The ePIC Computing Model and Calibrations

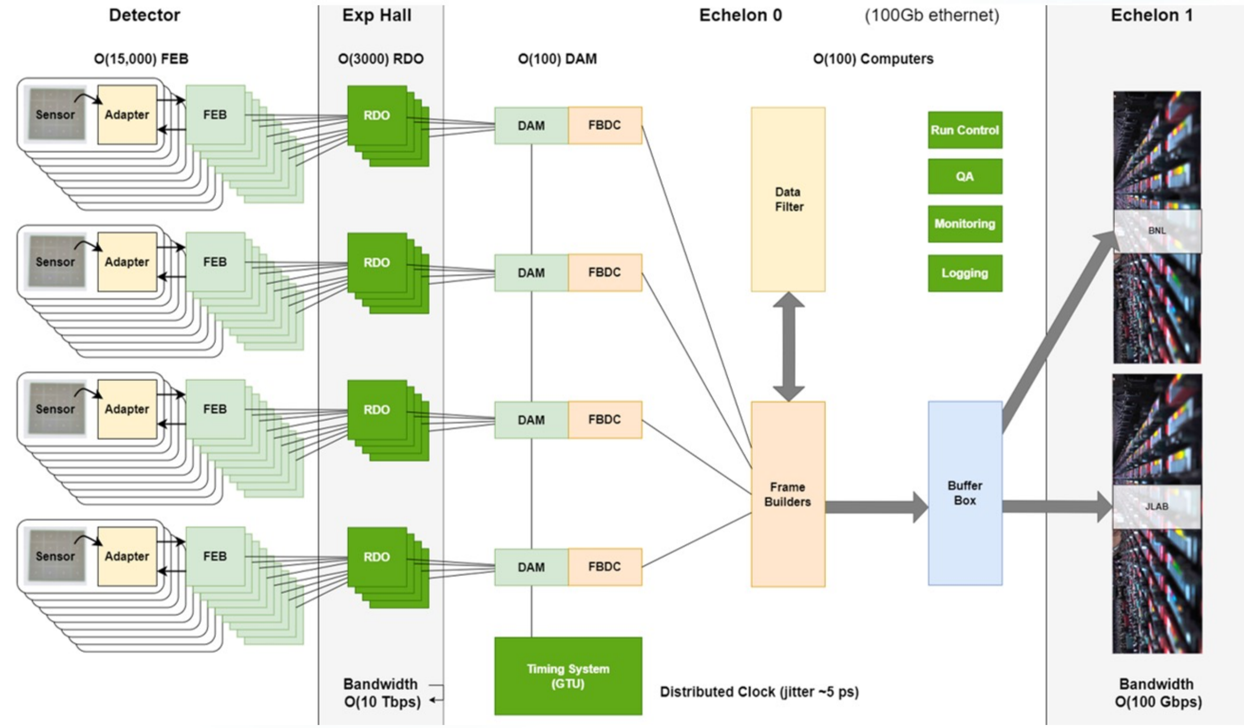
Marco Battaglieri (INFN-Genova), Markus Diefenthaler (JLab),
Taku Gunji (QNSI/U-Tokyo), Jeff Landgraf (BNL), Torre Wenaus (BNL)

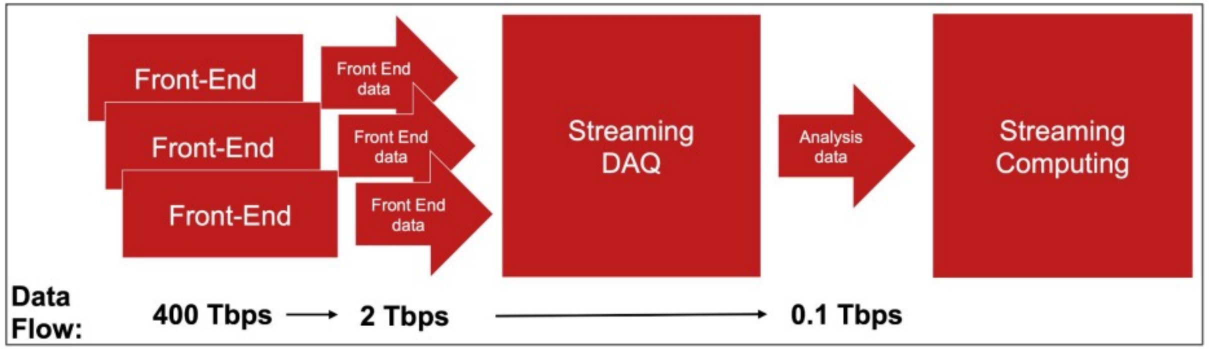
Outline

- ePIC Streaming DAQ and Computing
- Recent SRO WG activities
- Streaming Calibration and Alignment
- Milestones and needs of subsystem engagement

ePIC Streaming Readout and Computing

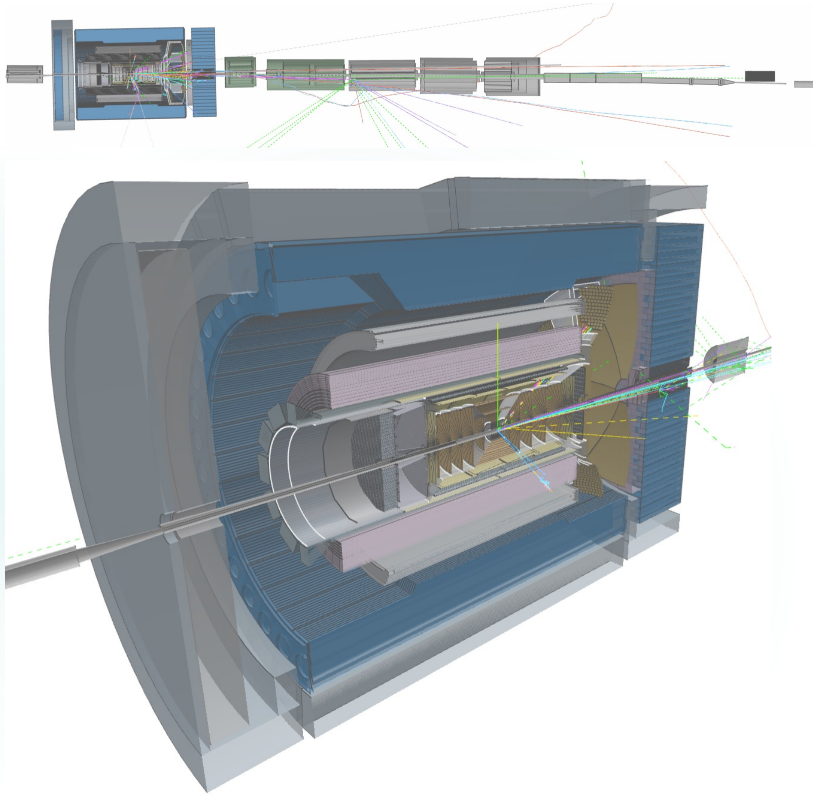
- All raw data (collision event + substantial background) from FEB is streamed continuously and streamed data is inspected by prompt holistic reconstruction to identify physics events.
- This needs seamless integration between subsystem readout, DAQ, and Computing.
  - Data filtering (ex, noise reduction), frame-building, reconstruction, calibration, analysis, monitoring
- Target : Rapid turnaround of 2-3 weeks for data for physics analyses
  - This turnaround time is constrained by the calibration timescale

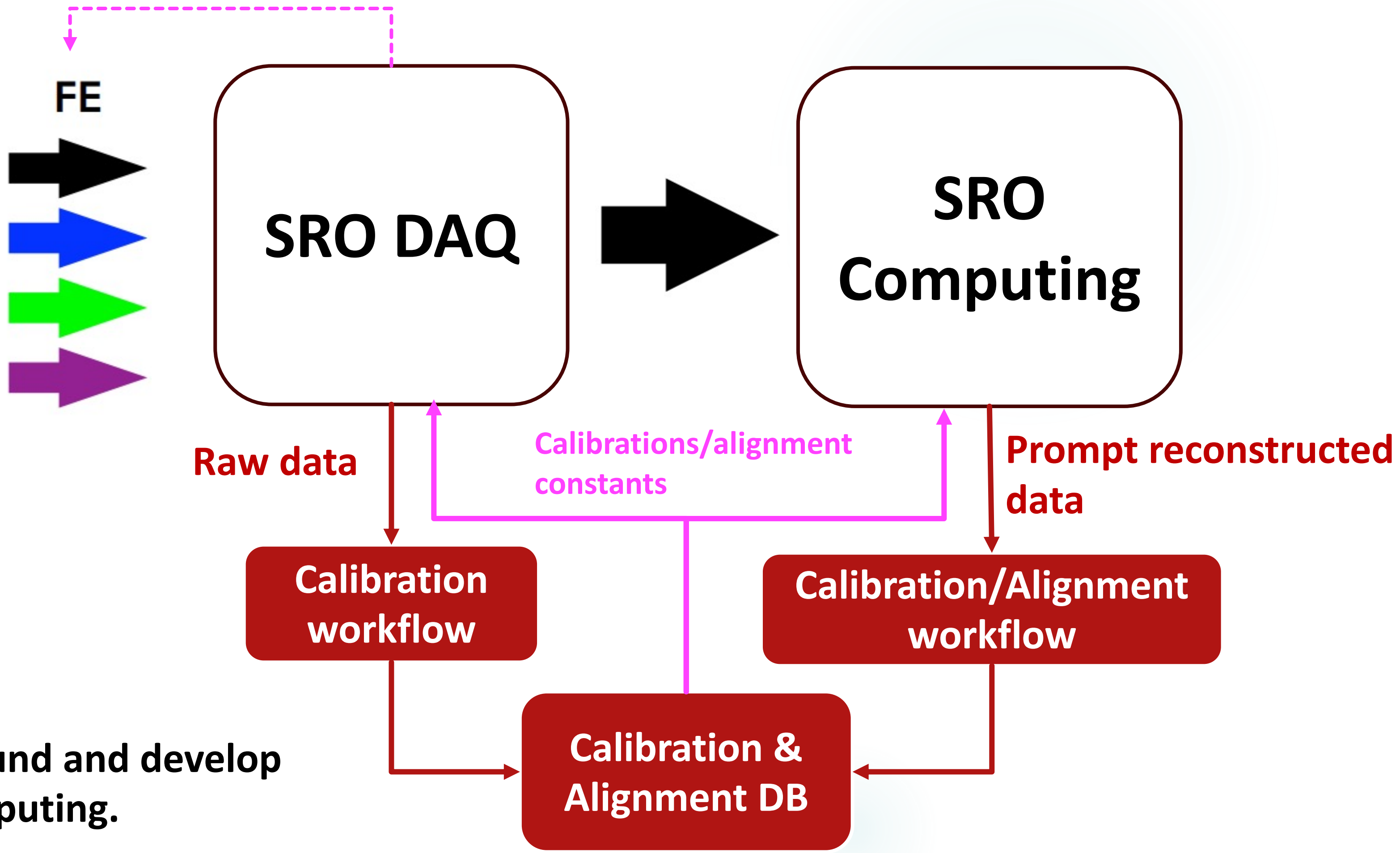




Streaming Calibration and Alignment

- Real-time calibration is challenging but essential for physics-quality full reconstruction in 2 weeks.
- This 2 weeks timescale is based on the statistics needed for reconstruction-level calibrations.

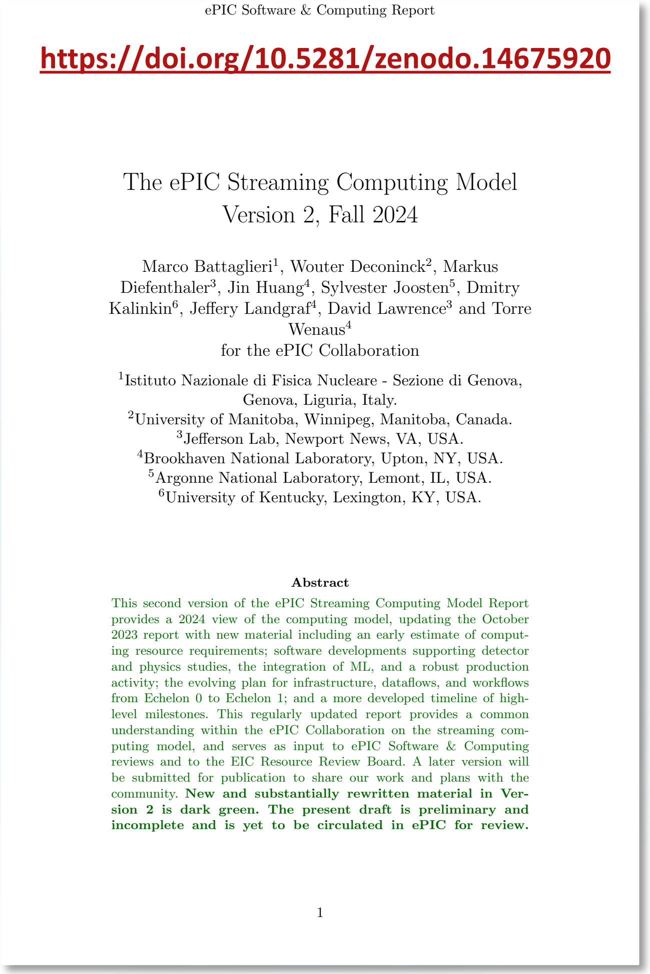




We need to validate two weeks turnaround and develop entire chains in streaming DAQ and computing.

ePIC Computing Model

- We developed the ePIC Streaming Computing Model to accommodate the requirements for streamed data processing, calibration, and streaming orchestration.


ePIC Software & Computing Report

https://doi.org/10.5281/zenodo.14675920

The ePIC Streaming Computing Model
Version 2, Fall 2024

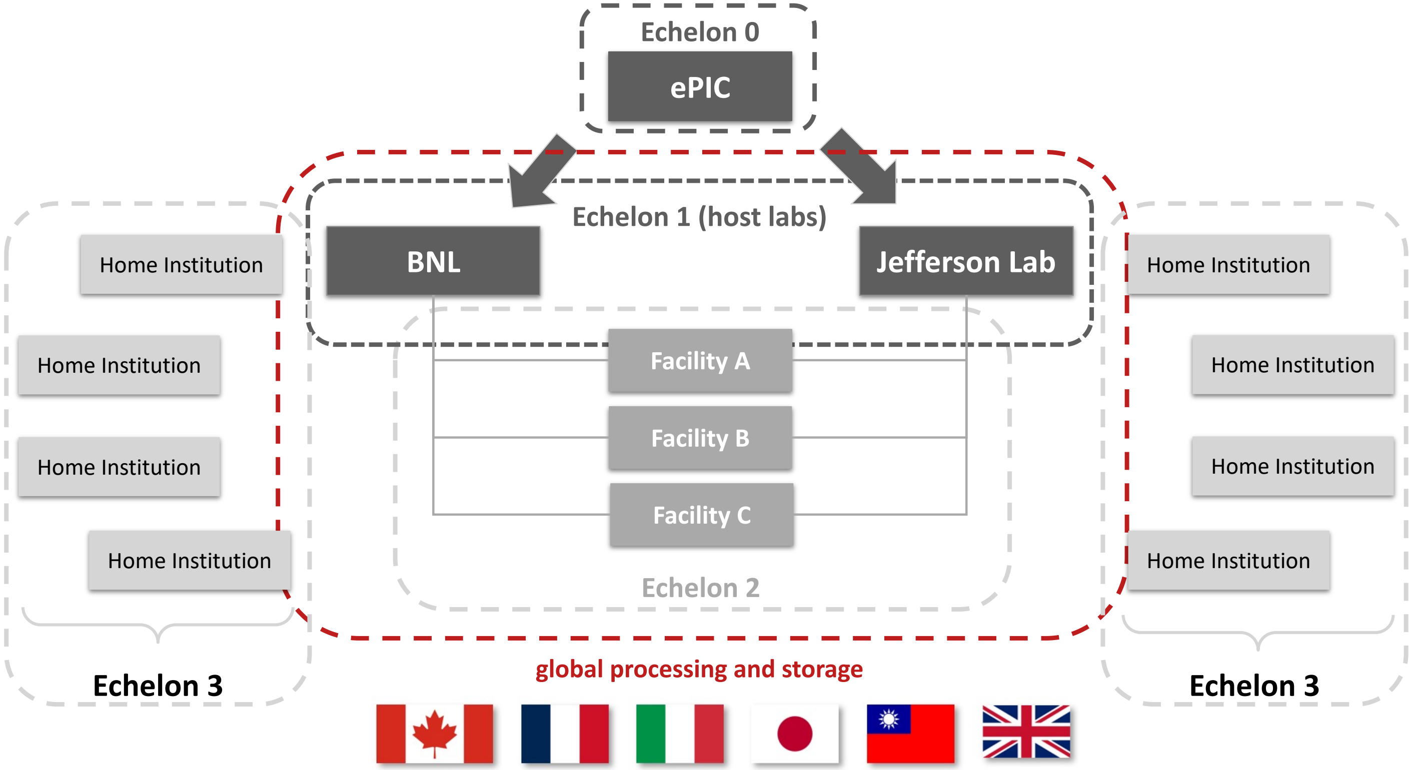
Marco Battaglieri[1], Wouter Deconinck[2], Markus Diefenthaler[3], Jin Huang[4], Sylvester Joosten[5], Dmitry Kalinkin[6], Jeffery Landgraf[4], David Lawrence[3] and Torre Wenaus[4]
for the ePIC Collaboration

[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



Echelon 0: ePIC experiment, DAQ system
Echelon 1: Two host labs, two primary ePIC computing facilities (prompt reconstruction)
Echelon 2: Global contributions leveraging commitments to ePIC computing
Echelon 3: Supporting the analysis community where they are at their home institutes

Recent SRO WG activities

- **Now the effort is moving from design to implementation. Our recent activities aim to define and test the interface between DAQ and computing by building several testbeds.**
- **Testbed plans** are taking concrete shape:
  - **Streaming orchestration using Rucio and Panda** — **On-going**
    - Developing E0-E2 streaming workflows and workload management system
  - **Streaming reconstruction using JANA2 and EICRecon** — **On-going**
    - Raw streamed data to collision event identification, reconstruction, and analysis.
  - **Alignment and Calibration workflows** — **Need to start now!**
    - Rapid data processing and execution of calibrations from standalone workflows to complicated workflows with subsystem dependencies (ex, alignment).
  - **Streaming analysis** — **Started**
    - Demonstrate simulation data production streaming to E2 site.

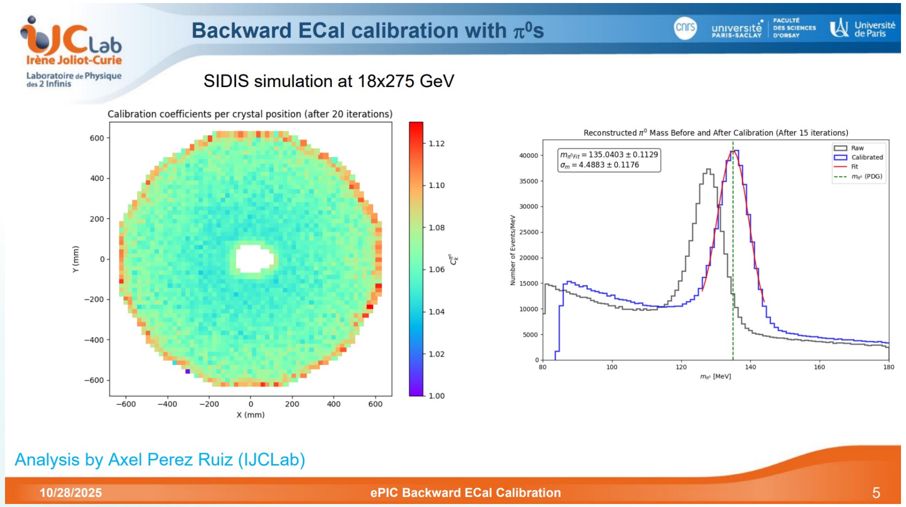
Level of Calibration

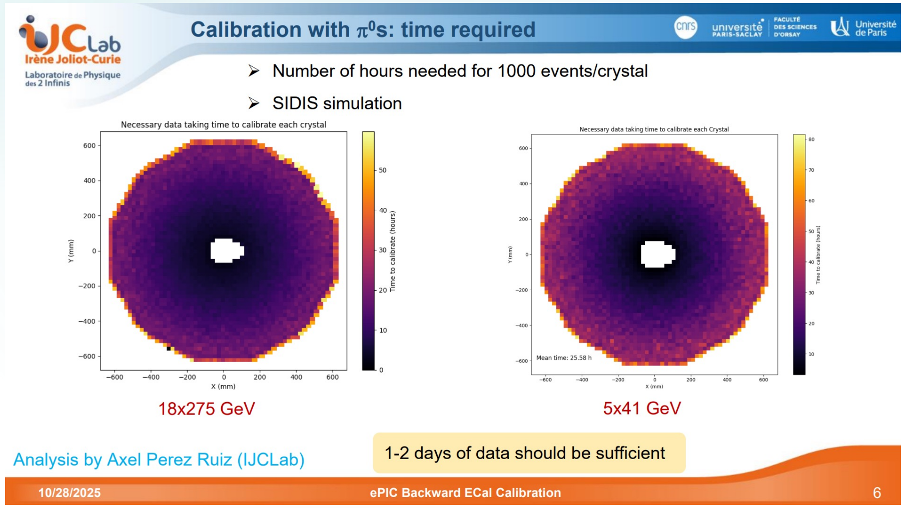
- **What is calibrated? (Calibration Content)**
  - **Detector physical parameters**
    - Bias voltages, gain settings, Temperature / radiation damage corrections, etc
      → Calibration affecting the *physical operation point* of detectors.
  - **Electronics and readout calibration**
    - Pedestals / offsets, Channel-to-channel timing, Amplifier gains, ADC linearity
      → Calibration of *electronics configuration parameters*.
  - **Reconstruction-level calibration**
    - T0 offsets, Energy calibration, Detector alignment
      → *Reconstruction-critical* calibration.
  - **Time-dependent corrections**
    - Clock drift, Temperature-induced slow drifts, Event-by-event T0 corrections via vertex
      → *Time-evolving* calibration parameters.

- **How calibration data is obtained (Calibration Methods)**
  - **Special runs**
    - Pedestal/noise runs, special bunch patterns, Low luminosity runs, Vernier scans
  - **Dedicated on-detector calibration systems**
    - Lasers, LEDs, Pulsers, Diodes
  - **Continuous monitoring during normal beam operations**
    - Built-in calibration features, Streaming pedestal monitoring, Online gain tracking
  - **Physics-based calibration using high-statistics events**
    - Calorimeter energy scale (π^0, MIP, electrons), Tracking alignment (residual-based)
  - **Time-dependent parameter estimation**
    - Clock correction, Slow thermal drift monitoring, Event-by-event T0 estimation

Prototyping

Carlos (Muñoz Camacho) presented on Oct. 28 on the **Backward ECal (EEEMCal) Calibration**.
The EEEMCal is a good example to start with. https://indico.bnl.gov/event/30349/





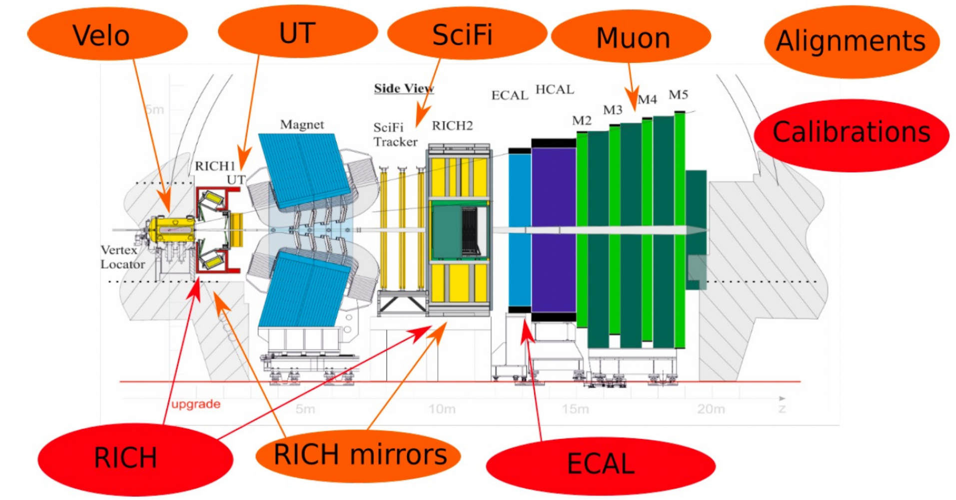
This prototype addresses reconstruction-level calibrations and physics-based calibration using high-statistics events.

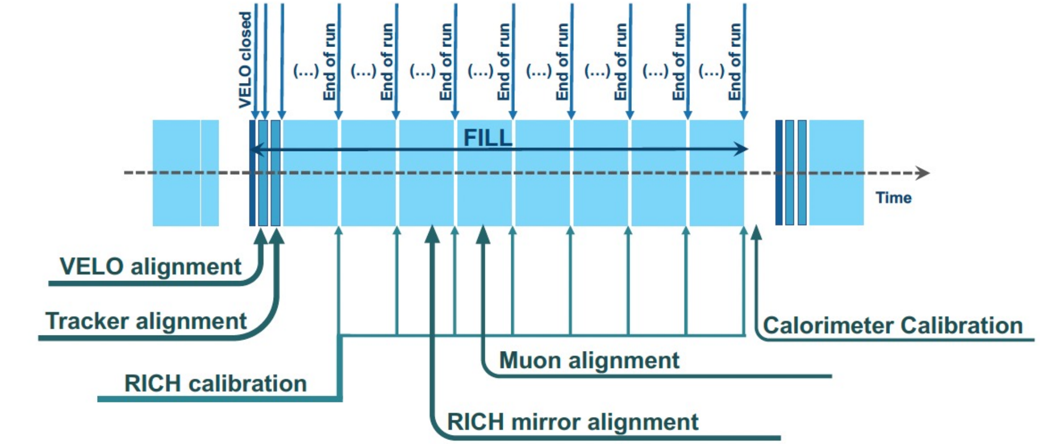
1. **Script Integration:** Carlos provides calibration scripts and integrates them into JANA2/EICrecon, defining data flow and required inputs.
2. **Workflow Implementation:** Implement file-based workflow first; then prepare for stream-based workflow
3. **AI-Driven Components:** Stepwise integration of calibration detection logic, automated validation, and selective human-in-the-loop checkpoints.
4. **Workflow Orchestration:** Proven workflows are then incorporated into the overall orchestration framework for automated operations.
5. **Milestones & Deliverables:** Prototype workflow (manual → semi-automated → AI-assisted), validated Conditions DB, documentation of APIs, state machine, ownership, and operational cycle; plan for scaling to full EIC detectors.

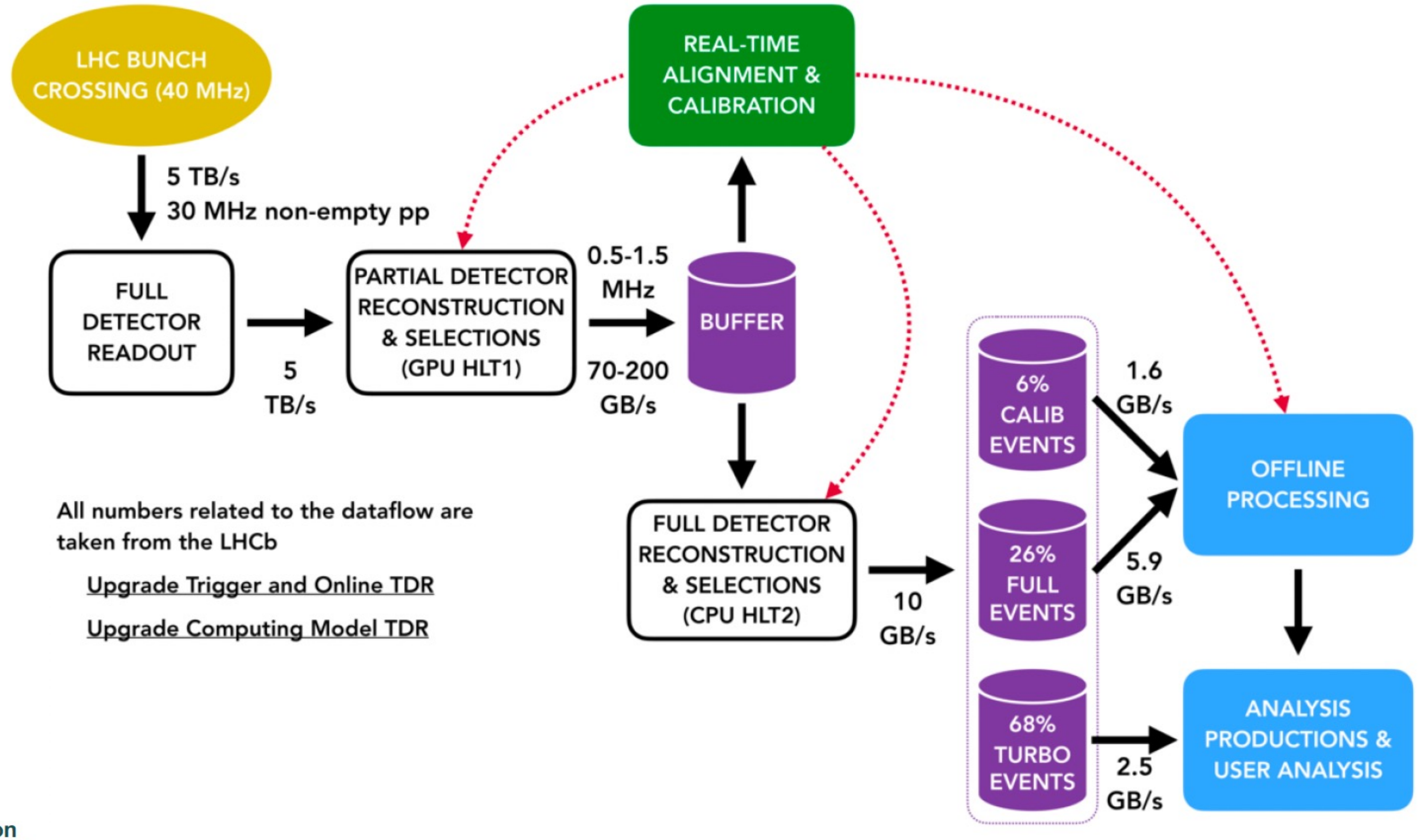
Other systems are welcome to join the prototype

Example of LHCb

Details will be given by Marco at the next talk.







We have to develop our plans for the alignment.

Toward (AI-Driven) Autonomous Calibration

- **Our primary objective is to build an *autonomous calibration system*, capable of detecting when calibrations are needed, executing them reliably, and integrating results into the reconstruction.**
- **AI/ML methods serve as powerful tools that can enhance selected components.**

Three Core Tasks for autonomy

1. Calibration Logic (AI-assisted decision engine)

- Software must detect when new calibration is required (e.g., change detection, drift detection) and update the state machine accordingly. **(AI can assist)**

2. Calibration Integration Into the State Machine

- The calibration workflow must connect to a calibration/conditions DB to track calibration status and link to calibration data, define who reads/writes constants, and manage workflow transitions.

3. Calibration Execution & Validation (AI-supported QC)

- Calibration scripts must compute new constants, validate them, and register them in the database.
(AI can assist in automated quality checks or validation scoring.)

Operational Boundary Conditions

1. Online Condition

- Must function during live data-taking
- File-based workflows as an initial stage but target is fully streaming, low-latency calibration loops

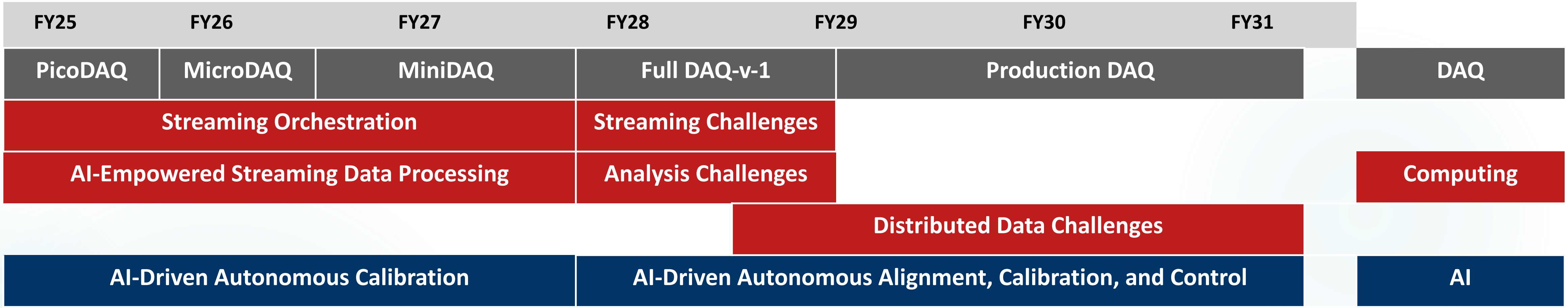
2. Human Condition

- Define necessary manual checkpoints
- Specify where human approval or override is required
- Aim for automation first → integrate human-in-the-loop later

3. Cybersecurity Condition

- Access control, signing of calibration constants
- Addressed in later implementation phase

Streaming DAQ and Computing milestones



Streaming DAQ Milestones and Deliverables

FY26Q1: PicoDAQ: Readout test setups

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

FY28Q1: MiniDAQ: Readout detector data using full hardware, timing chain, and orchestration systems

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

FY31Q3: Production DAQ: Ready for cosmics

Streaming DAQ and Computing milestones

FY25	FY26	FY27	FY28	FY29	FY30	FY31	
PicoDAQ	MicroDAQ	MiniDAQ	Full DAQ-v-1	Production DAQ			DAQ
Streaming Orchestration			Streaming Challenges				
AI-Empowered Streaming Data Processing			Analysis Challenges				Computing
			Distributed Data Challenges				
AI-Driven Autonomous Calibration			AI-Driven Autonomous Alignment, Calibration, and Control				AI

(AI-Driven) Autonomous Calibration

- Progress continues on understanding calibration workflows in collaboration with subsystem experts, with a focus on identifying timelines and interdependencies.
- The strategy for autonomy involves algorithms for change detection and agentic workflows.
- **FY28 Q1 Goal: Autonomous calibration of one detector system using simulated streaming data.**

Updating the Calibration Needs Table

- The previous calibration table (prepared one year ago by Jin) needs updating
- We must clarify **each calibration item's parameters, procedures, frequencies, and dependencies** —to ensure consistent workflows across subsystems.
- This will help **structure workflows, define connections, and highlight opportunities for (AI-assisted) automation**.

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 continous process														Post-reconstruction calibrations (applied at analysis stages)
			Task	Human intervention ?	Data Needed	Dependecy	T0 + 12hr	T0 + 24hr	T0 + 36hr	T0 + 48hr	T0 + 60hr	T0 + 72hr	T0 + 84hr	T0 + 96hr	Monitoring	Computing resource	
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	Processing							
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	Processing					LED		
AstroPix	Barrel																
ScifiPb	Barrel		SiPM gain		?												
fEMC	Forward	IV Scan	Pi0, eta->gg events energy scale	QA	Pi0 di-photon resonance ~1 day of production data		Data Acc.	Data Acc.	Processin	Processing					LED		High energy cluster non-linearity
			Second iteration pi0 (if needed)								Processing						
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	?	?	?	?	?			Final energy scale calibration (if needed)
low Q2 Tagger	Far Backward	Alignment?															
low Q2 Tagger (CAL)	Far Backward																
Pair Spec Tracker	Far Backward																
Par Spec Cal	Far Backward																
Direct Photon Cal	Far Backward																
B0 Tracking	Far Forward	Survey alignment/Cosmic	Alignment check		MIP		Processing										
B0 PbWO4	Far Forward	Survey alignment/Cosmic	SiPM gain		MIP/Gamma/Electrons		Processing								LED		
Roman (Pots)	Far Forward	laser/survey alignment Low lumi running	beam position monitors/fill by fill correction		MIP rate distribution in RP	Acc. BPM Potential use of vertex of central detector	Data Acc. Dependen	Processing									
Off Momentum	Far Forward						Data Acc. Dependen	Processing									
ZDC PbWO4	Far Forward	Survey alignment, timing delay	SiPM/APD gain, timing	QA	Photon		Processing								LED		
ZDC Sampling	Far Forward	Survey alignment, timing delay	SiPM gain	QA	Single neutron		Processing								LED		

Toward a Revised Calibration Table

Goals of the Reorganization

- Identify/Revise calibration items including parameters, procedures, frequencies, and dependencies
- Add missing tasks; eliminate obsolete entries
- Map each calibration task to subsystem, workflow category, calibration tier
- Clarify ownership and cross-detector dependencies

The responsibility-sharing proposal

Subsystem Responsibilities

- Review all calibration items relevant to the detector subsystem
- Confirm whether parameters, procedures, and update frequencies remain valid
- Provide missing calibration tasks or workflow changes introduced in the new structure
- Clarify dependencies on other subsystems (triggers, timing, alignment, etc.)
- Identify needs for new automation, tools, or monitoring
- **Update contact persons for calibration, software, and data-flow interfaces**

SRO Computing model WG Responsibilities

- Ensure each task is correctly categorized (method, tier, workflow type)
- Harmonize common tasks across subsystems and eliminate redundancies
- Maintain communication loop and track update status for each detector
- Identify items requiring further discussion or long-term development
- **Development of testbeds and algorithms of using typical use cases (ex, ECal energy calibration)**
- **Build the standardized calibration framework and integrate subsystem inputs**

Questions to the DSCs

- We will start progressing on understanding calibration plans and required workflows in collaboration with subsystem experts
- **Key points for DSC inputs:**
  - **Contacts: Update contact persons and clarify responsibilities**

  - **Review of Calibration Items:** Confirm which calibration parameters, procedures, and update frequencies remain valid. Identify missing tasks or obsolete ones.
  - **Workflow Alignment:** Define each subsystem's calibration workflows —inputs, processing steps, outputs, inter-subsystem dependencies (triggers, timing, alignment, shared detectors), and global calibration workflows such as global alignment.
  - **Automation & Tools:** Identify requirements for automation, monitoring tools, and AI-tools.
  - **Potential Bottlenecks:** Are any steps likely to delay calibration or prevent timely updates? **Are there any showstoppers that could prevent calibrating the ePIC data within two weeks?**
- **Next steps (short & long-term view)**
  - **Complete table update and align responsibilities**
    - **Our goal is to have the table updated by February 28**
    - **We will reach out to any DSCs we have not heard from by January 31**
  - Develop prototyping using Backward ECal and AI-driven prototype workflow
  - Prepare for streaming-based calibration integration by FY28.Q1
  - Coordinate with subsystem teams for full deployment

Reference

- Streaming computing model googledoc folder
- calibration workflow planning chart
- Computing resource estimates slides 20240904
- Computing resource requirements worksheet
- ePIC workflow management system requirements draft
- ePIC DAQ WG wiki
- ePIC detector digitization model spreadsheet

backup slides

Use cases and 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		✓	✓	

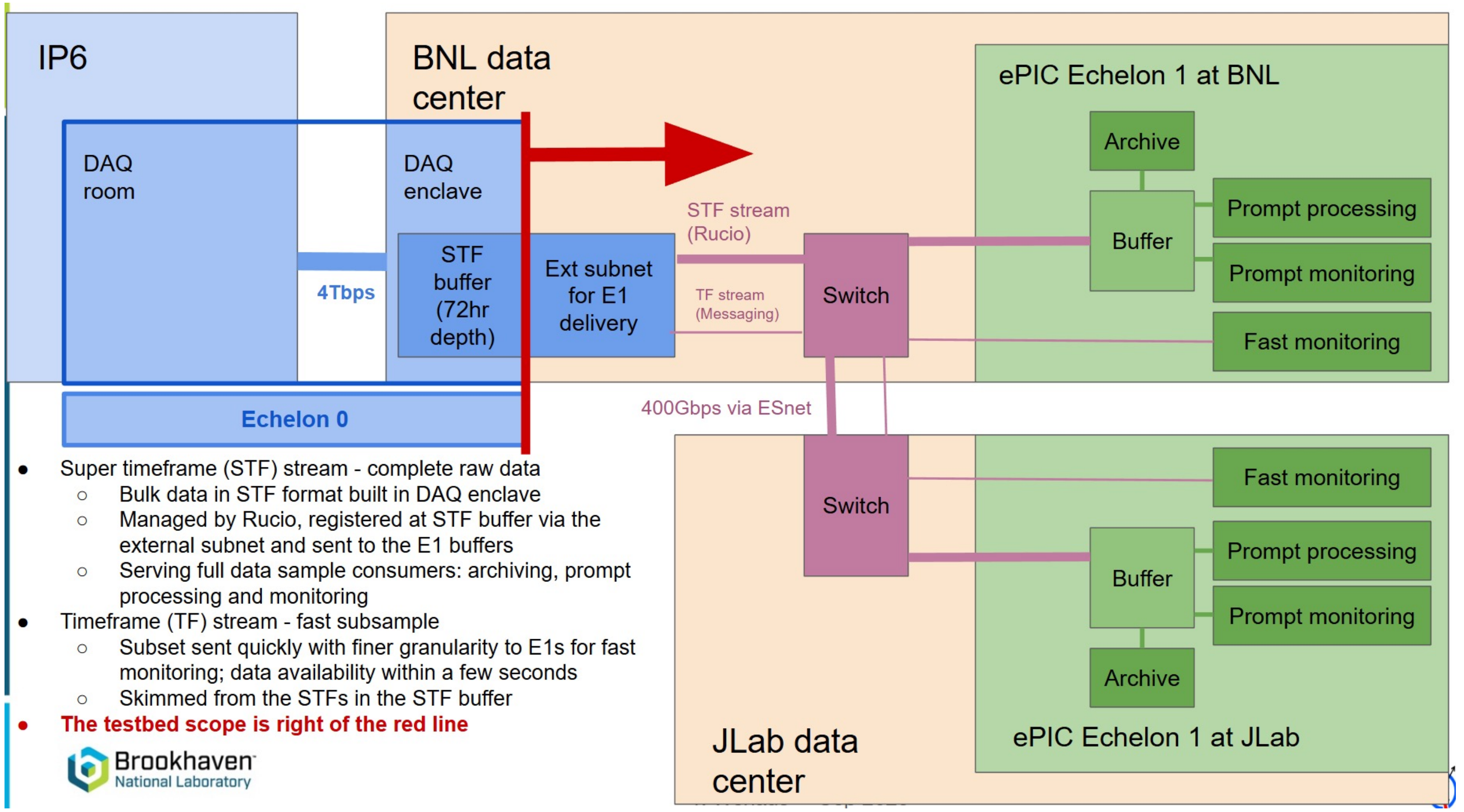
Prompt = rapid low-latency processing
Prompt processing of newly acquired data typically begins in seconds, not tens of minutes or longer

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

- Echelon 1s uniquely perform the low-latency streaming workflows consuming the data stream from Echelon 0
  - Archiving, monitoring, prompt reconstruction, rapid diagnostics
- Ensure the E1s have sufficient processing power for the low-latency workflows
- Apart from low-latency streaming, Echelon 2s are full participants in the use cases and accelerate them.

Echelon 0 (DAQ) to Echelon 1

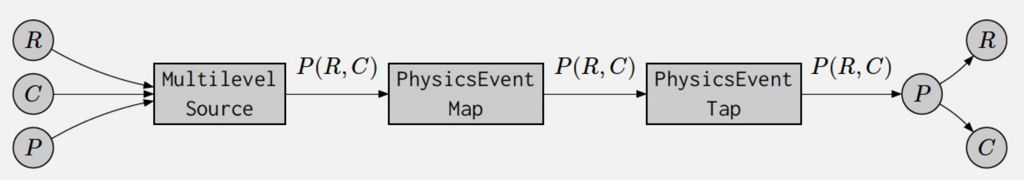
- What's in the data stream sent from DAQ?
  - Time frames, each containing all detector data within a time window of ~0.6ms, are built in DAQ
  - Time frames are aggregated in super time frames (STFs) which are sent out of DAQ to E1
    - Super Time Frame (STF) is a contiguous set of ~1000 time frames
    - Within a STF the TFs are time-ordered, as required for reconstruction
    - This O(1s), 2GB STF data unit is an appropriate granularity for Echelon 1 processing



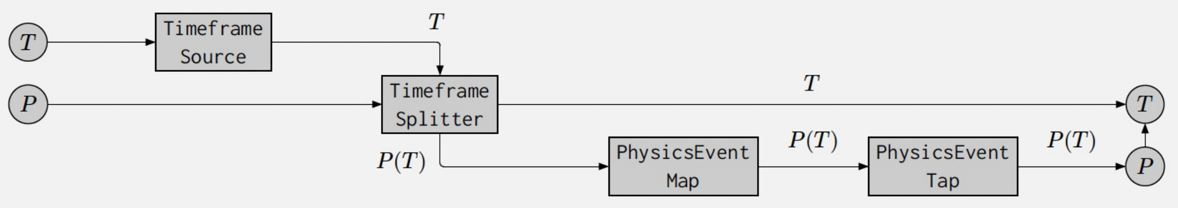
Streaming reconstruction : JANA2

- Multithreaded JANA2 framework provides a component-level hierarchical decomposition of data boundaries into **Run**, **Timeframe**, **PhysicsEvent**, and **Subevent** levels. This is essential for streaming processing.
- The **Folder** and **Unfolder** component interfaces enable traversal of this hierarchy by supporting operations such as splitting and merging data streams. This functionality has been tested and validated within **EICrecon**

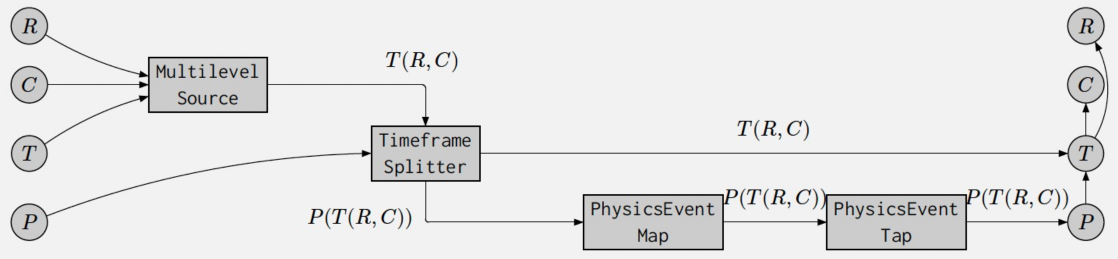
Introducing multilevel sources



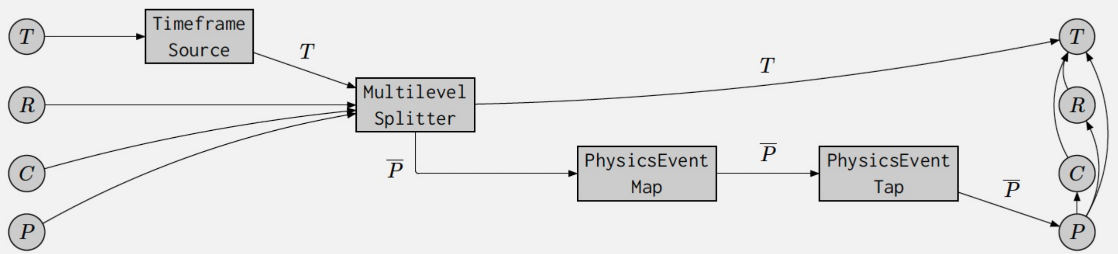
EICrecon timeframe splitting



Multilevel sources with timeframe splitting

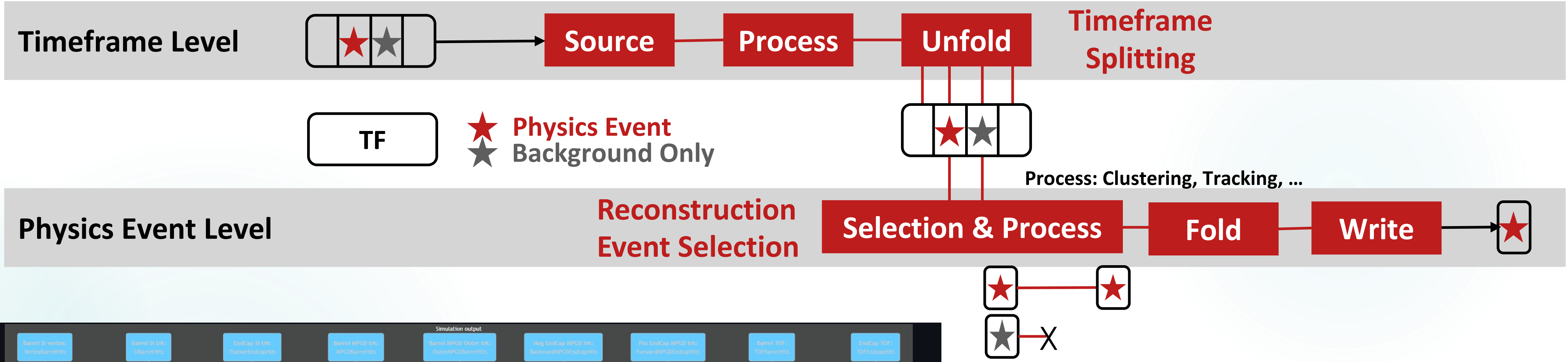


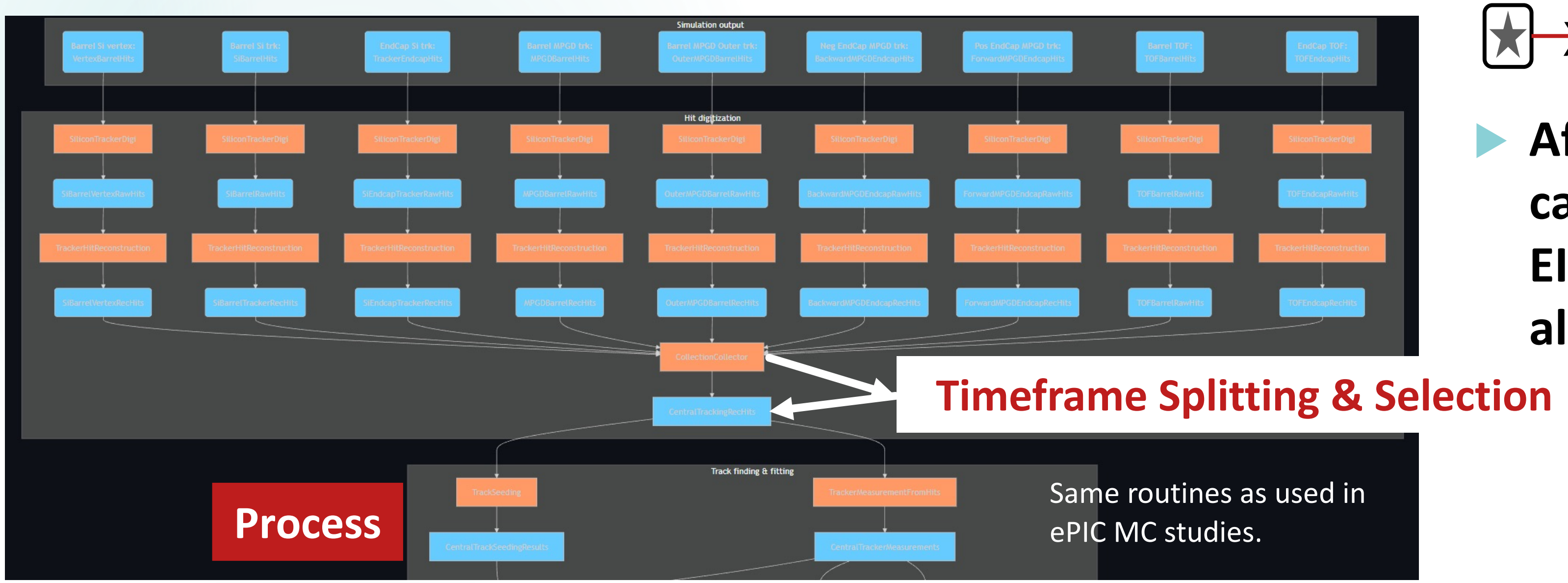
Timeframe sources with multilevel splitting



Streaming reconstruction : Prototyping

- TimeFrame data processing for event building based on JANA2 (TimeFrame Splitting and Selection)



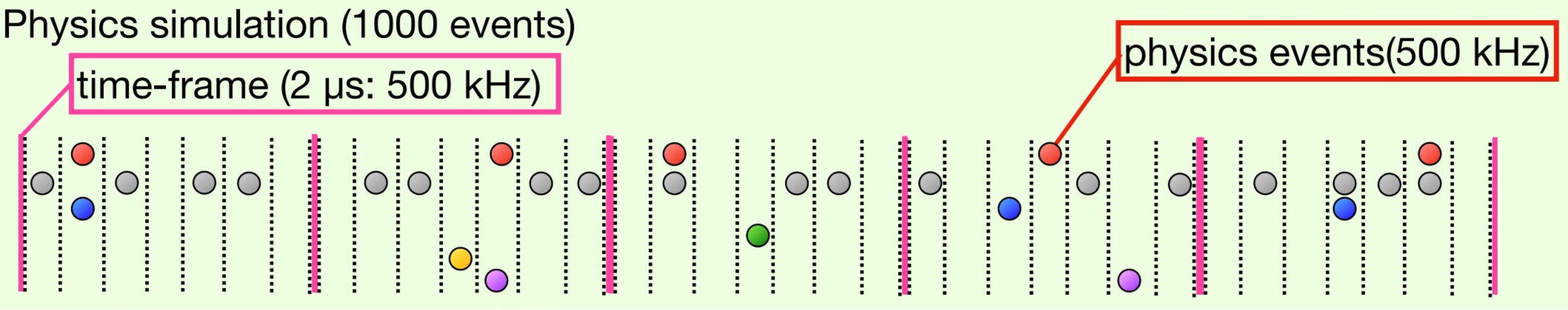


- After Selection, data (physics event candidates) is reconstructed in EICrecon using same tracking algorithms as used in MC studies.
- https://github.com/eic/EICrecon/blob/main/docs/design/tracking.md

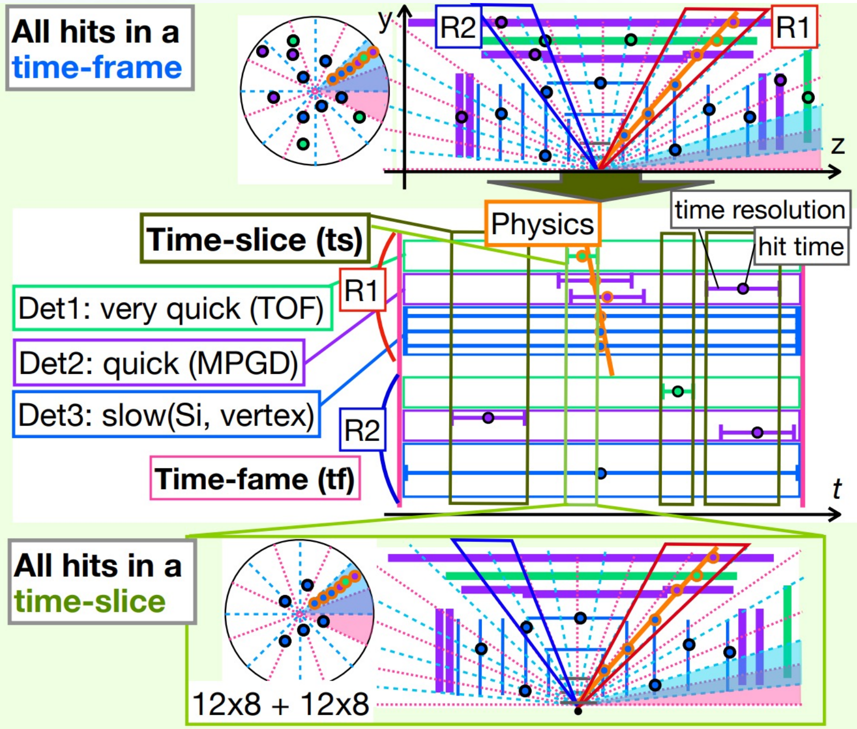
backup

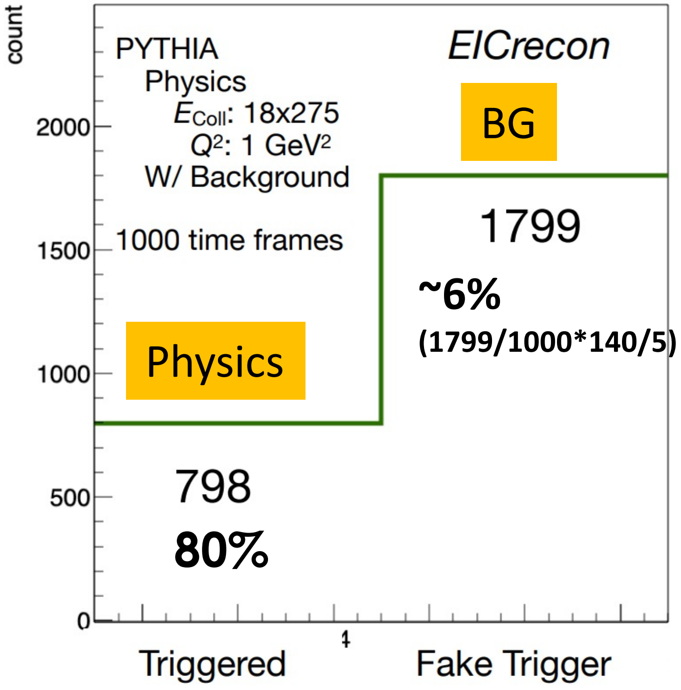
Streaming reconstruction : Performance

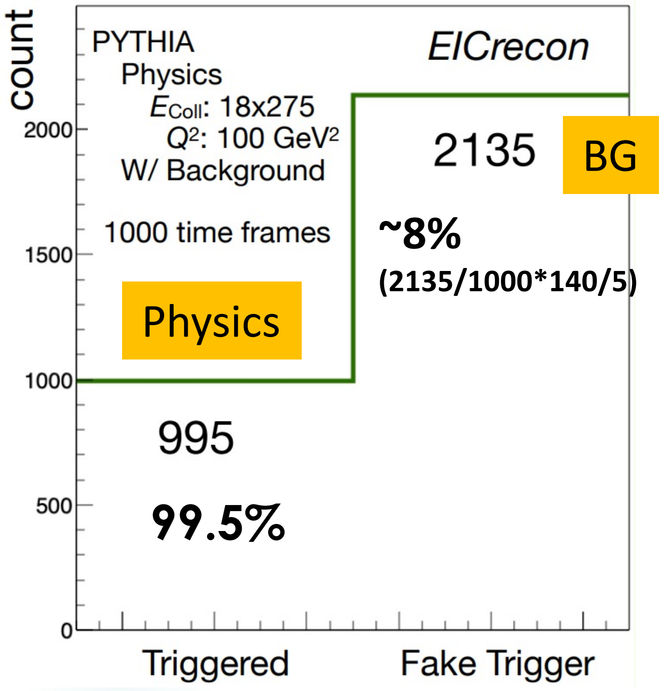
- **Algorithms to select physics event candidates are under development. Needs subsystem inputs!**



🔴	DIS NC 18x275 Q² > 1 (Deep inelastic scattering neutral current)	500 kHz
⚪	Synchrontron Radiation	14000 kHz
🔵	Electron bremsstrahlung radiation	317 kHz
🟢	Electron Touscheck scattering (intrabeam dcattering)	1.3 kHz
🟡	Electron Coulomb scattering processes	0.72 kHz
🟣	Proton beam gas interactions	22.5 kHz

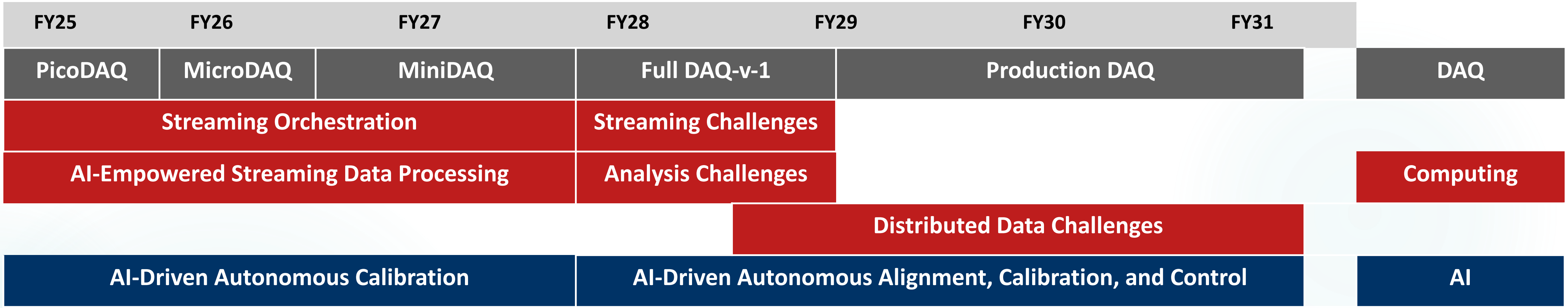






backup

Streaming DAQ and Computing milestones



Streaming Orchestration Milestones and Deliverables

- ✓ **Requirement documents** for streaming orchestration developed.
- **FY28 Q1 Goal**: Deliver a functional testbed for calibrating one detector system using simulated streaming data.
- Progress is ongoing in testbed development:
  - We are evaluating streaming orchestration using **PanDA + Rucio**.
  - We have demonstrated streaming data processing using **EJFAT**.
  - Additional prototypes under consideration: LHCb Allen, SPADI Alliance.

backup

Streaming DAQ and Computing milestones

FY25	FY26	FY27	FY28	FY29	FY30	FY31	
PicoDAQ	MicroDAQ	MiniDAQ	Full DAQ-v-1	Production DAQ			DAQ
Streaming Orchestration			Streaming Challenges				
AI-Empowered Streaming Data Processing			Analysis Challenges				Computing
			Distributed Data Challenges				
AI-Driven Autonomous Calibration			AI-Driven Autonomous Alignment, Calibration, and Control				AI

Streaming Data Processing Milestones and Deliverables

- ✓ **JANA2 enables data processing at the timeframe, event, and sub-event levels.**
- **FY28 Q1 Goal**: Achieve streaming data reconstruction with high efficiency in identifying physics collision events in simulations, including varying levels of background. This includes an AI/ML challenge focused on developing algorithms for distinguishing physics events from background.
- Progress is ongoing in streaming data reconstruction.

Testbed for Streaming Orchestration

Motivation:
- Evaluate how well existing distributed computing tools support streaming orchestration.
- Focus on practical deployment and performance in realistic environments.

- **Design Precepts**:
  - Robust geographical distribution across real-world networks
  - Full automation of data processing workflows
  - Complete exposure of system status and operational analytics
- **Approach**:
  - PanDA and Rucio align with the stated design precepts.
  - Both are deployed in live testbed instances at BNL:
    - Other sites can participate in collaborative testing and development: https://github.com/BNLNPPS/swf-testbed
  - Assume that data is delivered in STFs, each consisting of 1000 aggregated TFs, with a size of ~2 GB at a rate of ~1 Hz.
- **Streaming in Action (Testbed Observations)**:
  - Each STF contains approximately 45,000 events and takes approximately 19 hours to process on a single serial core
  - This latency is too high for timely detector status feedback
  - STF data is distributed across multiple workers at sub-file granularity
  - Sub-file fan-out and parallel processing enable true streaming behavior in the testbed
  - Data-driven logic automatically triggers E1 transfers and prompt processing upon file appearance.

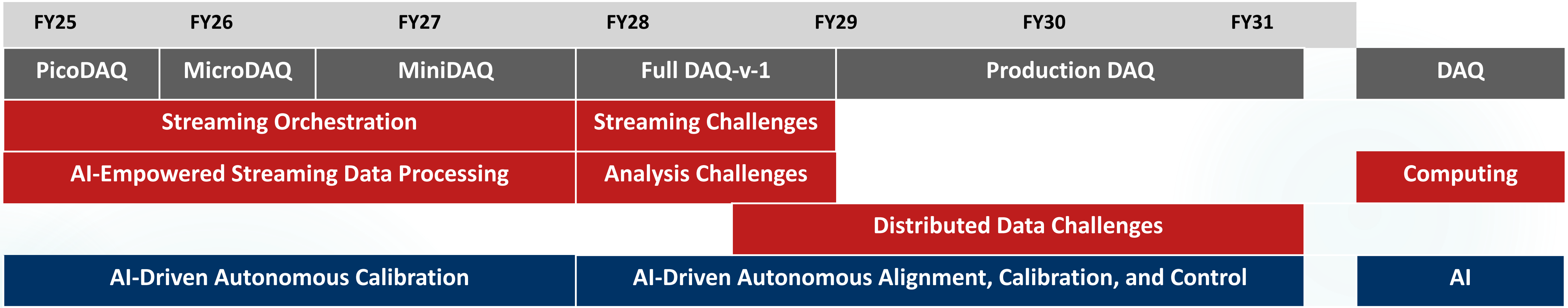
Levels of calibration:

Calibrations can be categorized by both the use of the calibration results, and by the method of producing the calibration results. We intend to focus on test setups for specific calibrations and gradually build up the infrastructure for defining and addressing each needed calibration scheme within the autonomous calibration framework.

- Use of calibration Results:
    - Calibrations that affect actual physical detector parameters (e.g. setting bias voltages)
    - Calibrations that affect electronics setup (e.g. pedestal values, electronics timing)
    - Calibration values that are sensitive within DAQ processing (e.g. gains, T0 offsets)
    - Calibrations values that are applied during reconstruction

- Method of calibration data acquisition
    - Special runs
        - Dedicated equipment on (e.g. lasers, pulsers, diodes)
        - Special electronics configuration (e.g. pedestals)
        - Special beam states (e.g. Reduced Bunch counts, Vernier Scans, low luminosity)
    - Monitoring of built in calibration features during normal beam time (e.g. lasers, pulsers, diodes)
    - Constant Time-Dependent Parameterization during beam operations (e.g. Clock corrections due to slow temperature drifts, event-by-event T0 corrections via vertex position)
    - High Statistics Parameterization during regular beam operations (e.g. Energy calibration in calorimeters)

Streaming DAQ and Computing milestones



- **Compute-Detector Integration**:
  - Joint deliverables between DAQ and **computing** to develop integrated systems for detector readout, data processing, and ultimately physics analysis.
  - **Key role of AI(/ML)**: Empowering data processing and enabling autonomous experimentation and control.