Markus Diefenthaler (Jefferson Lab)

ePIC Software & Computing Report

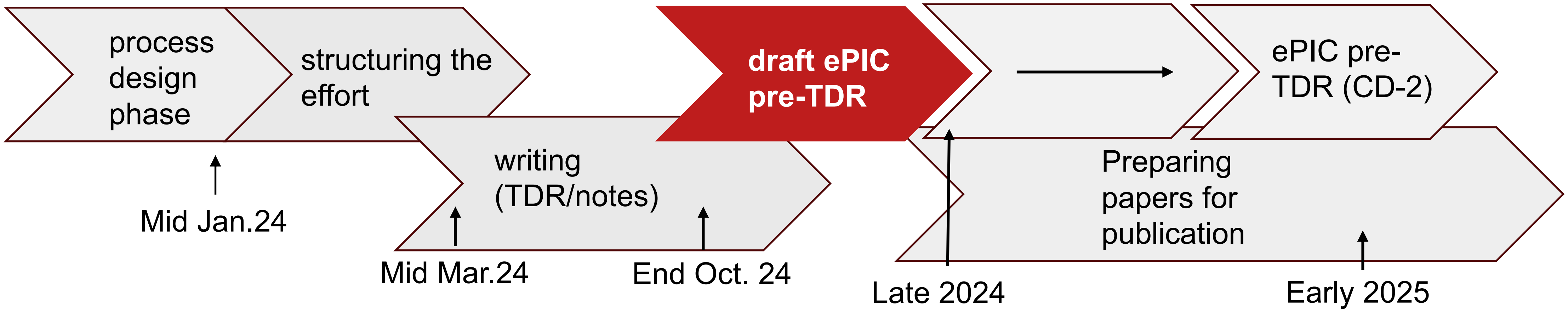
Software & Computing Priorities

Technical Design Report

Collaboration Engagement

Computing Model

TDR Strategy and Publications



- ePIC Software & Computing is essential to the TDR, providing **advanced software** and **simulation productions** that are the **input for detector and physics studies**:
  - ***"Software and Simulation Readiness for TDR"*** parallel session and plenary discussion at the ANL collaboration meeting, where we extensively defined the remaining development tasks, drawing on substantial and significant input from the collaboration at large.
  - **Good progress** since ANL collaboration meeting, improving the accuracy of the simulations and building up the reconstruction in shared priorities with DSCs and PWGs.
  - Utilized workfests, such as *"Path Towards Holistic Reconstruction",* to start defining development tasks for the remainder of the year.

Key Achievements and Milestones

Physics and Detector Simulations:

- Physics events and background merging at the HepMC level implemented.
- Geometry update and detailed DD4hep implementation for all sub-detectors - in agreement with geometry database.
- Initial support and service structure implementation.

Reconstruction Frameworks:

- Integrated cleanly with the PODIO-based data models and other layers of the key4hep stack.
- Enabling external configuration of existing components.
- Supporting timeframe splitting for streaming readout.

More throughout the talk. →

Reconstruction Algorithms in EICrecon:

- Developed a full chain of ACTS-based track reconstruction with auto material map generation, realistic seed finder, combinatorial track finding/fitting, ambiguity solver, and primary vertexing.
- Implemented initial prototype of electron-finder, which includes the implementation of track-calorimeter projections
- Made progress towards the implementation of particle flow in ePIC.
- Made other significant improvements across EICrecon, e.g. in the jet software.

Software Containers:

- CUDA-enabled containers, including PyTorch with CUDA support for GPU-accelerated AI/ML training.
- ONXX support.

Interplay of Software & Computing and Physics

Monthly meetings of the PWGs and Software & Computing

In addition, regular meetings of ACs and SCCs:

- Enhance communication.
- Specify the interface between reconstruction and physics analysis.
- Foster methods for rapid prototyping and integration of these prototypes into the ePIC Software Stack.
- Determine common priorities and implement agile strategies to achieve them efficiently.

AC/SCC MEETINGS

Reconstruction Tasks: Shared Priorities with PWGs

Electron Finder: Developing an algorithm for identifying electrons and the scattered electron of the DIS process.

- **Overall status**: Successfully closed in April release [report from General meeting].

Vertexing and PID: Enhancing the vertexing and PID capabilities to study heavy flavor physics.

- **Overall status**: **Primary** vertexing integrated [latest status update]. PID LUTs as interim solution for PID algorithms.
- **Next step**: Close out task by defining default parameters for primary vertexing for simulation campaigns.

Secondary Vertexing: Enhancing the vertexing and PID capabilities to study heavy flavor physics.

- **Overall status**: New reconstruction task on secondary vertexing led by Nicolas Schmidt (ORNL) [kick-off report].
- **Next steps:** Updates on secondary vertexing studies using ACTS and KFParticle.

Particle Flow: Improving the jet reconstruction using particle flow information. Not a priority for (pre?)TDR.

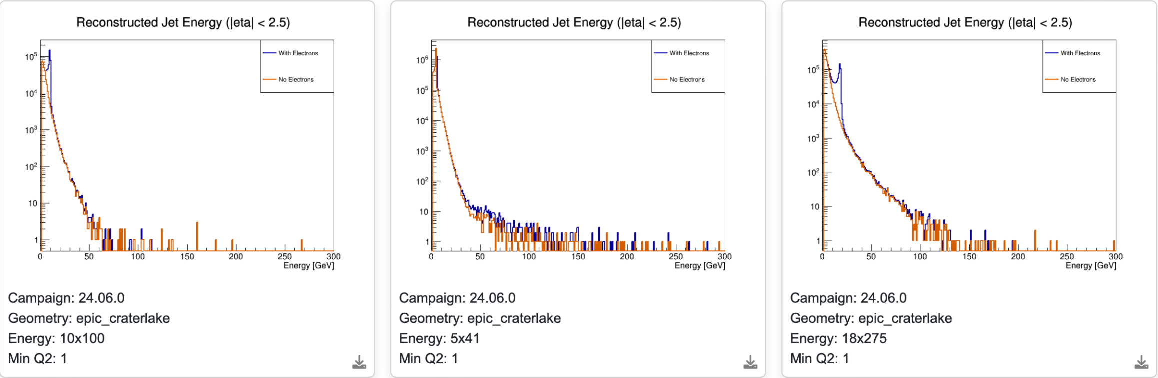
- **Overall status on calorimeter reconstruction**: Good progress on TDR priorities [report from CERN meeting].
- **Next steps**: Tuning parameters for track-based cluster merging [1406].

Low Q^2: Integration of the low-Q2 tagger for precise measurements of photo and vector mesons production.

- **Overall status**: Baseline implementation available, integrated ML model in EICrecon using TMVA.
- **Next steps**: Discuss remaining tasks before closing out reconstruction task.

Physics Studies on Simulation Campaigns (Selected Examples)

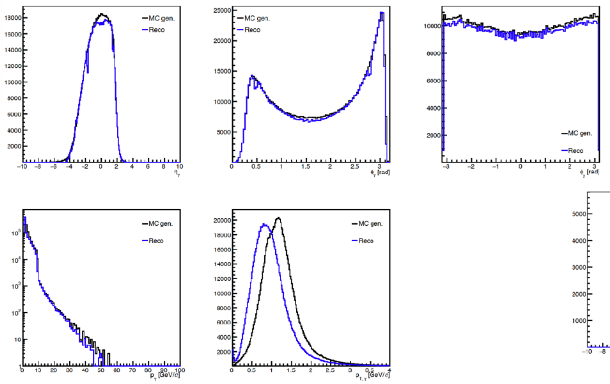
Progress report on jet and heavy flavor physics studies, presented by Brian Page (BNL).



Plots part of physics benchmarks and available in image browser (slide 11).

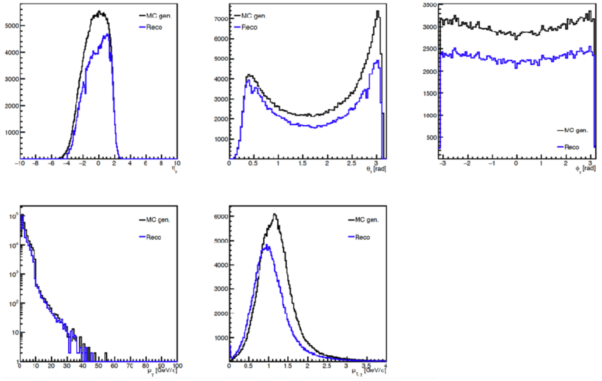
Progress report on exclusive physics studies, presented by Rachel Montgomery (University of Glasgow).

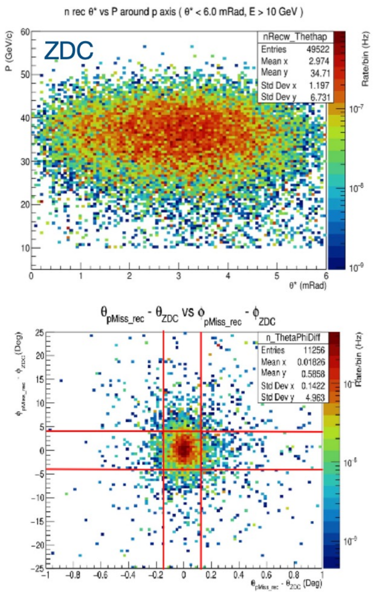
10 x 100, monthly production 24.04.0

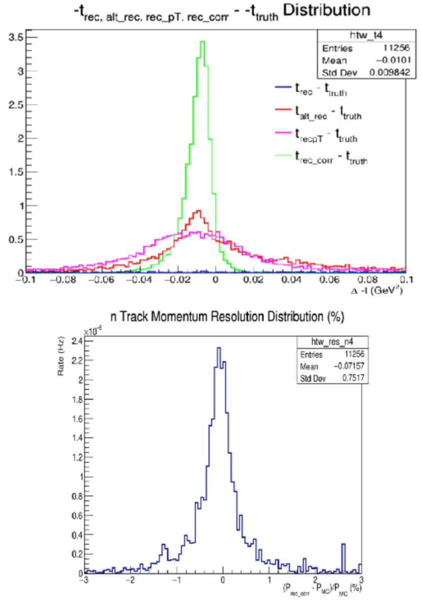


DVCS photons, studies by O. Jevons (University of Glasgow).

10 x 100, monthly production 24.06.0

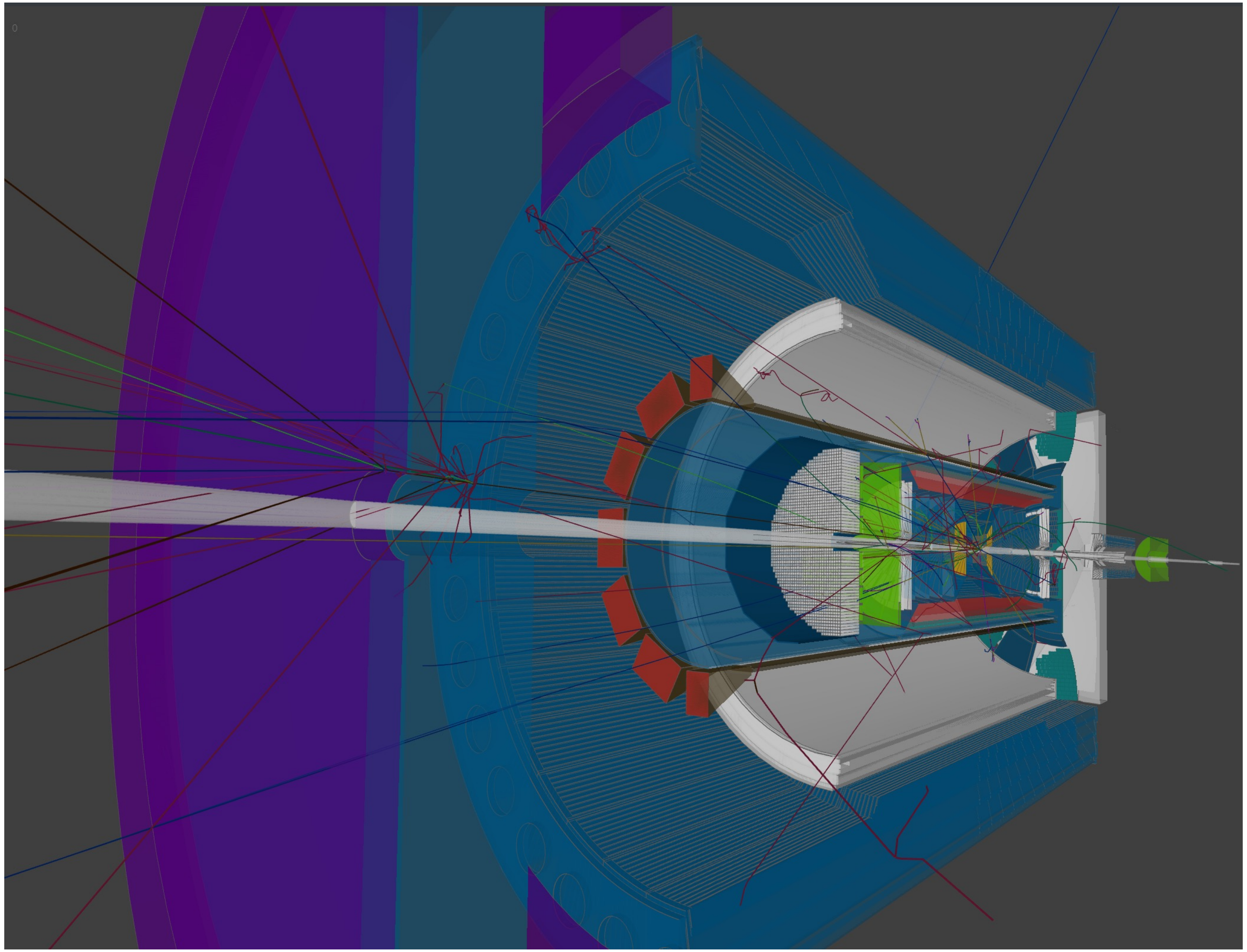






Meson form factors, Love Preet et al. (University of Regina and University of York).

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



Event Display for Reconstruction Development:

Available at:
https://eic.github.io/firebird/display

Features:

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

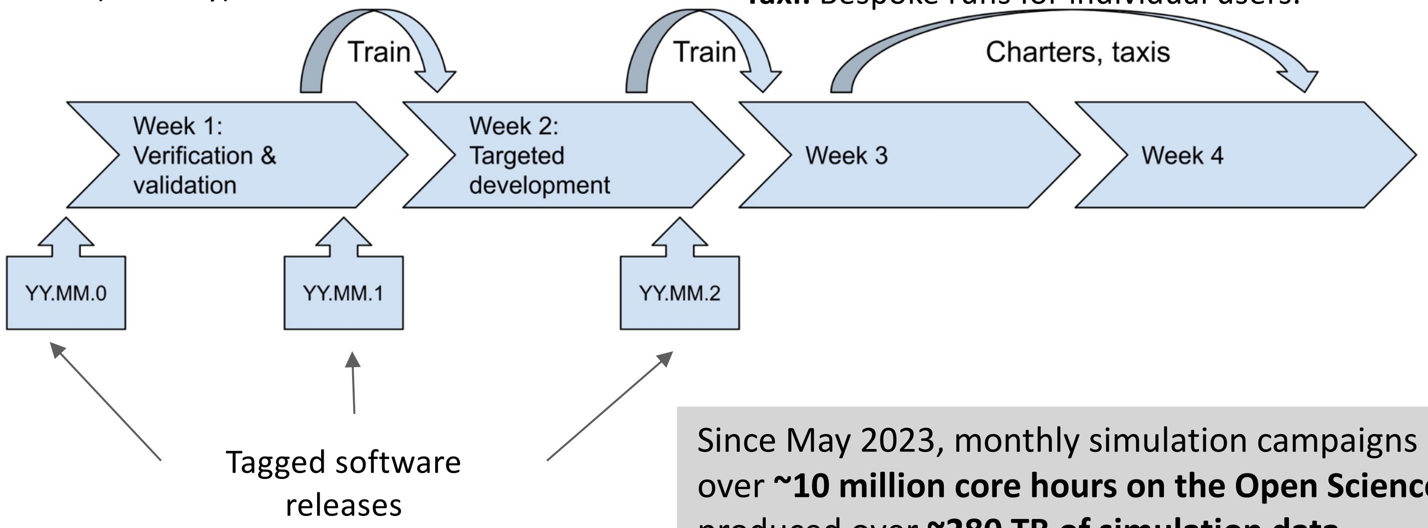
Monthly Simulation Productions

1. **Continuous deployment** of the software used for detector and physics simulations.
2. **Regular updates** of simulation productions for detector and physics studies in preparation for the TDR (and subsequent CD milestones).
3. **Timely validation and quality control** for simulation productions on datasets that require substantial time and resources. Focus on **benchmarks driven by Continuous Integration** (CI), a process that automates the testing and building of software.

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

Charter: Special interest runs for working groups.

Taxi: Bespoke runs for individual users.



Broad science program for the EIC: The selection of physics processes and associated MC simulations for the TDR has been finalized with the PWGs.

We will also re-add single productions.

Since May 2023, monthly simulation campaigns have used over **~10 million core hours on the Open Science Grid** and produced over **~280 TB of simulation data**.

Validation Thrusts

Continuous Integration: 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.

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.

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**.

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

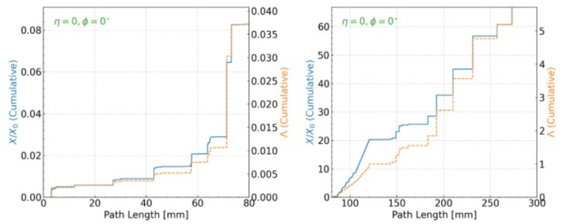
Geometry +

3 Images

Search...

Sort by

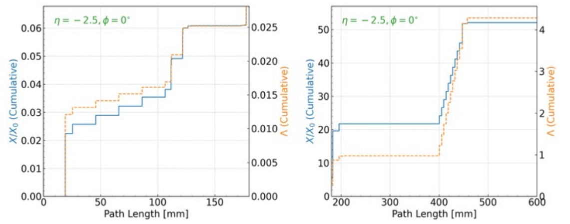
scan_raw_eta=0_phi=0



Campaign: 24.06.0
Geometry: epic_craterlake

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

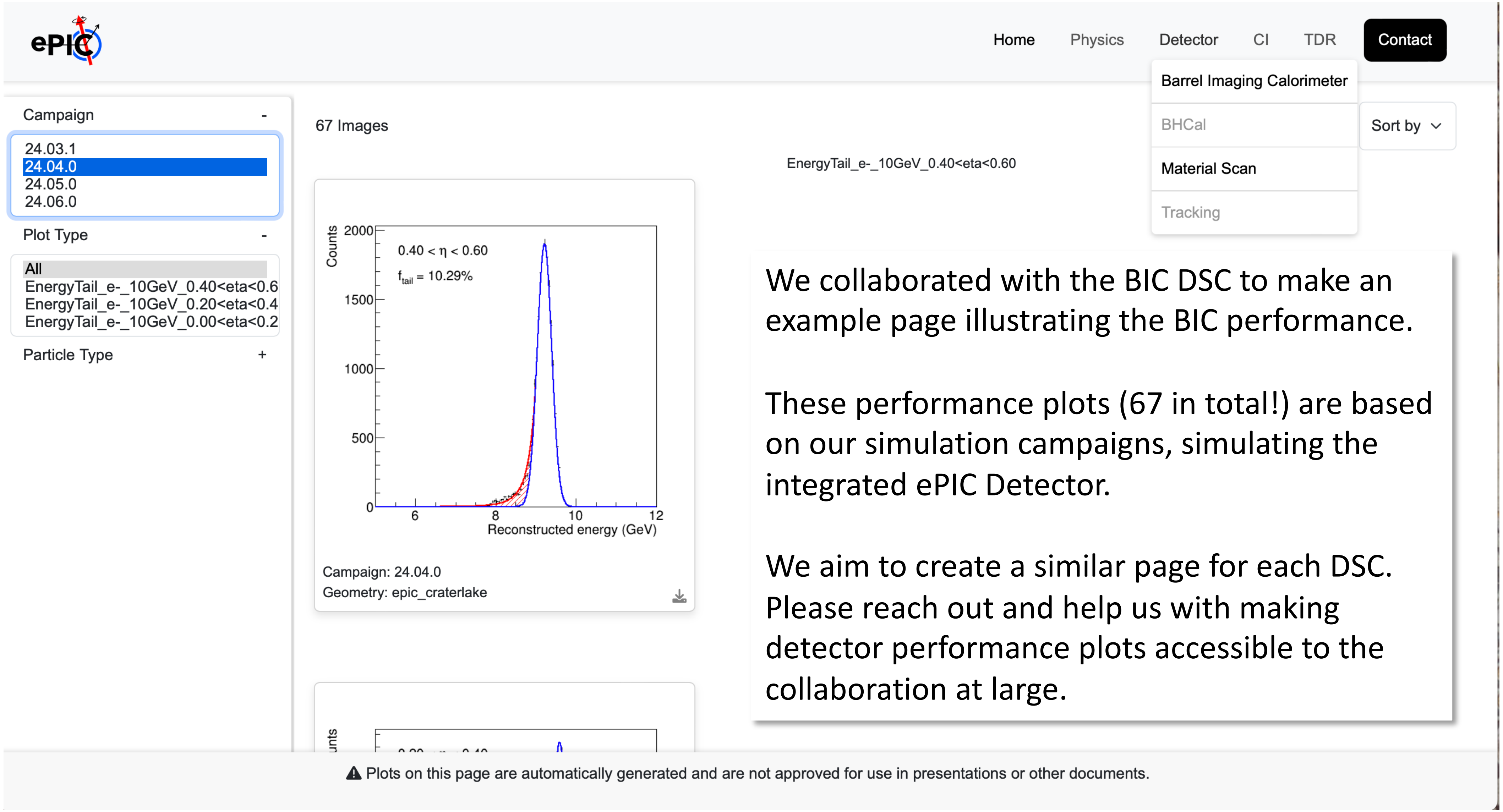
scan_raw_eta=-2.5_phi=0



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).

Standardized Sets of Plots



We collaborated with the BIC DSC to make an example page illustrating the BIC performance.

These performance plots (67 in total!) are based on our simulation campaigns, simulating the integrated ePIC Detector.

We aim to create a similar page for each DSC. Please reach out and help us with making detector performance plots accessible to the collaboration at large.

Software & Computing Priorities

Technical Design Report

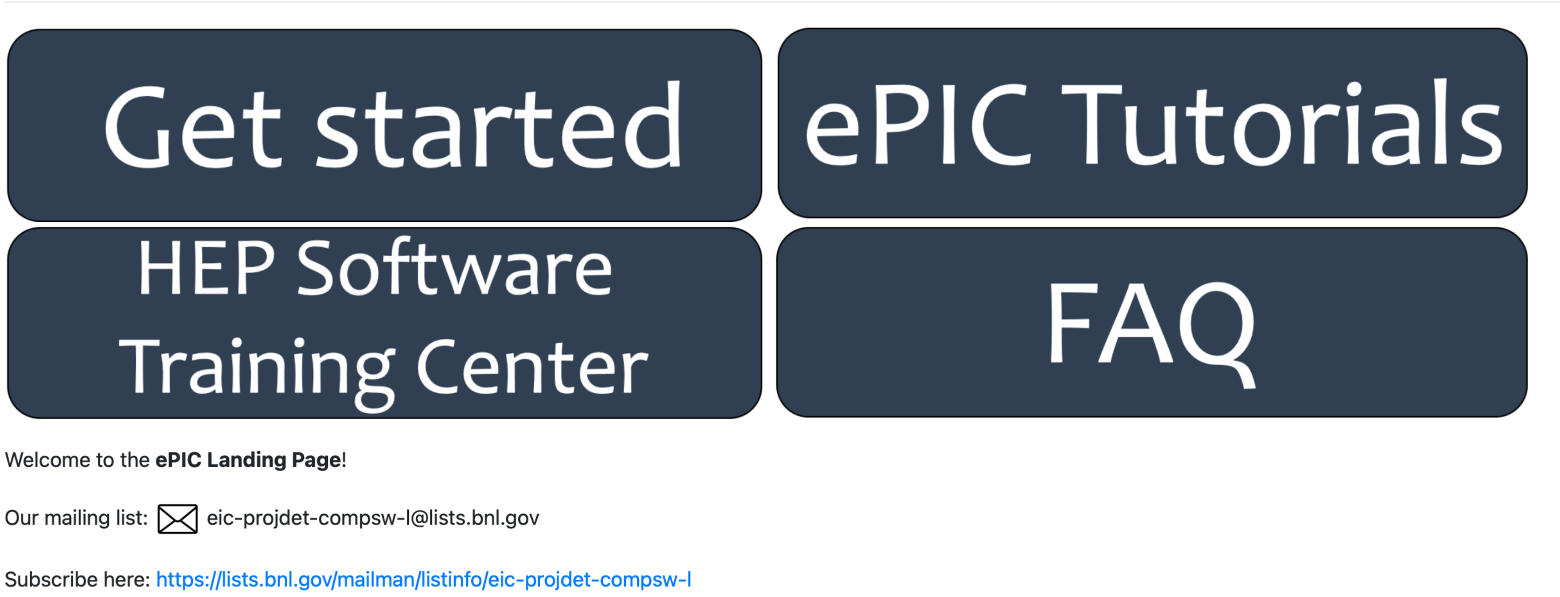
Collaboration Engagement

Computing Model

Onboarding

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

- Continuously updated and improved list of **useful links**, **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**.


Get started

ePIC Tutorials

HEP Software Training Center

FAQ

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.

Recent Success Stories: Collaboration members independently onboarding themselves using the Landing Page, immediately afterwards becoming actively involved in detector and physics studies for the TDR.

Software Tutorials

Ongoing tutorials series, most recent during CERN meeting, covering four key topics:

1. **Overview of ePIC Software** (Holly Szumila-Vance, Jefferson Lab)

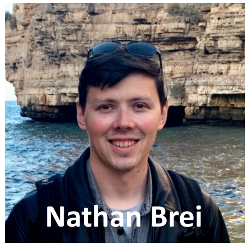
   Eic-shell Easy to get started locally… in only 1 line!

   ```
   curl -L get.epic-eic.org | bash
   ```

   Based on container images, the same images are used for simulation campaigns.

2. **Working with Simulation Output** (Stephen Kay, University of York)
3. **Simulating Detectors and Their Readout** (Simon Gardner, University of Glasgow)
4. **Reconstruction Algorithms** (Nathan Brei, Jefferson Lab)

45 participants from Africa, Asia, and Europe.



ePIC

Continuing our Tutorial Series

Next Tutorial: Working with Simulation Output
Presenter: Stephen Kay (University of York)
Date: Thursday, September 12
Time:

- 10:00 a.m. (BST) for Africa, Asia, Australia, and Europe.
- 4:00 p.m. (BST) for Africa, Europe, North America, and South America.

Upcoming Tutorials:

- Reconstruction Algorithms
- Validation
- "Working with Simulation Output" on a monthly basis,
  - With a focus on the changes to the software and simulation campaigns.

We **welcome suggestions for future tutorials**, including shorter tutorials on specialized topics, such as calorimeter reconstruction as suggested during the workfests.

Community Building

Regular meetings to drive forward priority targets and provide an avenue for new collaboration members to engage.

Summary: Meeting enabled software progress, pushed the review preparations, and informed our planning.

Summary: Status and plans; software and simulations for TDR, tutorials; streaming computing; software projects with HEP.

ePIC Software & Computing at CERN, April 22-26

DALL-E, could you please draw a picture of exciting discussions at CERN?

- May 1 marked the first anniversary of our WG structure and conveners being endorsed by the CC:
  - Essential review of the **WG status and plans**,
  - Yielding numerous concrete action items that have been and are being addressed.
- Critical review of **TRD tasks** and follow-up on priority items.
- **Software tutorials** with strong engagement from collaboration members in Africa, Asia, and Europe.
- Advancement of our activities in **streaming computing** and related to that distributed computing.
- Strengthen **collaboration with HEP**, specifically ACTS, CERN EP-SFT, HSF, key4hep, Rucio.

Detailed report in ePIC General Meeting on May 17.

Open, Collaborative Software Development

Encourage Upstream Contributions

- Requirements of well-formed HepMC as input has resulted in real improvements to multiple MCEGs used by EIC community.
- Various upstream contributions to DD4hep, ACTS, Spack, uproot,...

Encourage Social Coding

- CI platform provides the incentive for developers to commit code frequently: achieving data management and analysis preservation goals.
- Pull request reviews to ensure higher quality code and build developer skills.
    - **Current priority**: Expanding the review team and providing documentation on the purpose of code reviews.

Enable Access Without Restrictions

- ePIC collaboration members include over 170 institutions worldwide
- Data ‘publicly’ available through BNL S3 and publicly available through JLab xrootd.
- Flat data structures (i.e. could be a csv), stored as ubiquitous ROOT trees without need for data structure libraries.
- Support for uproot using numpy library.

Approaches Under Evaluation

- Rucio for data management (moving into production).
- Reproducible analysis workflow tools.
- Workflows for training ML models.

Software & Computing Priorities

Technical Design Report

Collaboration Engagement

Computing Model

Compute-Detector Integration to Accelerate Science

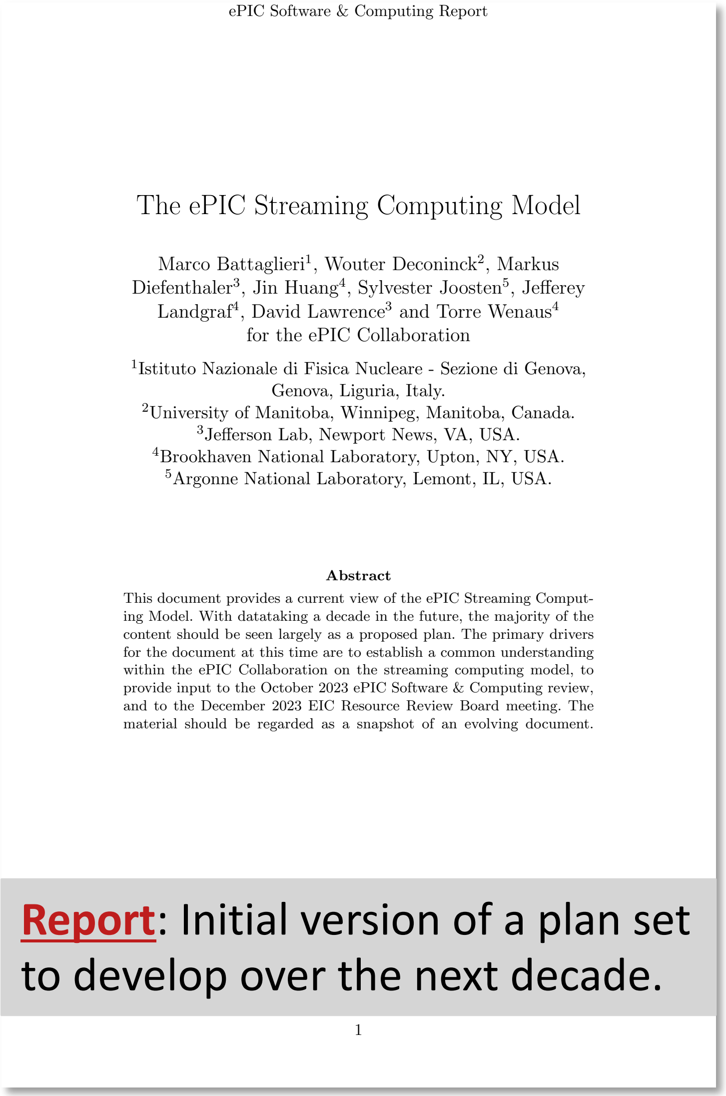
- **Problem** Data for physics analyses and the resulting publications available after O(1year) due to complexity of NP experiments (and their organization).
  - Alignment and calibration of detector as well as reconstruction and validation of events time-consuming.
- **Goal** **Rapid turnaround of 2-3 weeks for data for physics analyses**.
  - Timeline driven by alignment and calibrations.
  - Detailed discussion of alignment and calibration workflows with DSCs indicates that a **2-3 week timeline is feasible**; update on discussion as part of the Streaming Workfest.
- **Solution** Compute-detector integration using:

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

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

Heterogeneous computing for acceleration.

ePIC Streaming Computing Model



ePIC Software & Computing Report

The ePIC Streaming Computing Model

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

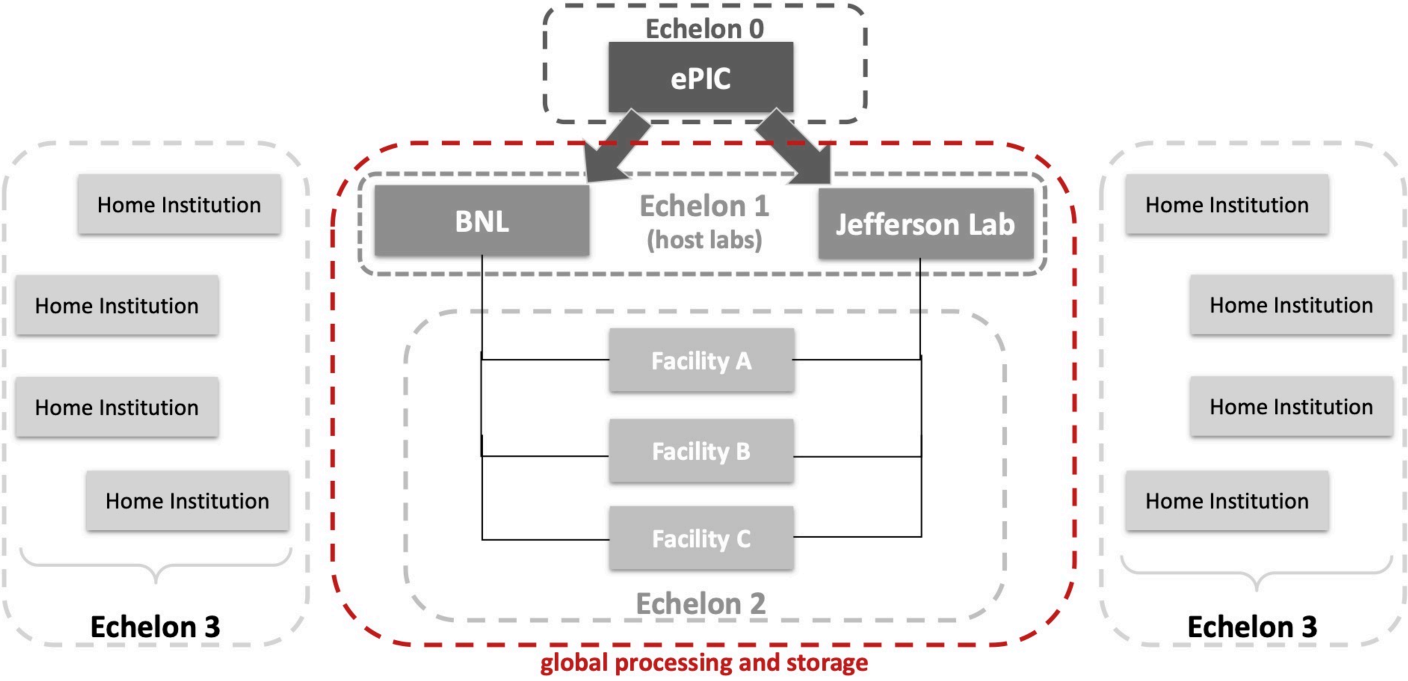
[1]Istituto Nazionale di Fisica Nucleare - Sezione di Genova, Genova, Liguria, Italy.
[2]University of Manitoba, Winnipeg, Manitoba, Canada.
[3]Jefferson Lab, Newport News, VA, USA.
[4]Brookhaven National Laboratory, Upton, NY, USA.
[5]Argonne National Laboratory, Lemont, IL, USA.

Abstract

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

1

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



Echelon 0: ePIC experiment.

Echelon 1: Crucial and innovative partnership between host labs.

Echelon 2: Global contributions.

Echelon 3: Full support of the analysis community.

Global Computing Infrastructure Contributions

From review: *"There are clearly very significant opportunities in in-kind computing infrastructure contributions."*

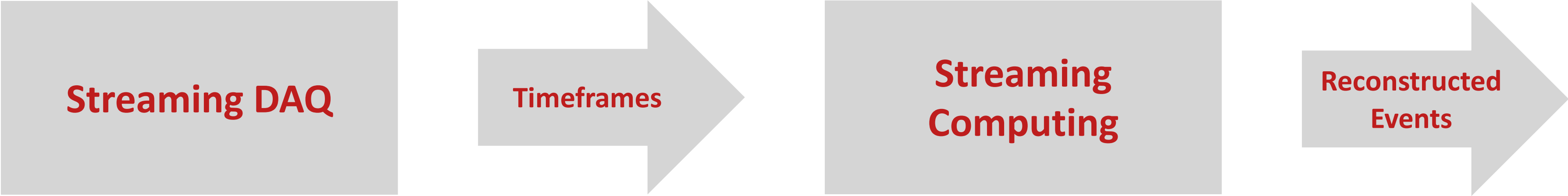
Use Case	Echelon 0	Echelon 1	Echelon 2	Echelon 3
Streaming Data Storage and Monitoring	✓	✓		
Alignment and Calibration		✓	✓	
Prompt Reconstruction		✓		
First Full Reconstruction		✓	✓	
Reprocessing		✓	✓	
Simulation		✓	✓	
Physics Analysis		✓	✓	✓
AI Modeling and Digital Twin		✓	✓	

- Design of computing model aims for **effective integration and management** of **global computing infrastructure contributions**.
- **Canada**, **Italy**, and **United Kingdom** engaged as a proof of concept in this context:
  - Integration of resources from international partners into simulation campaigns on Open Science Grid.
- Representatives from ePIC and international partners will manage the **EIC International Computing Organization** (EICO) under the auspices of the EIC Computing and Software Joint Institute (ECSJI).

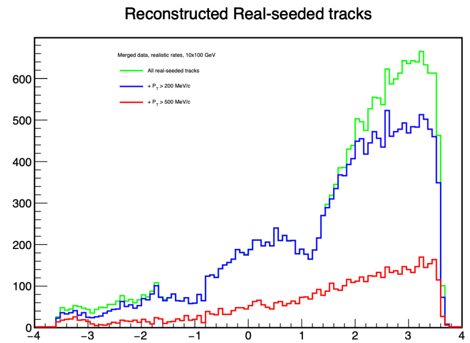
Prototype of Event Reconstruction from Realistic Timeframes

Discussed in part at [Streaming Workfest](#)

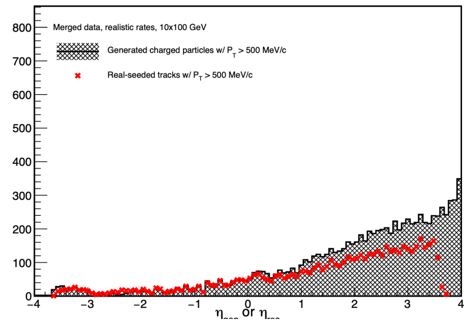
Scope of the first prototype: Track reconstruction only. Demonstrated that we can correlate hits in a realistic time frame to the various events in the time window of the MAPS of 2µs.



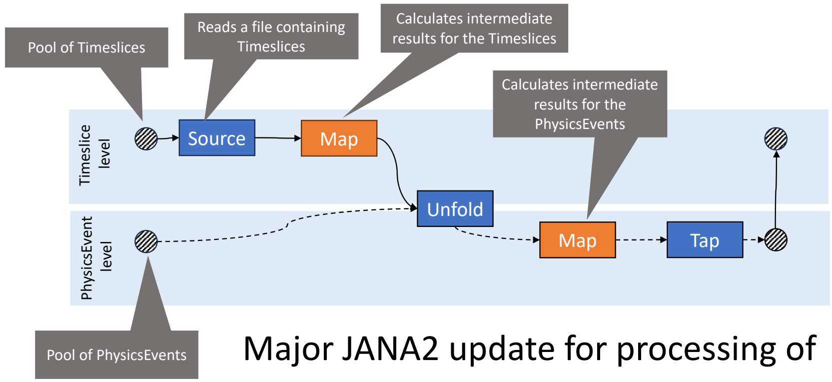
0.6ms timeframes



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 Computing Model Milestones

Milestones for TDR:

- Prototype of event reconstruction from realistic frames (**done**).
- Quantitative computing model (**work in progress**).
- Publication of the existing report on the "ePIC Streaming Computing Model":
  - Currently under revision for the ePIC Software & Computing review on Sept. 26–27.

Milestones During Detector Construction Phase:

Discussed in part at Streaming Workfest

- **Ongoing planning with Electronics and DAQ WG to align on shared priorities**:
  - Provisioning DAQ and software sufficient for test beams:
    - Small scale real-world testbeds for the developing DAQ and software.
    - We welcome the use of EICrecon in test beams and have begun developing the first plugin for reading data.
    - Streaming challenges exercising the streaming workflows from DAQ through offline reconstruction, and the Echelon 0 and Echelon 1 computing and connectivity. We have successfully streamed EDM4eic data.
- Data challenges exercising scaling and capability tests as distributed ePIC computing resources at substantial scale reach the floor, including exercising the functional roles of Echelon 1 and 2, particularly Echelon 2, the globally distributed resources essential to meeting ePIC's computing requirements.
- Analysis challenges exercising autonomous alignment and calibrations.
- Analysis challenges exercising end-to-end workflows from (simulated) raw data to exercising the analysis model.

Summary

- The ePIC software stack is a **modern** and **modular toolkit** built from NP/HEP community tools and components from HPC and Data Science; ePIC is an **active** member of the **software & computing community** in NHEP.

- **High level milestones** ensures 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. 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.

Many thanks to the deputy SCCs (Wouter Deconinck, Dmitry Kalinkin, Torre Wenaus)**, WG conveners** (Derek Anderson, Marco Battaglieri, Thomas Britton, Torri Jeske, Kolja Kauder, Stephen Kay (nominated) Jeffery M. Landgraf (nominated), Shujie Li, Jin Huang, Chao Peng, Sakib Rahman, Holly Szumila-Vance)**, and the many other contributors to ePIC Software & Computing!!! Many thanks to the ACs** (Salvatore Fazio, Rosi Reed) **for rocking the AC/SCC coordination!**