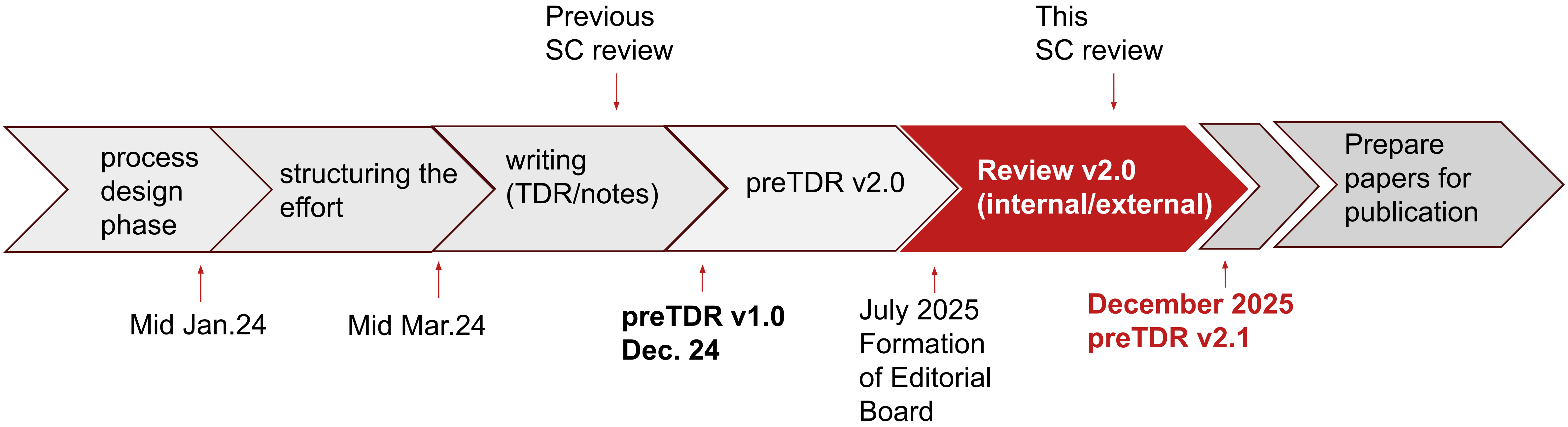
Supporting preTDR and TDR Preparations

Dmitry Kalinkin (Brookhaven National Laboratory)

Charge questions

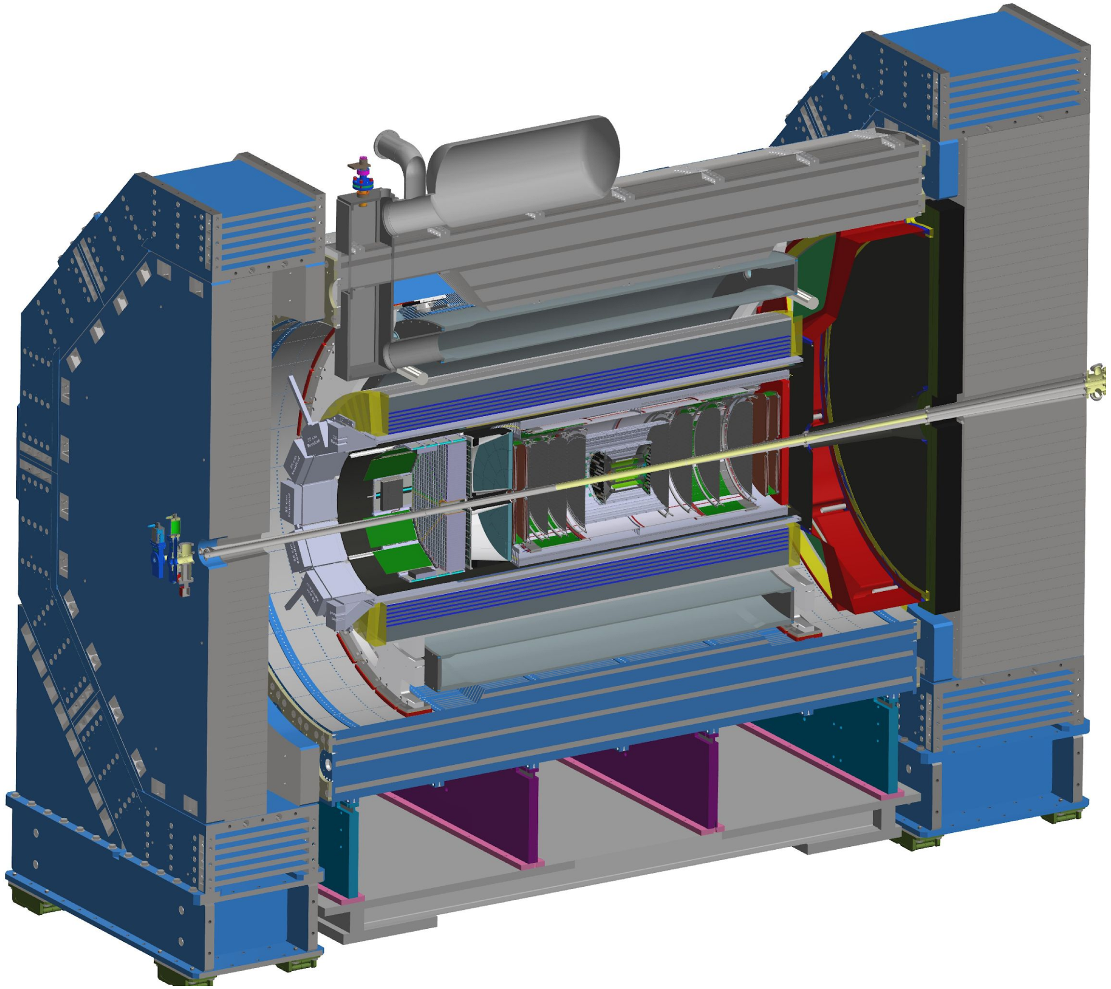
1. **ePIC Computing and Software: Assess the progress towards TDR readiness and implementing the ePIC computing model.** As appropriate, assess the plans and outcomes for data challenges and otherprototypes/testbeds. Assess the interfaces between the EIC detector DAQ and the ePIC computing and software organization as appropriate.
2. **Resources: Assess the short-term and long-term resource planning relative to the near-term outcomes and long-term goals. As appropriate, comment on the delivery of resources by international partners.**
3. **ECSJI**: Evaluate short and long terms support resource planning and delivery by the host laboratories through the ECSJI. As appropriate, evaluate the collaboration and division of responsibilities between the two labs.
4. **Broader Community**: Assess as appropriate engagements with aligned Software and Computing organizations across nuclear physics and particle physics communities and other relevant communities.
5. **Overall**: Assess the ePIC and ECSJI communication and collaboration. Assess the response to previous review recommendations.

pTDR and TDR timeline



- Relevant chapters for experimental systems and detector performance for EIC science program
- Software deliverables:
  - coordinate development of simulation and reconstruction software and integration into ePIC software stack
  - continued support with monthly software releases and simulation campaigns

"Software is the Soul of the Detector"

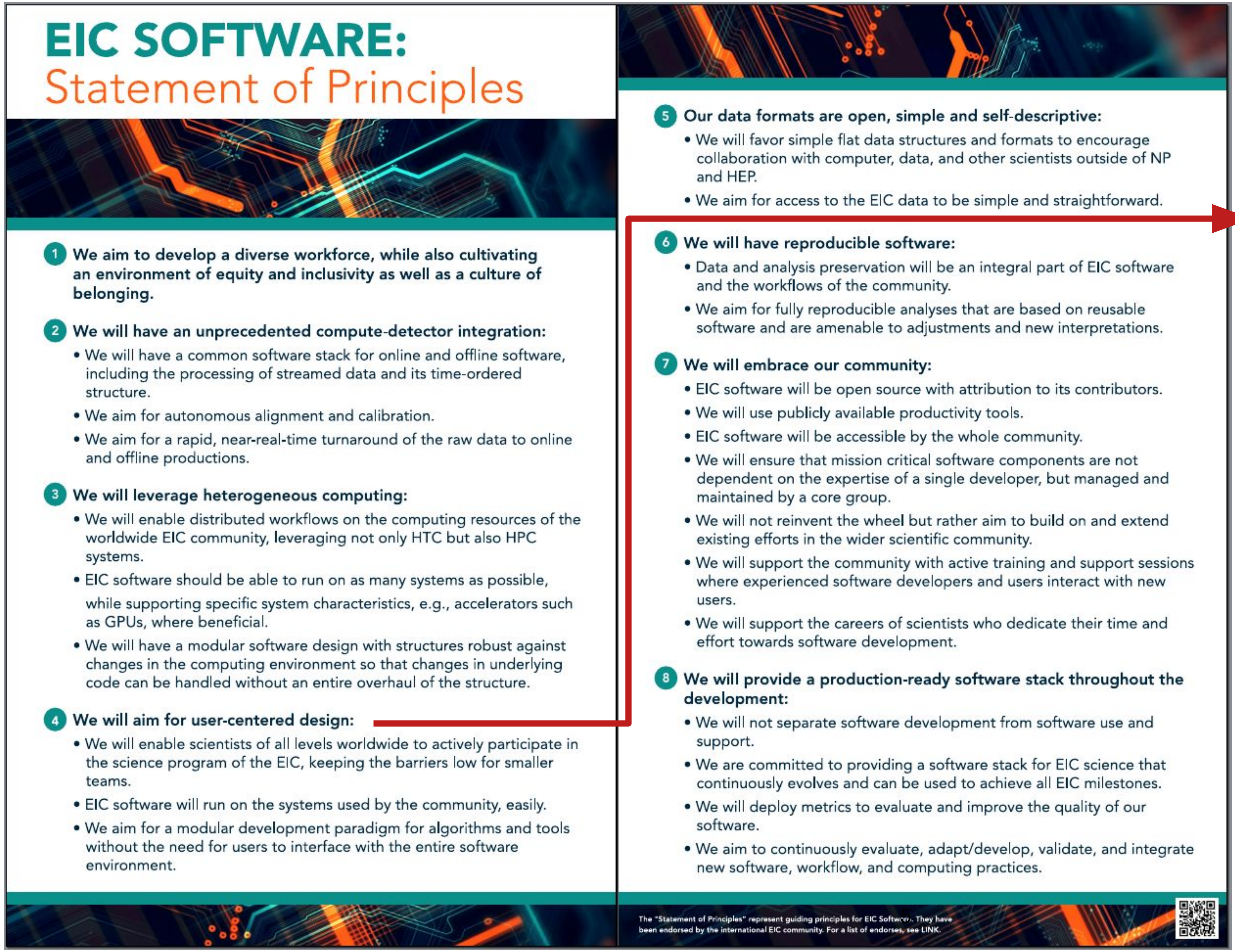


Great Software for Great Science:

- **Design and Construction:** Integrated and validated simulations are essential for evaluating detector performance and determining physics reach.
- **Operation:** Rapid processing of streamed data using streaming readout, AI, and distributed computing. Autonomous experimentation and control.
- **Physics Analysis:** 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**.

User-Centered Design at EIC



EIC SOFTWARE:
Statement of Principles

1. **We aim to develop a diverse workforce, while also cultivating an environment of equity and inclusivity as well as a culture of belonging.**

2. **We will have an unprecedented compute-detector integration:**
   - We will have a common software stack for online and offline software, including the processing of streamed data and its time-ordered structure.
   - We aim for autonomous alignment and calibration.
   - We aim for a rapid, near-real-time turnaround of the raw data to online and offline productions.

3. **We will leverage heterogeneous computing:**
   - We will enable distributed workflows on the computing resources of the worldwide EIC community, leveraging not only HTC but also HPC systems.
   - EIC software should be able to run on as many systems as possible, while supporting specific system characteristics, e.g., accelerators such as GPUs, where beneficial.
   - We will have a modular software design with structures robust against changes in the computing environment so that changes in underlying code can be handled without an entire overhaul of the structure.

4. **We will aim for user-centered design:**
   - We will enable scientists of all levels worldwide to actively participate in the science program of the EIC, keeping the barriers low for smaller teams.
   - EIC software will run on the systems used by the community, easily.
   - We aim for a modular development paradigm for algorithms and tools without the need for users to interface with the entire software environment.

5. **Our data formats are open, simple and self-descriptive:**
   - We will favor simple flat data structures and formats to encourage collaboration with computer, data, and other scientists outside of NP and HEP.
   - We aim for access to the EIC data to be simple and straightforward.

6. **We will have reproducible software:**
   - Data and analysis preservation will be an integral part of EIC software and the workflows of the community.
   - We aim for fully reproducible analyses that are based on reusable software and are amenable to adjustments and new interpretations.

7. **We will embrace our community:**
   - EIC software will be open source with attribution to its contributors.
   - We will use publicly available productivity tools.
   - EIC software will be accessible by the whole community.
   - We will ensure that mission critical software components are not dependent on the expertise of a single developer, but managed and maintained by a core group.
   - We will not reinvent the wheel but rather aim to build on and extend existing efforts in the wider scientific community.
   - We will support the community with active training and support sessions where experienced software developers and users interact with new users.
   - We will support the careers of scientists who dedicate their time and effort towards software development.

8. **We will provide a production-ready software stack throughout the development:**
   - We will not separate software development from software use and support.
   - We are committed to providing a software stack for EIC science that continuously evolves and can be used to achieve all EIC milestones.
   - We will deploy metrics to evaluate and improve the quality of our software.
   - We aim to continuously evaluate, adapt/develop, validate, and integrate new software, workflow, and computing practices.

The "Statement of Principles" represent guiding principles for EIC Software. They have been endorsed by the international EIC community. For a list of endorses, see LINK.

User-Centered Design

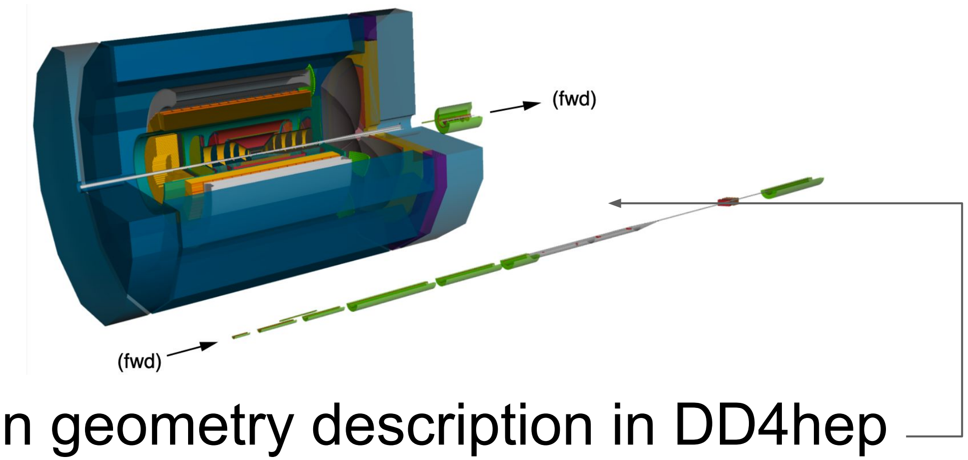
- State of Software Survey
- Follow-up Focus Groups
- Develop Testing Community

Concrete aims

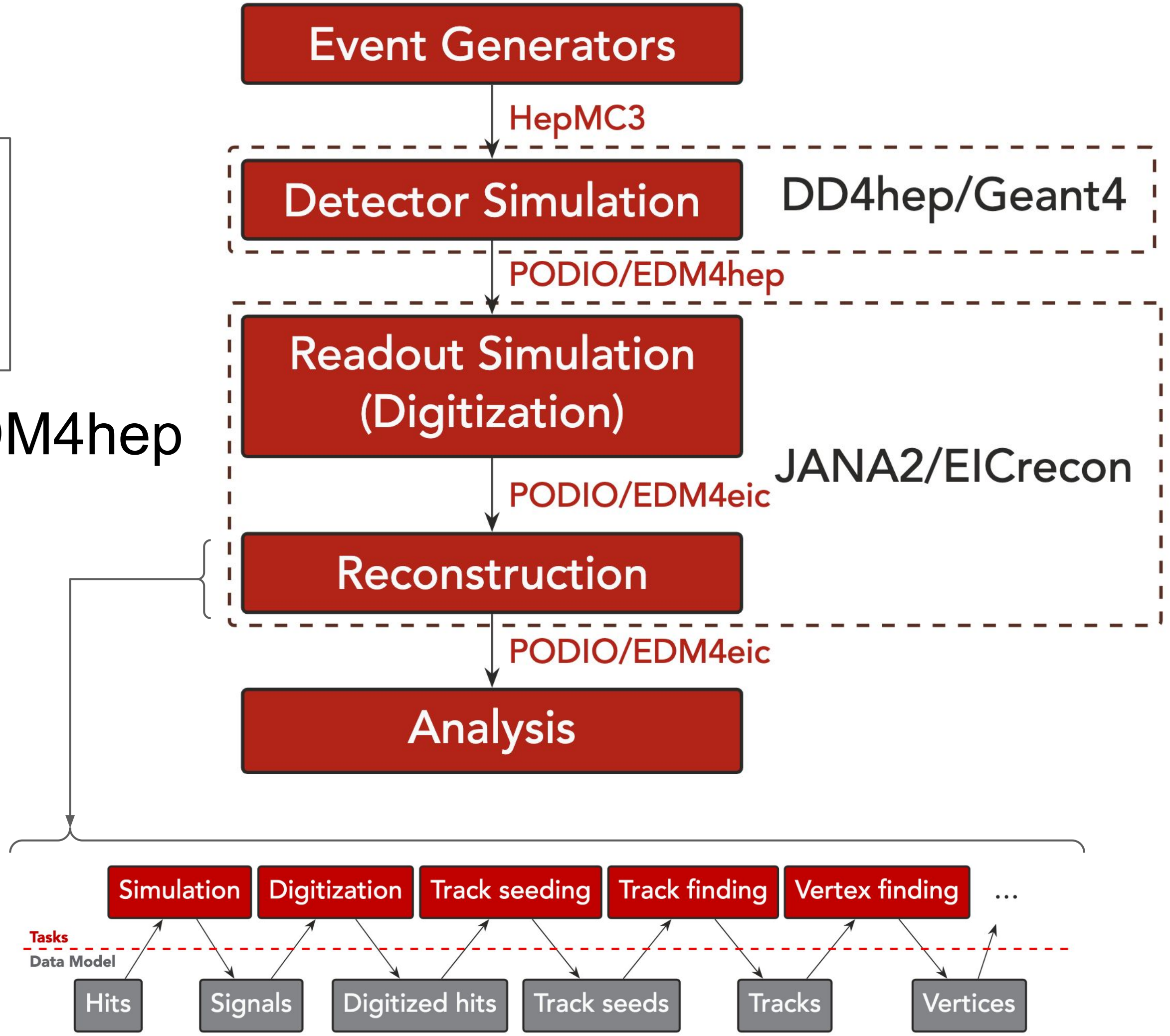
- Discoverable Software
- Reusable Workflows
- Data and Analysis Preservation

… all to provide a highly advanced software stack!

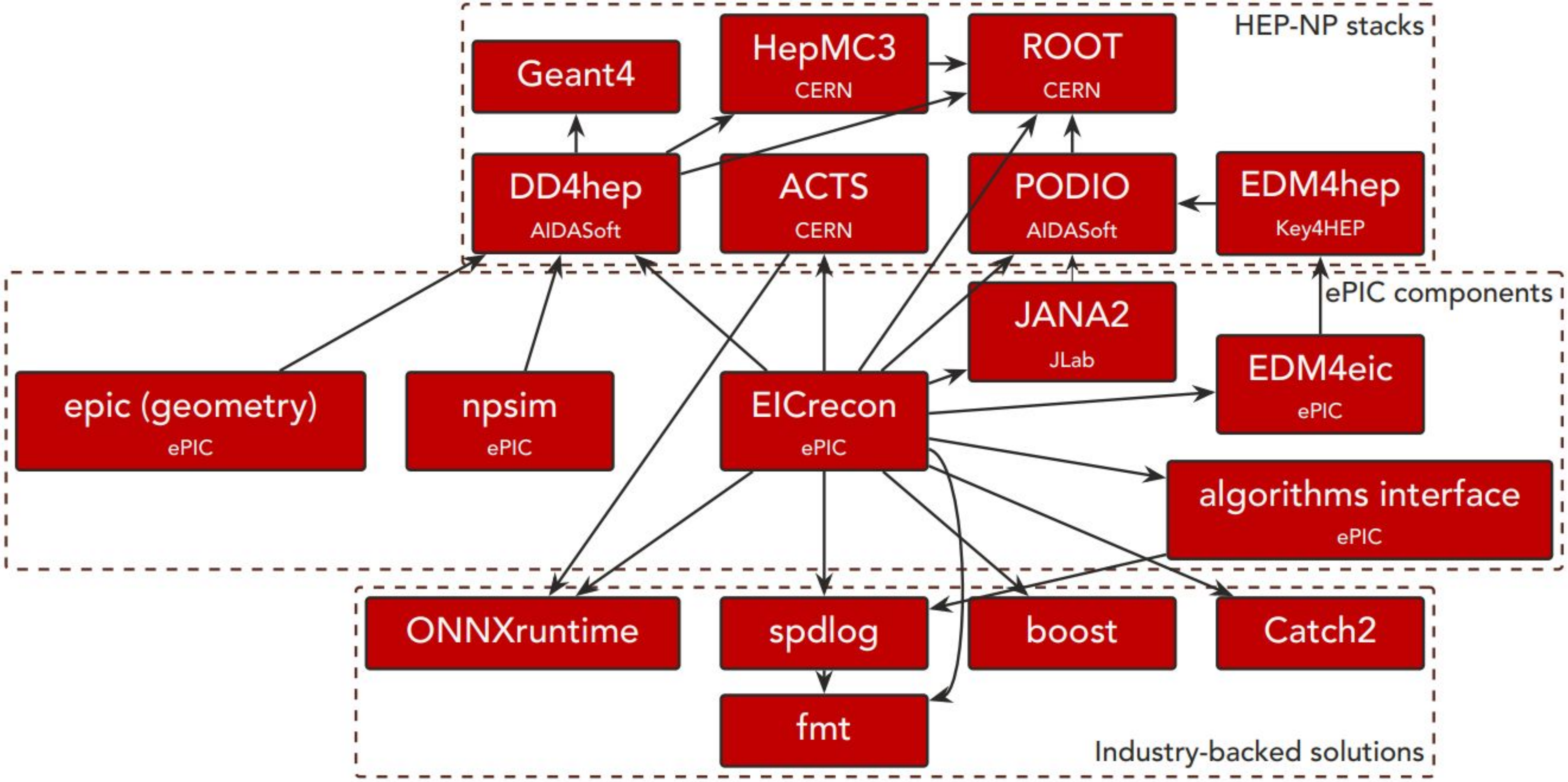
Simulation and Reconstruction Framework



- Common geometry description in DD4hep
- Data model is EDM4eic - an extension of EDM4hep
- Framework requirements include
  - Multithreading support (now in EICrecon)
  - Heterogeneous computing
  - Streaming workflow support
- Analysis capability requirements
  - Holistic reconstruction
  - Test beam data analysis

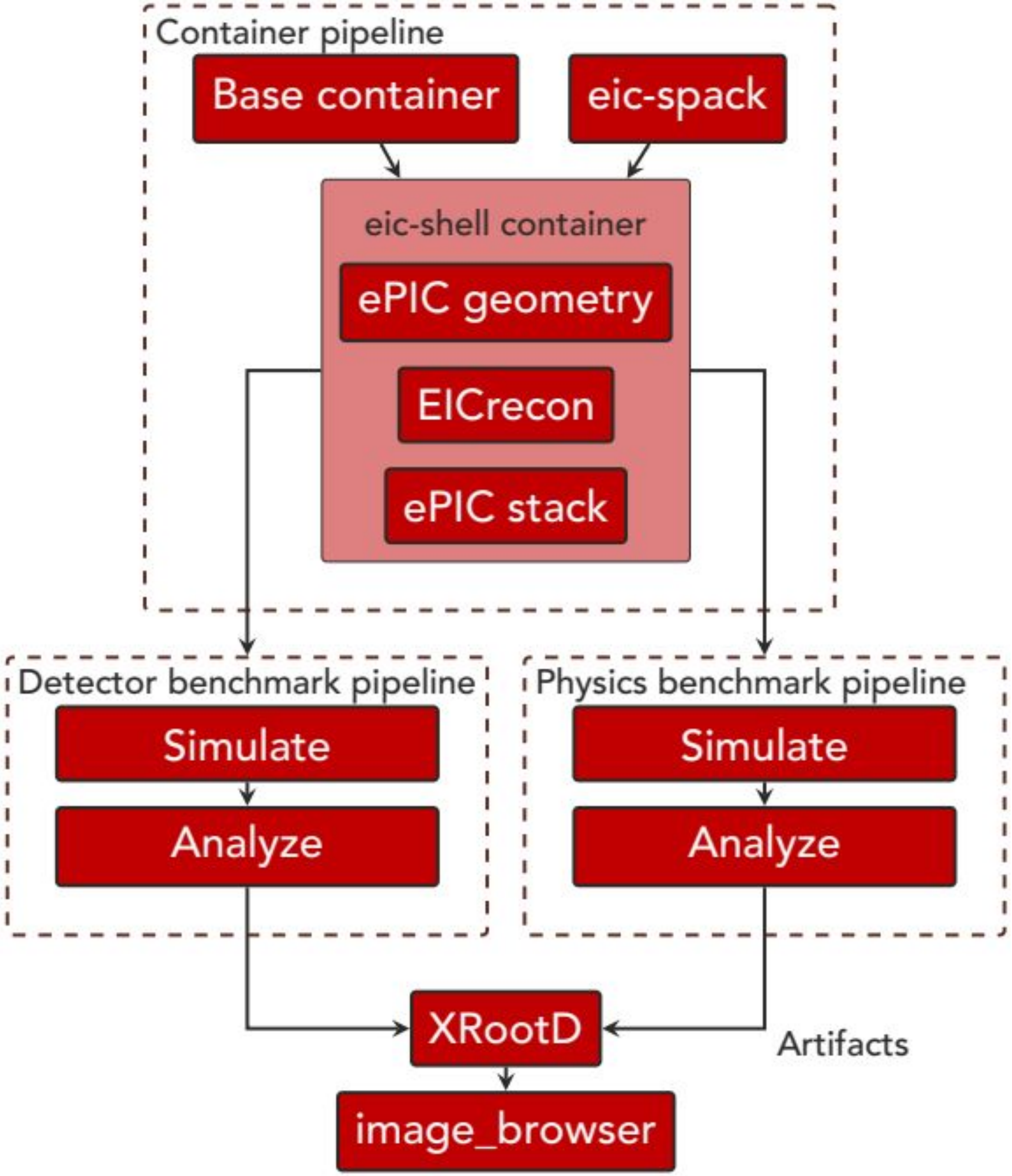


The ePIC Software Stack and Key Dependencies



A **modular software stack** with interdependent components requires more than CI/CD testing of individual components. We utilize **full-stack** testing, validation, benchmarking.

Full-Stack Testing, Validation, Benchmarking Workflow



This workflow is run on every commit in any component developed by ePIC, and on every container build recipe change.

Requires simulations and reconstruction, with **sufficiently granular caching** to speed up workflows.

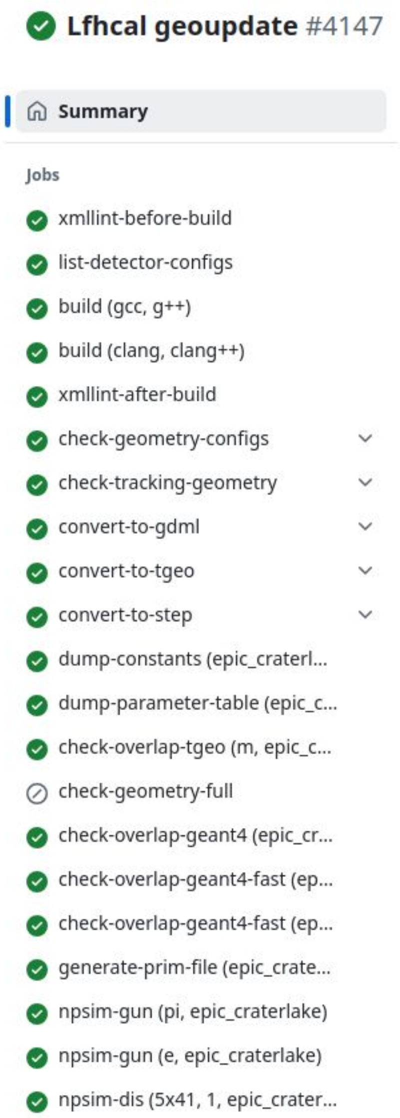
Three key phases:

- **Build** temporary container, push to registries (tag is sequential id)
- **Run** testing, validation, benchmarks inside the temporary container
- **Publish** testing results to XRootD, from where the visualizer pulls artifacts for comparison

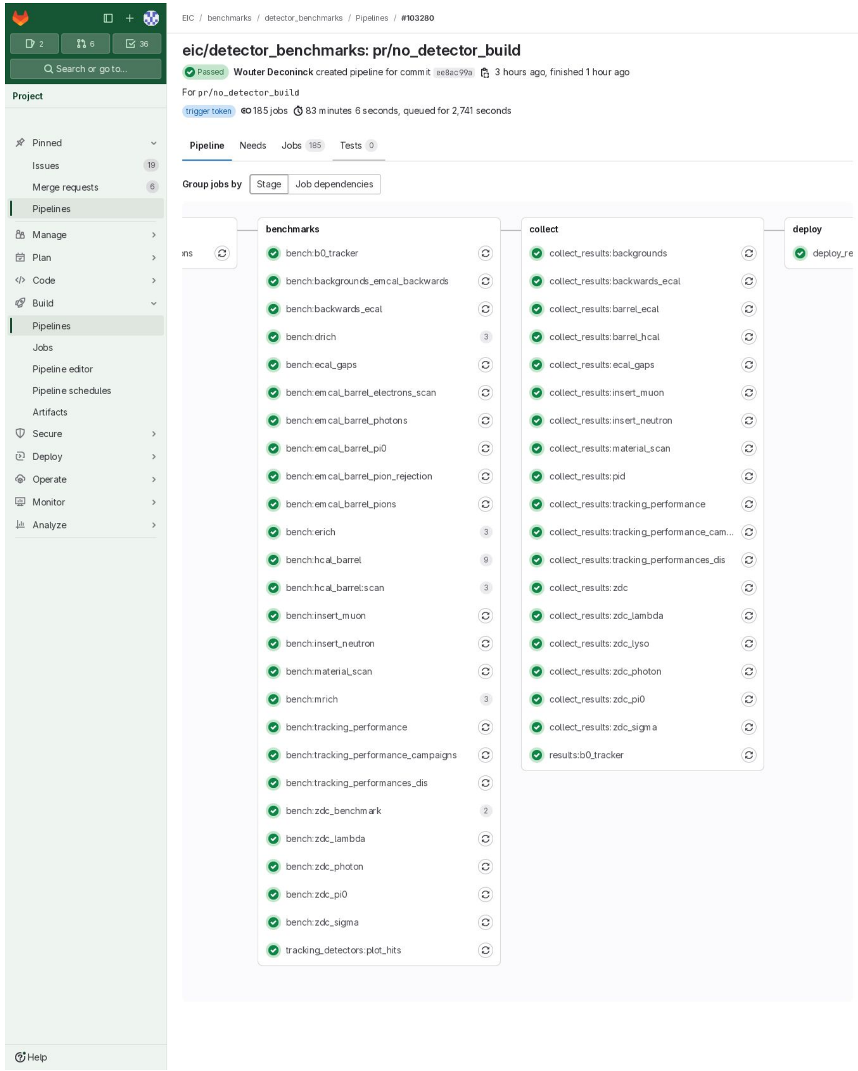
Great effort goes into educating our users about it.

Full-Stack Testing, Validation, Benchmarking Workflow

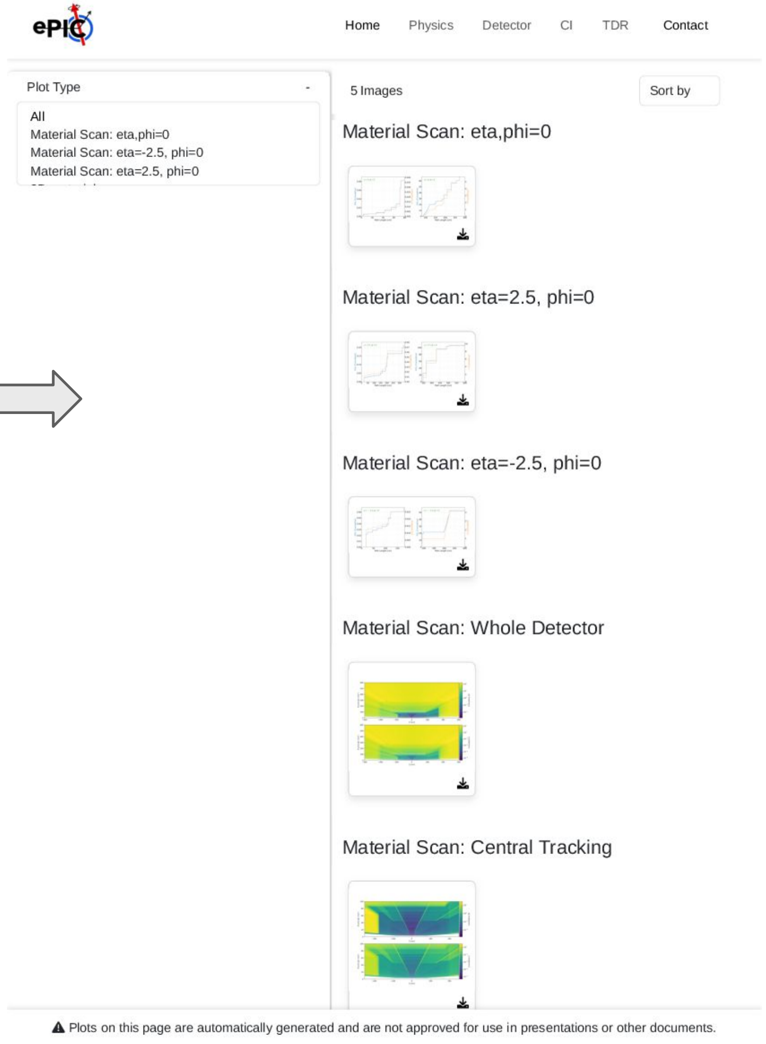
Code change on GitHub



Benchmarking on Gitlab

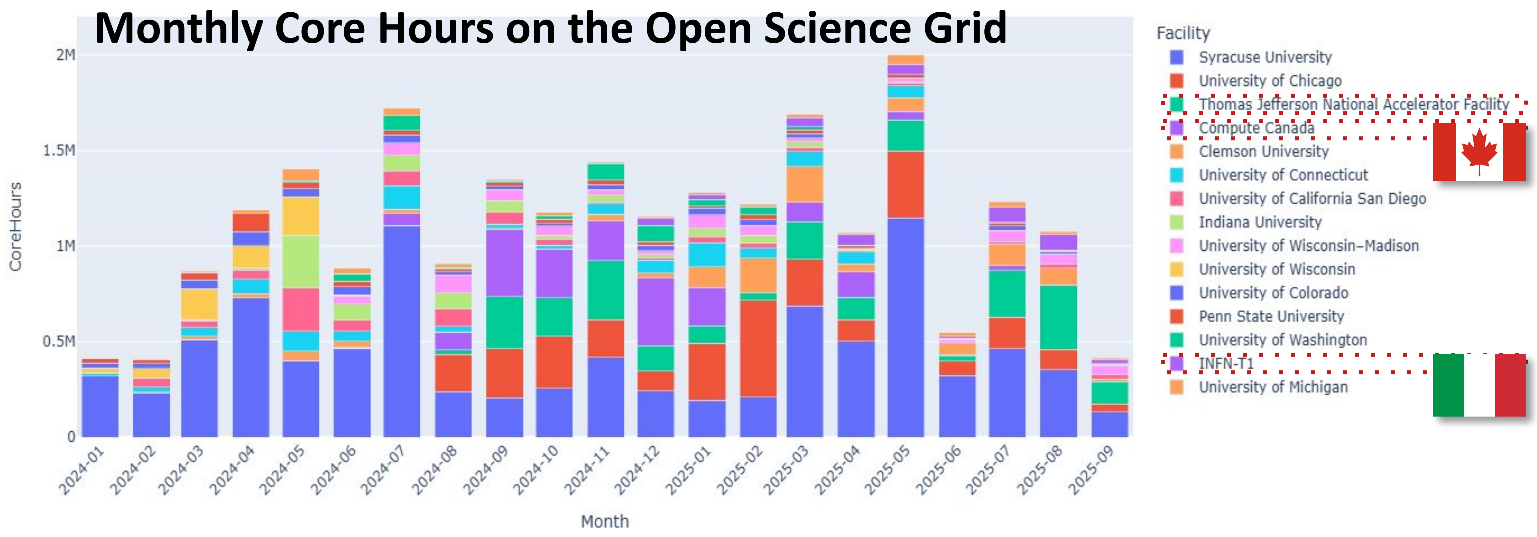


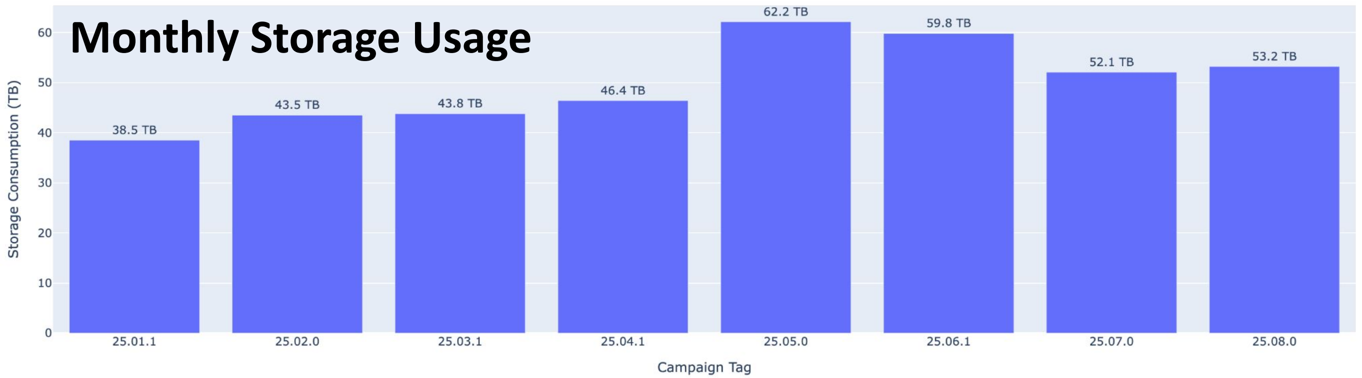
Presentation Interface



Simulation Campaigns

- We provide simulation productions tailored to the needs of the collaboration, as defined by the DSCs and PWGs.
- Simulation campaigns are conducted monthly, based on the software release for the corresponding month.
- These simulations serve as the standard for detector and physics studies for the preTDR and also the Early Science Program.
- In the past year, monthly simulation campaigns consumed approximately 15 million core hours on the OSG, generating over 500 TB of simulation data. We anticipate a ~2x storage increase, and are in talks with institutions in Japan, Taiwan, Italy and UK for additional resources.





We are capable of integrating new detector geometry and algorithms within a month, processing millions of events needed to assess scientific impact.

Priorities for 2025

- **Improved Communication**
  - **(Almost) Weekly Software News**: Regular reports highlighting major changes to the software stack, notable code merges, and updates from the WGs.
  - **Meeting Notes**: Summaries of WG meetings, including outcomes and next steps, are provided to enable asynchronous participation.
- **Updated WG Charges and Priorities**
  - Reflect outcomes of the collaboration meeting in Frascati, Italy. (for background see in John's slides)
  - Example priorities include background integration in simulations and resolving discrepancies between engineering and simulation designs.
- **Collaboration with Physics WGs**
  - **Coordination** on **simulation targets**, **reconstruction**, and **analysis tools**.
- **Coordination with DSCs**
  - Requested **updates on software priorities** to facilitate the integration of DSCs simulation efforts into ePIC software and simulation campaigns.
  - Coordination **improved shared development**, e.g., by aligning four DSCs under a common digitization work plan.

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.

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.

The specific charges and priorities for each WG are provided on slides 25 to 29.

Geometry validation

A major task of resolving discrepancies between engineering and simulation designs requires as much coordination as technical work!

DSC

Work with project engineers to converge on subsystem design and files issue and PR for simulation geometry to implement the change

Project

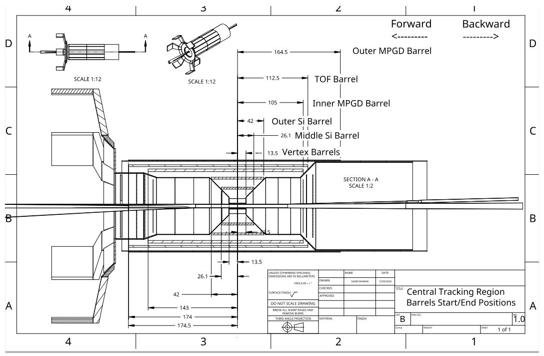
Provide project CAD files and comparison log back

Physics and Detector Simulation WG

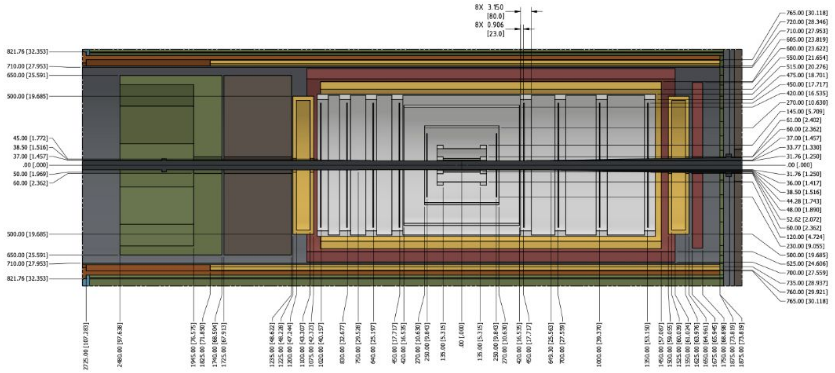
Generate drawings using OnShape and provide them to the EIC project together with STEP files

Convert subsystem geometry to step using npdet_to_step

Synchronizing DSC Design, ePIC Simulation Geometry and Project CAD (S. Rahman)



ePIC simulation geometry converted to CAD

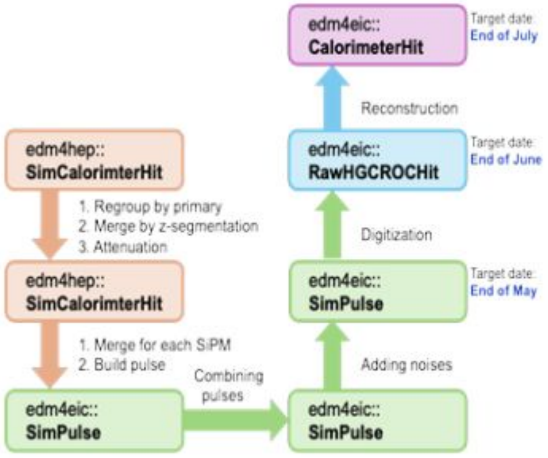


Project detector CAD design

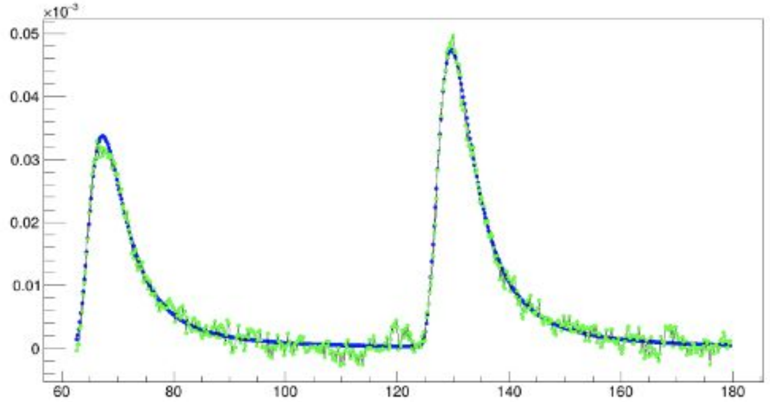
As part of the ongoing design process, DSCs are responsible for addressing discrepancies between engineering and simulation designs. A first round initiated last March is near completion this October.

Digitization and noise in simulation

Digitization is key to understanding detector response for timing and resolution



Data flow for HGCROC digitization by Derek Anderson



Example simulated pulse shapes with added noise by Simon Gardner

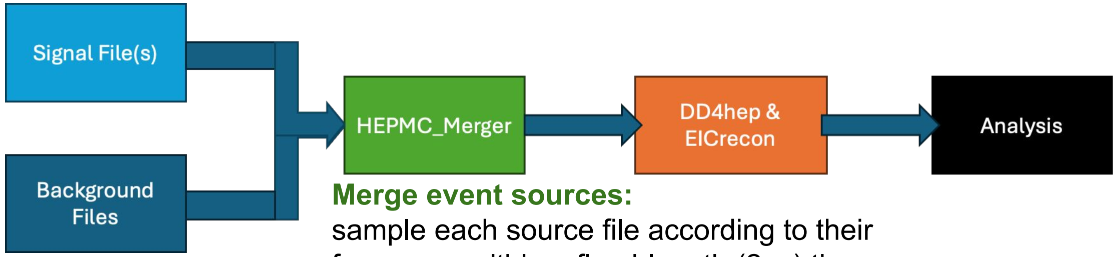
Rapid advancements in digitization driven by various subsystems

- **BTOF** for AC-LGAD
- **Low-Q^2** tagger for Timepix4 (potential application for the Allpix^2 package)
- **LFHCAL** & **BIC** CALOROC/HGCROC (potential application for the SimSiPM package)

Strong emphasis on finding commonalities to establish **co-review** and improve **reuse of work**.

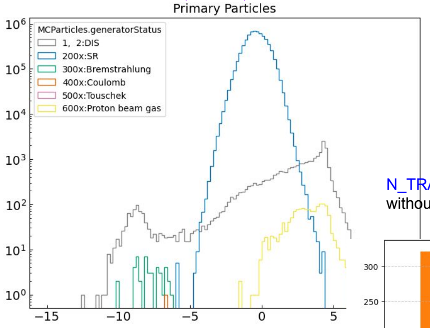
Supporting analysis and comparisons on test beam measurements is key to **validation**.

Background embedding and tracking performance

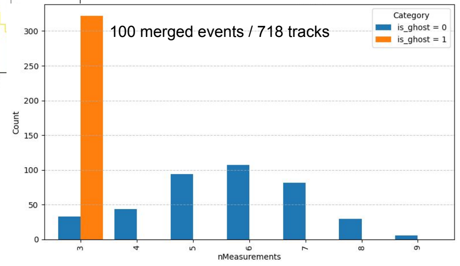


Merge event sources:
sample each source file according to their frequency within a fixed-length (2us) time window

- Merging of full set of backgrounds for various beam energies into physics events implemented in the campaign simulation chain – **enabling highly advanced background studies**
- Working with EIC project on integrating externally-simulated (at beamline volume) Synchrotron Radiation sample – **a crucial but non-trivial optimization**
- Especially important for studies on tracking detectors (on the right) and time-frame based reconstruction

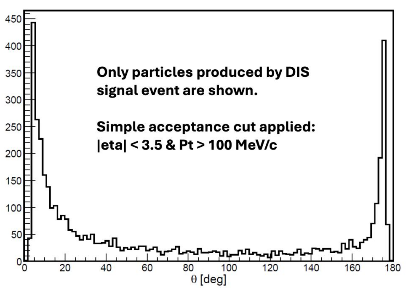


N_TRACK_HIT_MIN = 4 would reject most ghost tracks without hurting the efficiency significantly

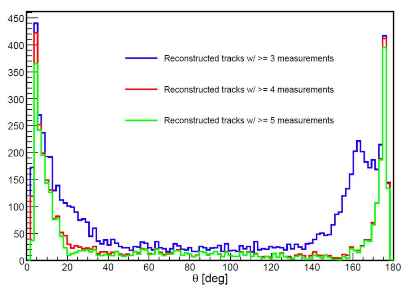


Work in progress by S. Li

Generated charged particles from DIS signal event

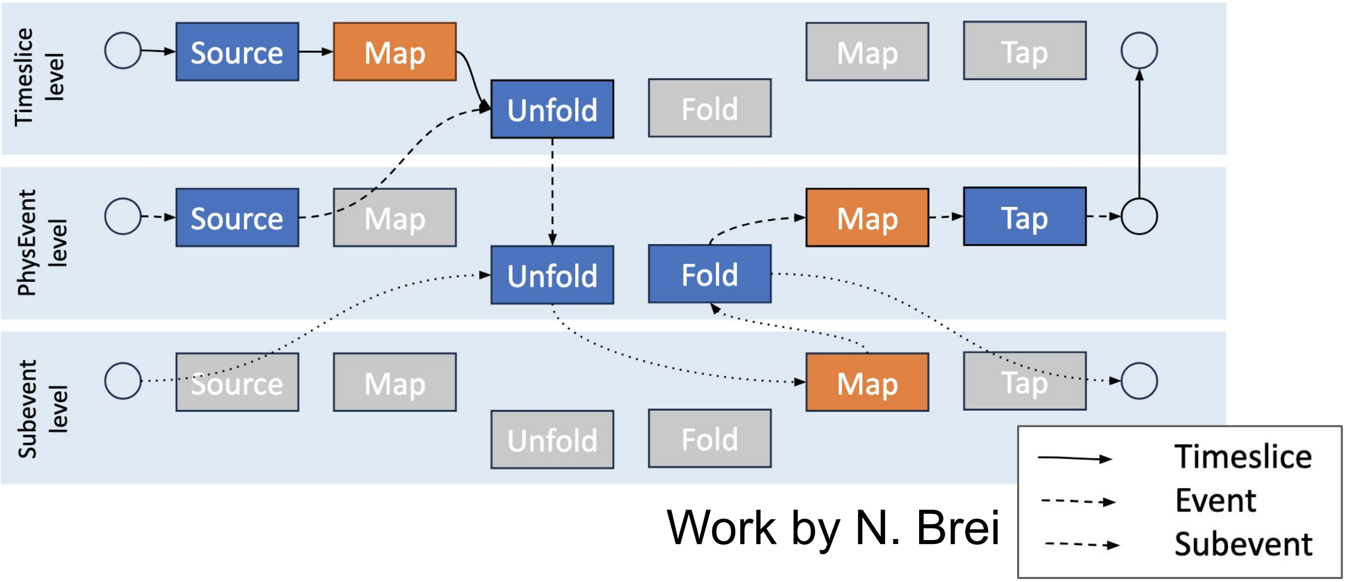


With synchrotron radiation background included



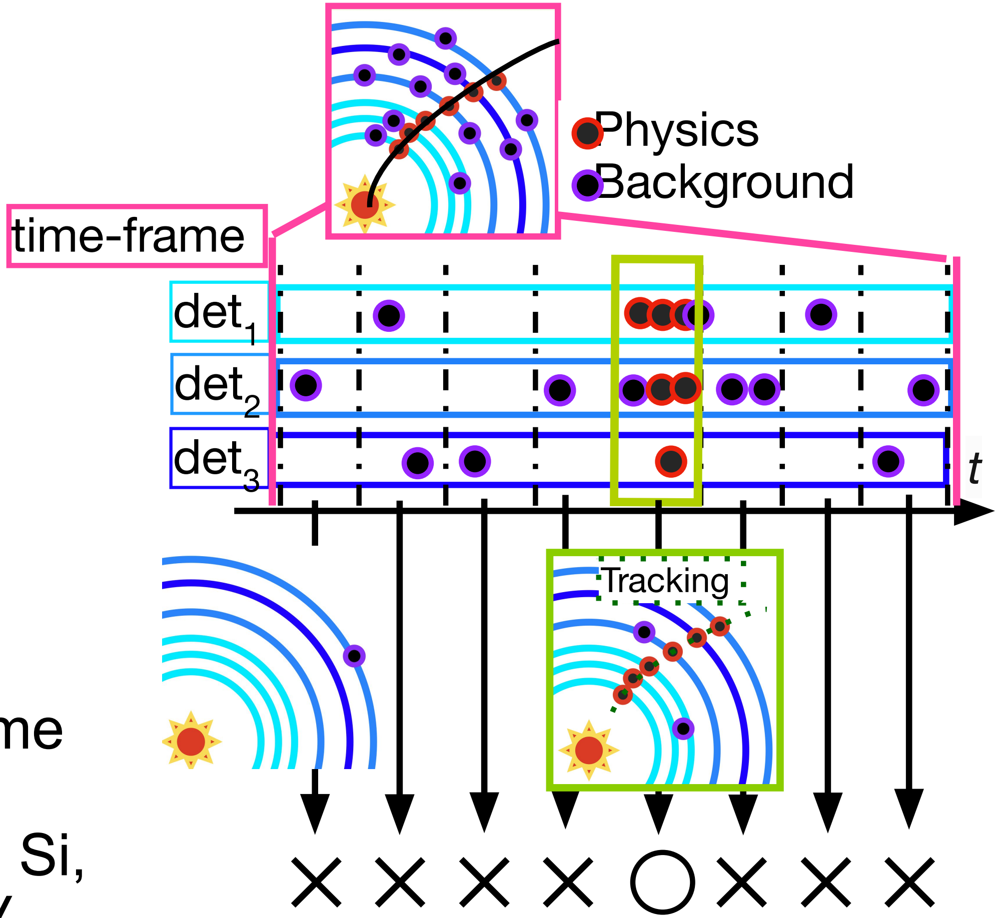
Work in progress by B. Schmookler

Reconstructing events from Time-frames



Work by N. Brei

- Utilizes **advanced-topology event processing functionality** from JANA2
- A **prototype algorithm** based on slicing a time frame by ~20 ns window of a 2 μs time-frame
- A simple "trigger" condition on raw simulated hits in Si, TOF and MPGD **achieves >99% efficiency at <1% background**

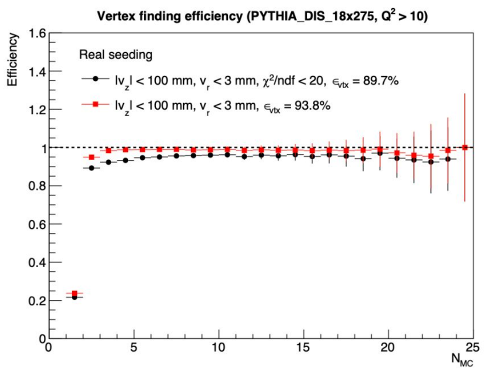


Work by T. Kumaoka

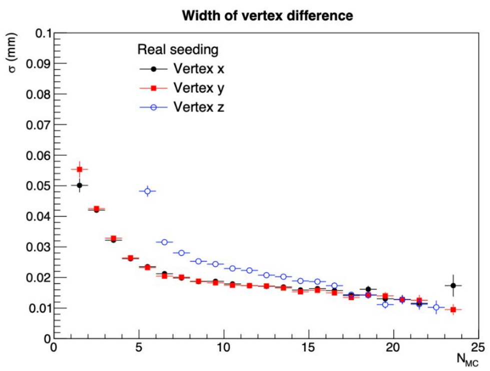
Developments in Vertex Finding

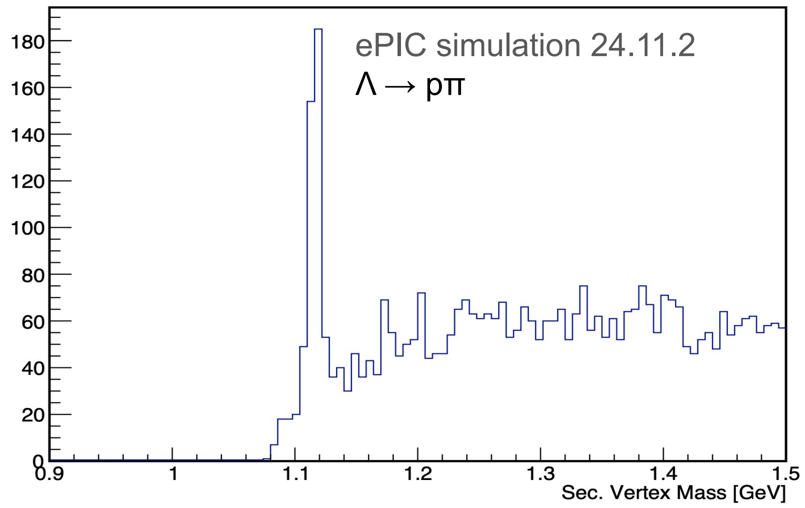
- The **primary vertex finder** using Acts' IterativeVertexFinder was demonstrated to have a reasonable performance
- **Secondary vertexing** using AdaptiveMultiVertexFinder implemented and will be integrated soon
- The collaboration explores possibility of applying other packages for Heavy-Flavor physics

Efficiency vs. N_{MC}



Resolution vs. N_{MC}





- High vertex finding efficiency at high multiplicity
- About 15 μm resolution at high multiplicity

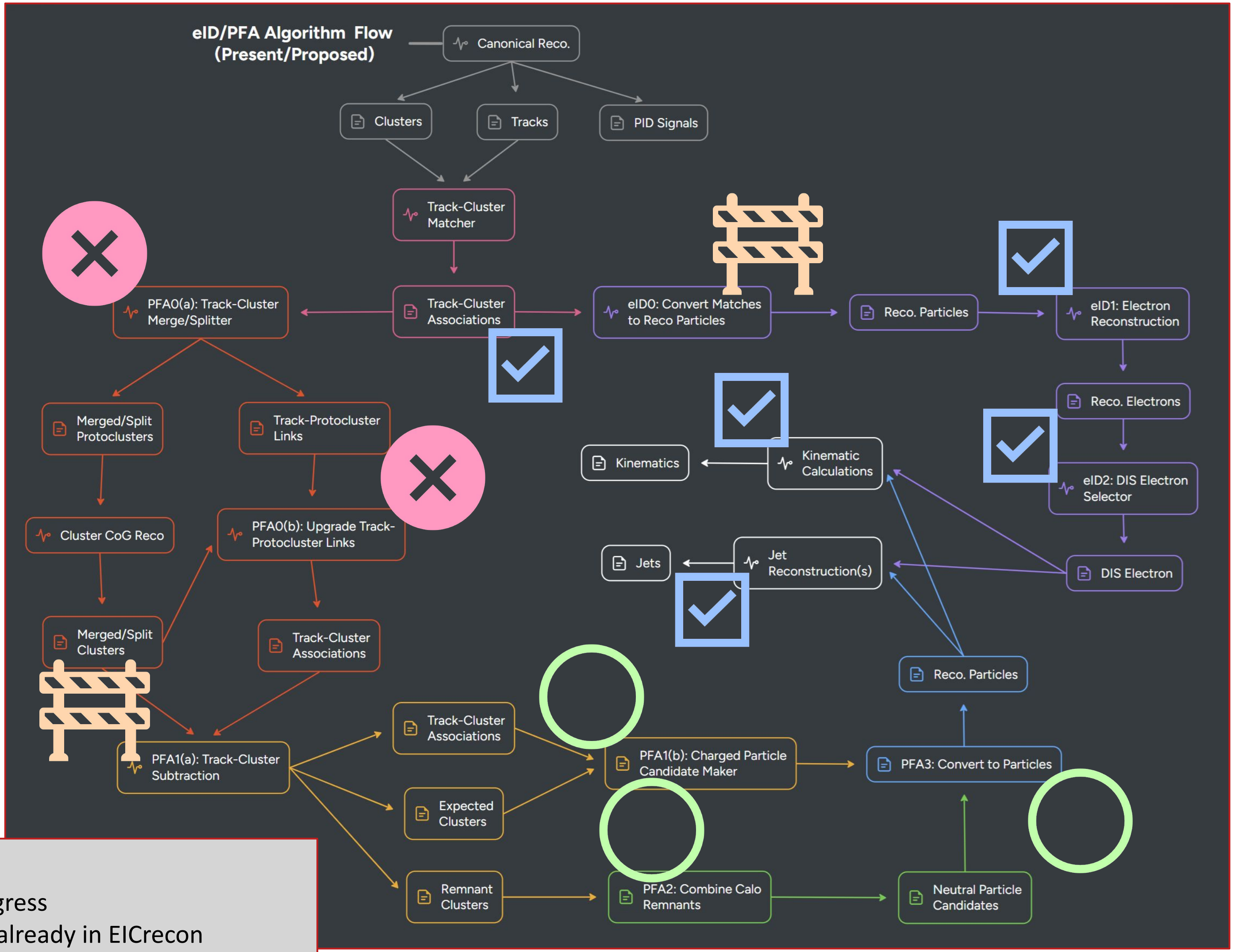
Slide by Rongrong Ma

Work in progress by Bishoy H. Dongwi

More on interaction with Acts in Wouter's talk.

Particle Flow: Development and Planning

- Initially deemed non-essential for TDR, but later **identified as a need by hadronic calorimetry DSCs as well as Jet/HF PWG**.
- Planning organized according to a staged approach with a goal for modularity in mind.
- Work on Particle Flows also advances implementations for electron ID and holistic reconstruction at ePIC.



= To-do
= In progress
= Done/already in EICrecon
= Blocked

Chart by D. Anderson

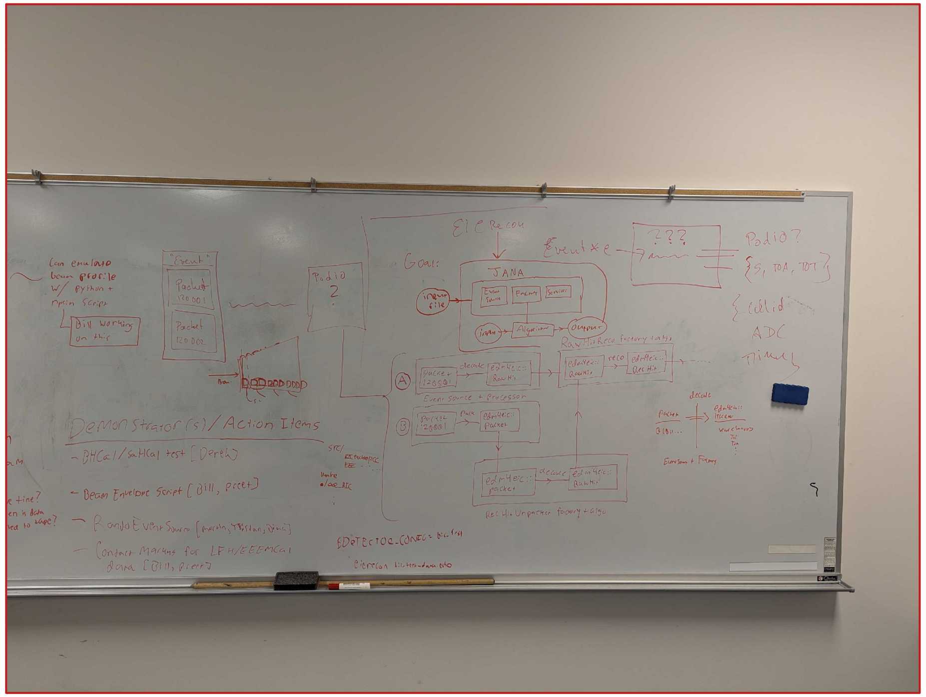
Test Bench and Test Beam: Integration with DAQ and Data

Key priorities

- Getting Test Data into reliable common storage and Rucio tracking
- Planning additions to our PODIO Event Data Model needed to support processing in EICrecon
- Infrastructure for alternative test beam geometries (example implementation provided)

Presentations during last collaboration meeting from DAQ frameworks to enable Test Data collection:

- **RCDAQ** - Martin Purschke (BNL)
- **CODA+ERSAP** - Vardan Gyurjyan (JLab)
- **nestDAQ** - Noboyuki Kobayashi (OU-RCNP)



From brainstorming session for Test Beam interfaces.

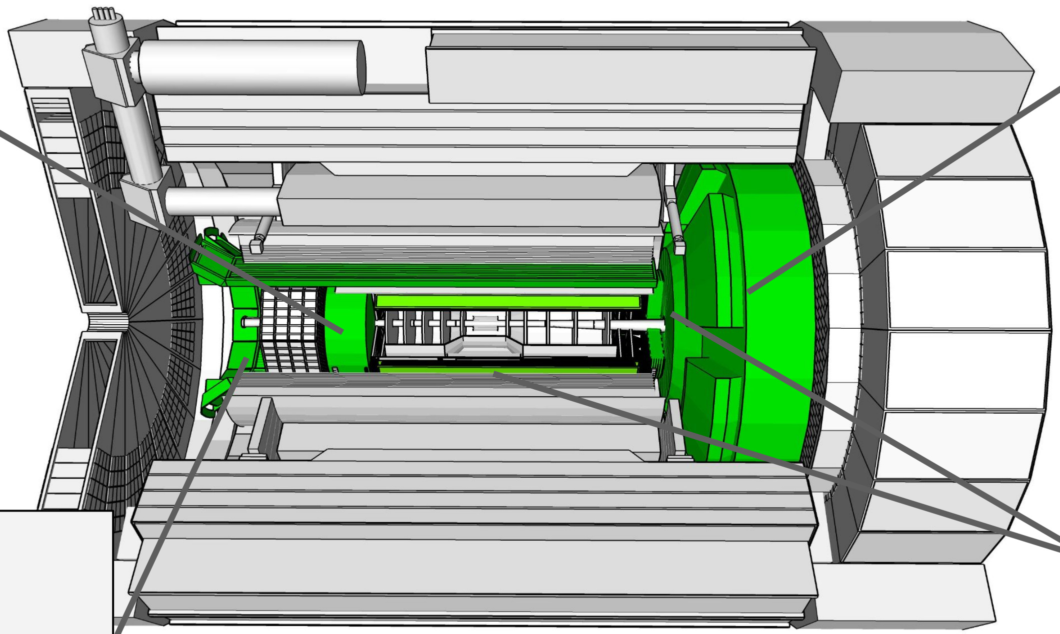
Supports DSCs and advances simulations

Particle Identification at ePIC

Immediate needs for pTDR are addressed by providing external (standalone) simulations via LUTs. A more integrated approach is expected for the TDR studies.

pfRICH

- “IRT2” integration effort underway towards TDR
- Out-of-tree integration into ePIC stack is available



dRICH

- Relies and collaborates on the IRT2 effort

DIRC

- Promising AI/ML solutions available from several groups

Barrel & Encap TOF

- Driving force behind work on waveform-based digitization simulation
- Reconstruction to be implemented

Recognized by the collaboration as a problem to be addressed for CD-2. It is our major priority to enable development and integration of PID reconstruction. To strengthen PID efforts, we will add a third convener to the Reconstruction WG.

Tutorials, Helpdesk and User Surveys

Online materials

- Many new collaborators successfully onboarded themselves using only the landing page
- Recent collaboration survey by User Learning received a testimony: “Having this talk/form [survey] already sets the EIC above lots of other places I've worked with”

Any member of the collaboration can **directly contribute to any of the websites by submitting Pull Requests**.

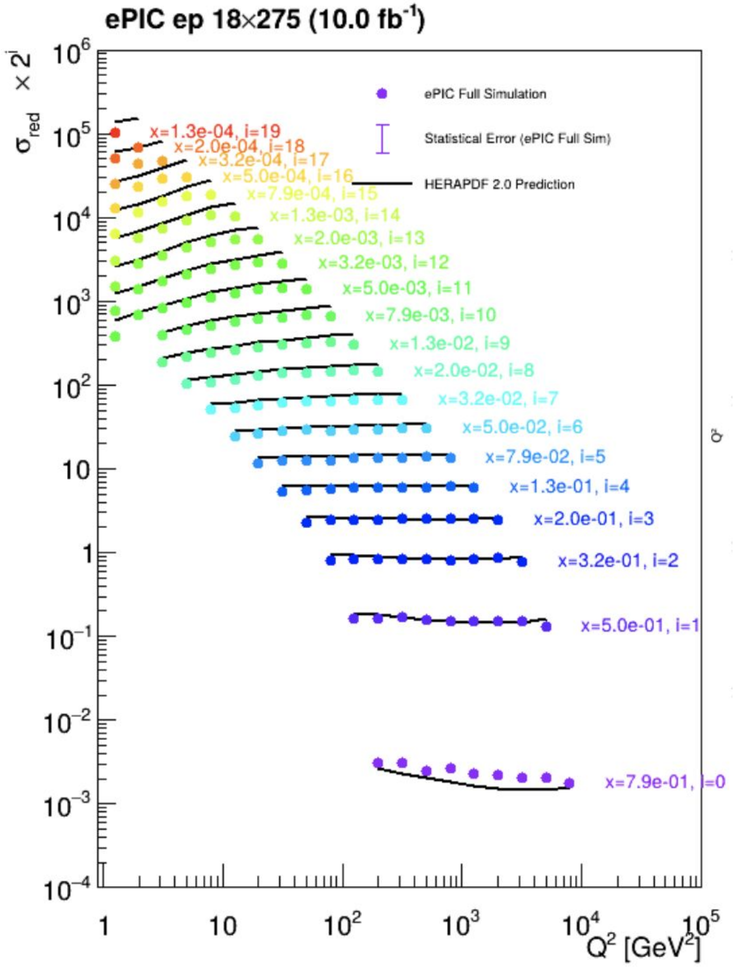
Tutorial Series

- 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)

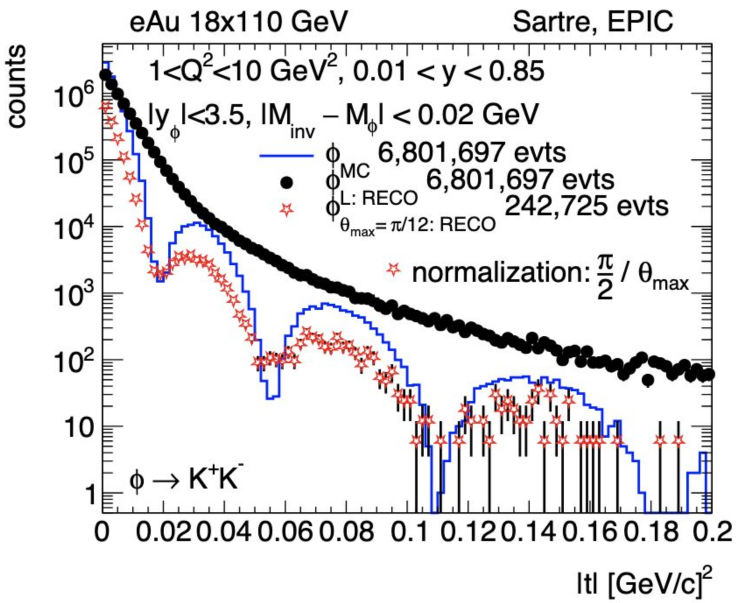
Helpdesk on Mattermost

- A place to ask questions without judgement
- Experts are readily available, and questions are addressed efficiently
- Well appreciated by users

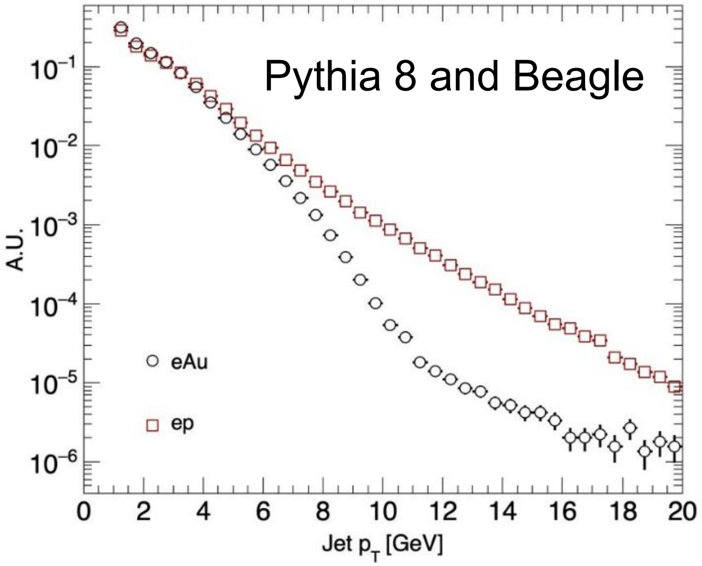
Enabling pTDR and Early Science Physics Studies



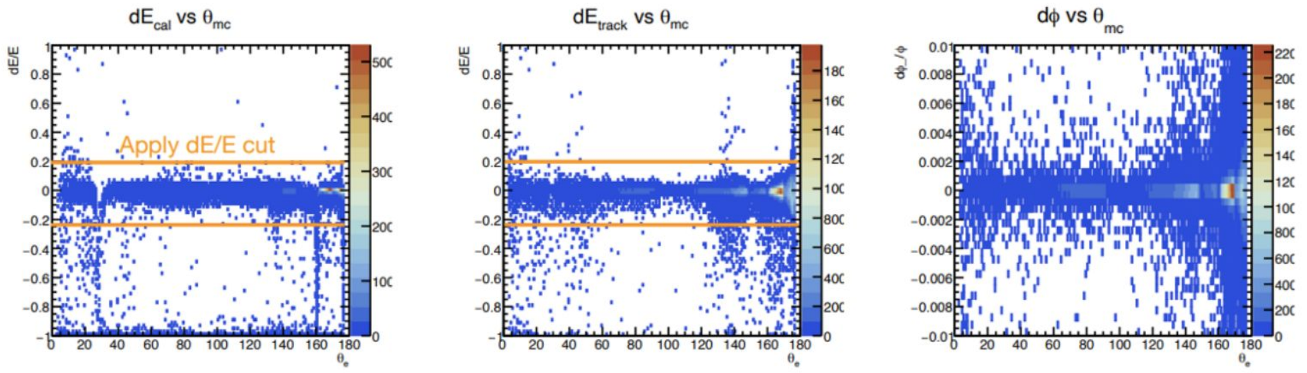
Plot by S. Maple



Plot by M. Kesler

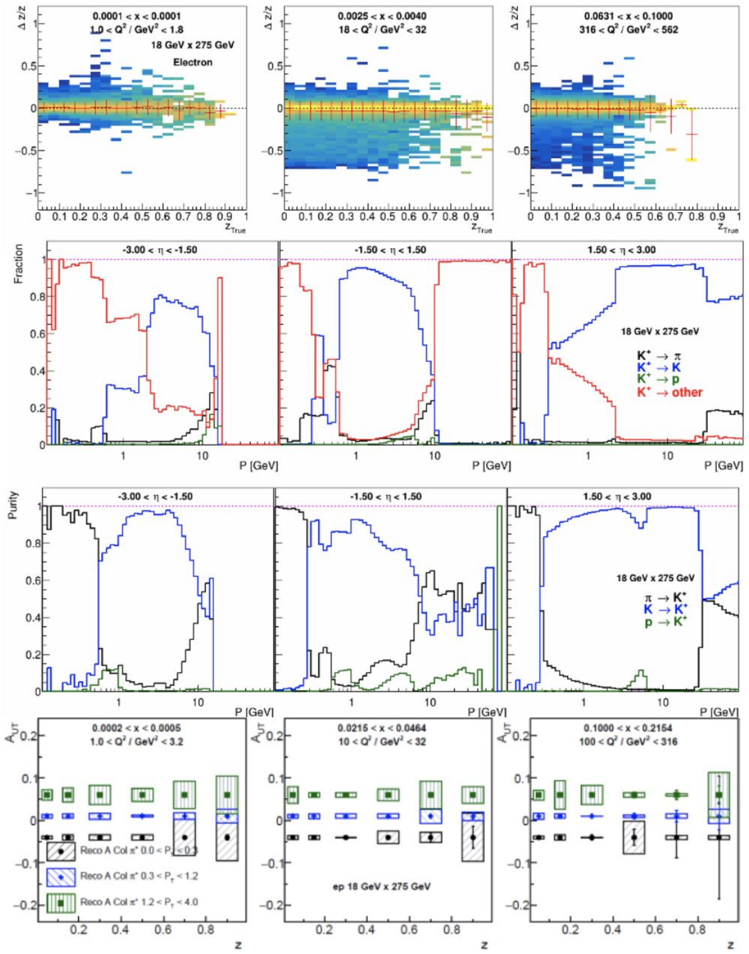


Plot by B. Page and D. Lemos



Plot by W. Lin

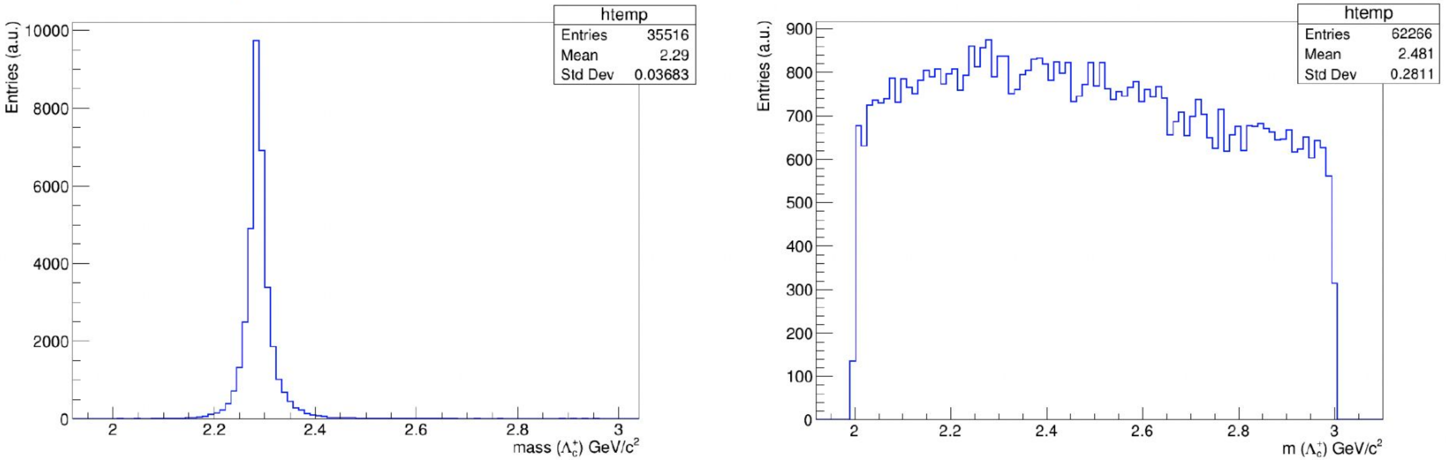
Enabling pTDR and Early Science Physics Studies



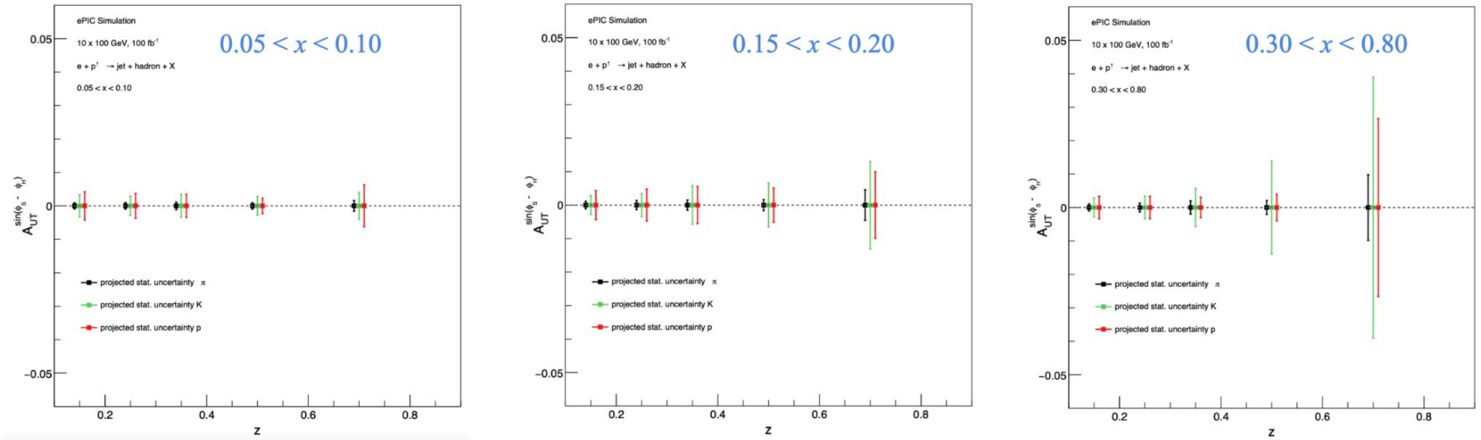
Plot by R. Siedl

Signal: Signal from Λ_c^+ Sample + Signal from DIS Sample

Background: Background from DIS Sample



Plots by S. Kumar



Plots by K. Adkins

Summary

- **In 2024, we defined preTDR readiness for software and simulations with the collaboration—and successfully met those goals. In 2025, we renewed our priorities at the Frascati Collaboration Meeting, with many topics under our charge already reaching an advanced stage and positioning us well for the next phase.**
- ePIC Software is successfully being used to
  - deliver crucial results needed for finalizing detector design (pTDR)
  - validating its fitness for the purposes of the EIC science program
- Ongoing software developments are kept aligned with ePIC computing model
- Strong emphasis is placed on **coordination with Physics WGs** on simulation targets, reconstruction
- Surveys of DSCs allowed us to have a clear picture for ongoing and planned work
- Our advanced software and large-scale simulations serve as the backbone of the pre-TDR and Early Science Program efforts. We are coordinating effectively with the DSCs and PWGs, aligning our priorities with the deliverables and milestones of the EIC Project.

Backup

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.

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.

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.

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.

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.

DSC or PWG	Simulation Path	Generator Config	Background	Events	New Request	Pre-TDR Use	Early Science Use
Exclusive/Diffractive/Tagging	Input Files: /w/eic-scshelf2104/users/sjdkay/Jul2025_Campaign_Input/Afterburner_Output/pion	DEMPgen1.2.4	No		Yes	Maybe	Yes
Exclusive/Diffractive/Tagging	/w/eic-scshelf2104/users/ gbxalex/SimCampaign_Input	lAger3.6.1	No		Yes	Yes	Yes
Exclusive/Diffractive/Tagging	/w/eic-scshelf2104/users/ gbxalex/SimCampaign_Input	lAger3.6.1	No		Yes	No	Yes
Exclusive/Diffractive/Tagging	/gpfs02/eic/wlin/docu/eHe3_10x166_1M	BeAGLE	No		Yes	No	Yes
Exclusive/Diffractive/Tagging	/gpfs02/eic/wlin/docu/eHe3_18x110_1M	BeAGLE	No		Yes	Yes	No
Exclusive/Diffractive/Tagging	/gpfs02/eic/wlin/docu/eHe3_10x110_1M	BeAGLE	No		Yes	Yes	No
Exclusive/Diffractive/Tagging	/gpfs02/eic/wlin/docu/eHe3_5x41_1M	BeAGLE	No		Yes	Yes	No
Exclusive/Diffractive/Tagging	/gpfs02/eic/jkim/DVpi0P/sim_data/	EpIC	Yes		Yes	No	Yes
Exclusive/Diffractive/Tagging	/gpfs02/eic/jkim/DVpi0P/sim_data/	EpIC	No		Yes	Yes	Yes
Exclusive/Diffractive/Tagging	/gpfs02/eic/jkim/DVpi0P/sim_data/	EpIC	Yes		Yes	Yes	No
Exclusive/Diffractive/Tagging	/gpfs02/eic/gpenman/DDVCS_18x275	EpIC	Yes		Yes	Yes	No
Exclusive/Diffractive/Tagging	/gpfs02/eic/gpenman/DDVCS_18x275	EpIC	No		Yes	Yes	No

… 50+ simulation requests processed