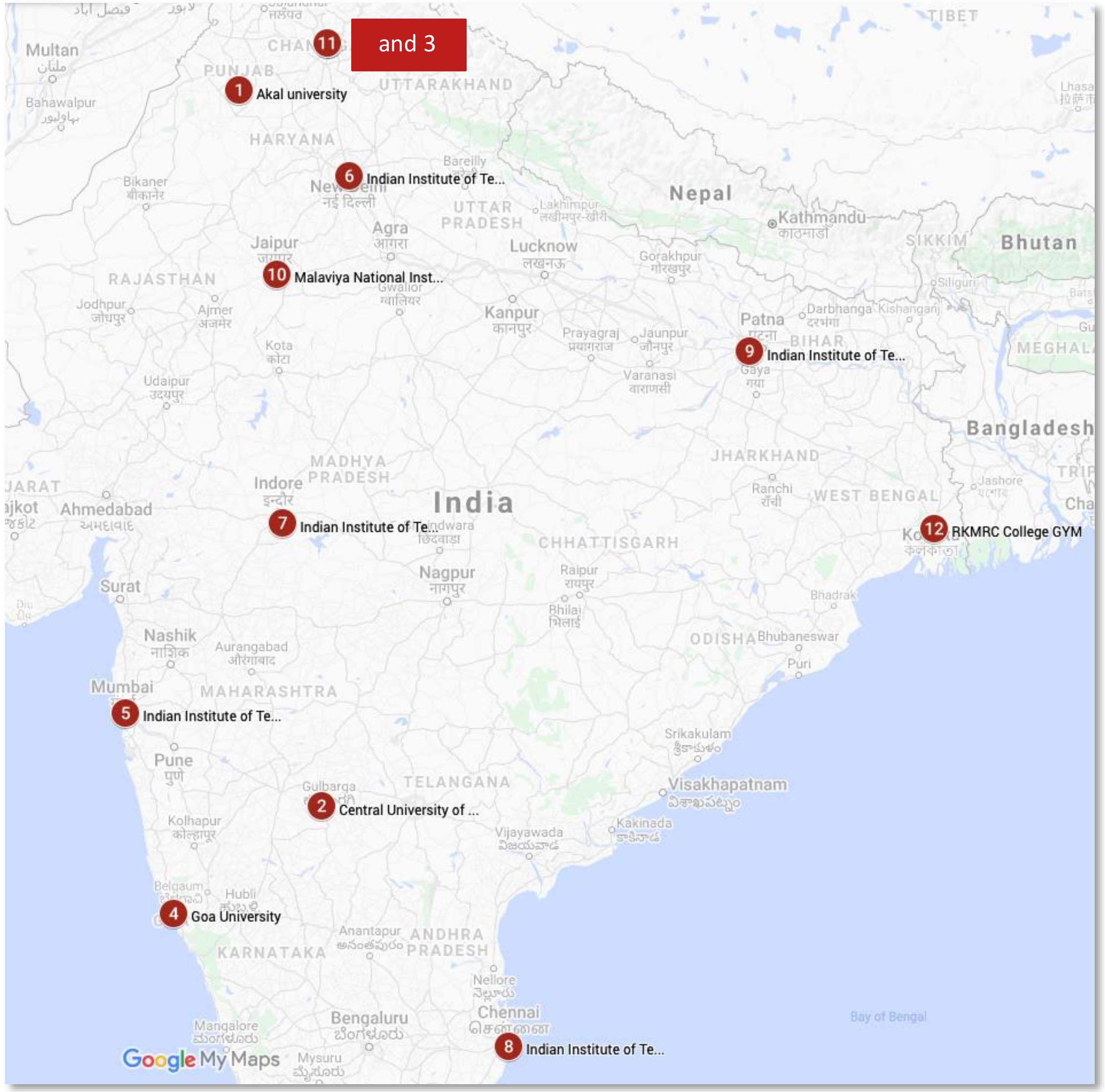
# Contributing to ePIC Software & Computing

Markus Diefenthaler (ePIC Software & Computing Coordinator)

# EIC Software Involvement by Indian Institutions Starting from 2020




1. **Akal University** Ramandeep Kumar
2. **Central University of Karnataka** Deepak Samuel
3. **DAV College** Monika Bansal
4. **Goa University** Prabhakar Palni
5. **Indian Institute of Technology Bombay** Sadhana Dash and Basanta Nandi
6. **Indian Institute of Technology Delhi** Tobias Toll
7. **Indian Institute of Technology Indore** Ankhi Roy
8. **Indian Institute of Technology Madras** Prabhat Pujahari
9. **Indian Institute of Technology Patna** Neha Shah
10. **MNIT Jaipur** Kavita Lalwani
11. **Panjab University** Lokesh Kumar
12. **RKMRC College** Amal Sarkar

# Lessons Learned

- There are enormous opportunities for India to play a leading role in software, computing, and associated physics studies.
- ePIC Software & Computing is excited about these opportunities.

- Engagement from India is currently limited. We are eager to explore how we can help strengthen participation. This meeting serves as a key step in that direction.
- One major challenge is the time difference between India and the United States (9.5 –15.5 hours). This presents significant obstacles for onboarding, collaboration, and mentorship.

- It is essential for India to establish a strong software & computing hub that can coordinate activities across the subcontinent and provide local mentorship and support.
- ePIC Software & Computing is enthusiastic about collaborating to realize these goals and expand involvement from India.

# ePIC Software & Computing Organization

Spokesperson's Office

Detector | **Software & Computing** | Physics

***Guiding Principles:***

- *Statement of Software Principles*
- *Sustainability*

**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 (not yet activated)*

**Deputy Coordinator (Development)**
Dmitry Kalinkin (U Kentucky)

**Development Working Groups**:

- Physics and Detector Simulation
- Reconstruction Framework and Algorithms
- *Analysis Tools (not yet activated)*

**Deputy Coordinator (Infrastructure)**
Torre Wenaus (BNL)

**Infrastructure Working Groups**:

- Streaming Computing Model
- *Heterogeneous Computing (not yet activated)*
- *Distributed Computing (not yet activated)*

# Development Priorities

Chao Peng (ANL), Sakib Rahman (BNL)

## Physics and Detector Simulation

- **Charge**:
  - Development of accurate MC simulations using a suite of physics and background generators and detector simulation based on Geant4 and DD4hep
- **Priorities for 2025**:
  - Continue to support the **detector design** and integration with services.
  - Collaborate with the EIC Project to evaluate the **differences between the engineering and simulation designs**, and lead discussions with the DSCs on how to address these differences.
  - Continue to support the development of **background modeling** and implement its timing structure in physics and detector simulations, together with the Background TF.
  - Enable **simulation of streaming readout** by providing the option to switch between streaming data and event data modes.
  - Coordinate the **development of digitization and noise models** with the DSCs and the Electronics and DAQ WG.

# Engineering Design Cross-Checks and New Beam Pipe Implementation

A major task of resolving discrepancies between engineering and simulation designs (in works since past year)

- Setup a conversion routine from DD4hep geometry to STEP CAD files.
- Working together with EIC Project to identify discrepancies. Meeting with engineers was highly productive.
- Working with DSCs to evaluate issues and implement changes as needed.

**DSC**

Works with project engineers to converge on subsystem design. Implements design ePIC geometry repository.

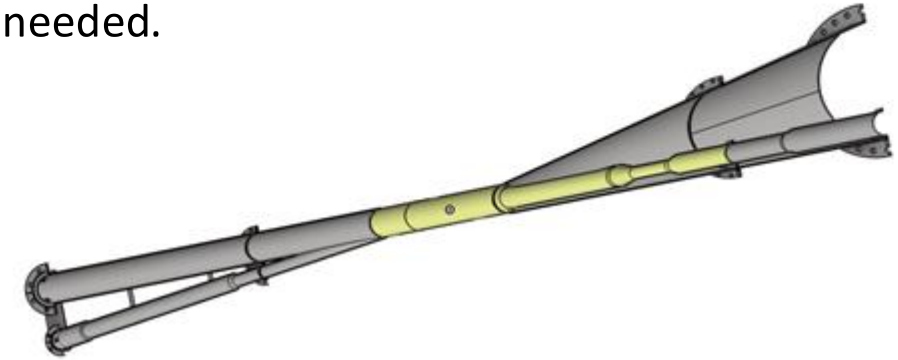
**Physics and Detector Simulation WG**

Converts subsystem geometry to STEP files using npdet_to_step.

Generate cross-section drawings using OnShape and provide STEP files and drawings to EIC Project.

Provide CAD files and comparison log back..

**EIC Project**

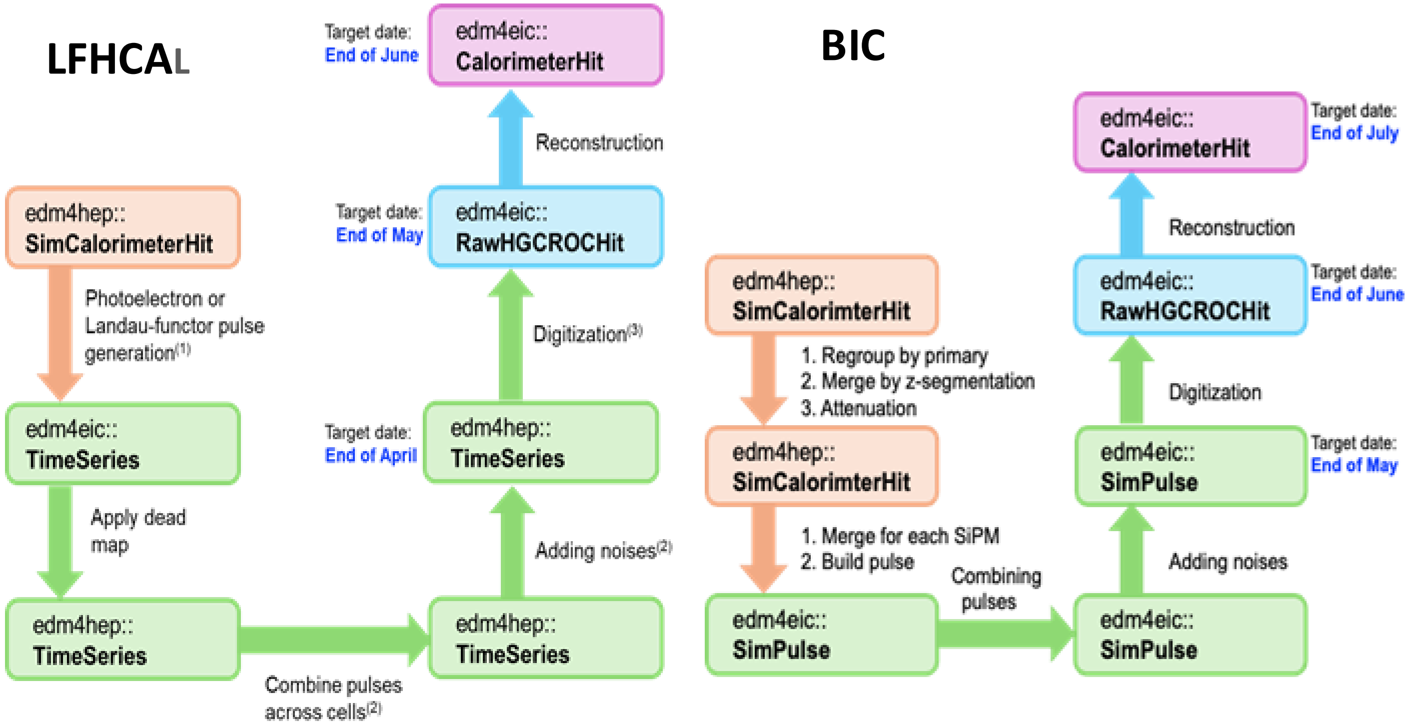


**New beampipe design**

Essential for background estimates due to the synchrotron radiation. Implemented in DD4hep geometry and validated.

Discrepancies between the engineering and simulation designs are being actively addressed by the DSCs. Physics and Detector Simulation WG continues to coordinate with the DSCs on ongoing updates to the ePIC geometry.

# Digitization Implementation: Progress on LFHCAL and BIC




- Proposed algorithmic flows for (more) realistic digitization chains in LFHCAL (Derek Anderson) and BIC simulations (Minho Kim).
- Significant overlap with work by Simon Gardner, **Tommy Chun Yuen Tsang**, and more in context of tracking digitization.

| Tag | Task/Work | Assignee | Status |
|---|---|---|---|
| DAT0 | edm4eic::SimPulse | Sylvester | Done |
| PUL0 | In-cell pulse combining | Simon | Done |
| PUL1 | Pulse 1/f noise application | Simon | Done |
| PUL2 | Attenuation + regrouping | Minho | PR under review |
| PUL3 | Pulse generation | Minho | To-do |
| PUL4 | Integrate pulse combining/noise | Minho | To-do |
| PUL5 | Detector variation in pulse | Simon | Issue opened |
| DAT1 | edm4eic::RawCALOROCHit | Derek | PR under review |
| DIG1 | CALOROC digitization | Minho | To-do |
| REC1 | CALOROC reconstruction | Afnan | To-do |

Tasks towards implementation.

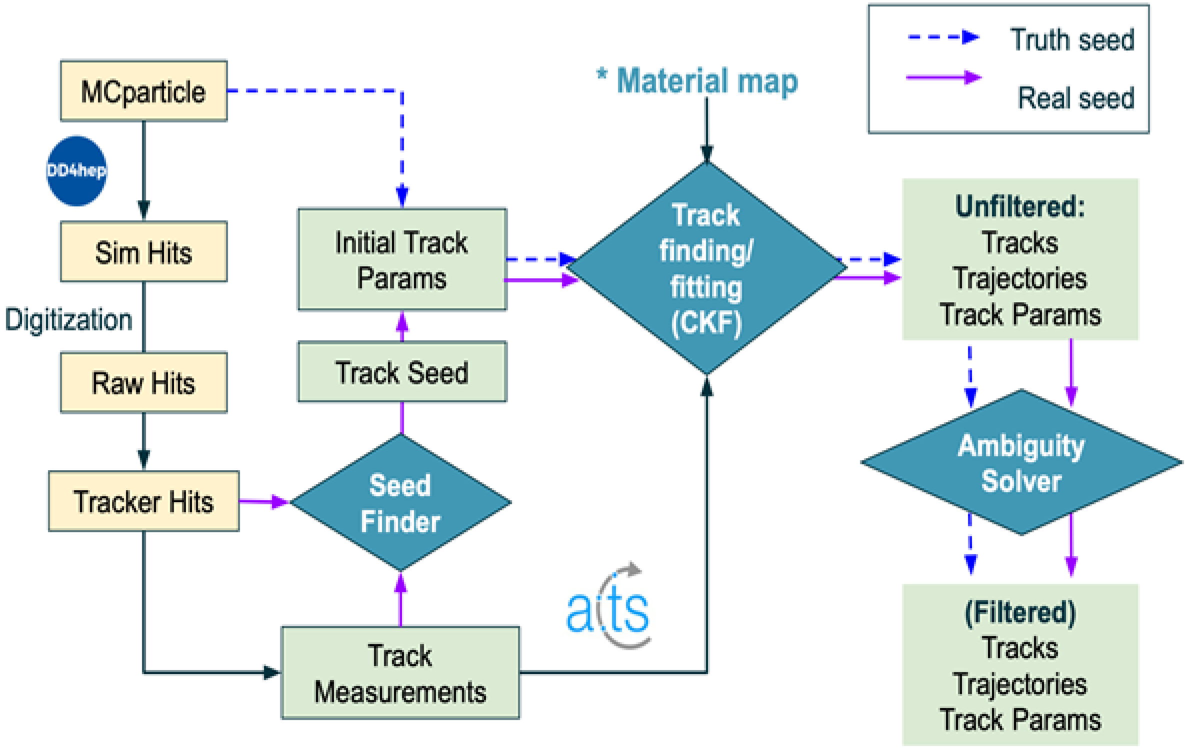
# Development Priorities

Derek Anderson (ISU), Shujie Li (LBNL)

## Reconstruction Framework and Algorithms

- **Charge**:
  - Development of a holistic and modular reconstruction for the integrated ePIC detector.
- **Priorities for 2025**:
  - Drive the **development of the reconstruction framework to meet ePIC needs**, e.g., on modularity or streaming data processing.
  - Host collaboration-wide discussions on all aspects of reconstruction, driving the **work toward holistic reconstruction**.
  - Enable reconstruction algorithms to **handle physics events with background**.
  - Collaborate with **PWGs** on **shared reconstruction priorities**, which currently include:
    - Secondary vertexing
    - Hadron identification
    - Particle flow algorithms for jet reconstruction
    - Event kinematics
  - Integrate continued development of **web-based event display** in reconstruction efforts.

# Reconstruction Updates | Tracking and Vertexing




**Algorithms:**

- The full chain of track reconstruction (seeding+CKF+ambiguity resolution) is set up in EICrecon.
- Primary vertexing ready, with analysis script included in benchmarks.
- 2nd vertexing development ongoing:
  - With ACTS::AdaptiveMultiVertexFinder, See Bishoy's talk
  - With KFParticle package
    - Stand-alone tests
    - Setup interface b/w EICrecon object and KFParticle
- Attempts to Insert pixel noise to silicon detector signals.

**Performance:**

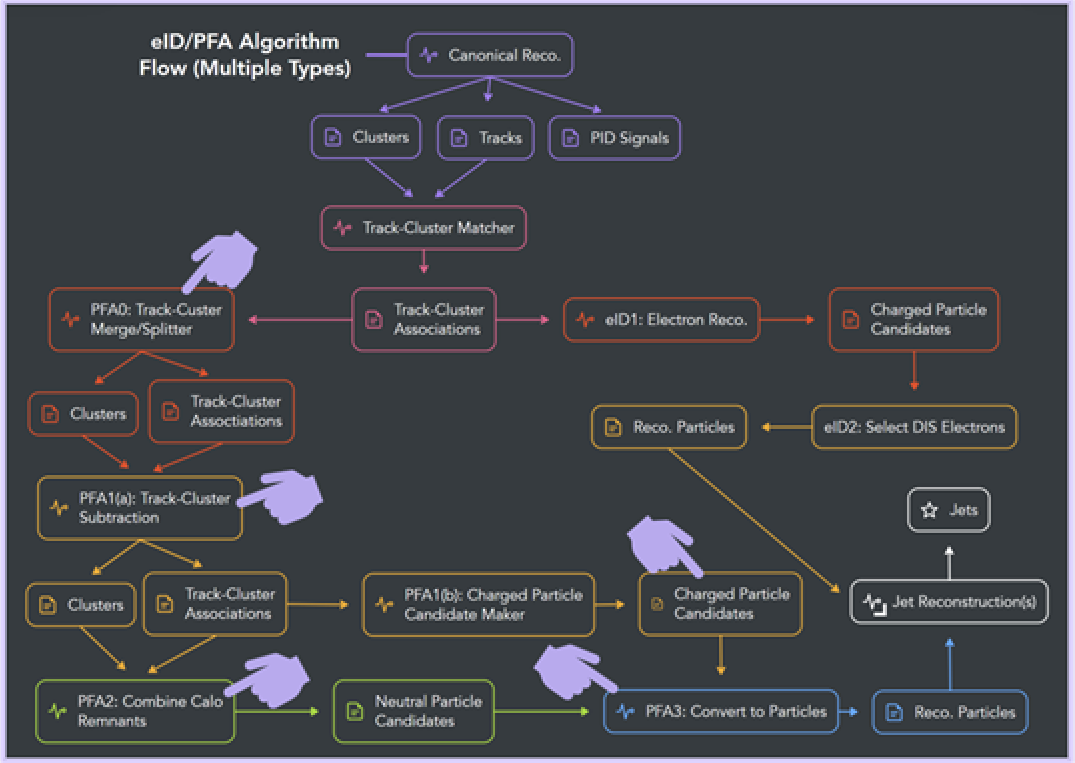
- Improve 2D digitization for MPGD.
- Identify and label signal v.s. background particles, followup in next week's ePIC Software & Computing meeting.

**ACTS4NP workshop@LBNL, May 12-15: ongoing**

- Tutorials and hands-on
- Geometry, data model
- AI/ML, GPU tracking…

# Reconstruction Updates | Particle Flow and Jet Reconstruction

Algorithmic flow of proposed PF baseline, components marked by purple hand, see also Discussions in Frascati.




- **Goal**: Implement baseline by first half of this year.
- Effort identified and tasks assigned.

| Items on Critical Path | Status |
|---|---|
| PFA0: update merge/splitter | PR under review |
| PFA0(a): track-protocluster link | PR under review |
| PFA0(b): track-protocluster match promoter | To-Do |
| PFA1(a): track-cluster subtractor | PR ready* |
| PFA2: remnant combiner | Assigned |
| Charged/neutral candidates | PR drafted |
| PFA1(b): charged candidate maker | To-Do |
| PFA3: particle regression | To-Do |

| Items Not on Critical Path | Status |
|---|---|
| Track-cluster matcher | PR merged |
| Cross-calo topocluster maker | Closed* |

# Reconstruction Updates | Hadron Identification and Event Kinematics

- **RICH team making steady progress towards integrating IRT2.0 into EICrecon**:
  - Regular updates given by Brian Page in Recon WG meetings
  - **Right:** slide from Brian's most recent update
- **Select highlights from recent update**:
  - Unified RICH (Q/PF/D) plugin available for debugging
  - Initial IRT2.0 algorithm & factory interfaces in progress

## Next steps [reordering the focus]

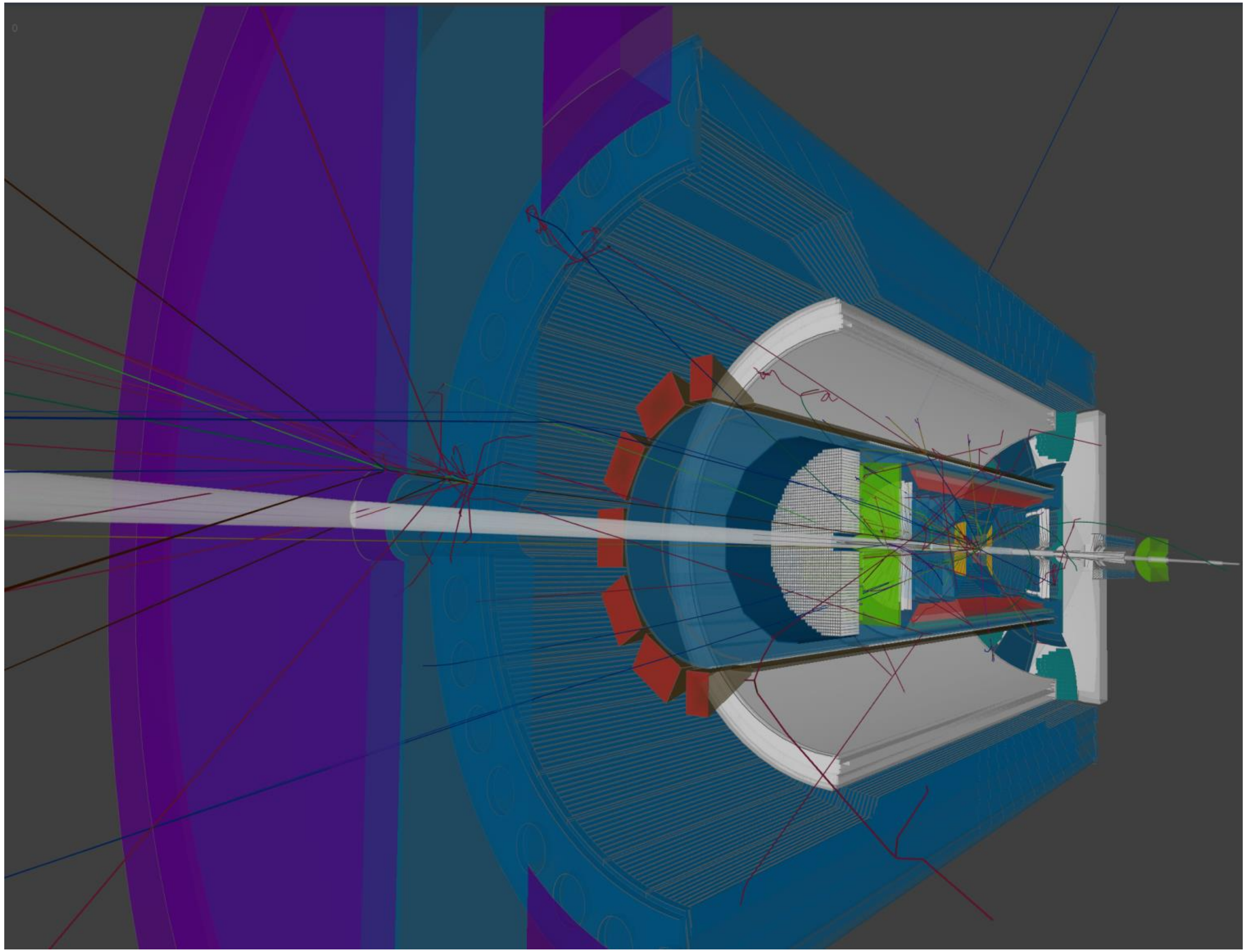
- Implement a single spherical mirror in QRICH
  - To confirm that IRT 2.0 (no sampling along the trajectory) works for dRICH gas radiator
- Confirm that using physical volumes (dd4hep implementation) rather than {logical volume & parent physical volume copies} in the geometry tree (pfRICH standalone code implementation) for sensitive volume elements (cell IDs) does not pose an issue for more complicated geometries (like dRICH sectors) with IRT algorithm
- Back to the "Next steps" items shown 2025/04/03 slides
  - pfRICH conical & pyramid mirrors implementation
  - Calibration pass (and calibration data import in EICrecon)
  - Standalone post-processing scripts -> integrate into IrtDebugging.cc plugin
  - Represent IRT 2.0 digitization as an intermediate plugin
  - Start thinking of a data model

3

**Kinematics (Charlotte Van Hulse, Rachel Montgomery)**: Establishing an initial dictionary of kinematics for physics processes and supporting physics analyses.

- **Overall status**: Finalizing cross check for the reconstruction of the SIDIS variables. Exclusive PWG has agreed upon two initial methods for t-reconstruction. Implementation started. More method added later.
- **Next Steps**: Following the plan described in the last ePIC Physics, Software & Computing Discussions.

# Visualization via Firebird: Web-Based Event Display Based on Phoenix



**Event Display for Reconstruction Development**:

**Available in eic-shell and at**:
https://eic.github.io/firebird/display

**Features:**

- Supports ePIC geometries.
- Integrates with simulation output for comprehensive studies and debugging.

# Infrastructure Priorities

Marco Battaglieri (INFN Genoa), **Taku Gunji (Tokyo)**, Jeff Landgraf (BNL)

## Streaming Computing Model

- **Charge**:
  - Development of the computing model for the compute-detector integration using streaming readout, AI/ML, and heterogeneous computing, in collaboration with the Electronics and DAQ WG.
- **Priorities for 2025**:
  - Define **requirements for streaming orchestration** and **set up corresponding testbeds**:
    - Develop a testbed for event reconstruction from streamed data in EICrecon, separating signal from background events and demonstrating how we will reconstruct physics events.
    - Establish an initial testbed for super time frame building and processing, and deliver a corresponding requirements document.
  - Document alignment and calibration workflows jointly with the DSCs and identify **requirements for autonomous alignment and calibration**.
  - **Publish the ePIC Streaming Computing Model report**, and the related section in the (pre)TDR.

# Compute-Detector Integration to Maximize and Accelerate Science

- Maximize Science **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.
- Accelerate Science **Rapid turnaround of 2-3 weeks for data for physics analyses**.
  - Timeline driven by alignment and calibration.
  - Preliminary information from DSCs indicates that 2-3 weeks are realistic.
- Technologies 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). |

# The Role of AI

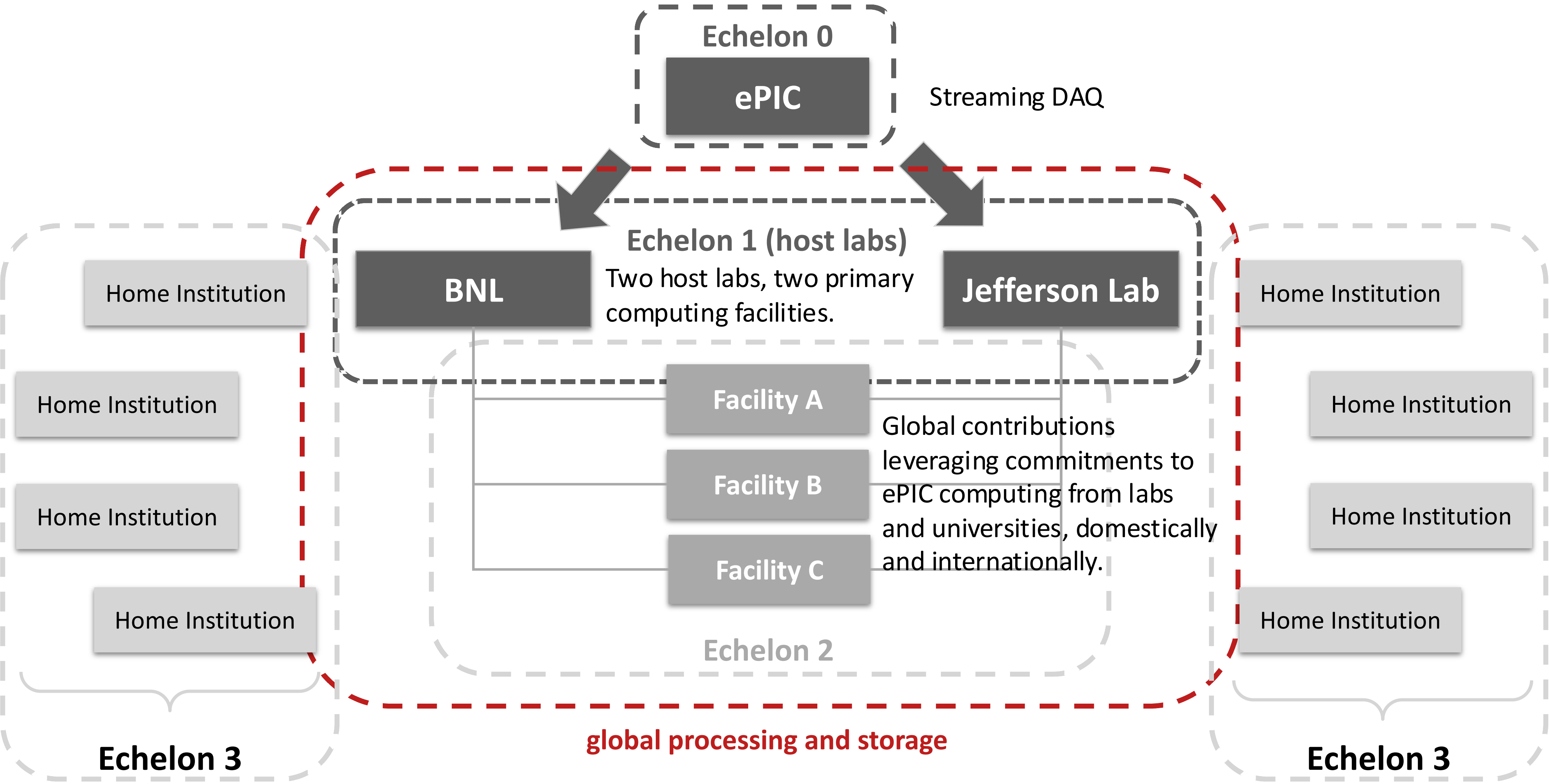
- **Compute-detector integration** using:

| | | |
|---|---|---|
| **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.

# 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.

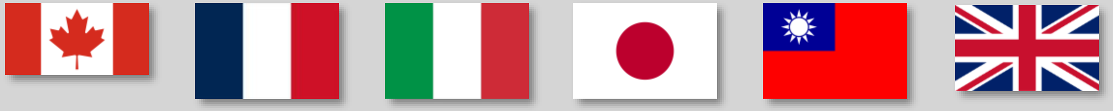
# Computing Use Cases and Their Echelon Distribution

| 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 | | ✓ | ✓ | |

**Substantial role for Echelon 2 in tentative resource requirements model**

| 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 and **accelerates** them:
  - Tentative resource requirements model assumes a **substantial role for Echelon 2**.
  - Capabilities and resource requirements for Echelon 2 resources developed jointly with the community.
  - Forming EIC International Computing Organization (EICO):



  - The power of distributed computing lies in its flexibility to shift processing between facilities as needed.

# Computing Resource Needs (EIC Phase I) 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** |

**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 WLCG 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 at a scale of ATLAS and CMS today.

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

# 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**:
- 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.**

**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.
**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.

# Streaming Computing: Getting to the Specifics

- We have our streaming computing model document V2 (Oct 2024), and an evolving conception of E0-E1 dataflow and workflows, developed in an active streaming computing model meeting series and expressed in the schematic (previous slide).
- Emphasis now is moving from reports and schematics to the specifics.
- Prototyping ideas and tools in testbeds, guided by requirements.
  - Gathering input on a requirements document for streaming orchestration.
  - Testbed plans are taking concrete shape:

**Developing E0-E1 streaming workflows in a testbed utilizing Rucio and PanDA**:
- R&D instances of Rucio and PanDA are operating at BNL for this.
- Effort identified and ramping, preliminary plan taking shape.

**Calibration workflows** are to be covered as well in the testbeds:
- Describing and executing complex calibration workflows with their dependencies.
- We need a volunteer to gather calibration info across the DSCs and help establish test workflows.

**Streaming reconstruction**:
- Raw data stream to collision event identification to reconstruction and analysis.
- Effort identified, ongoing work by Jefferson Lab and University of Tokyo.

**Data handling, storage and archiving**:
- Object stores? XRootD roles? Handling fine-grained low-latency data distinct from bulk stream?

# Operation Priorities

Thomas Britton (JLab), Sakib Rahman (BNL)

## Production

- **Charge**:
  - Responsible for the coordination and production of simulation campaigns based on priorities from the Technical and Analysis Coordinators.
  - Develop automated production workflows that scale with the needs of the collaboration.
- **Priorities for 2025**:
  - **Automation Priorities**:
    - Improve the exposure and organization of monitoring so that no one needs to be an OSG expert to track progress, thereby enabling more individuals to participate in operating the monitoring.
    - Explore workflow and workload management tools.
  - **Simulation Campaign Priorities**:
    - Roll out Rucio to the collaboration as the default method for finding and accessing simulation productions.
    - Establish liaisons with DSCs and PWGs to actively participate in the simulation campaigns.

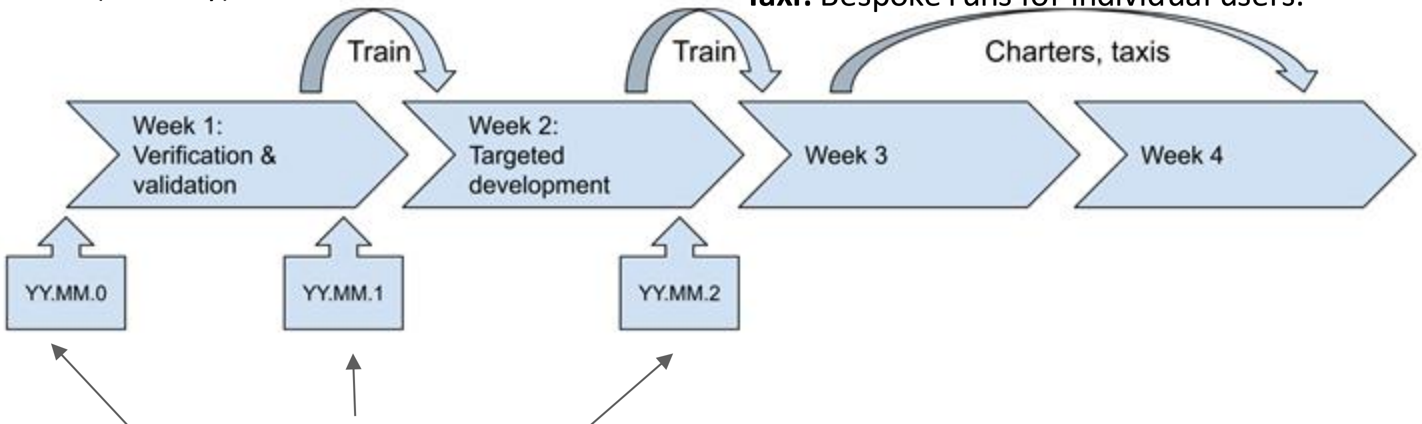
# Monthly Simulation Productions

1. **Regular updates** of simulation productions for detector and physics studies in preparation for the **preTDR**, subsequent CD milestones, and development of **early science program**.
2. **Timely validation and quality control** for simulation productions on datasets that require substantial time and resources. Focus on **benchmarks driven by Continuous Integration / Continuous Deployment**, a process that automates the testing and building of software.
3. **Continuous Deployment** of the software used for detector and physics simulations.
    - Latest merged features available the next day in nightly containers.

**Train:** Major central campaign at a fixed (monthly) schedule.

**Charter:** Special interest runs for working groups.

**Taxi:** Bespoke runs for individual users.




Tagged software releases

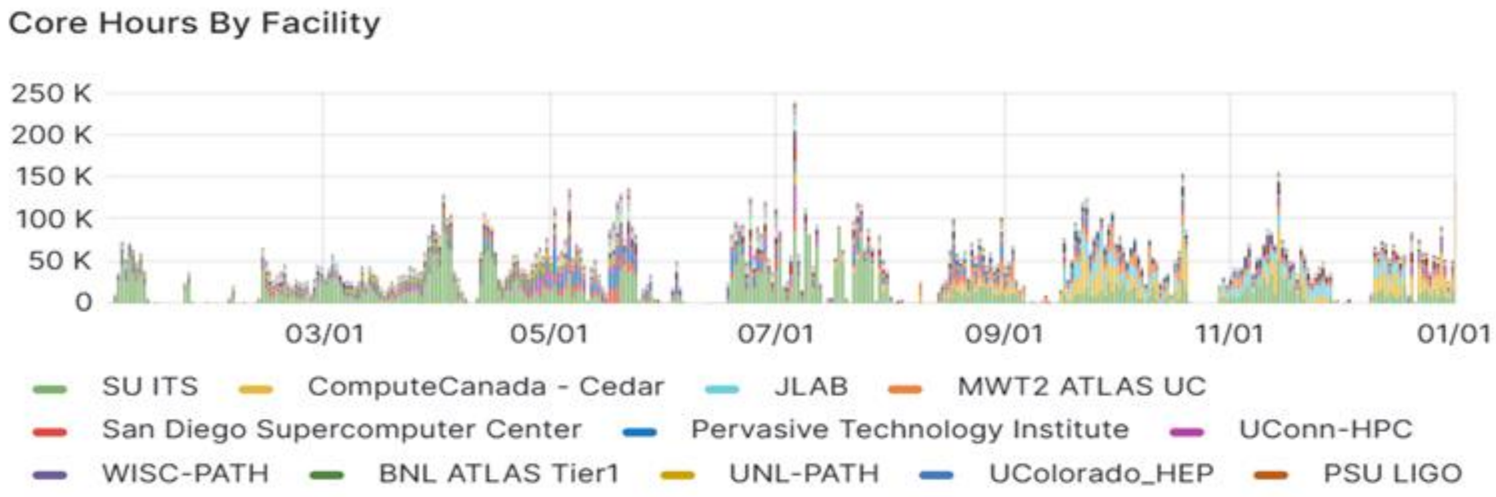
**Broad science program for the EIC**: The selection of physics processes and associated MC simulations for the preTDR has been finalized with the PWGs; however, it will continue to evolve over the coming years.

We can turn around new geometry and new algorithms within a month on the millions of events necessary to assess the impact.
In 2024, monthly simulation campaigns have used **15 M core hours on the Open Science Grid** and produced over **~500 TB of simulation data**.

# Production WG Highlights

**Infrastructure improvements**

- Improved throughput from Jefferson Lab
- Lustre disks made available at Jefferson Lab to improve performance
- Set up a cloud RSE for logging
- **Rucio instance** at Jefferson Lab
- Initiated and started testing new BNL submit host
- Added 200TB of XRootD storage at BNL with token-based read-write access
- **Successful integration of international resources**: Digital Research Alliance of Canada and INFN




**Rucio in Production**

- Successful test jobs running on OSG and writing back to XRootD storage.
- January, 2025 campaign expected to be the first campaign with Rucio.

**Simulation Production Links**

- Live campaign updates
- Input Preprocessing Policy
- Default List of Datasets
- File access instructions
- **NEW** Dataset Request Form

# Simulations Campaigns with Background

**Background Mixed Running Plan**

**Signal**:
Case 1: pythia8 DIS NC q2>1 (1 signal per integration frame)
Case 2: pythia6 SIDIS q2=0 to 1 (realistic signal frequency)
Case 3: No signal
+
Electron Beamgas (realistic frequency at 10000 Ahr)
+
Proton Beam gas (realistic frequency at 10000 Ahr)
+
Electron Synchrotron Radiation (realistic frequency)

**Directory Nomenclature**

Case 1:
/volatile/eic/EPIC/RECO/25.04.1/epic_craterlake/**Bkg_1SignalPer2usFrame/Egas_18GeV_Hgas_275GeV_Vac_10000Ahr_Synrad_18GeV**/DIS/……

| Dataset | Estimated Completion Date | Campaign |
|---|---|---|
| Case 1 18x275 | 10 May 2025 | 25.04.1 |
| Case 3 18x275 | 10 May 2025 | 25.04.1 |
| Case 2 18x275 | 10 June 2025 | 25.05.X |
| Case 1 10x275 | 10 June 2025 | 25.05.X |
| Case 2 10x275 | 10 June 2025 | 25.05.X |
| Case 3 10x275 | 10 June 2025 | 25.05.X |

# Operation Priorities

Stephen Kay (York), Holly Szumila-Vance (FIU)

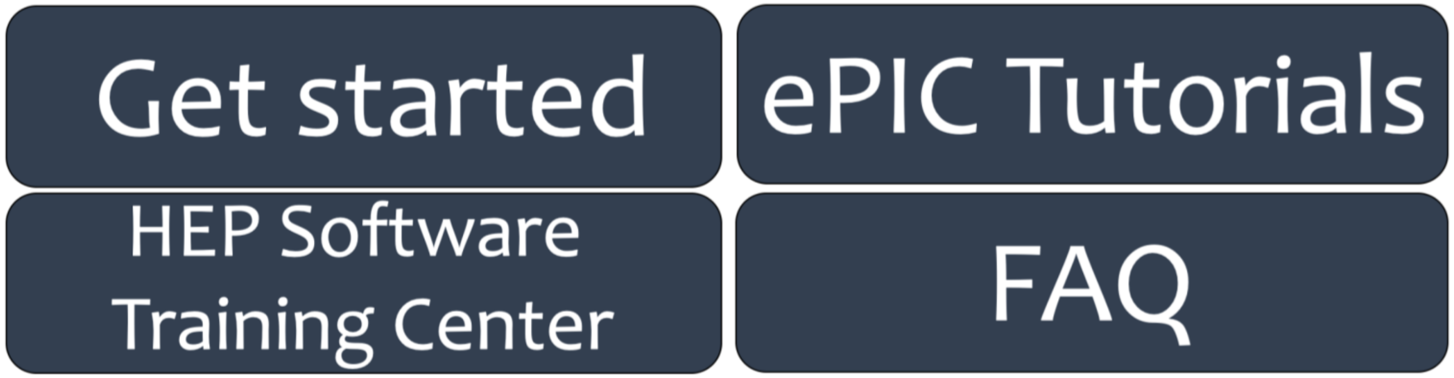
## User Learning

- **Charge**:
    - Responsible for onboarding via a landing page for new collaboration members and additional appropriate mechanisms.
    - Responsible for support via documentation, help desk, and training.
    - Ensure that software is discoverable (easy to use with only minimal instructions) and simulated data and metadata is findable.
- **Priorities for 2025**:
    - New initiative: **Roadmap towards discoverable software**.
    - **Revised and frequently updated FAQs**.
    - **Rolling schedule of software tutorials** that incorporates updated versions of existing tutorials, new material, and relevant resources from the HSF Training WG.

# User Learning Highlights

**Landing Page on https://eic.github.io/documentation/landingpage.html**:

- Full toolkit to get new users started: **“Getting Started”, software tutorials**, and **FAQs**.
- Lots of new documentation accessible to users in mostly uniform format.
- Any member of the collaboration can **directly contribute by submitting change requests**.
- Goal to fully integrate with new ePIC website.




Welcome to the **ePIC Landing Page**!

Our mailing list: eic-projdet-compsw-l@lists.bnl.gov

Subscribe here: https://lists.bnl.gov/mailman/listinfo/eic-projdet-compsw-l

**HelpDesk on Mattermost** consistently provides software support.

**New initiative for 2025** – Discoverable Software.

# User Learning Activities

**Tutorial Series Continues**:

- Understanding the Simulation Output – Shujie Li
- Analysis and Working with the Simulation Output – Stephen Kay
- Getting Started with a Physics Analysis – Alex Jentsch
- Inclusive Kinematics Reconstruction – Stephen Maple

**HSF-India/ePIC Workshop ongoing!**

- Tutorial material preparation: New Github repo with materials
- Tailoring material to audience: Pythonic focus

**User Learning is supporting PWGs with analysis tutorials**.

# Operation Priorities

## Validation Deactivation in 2025 Due to Lack of Workforce

Over the past year, the Validation WG has focused on detector and physics benchmarks. There are two primary use cases for these benchmarks: CI and data quality plots.

While there has been some progress in developing benchmarks, participation in the Validation WG meetings has been lower than in other WGs, to the point that we have not found a successor for Dmitry as convener.

Given this situation, we are deactivating the Validation WG for now, moving CI responsibilities to the Development, where they fit very well, and data quality plot responsibilities to the Production WG.

# Validation WG

**Continuous Integration (CI)**: Automated benchmarks are run with each code integration, ensuring that the code adheres to quality standards and functions as intended. These benchmarks provide immediate feedback to developers about the impact of their changes, enabling them to address issues promptly.

**Approach for CI**: **capybara** boosted code review productivity; successful adoption of Snakemake-based workflows.

**Data Quality Plots:** A standardized set of plots assists in visualizing and understanding the integrity of the data, identifying missing values or unexpected results. These plots facilitate the tracking of data quality changes over time, helping to pinpoint trends or shifts that may necessitate intervention. Additionally, making these plots accessible to everyone in the collaboration enhances communication and engagement.

**Approach for Data Quality**: **Image browser** with automated upload of CI detector benchmarks; underlying database is currently managing 250k+ images.

We benefit from excellent Continuous Integration workflows in the development of our software, and we are using the **TDR efforts** to **establish a standardized set of detector and physics performance plots**.

Here, the **collaboration-wide commitment** to ensuring that the **plots from the TDR** are **reproducible** makes a key difference.

# Image Browser at https://eic.jlab.org/epic/image_browser.html#

Home | Physics | Detector | CI | TDR | Contact

Campaign -

24.03.1
24.04.0
24.05.0
24.06.0

Plot Type -

All
scan_raw_eta=0_phi=0
scan_raw_eta=-2.5_phi=0
scan_raw_eta=2.5_phi=0

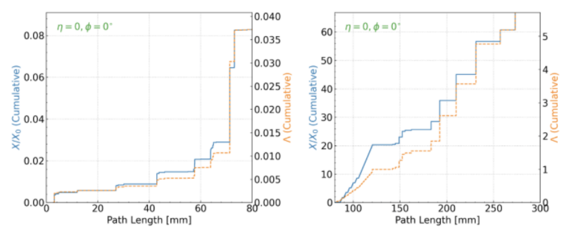
Geometry +

3 Images

Search...

Sort by

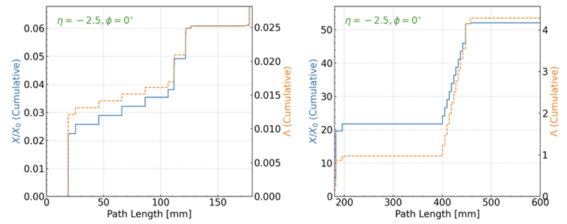
scan_raw_eta=0_phi=0




Campaign: 24.06.0
Geometry: epic_craterlake

Plots, here material scans, generated for each simulation campaign.

scan_raw_eta=-2.5_phi=0




Campaign: 24.06.0
Geometry: epic_craterlake

We reply on your input to get the user interface / user experience experience right. Please provide feedback via the **Contact** section (or GitHub issues).

# ePIC Computing & Software Reviews

**EIC Computing and Software Advisory Committee (ECSAC)**
advise host laboratories on the progress and status of computing and software for the ePIC collaboration.

**Mohammad Al-Turany (GSI)**

**Simone Campana (CERN)**

**Christoph Pauss (MIT)**

**Verena Martinez Outschoorn (UMass Amherst)**

**Frank Würthwein (chair, UCSD)**

- **Annual reviews** with a **charge reflective of** the **EIC schedule**, the **stage of the ePIC experiment**, and **impending deadlines**.
- **Recent review** on September 26–27, 2024 included: **Assessment of the ePIC Computing Model**.

# ECSAC Review

**A lot of engagement. Intense, productive discussions.**

| | |
|---|---|
| **Is there a comprehensive and cost-effective short and long-term plan for the software and computing of the experiment?** | **Yes** |
| • The pre detector technical design report (TDR) is scheduled to be delivered in 2025. Are the resources for software and computing sufficient to deliver the TDR? | **Yes** |
| • Is the design of the ePIC computing model and resource needs assessment adequate for this stage of the project? | **Yes** |
| • Is the ePIC computing and model flexible? Can it evolve and integrate new technologies in software and computing? | **Yes** |
| **Are the plans for software and computing consistent and integrated with standard practices across nuclear physics and particle physics communities, especially given technical evolution over the next decade?** | **Yes** |
| **Are the ECSJI plans to integrate into the software and computing plans of the experiment sufficient?** | **Yes** |
| **Are the plans for the integrating international partners' contributions flexible and adequate at this stage of the project?** | **Yes.** |

**Two Recommendation, to the Host Labs and ePIC:**

- Provide a detailed plan and timeline before the next ECSAC meeting for creating dedicated effort to ePIC Software & Computing team.
- Investigate how U.S. universities can contribute to the software and computing needs of the experiment, and present a plan at the next ECSAC review.

# Strategic Resource Allocation Plan

**ECSAC**: *"If you were given 4 dedicated experts, make a prioritized list of the tasks/areas where you would employ them for your short/medium/long term needs."*

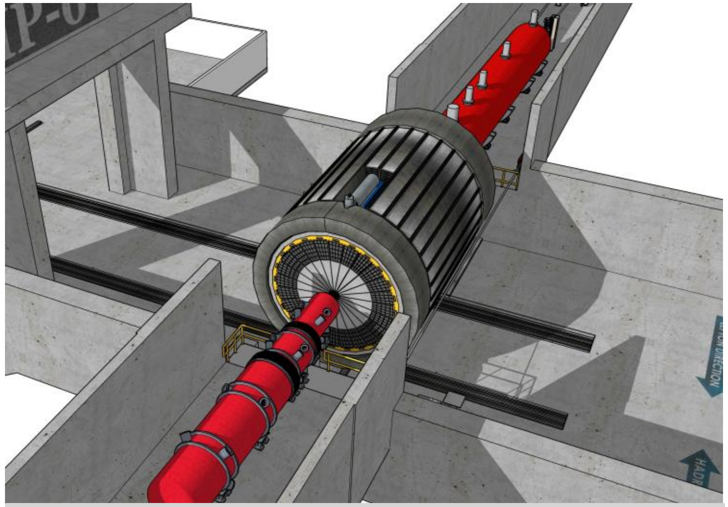
**Short/Medium-Term (next 3 years)**

- Establish a dedicated effort in collaboration with Electronics & DAQ to develop integrated DAQ-computing workflows, working towards a full streaming DAQ chain test.
- Holistic full PID full reconstruction (lepton-hadron separation, lepton ID, hadron ID) implementation in the ePIC software stack utilizing the full capabilities of the integrated detector (PID, calo, tracking, etc.).
- Support AI/ML workflow integration in full simulation and reconstruction algorithms.
- ACTS expert for track seeding, track fitting, vertex finding algorithm development, tuning, and evaluation.

**Long-Term (4+ years)**

- Continued support for streaming DAQ workflows in collaboration with Electronics & DAQ
- Expert in fast simulations to reduce the computational cost of the simulation campaigns to interpret data
- Expert in hardware accelerators to develop collaboration expertise to speed up simulation and reconstruction and leverage HPC platforms
- Distributed computing expert to develop operations between Echelon-1/2 and support progressively scaled up challenges

# Outlook



***"Software is the soul of the detector"***
(Ian Shipsey, University of Oxford)

**Great software for great science**:

- **Design and Construction:** Integrated and validated simulations are essential for evaluating performance and determining physics reach.
- **Operation:** Autonomous experimentation and control using software/AI. Distributed and rapid processing of streamed data.
- **Research:** Software and data enable discovery.

- We **work together**, on a global scale and with other fields, on great software for great science.
- We focus on **modern scientific software & computing practices** to ensure the **long-term success of the EIC scientific program** throughout all CD milestones.

**Discussion Topics**:

**There are enormous opportunities for India to play a leading role in software, computing, and associated physics studies. ePIC Software & Computing is excited about these opportunities.**

# How to Engage

## How to Engage

- Use the **landing page** to begin using the ePIC software and simulations.
- Join the **HelpDesk on Mattermost** for questions related to all aspects of software and simulations.
- Attend our **weekly meetings** on Wednesdays at 11:00 a.m. (EDT).
  - Follow our Software News for updates from WGs.
- Participate in **WG discussions** based on your interests.
  - Refer to our meeting notes for discussion outcomes and action items.
  - Collated meeting notes for each WG are available through links in the corresponding Indico events.

# Backup

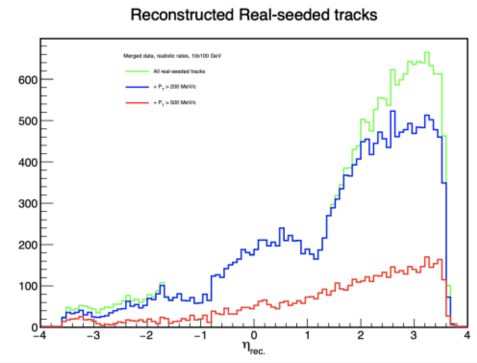
# Prototype of Event Reconstruction from Streaming Data

**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.

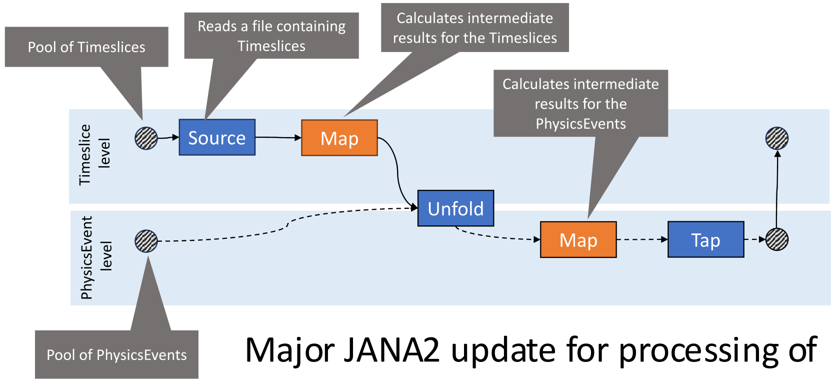
**Streaming DAQ** → **Timeframes** → **Streaming Computing** → **Reconstructed Events**

Continuous stream of data

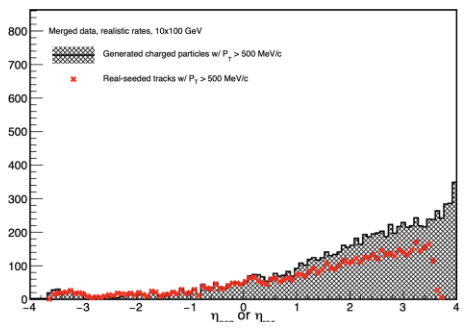
0.6ms timeframes



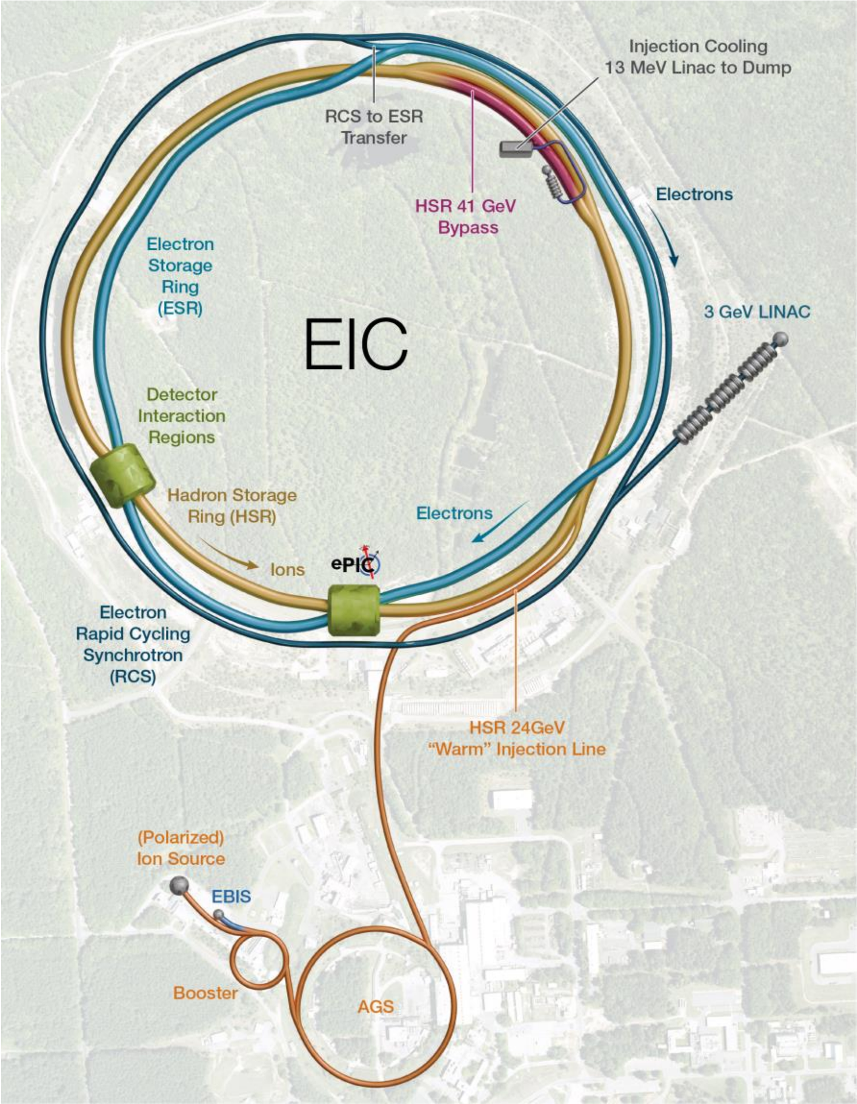

- Data transferred in collections called *timeframes*.
- Each timeframe includes:
  - Data read from detectors over a time window of $2^{16}$ cycles of the beam RF, equivalent to 0.6 ms.
  - Channel information and corresponding timing data




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



# 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:
  - For the computing resource estimate, we assume a luminosity scenario of L = $10^{33}$ $cm^{-2}$ $s^{-1}$ = 1 kHz/μb in 2034.

# 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 $710 \times 10^9$ events per year.
  - The data volume of 35.5 PB per year will be replicated between Echelon 1 facilities (71 PB in total).

# 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 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.