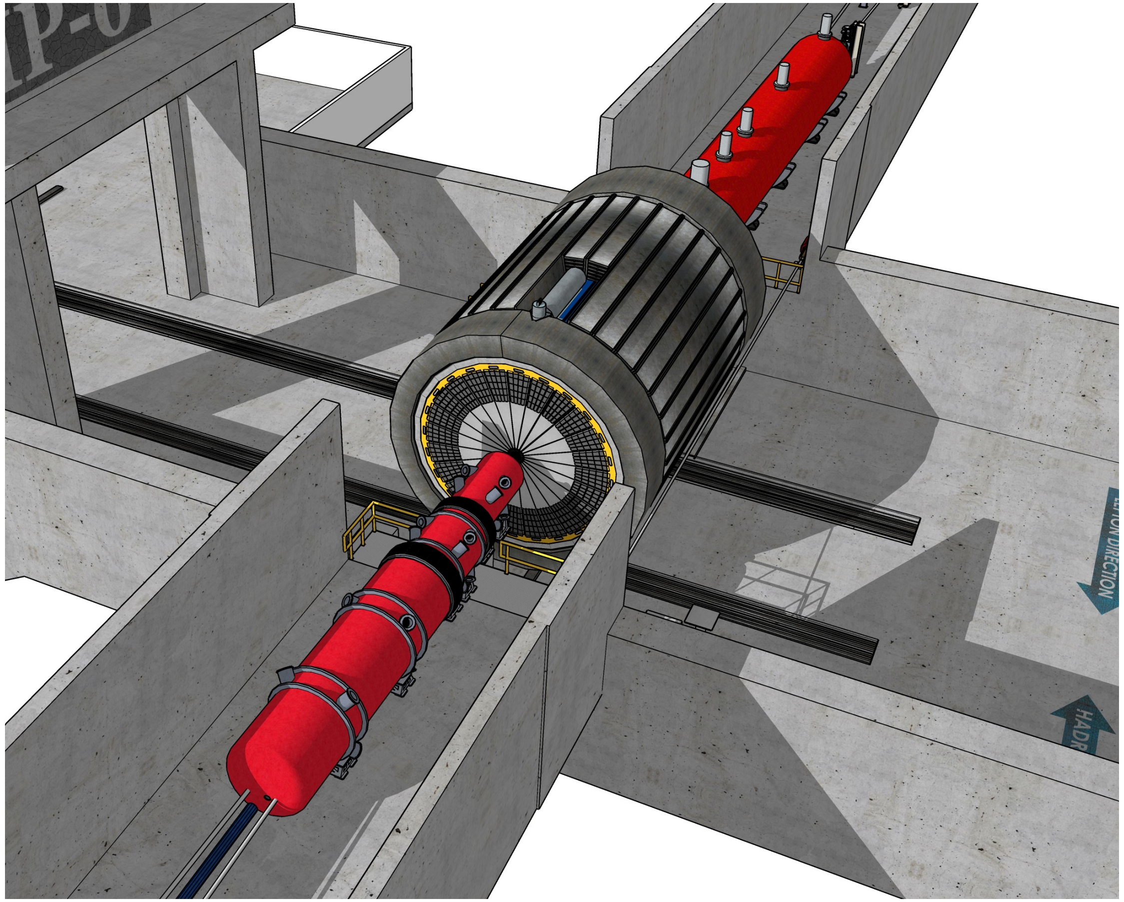
ePIC Streaming Computing Model

Markus Diefenthaler (Jefferson Lab)

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 computing.

Advantages of Streaming Readout
- Simplification of readout (no custom trigger hardware and firmware) 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.
  - 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 for rapid processing.

Heterogeneous computing for acceleration (CPU, GPU).

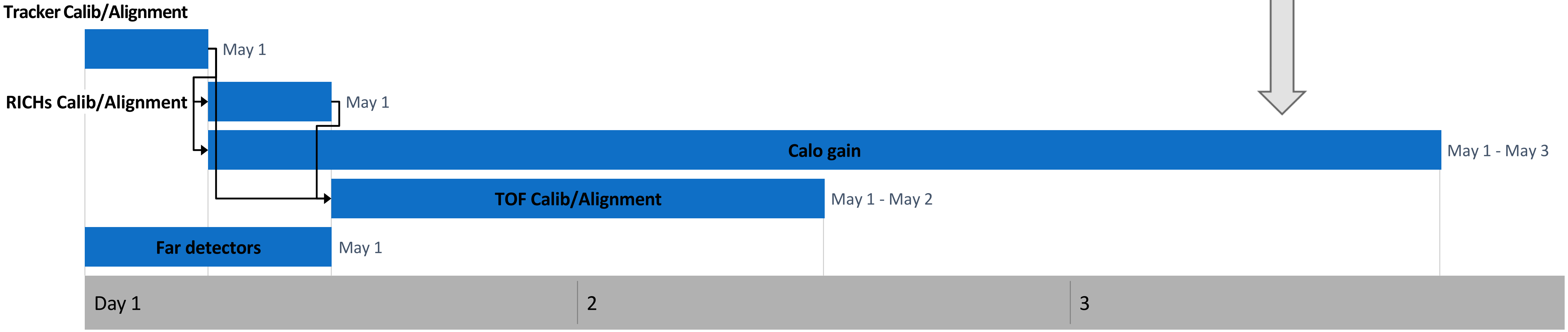
Alignment and Calibration Planning

Spreadsheet

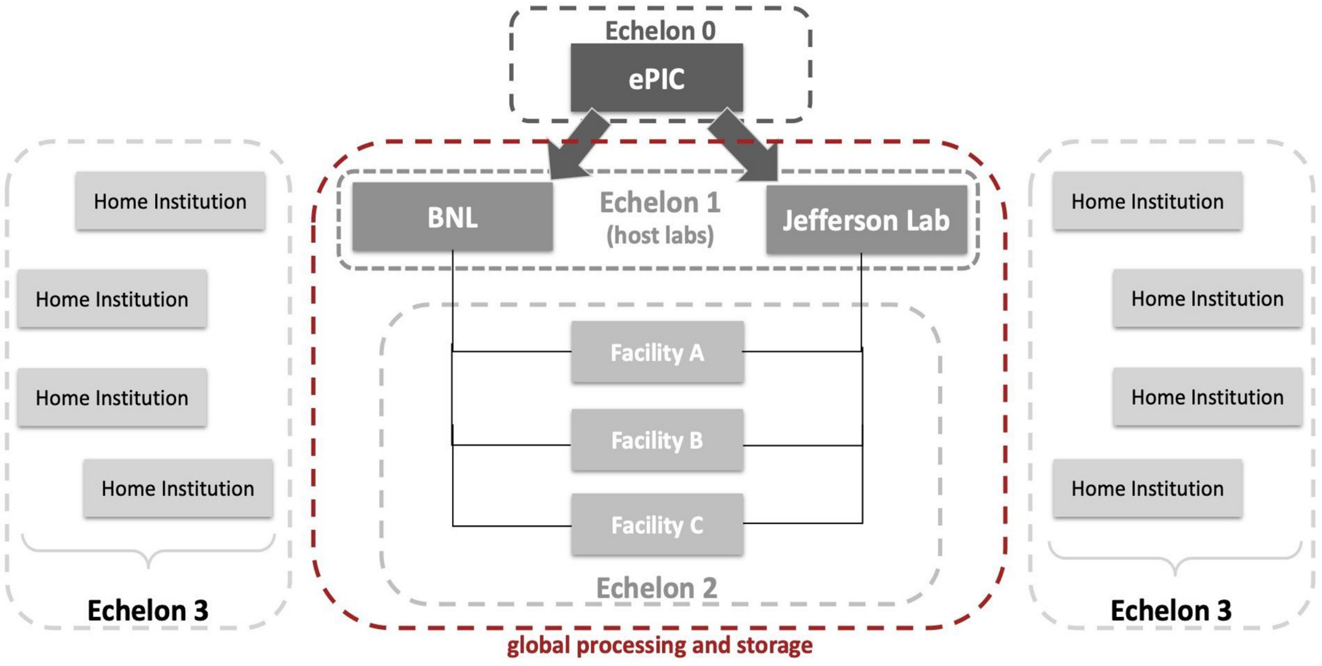
Alignment and Calibration of the ePIC Detector

- Series of meetings with detector experts to discuss alignment and calibration procedures and requirements for each detector subsystem.
- Summarized in a spreadsheet outlining alignment and calibration workflows.

Subsystem	Region	Pre-physics-operation calibrations (Cosmic, no-beam calibration, commissioning)	Steady State calibrations: aim to produce final reconstruction-ready calibration within few days of physics data taking in a conti							
			Task	Human intervention ?	Data Needed	Dependecy	T0 + 12hr	T0 + 24hr	T0 + 36hr	T0 + 4
MAPS	Barrel+Disk	Threshold Scan / ALICE=20min Fake rate scan/noisy pixel masking	(See Alignment)							
MPGD	Barrel+Disk	?	?							
bTOF, eTOF (ac-lgad)	Barrel/Forward	Bias voltage determination ASIC baseline, noise, threshold Clock sync Time walk calibration	Gain calibration TDC bin width determination Clock offset calibration Hit position dependency (intrinsic and c-by-c)	QA	High p tracks ~1hr of production data?	Tracking, pfRICH	Data Acc. Dependen	Dependen	Processin	Proces
Central Detector Tracker Alignment		Initial alignment	Alignment Check/Update (if needed)	QA	Prodcution data		Processing			
pfRICH	Backward	Thresholds (noise dependent), dynamic range adjustments, timing offsets, synchronization Initial alignment	Alignment Check/Update (if needed) Time dependencies (Aerogel transparency, mirror reflectivity, Gas pressure)	?	Prodcution data		Data Acc.	Processing		
DIRC	Barrel	Laser data?	?	?						
dRICH	Forward	Bunch timing offset scan Threshold scan Noise masking	Track based alignment	?	High p tracks ~1hr of of production data?	Tracking	Data Acc. Dependen	Processin	Processing	
bEMC	Backward	Cosmic and LED for the initial gain balancing	DIS Electron Pi0->gg events energy scale	QA	DIS electron Pi0 di-photon resonance ~1 day of production data	Tracking	Data Acc. Dependen	Data Acc.	Processin	Proces
AstroPix	Barrel									
ScifiPb	Barrel		SiPM gain		?					
fEMC	Forward	IV Scan	Pi0, eta->gg events energy scale Second iteration pi0 (if needed)	QA	Pi0 di-photon resonance ~1 day of production data		Data Acc.	Data Acc.	Processin	Proces
bHCAL	Backward	LED	?							
cHCAL	Barrel	MIP calibration Gain calibration	(See hadronic e-scale calib)							
fHCAL	Forward									
fHCAL insert	Forward									
Hadronic energy scale calibration		?	Set full calo stack energy scale for hadroinc shower and jets	?	High energy hadronic showers and jets	Tracking h-PID	Data Acc. Dependen	Data Acc. Dependen	Data Acc. Dependen	?
low Q2 Tagger	Far Backward	Alignment?								
low Q2 Tagger (CAL)	Far Backward									



The ePIC Streaming Computing Model



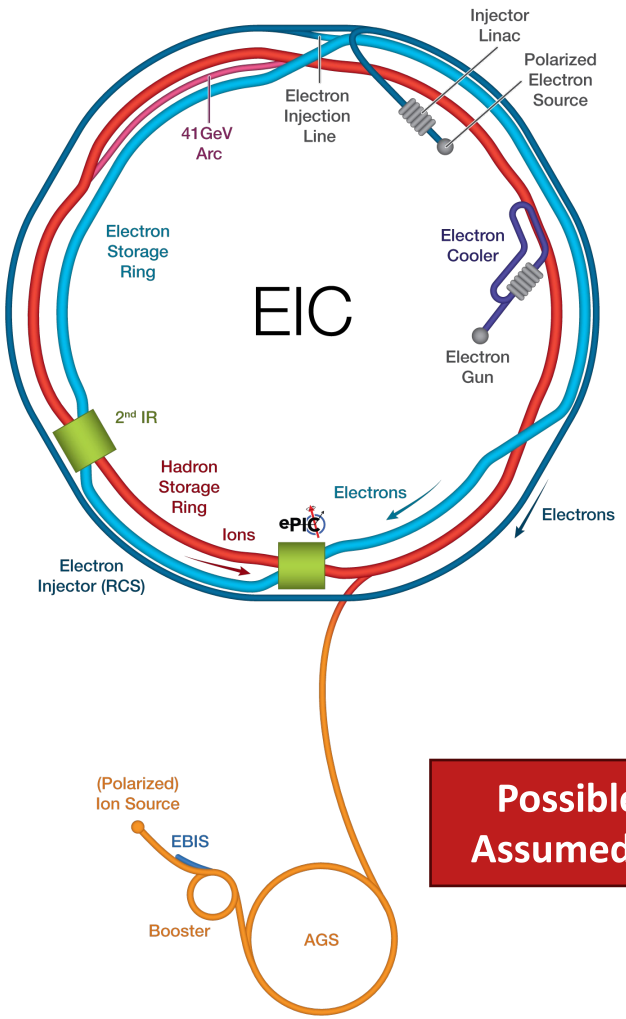
Echelon 0: ePIC Streaming DAQ.

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.

Towards a Quantitative Computing Model: The EIC and Event Rates



Possible Scenario, Assumed in Estimate

- **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 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).

Computing Use Cases and Their Echelon Distribution

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		✓	✓	

Prompt := rapid low-latency processing.

Prompt processing of newly acquired data typically begins in seconds, not tens of minutes or longer.

Assumed Fraction of Use Case Done Outside Echelon 1	
Alignment and Calibration	50%
First Full Reconstruction	40%
Reprocessing	60%
Simulation	75%

- **Echelon 1** sites uniquely perform the **low-latency streaming workflows** consuming the data stream from Echelon 0:
  - Archiving and monitoring of the streaming data, prompt reconstruction and rapid diagnostics.
- Apart from low-latency, Echelon 2 sites fully participate in use cases:
  - Tentative resource requirements model assumes a **substantial role for Echelon 2**.

Computing Resource Needs and Their Implications

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**

Computing Resource Needs in 2034 for EIC Phase I

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.

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

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.

Technology Choices and Evolution

- **Modularity is Key**: We will have a modular software design with structures robust against changes in the computing environment so that changes in underlying code can be handled without an entire overhaul of the structure.
- **Lessons Learned from the NHEP Community** informed the ePIC Streaming Computing Model:
  - Our software is deployed via containers.
  - Our containers are distributed via CernVM-FS.
  - We run large-scale simulation campaigns on the Open Science Grid.
  - Access to our simulations is facilitated through XRootD.
  - We are in the process of deploying Rucio for distributed data management, improving access for collaborators to specific simulation files.
- **Engagement in Advanced Scientific Computing Discussions,** including:
  - DOE SC ASCR's Integrated Research Infrastructure (IRI) program. Data-integration-intensive and time-sensitive patterns highly relevant for ePIC.
  - DOE SC ASCR's High Performance Data Facility (HPDF) project, not only enabling these patterns but also potential partnership on data and analysis preservation.
  - HEP Software Foundation's discussions on analysis facilities, analysis use cases, and analysis infrastructure.
- **Software Stewardship by the NHEP Community**:
  - Participation in workshop committee for the *“Software Infrastructure for Advanced Nuclear Physics Computing.”*
  - Engaging in discussions about HSF Affiliated Projects and Software.
- **Data and AI**:
  - AI has a strong presence in ePIC. Dmitry will outline our initiatives to integrate existing AI solutions into our production workflows.
  - We will help with guiding the DOE SC Round Table on *“Transformational Science Enabled by Artificial Intelligence,”* which will shape ePIC's approach to leveraging AI.

The Role of AI

- **Compute-detector integration** using**:**

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

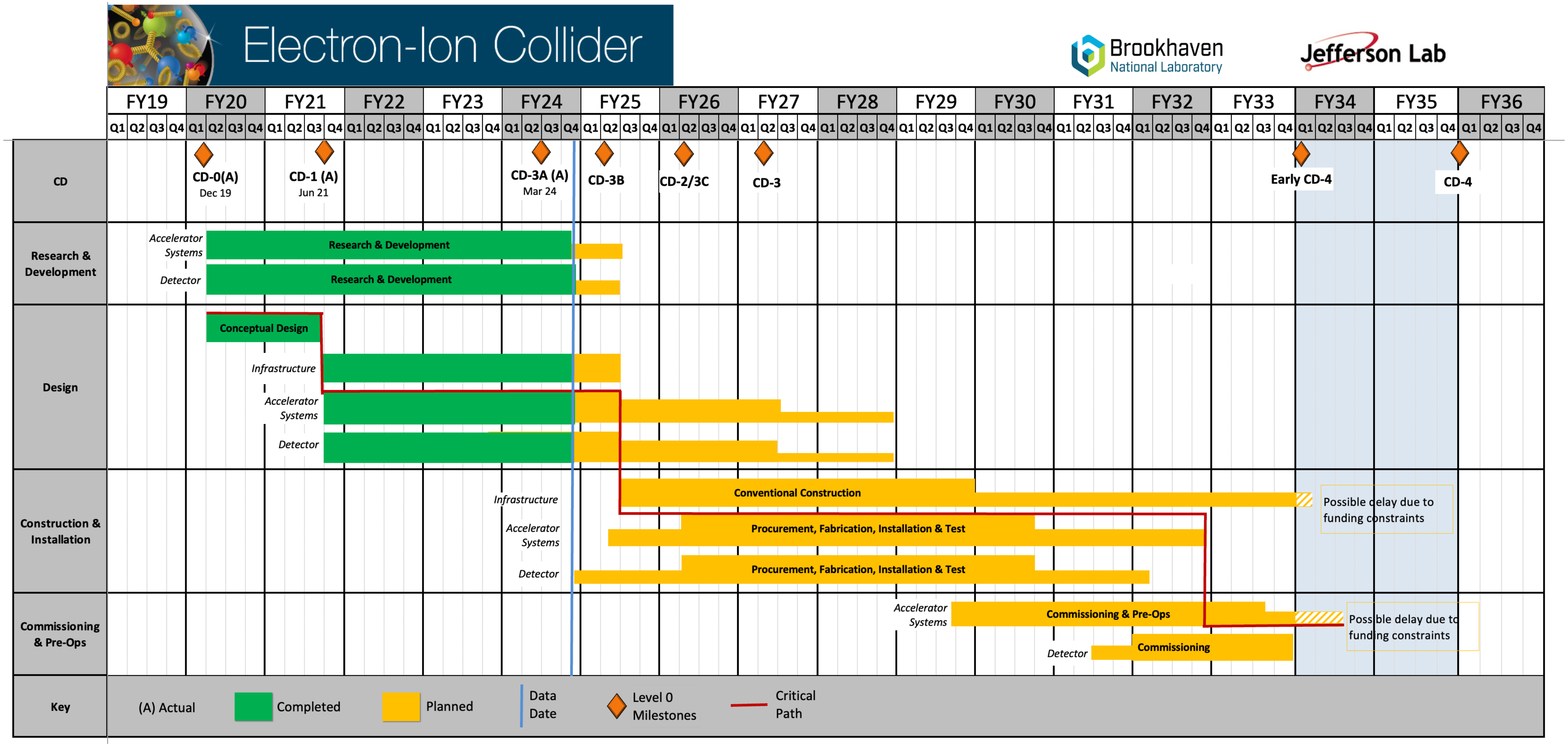
AI for autonomous alignment and calibration as well as reconstruction and validation for rapid processing.

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.

High Level Installation Schedule



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.
Start streaming and processing streamed data between BNL, Jefferson, DRAC Canada, and other sites.

Support of test-beam measurements, using variety of electronics and DAQ setups: (FY26Q4)
- 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.

Establish autonomous alignment and calibration workflows that allows for validation by experts. (FY28Q1)

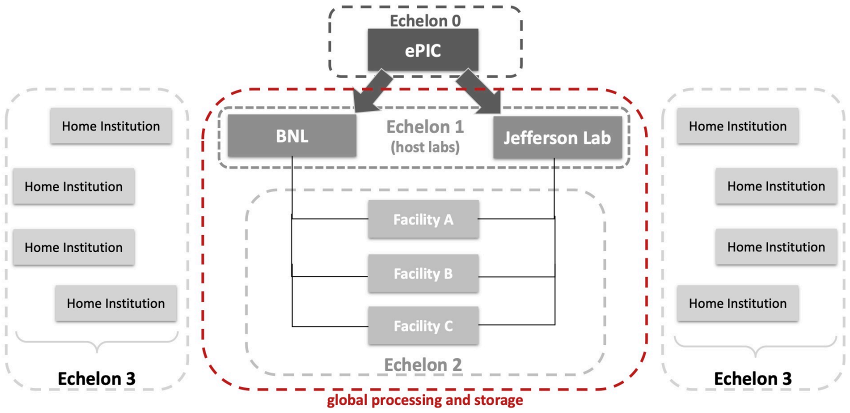
Analysis challenges exercising end-to-end workflows from (simulated) raw data.

Streaming challenges exercising the streaming workflows from DAQ through offline reconstruction, and the Echelon 0 and Echelon 1 computing and connectivity. (FY29Q2)
Analysis challenges exercising autonomous alignment and calibrations.

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. (FY31Q3)

Summary

- **Streaming Readout of the ePIC Detector to maximize and accelerate science:**
  - ePIC aims for **rapid turnaround of 2-3 weeks for data for physics analyses**.
  - Timeline driven by alignment and calibration.
- **Four tiers of the ePIC Streaming Computing Model computing fabric:**
  - **Echelon 0**: ePIC experiment and its streaming readout.
  - **Echelon 1**: Crucial and innovative partnership between host labs.
  - **Echelon 2**: Essential global contributions.
  - **Echelon 3**: Full support of the analysis community.

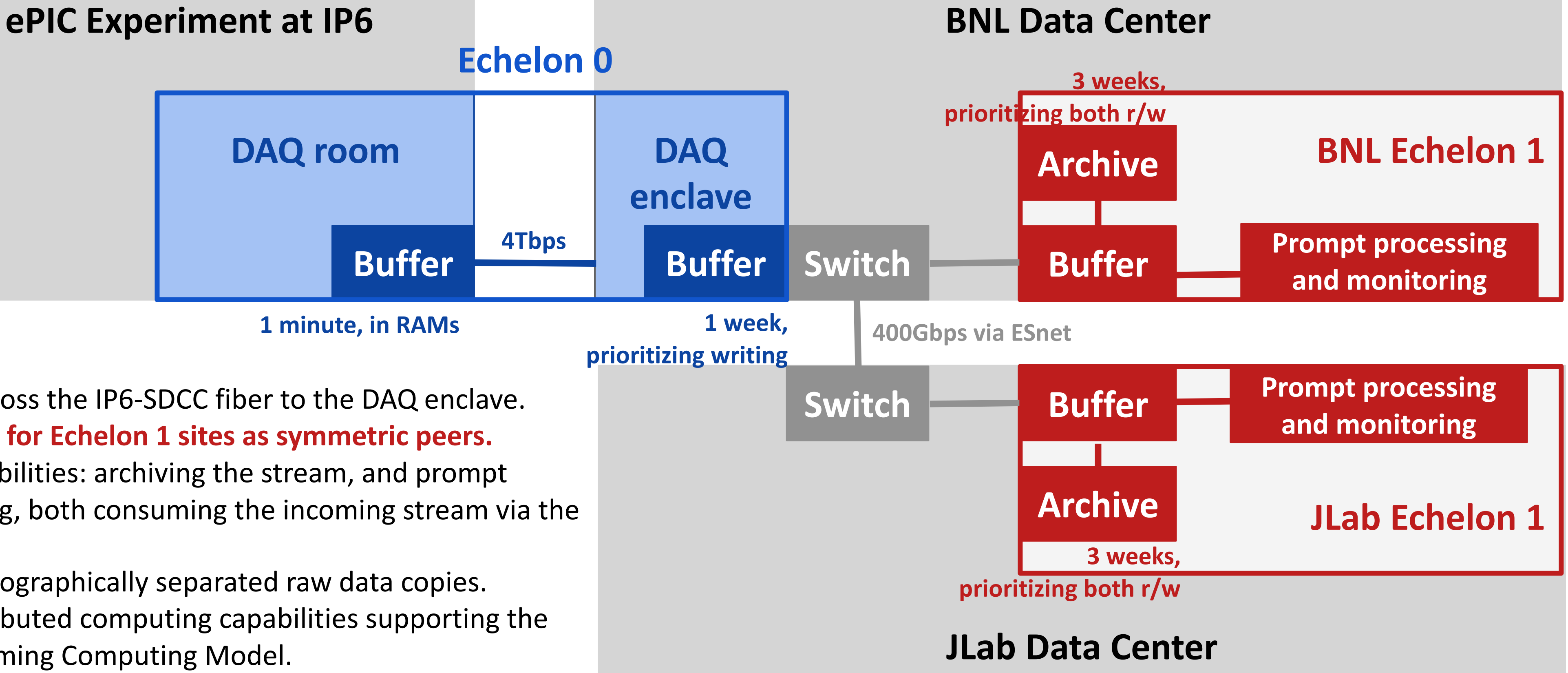


- **High level milestones** ensure that the agile development process is continuously confronted with real world exercising of the software and the developing realization of the computing model:
  - Priority always given to meeting near-term needs. Dmitry will report on how ePIC leverages monthly production campaigns, CI-driven benchmarks, and timeline-based prioritization to **ensure timely completion of the simulation studies for the TDR**.
  - **Longer range timeline progressively exercising the streaming computing** model to deliver for the needs of the CD process, for specific applications, e.g. test beams, for scaling and capability challenges, and ultimately for the phases of data taking.

Backup

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?

Sec. 3.11

- 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.

- **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.

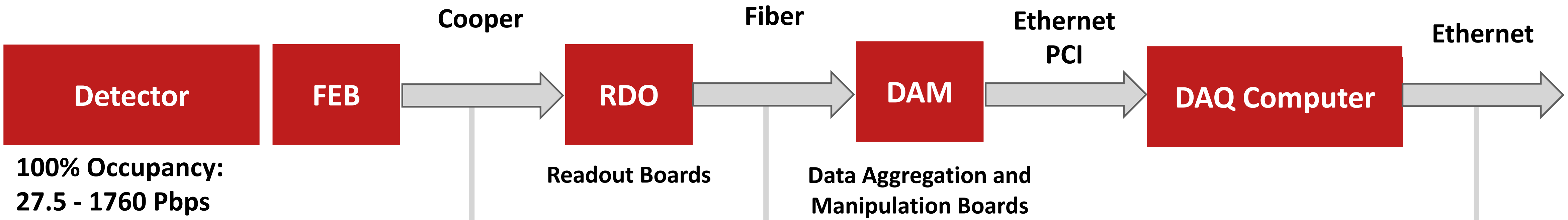
- The event data stream is in the form of **timeframes of 0.6ms** (defined by 2^{16} cycles of the EIC Clock):
  - DAQ inserts file and run markers into the stream.
  - It is files that hit the switch, the full dataset delivered identically to the two Echelon 1 sites.
  - Optimal file granularity is TBD.
  - Consistent with low-latency real-time processing, storage system efficiency, etc.

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%	

Streaming DAQ: Expected Worst-Case Data Rate Contributions



Aggregate	2.0 Tbps
Signal + Background	400 Gbps
Noise	1.6 Tbps

Aggregate	2.0 Tbps
Per RDO (Avg.)	0.7 Gbps

Aggregate	96 Gbps
Signal	38 Gbps
Synchrotron Radiation	0.1 Gbps
Electron Beam	22 Gbps
Hadron Beam	4 Gbps
Noise	32 Gbps

Data reduction limited to **low-level** (e.g., zero suppression) and **implemented only as necessary**.
The impact of data reduction on systematic uncertainties must be fully understood.

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.

The Role of AI

- **Compute-detector integration** using**:**

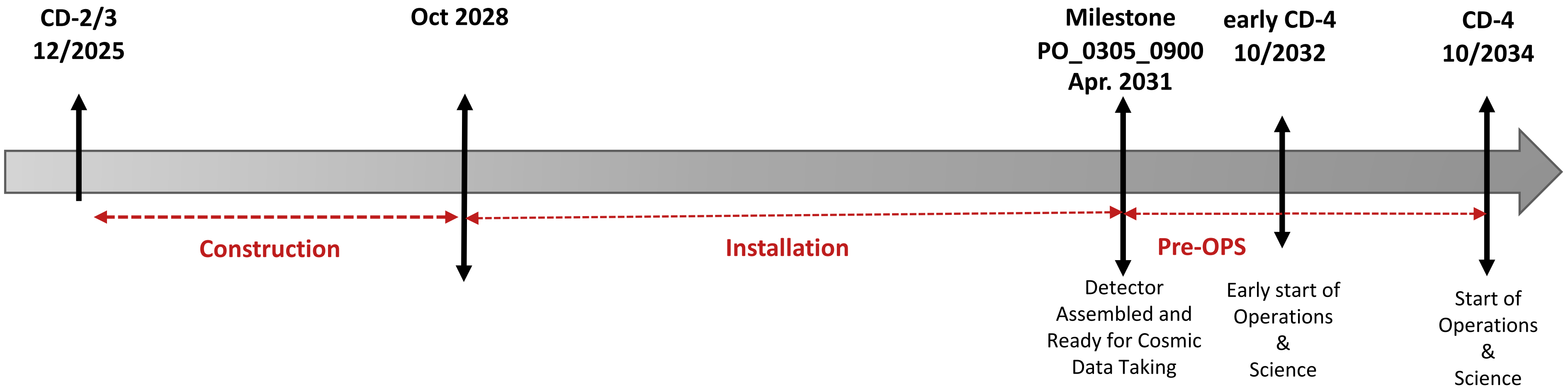
Streaming readout for continuous data flow of the full detector information.	**AI** for autonomous alignment and calibration as well as reconstruction and validation for rapid processing.	**Heterogeneous computing** for acceleration.

- 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.

Milestones Prior to CD2/3 and TDR

Sec. 9.2.1



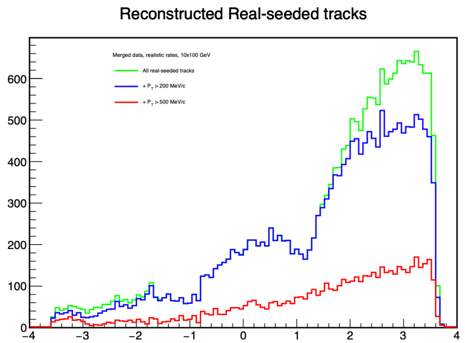
- **Software and simulation readiness for TDR and subsequent EIC Project phases**:
  - Prototype for streaming reconstruction: Update in the following slide.
  - Prototype on alignment and calibration workflow: Work in progress.
  - Dmitry will report on TDR software and simulation readiness.
- **Quantitative Computing Model**:
  - **This talk in this review**: Overview and current status.
  - **Report on "ePIC Streaming Computing Model"**: Under revision, see **sections highlighted in green**.
  - **Summary and publication plans**:
    - Summary to be included in the TDR.
    - Full version intended for publication in Computing and Software for Big Science.

Prototype of Event Reconstruction from Realistic Timeframes

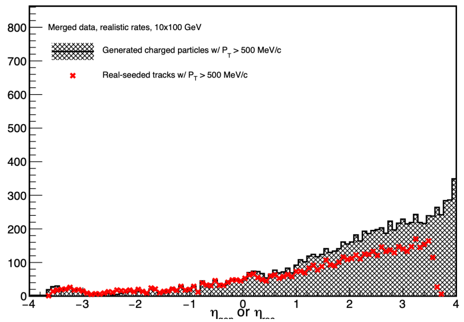
Scope of the first prototype: Track reconstruction only. Demonstrated that we can correlate hits in a realistic time frame to the various events in the time window of the MAPS of 2µs.

Streaming DAQ → **Timeframes** → **Streaming Computing** → **Reconstructed Events**

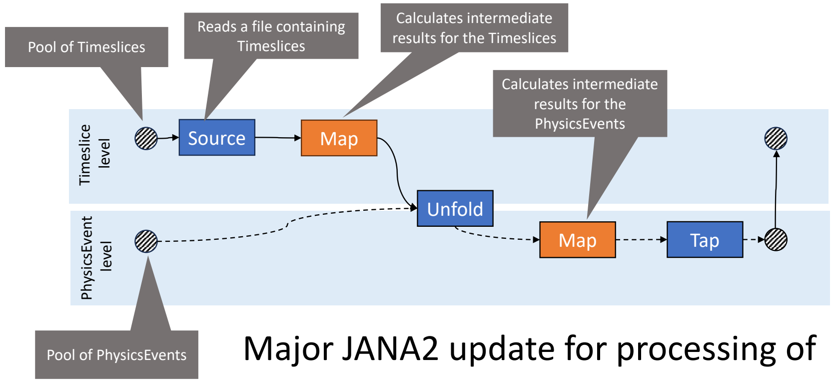
0.6ms timeframes



Preliminary results by Barak Schmookler (UCR).



- Reached consensus on composition of *realistic* time frames in terms of signal and background.
- Started to implement infrastructure for building timeframes instead of events in detector simulations. Needs to be changed to post-Geant4 and post-digitization.



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

Streaming Data Processing

Traditional Workflow Characteristics in NP and HEP Experiments:

- Data is acquired in online workflows.
- Data is stored as large files in hierarchical storage.
- Offline workflows process the data, often with substantial latency.
- Batch queue-based resource provisioning is typical.
- Key features: discrete, coarse-grained processing units (files and datasets) and decoupling from real-time data acquisition.

ePIC Streaming Data Processing Characteristics

- Quasi-continuous flow of fine-grained data.
- Dynamic flexibility to match real-time data inflow.
- Prompt processing is crucial for data quality and detector integrity.
- Processing full data set quickly to minimize time for detector calibration and deliver analysis-ready data.

Challenging Characteristics of Streaming Data Processing:

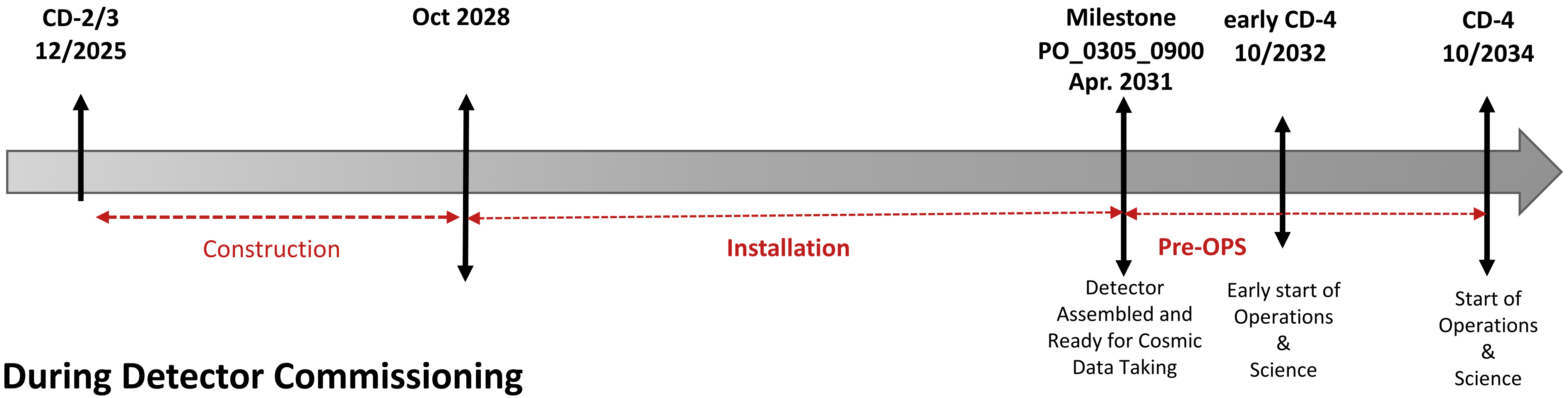
- **Time critical**, proceeding in near real time.
- **Data driven**, consuming a fine-grained and quasi-continuous data flow across parallel streams.
- **Adaptive and highly automated**, in being flexible and robust against dynamic changes in data-taking patterns, resource availability and faults.
- **Inherently distributed** in its data sources and its processing resources.

Assumptions for Infrastructure:

- Existing batch-style processing likely to remain.
- Dynamic processing, e.g. Kubernetes, may displace the batch model.
- Design the system for both batch and dynamic processing to ensure resilience against technology evolution.
- Accommodate but effectively hide these underlying infrastructure characteristics.

Milestones During Detector Commissioning and Early Datataking

Sec. 9.2.3



Milestones During Detector Commissioning

- This phase has unique expectations and requirements compared to steady-state operation:
  - Utilization of semi-triggered data-taking modes.
  - Initial calibrations.
  - Gradual extension of first pass processing from Echelon 1 to Echelon 2.
- Careful planning of software & computing efforts and leveraging experience from data and analysis challenges during the detector construction phase essential.

Milestones During Early Datataking Phase

- Simpler and more conservative approaches will be used during initial data-taking phase.
- The ePIC Streaming Computing Model will be gradually deployed and validated.