Gbit/s. Consequently, Echelon 0 requires an outgoing network connection of at least 200 Gbit/s.

Echelon 1: Each Echelon 1 facility has similar requirements, as it will receive up to 100 Gbit/s of raw data and will share this data with Echelon 2. In addition, Echelon 1 will send a small amount of monitoring data, approximately 1 Gbit/s, back to Echelon 0. Echelon 1 will also receive calibration and analysis data from various Echelon 2 nodes at a comparable rate of about 1 Gbit/s.

Echelon 2: The network connection requirements for Echelon 2 facilities will depend on the proportion of raw data they intend to process. For the 10% of Echelon 1 scenario, a network connection of 20 Gbit/s would be required.

Streaming DAQ and Computing Milestones

Streaming DAQ Release Schedule:

PicoDAQ **FY26Q1**
- Readout test setups

MicroDAQ: **FY26Q4**
- Readout detector data in test stand using engineering articles

MiniDAQ: **FY28Q1**
- Readout detector data using full hardware and timing chain

Full DAQ-v1: **FY29Q2**
- Full functionality DAQ ready for full system integration & testing

Production DAQ: **FY31Q3**
- Ready for cosmics

Streaming Computing Milestones:

Start development of streaming orchestration, including workflow and workload management system tool.
Streaming and Data Challenges: Start streaming and processing streamed data between BNL, Jefferson, DRAC Canada, and other sites.

↔ FY26Q4: **Support of test-beam measurements, using variety of electronics and DAQ setups**:
- Digitization developments will allow detailed comparisons between simulations and test-beam data.
- Track progress of the alignment and calibration software developed for detector prototypes.
- Various JANA2 plugins for reading test-beam data required. Work started on an example.

↔ FY28Q1: **Autonomous Experimentation and Control:** Establish autonomous alignment and calibration workflows that allows for validation by experts.

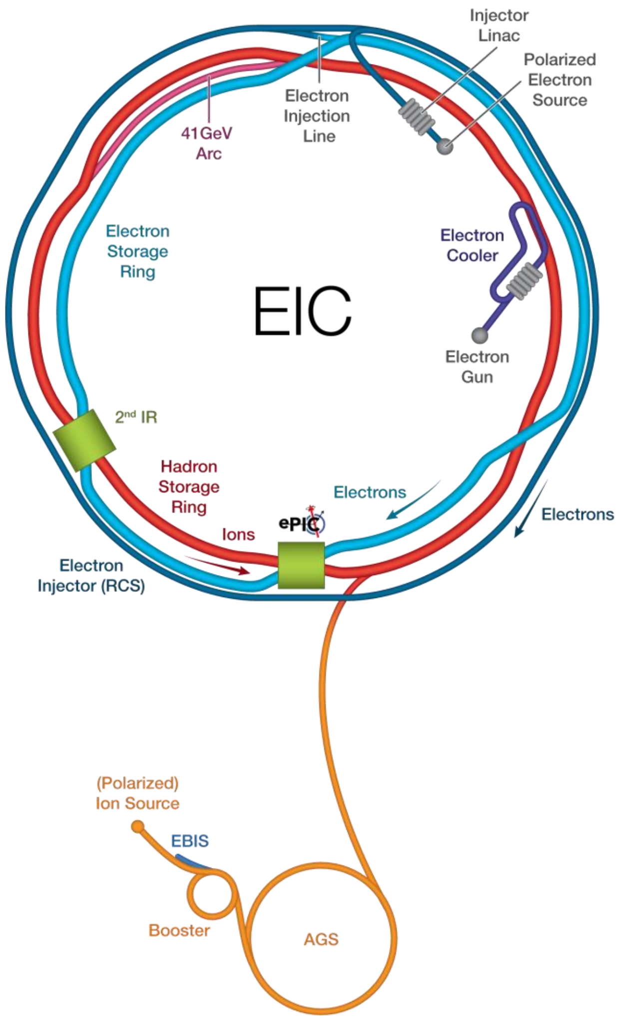
Analysis challenges exercising end-to-end workflows from (simulated) raw data.

↔ FY29Q2: **Streaming challenges** exercising the streaming workflows from DAQ through reconstruction, and the Echelon 0 and Echelon 1 computing and connectivity.
Autonomous Experimentation and Control: Exercising autonomous alignment, calibrations, and validation.

↔ FY31Q3: **Data challenges** exercising scaling and capability tests as distributed ePIC computing resources at substantial scale reach the floor, including exercising the functional roles of the Echelon tiers, particularly Echelon 2, the globally distributed resources essential to meeting computing requirements of ePIC.

Backup

Towards a Quantitative Computing Model: The EIC and Event Rates



- **Versatile machine**: versatile range of beam polarizations, beam species, center of mass energies.
- **High luminosity** up to L = 10^{34} cm^{-2} s^{-1} = 10 kHz/μb.
  - The e-p cross section at peak luminosity is about 50 μb. This corresponds to a signal event rate of about 500 kHZ.
- The **bunch frequency** will be **98.5MHz**, which corresponds to a **bunch spacing** of about **10ns**.
  - For e-p collisions at peak luminosity, there will be in average 200 bunches or about 2μs between collisions (98.5MHz / 500 kHz).
- The EIC Project and ePIC are currently discussing the early science program of the EIC given the phaseout operations with strong hadron cooling (SHC), required for high luminosity:
  - 2034: L = $1x10^{33}$ cm^{-2} s^{-1} = 1 kHz/μb EIC Phase I
  - 2038: L = $4x10^{33}$ cm^{-2} s^{-1} = 4 kHz/μb EIC Phase II
  - 2041: L = $8x10^{33}$ cm^{-2} s^{-1} = 8 kHz/μb SHC
- For the computing resource estimate, we consider the EIC Phase I luminosity scenario of L = 10^{33} cm^{-2} s^{-1} = 1 kHz/μb.

Towards a Quantitative Computing Model: Rate Estimates from MC

Details on rate estimates are available on the Wiki pages of the background task force:
The rate estimates are based on a scenario with a peak luminosity of L = 10^{34} cm^{-2} s^{-1}.

E_e x E_p [GeV x GeV]	5 x 41	5 x 100	10 x 100	10 x 275	18 x 275
Signal Rates					
e-p cross section [μb]	28	35	41	50	54
e-p rates [kHZ]	12.5	129	184	500	83
Background Rates					
e-beam gas rates [kHZ]	2182.0	2826.4	3177.3	3177.3	316.9
p-beam gas rates [kHZ]	12.2	22	31.0	32.9	22.5

Bounds for signal and background event numbers, assuming running 60% up-time for ½ year = 9,460,800 s:

E_e x E_p [GeV x GeV]	Signal Events	Background Events
18 x 275	0.79×10^{12}	3.21×10^{12}
10 x 275	4.73×10^{12}	30.38×10^{12}

Towards a Quantitative Computing Model: Rate Estimates from Streaming DAQ

- **Event size of in average 400 kbit**,
  - Including signal and background apart from detector noise,
  - Assuming that detector noise can be substantially reduced in early stages of processing.
  - Event sizes will decrease in later stages of data taking as detector thresholds are raised.

- **Data rate of in average 30 Gbit/s**,
  - Estimate of upper limit: 10Gbit/s for detector noise + event rate * event size.
  - Event rate = 50 KHz for EIC Phase 1 luminosity and maximum e-p cross section of *50 μb.*

- **Running 60% up-time for ½ year = 9,460,800 s**:
  - Data rate of 30 Gbit/s results in 710 × 10^9 events per year.
  - The data volume of 35.5 PB per year will be replicated between Echelon 1 facilities (71 PB in total).

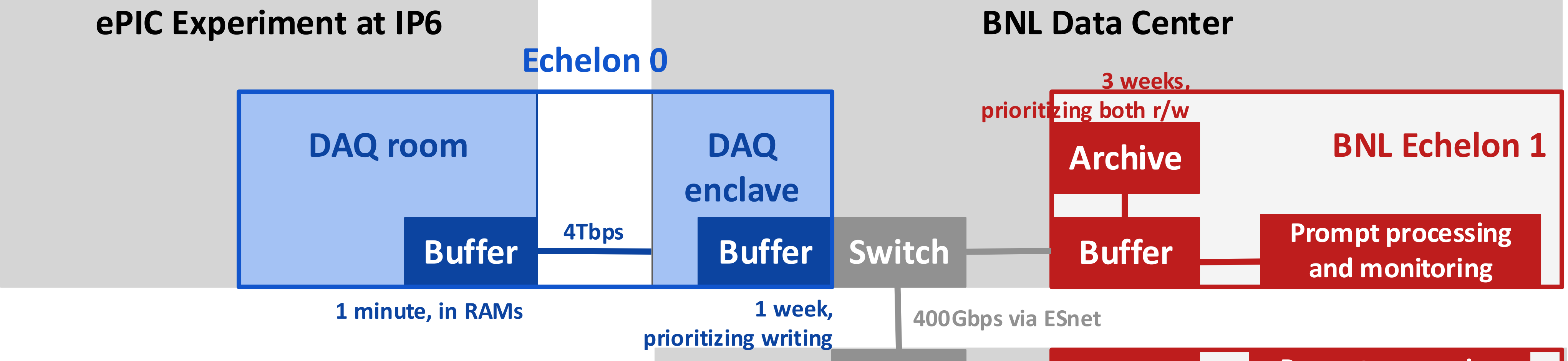
Towards a Quantitative Computing Model: Reconstruction and Simulation

Reconstruction and Simulation Times	Times based on current software on modern cores
Reconstruction event processing time with background **[s]**	2
Reconstruction algorithmic speedup factor 10yrs out	1.5
Simulation event processing time with background **[s]**	15
Full simu speedup factor 10yrs out	1.5
Combined time with background, with speedup **[s]**	11

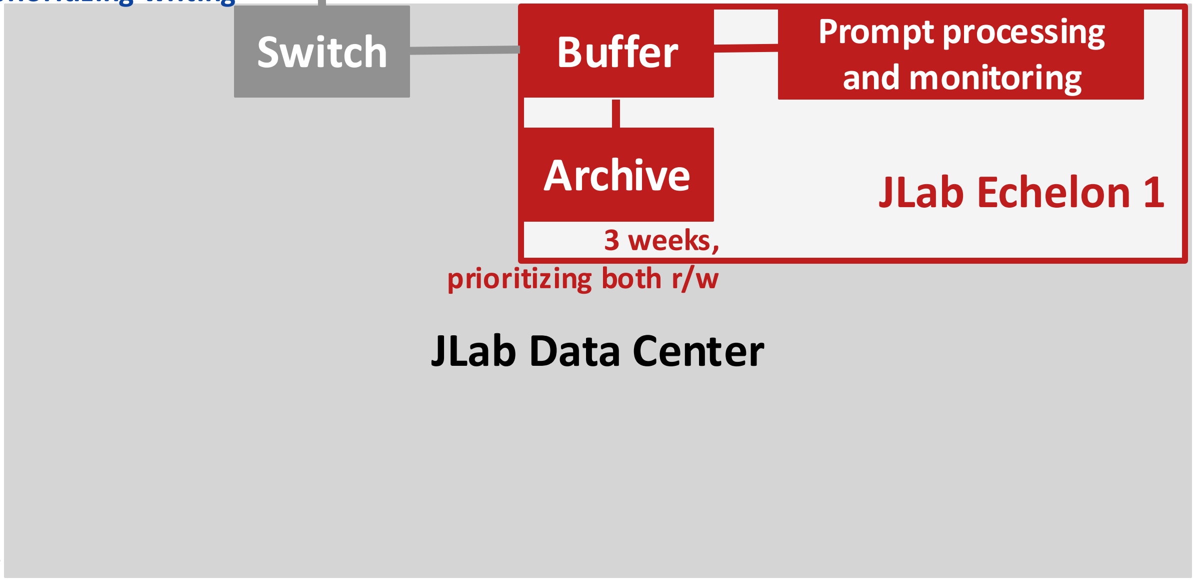
Simulation Use Cases		
Number of simulated events per event of interest	10	The canonical 10x more.
Optimized simu events per physics event	4	~40% of measured events will be signal.
Fast simulation speedup relative to full simulation	4	
Proportion of simulation events using fast simulation	70%	

Data Transfer Between Echelon 0 and Echelon 1

Buffers in DAQ and **Echelon 1** sites ensure latency tolerance to avoid deadtime, smooth streaming operation and robustness against data flow interruptions.



- Echelon 0 extends across the IP6-SDCC fiber to the DAQ enclave.
- **Deliver the capability for Echelon 1 sites as symmetric peers.**
- Foremost E1 responsibilities: archiving the stream, and prompt processing/monitoring, both consuming the incoming stream via the buffer:
  - Delivers two geographically separated raw data copies.
  - Uses ePIC distributed computing capabilities supporting the E0/1/2/3 Streaming Computing Model.
  - Will be up to the ePIC collaboration together with sites to determine the E1 roles in detail.



What Data is Being Transferred Between Echelon 0 and 1?

- The **definition of raw data** is up to ePIC and to-be-defined in the ePIC Data Management Plan.
- It is the **data that flows**, during data taking, **out of the Streaming DAQ** (Echelon 0) to the switch and is distributed to the two Echelon 1 sites:
    - When ePIC starts, and for at least *n* years, the data stream arriving from the Streaming DAQ will be archived in full, untouched.
    - In year *n+1* ePIC may be confident enough to do immediate processing at the two Echelon 1 sites to reduce the data before archiving, i.e. archive only events of interest.
    - Regardless, our responsibility is to deliver a system designed to archive 100% of the Echelon 0 stream.
- There will be other data in the stream with continuous relevance during accelerator and detector operations that will stream continuously but is not part of the raw data, e.g. slow control monitoring or collider-experiment feedback that will be archived in databases.

- **Data reduction in the Streaming DAQ will be strictly limited**:
    - From the beginning, irreversible data reductions will be recorded for event subsamples, to develop and debug.

- The event data stream is in the form of **timeframes of 0.6ms** (defined by 2^{16} cycles of the EIC Clock):
    - DAQ inserts file and run markers into the stream.
    - It is files that hit the switch, the full dataset delivered identically to the two Echelon 1 sites.
    - Time frames will be aggregated into **supertimeframes**, e.g. ~1000 TFs, corresponding to .6sec, ~2GB.
    - Files should be identical at the two E1s (as in the scheme discussed up to now, with DAQ inserting file markers),

Support of Test-Beam Measurements

- **Goal**: CODA and rcdaq will be used for detector test-beam measurements. We want to make sure that the data from CODA and rcdaq can be processed in EICrecon.
- **Tasks**:
  - Example for conversion from detector electronics data to ePIC data model / podio.
  - Define interface, e.g., using podio or working with CODA and rcdaq formats.
  - Implement interface in EICrecon (JANA2 plugins) and/or CODA and rcdaq.

- **Goal**: Track progress of the alignment and calibration software developed for detector prototypes.
- **Tasks**:
  - Define how we track alignment and calibration software developments.
  - Implement calibration example in EICrecon to show how EICrecon can be used for calibrations.

- **Goal**: Digitization developments will allow detailed comparisons between simulations and test-beam data.
- **Task for Physics and Detector Simulation WG.**

Streaming Orchestration

- **Goal**: The distributed workflow and workload management for streaming data will keep us busy for many years. We need to define how we approach this problem.
- **Tasks**:
  - Requirements document for workflow and workload management for streaming data at Echelon 1 and 2.
    - Torre will start requirements document on Overleaf.
    - Initial authors: Francesco, Jeff, Markus, Taku, and Torre.
    - Reviewers: David.
    - Discuss first version on March 18
  - Meetings on lessons learned from experiments, e.g., GRETA (Markus) and Vera C. Rubin Observatory (Torre), and existing tools, e.g., ERSAP (David), ALICE/O2 (Markus), PanDA (Torre), XRootD (Torre).
  - Define and setup testbeds for streaming orchestration, initially based on simulations
    - Depends on first version of requirements document, first discussion on test beds on March 18
  - Discuss opportunities for realistic test cases.

Streaming Challenges (E0+E1)

- **Goal**: We will develop and improve the streaming workflows from DAQ through reconstruction, and the Echelon 0 and Echelon 1 computing and connectivity over the coming years. We need to define how we approach this problem.
- **Tasks**:
  - **Testbed for event reconstruction from streamed data** (a task also referred as event building).
    - Proposed next steps after the first successful prototype.
    - Documentation on how to run JANA2 in event and streaming mode.
    - Working example on how to change in JANA2 between timeframe and event level.
    - Simple algorithm that separates signal from background events and demonstrated how we will reconstruct physics events in EICrecon. Developing and optimizing these algorithms for efficiency and performance will keep us busy over the next decade.
    - Workforce: Nathan, Taku, Takuya.
  - **Test bed for supertimeframe building**:
    - Work with Physics and Detector Simulation WG on timeframes.
    - Develop tool for aggregating these timeframes.
    - Exercise the timeframe building and processing in Rucio.
    - Exercise the timeframe building and processing in object store.
    - Write requirements document based on initial testbed.
    - Workforce: Unknown.
  - Lessons learned from sPHENIX (Joe), after Quark Matter 2025

Streaming Challenges (E0+E1)

- **Goal**: We will develop and improve the streaming workflows from DAQ through reconstruction, and the Echelon 0 and Echelon 1 computing and connectivity over the coming years. We need to define how we approach this problem.
- **Tasks**:
  - **Testbed for event reconstruction from streamed data** (a task also referred as event building).
    - Proposed next steps after the first successful prototype.
    - Documentation on how to run JANA2 in event and streaming mode.
    - Working example on how to change in JANA2 between timeframe and event level.
    - Simple algorithm that separates signal from background events and demonstrated how we will reconstruct physics events in EICrecon. Developing and optimizing these algorithms for efficiency and performance will keep us busy over the next decade.
    - Workforce: Nathan, Taku, Takuya.
  - **Test bed for supertimeframe building**:
    - Work with Physics and Detector Simulation WG on timeframes.
    - Develop tool for aggregating these timeframes.
    - Exercise the timeframe building and processing in Rucio.
    - Exercise the timeframe building and processing in object store.
    - Write requirements document based on initial testbed.
    - Workforce: Unknown.
  - Lessons learned from sPHENIX (Joe), after Quark Matter 2025

Distributed Data Challenges (E1+E2)

- Discussion on deliverables and milestones deferred to later this year.

Analysis Challenges (E3, using services at E1+E2)

- Discussion on deliverables and milestones deferred to later this year.

Autonomous Experimentation and Control (Alignment, Calibration, Validation)

- **Goal**: Establish autonomous alignment and calibration workflows that allows for validation by experts.
- **Tasks**:
  - **Discuss status of documentation on alignment and calibration workflows**:
    - Do we have any gaps?
    - How to share the information with the collaboration?
    - Have we demonstrated that – based on current understanding of workflows and their dependence – alignment and calibration workflows can be done in less than two weeks?
    - Workforce: Unknown.
  - **Meetings on lessons learned on autonomous calibrations, taking our unique requirements into account:**
    - Experiments: ALICE (Chiara Zampolli), LHCb (Markus), GlueX (Torri), sPHENIX (Joe)
    - Timescale depends on documentation on ePIC alignment and calibration workflows
  - **Example - Alignment**
    - Framework within Acts exists to apply alignment calibrations on the fly. Needs some "hooks" developed in EICrecon
    - Nontrivial effort needed to decide on scheme, however. Will we use millepede? Acts has initial development of alignment software, but not close to production ready. Do we develop something ourselves, or put effort into Acts algorithm?
      - In any case, no software currently exists to prepare our data for any alignment minimization scheme
    - Can use alignment as a calibration example for advocating for FTEs. Much work needed to develop software let alone making it autonomous