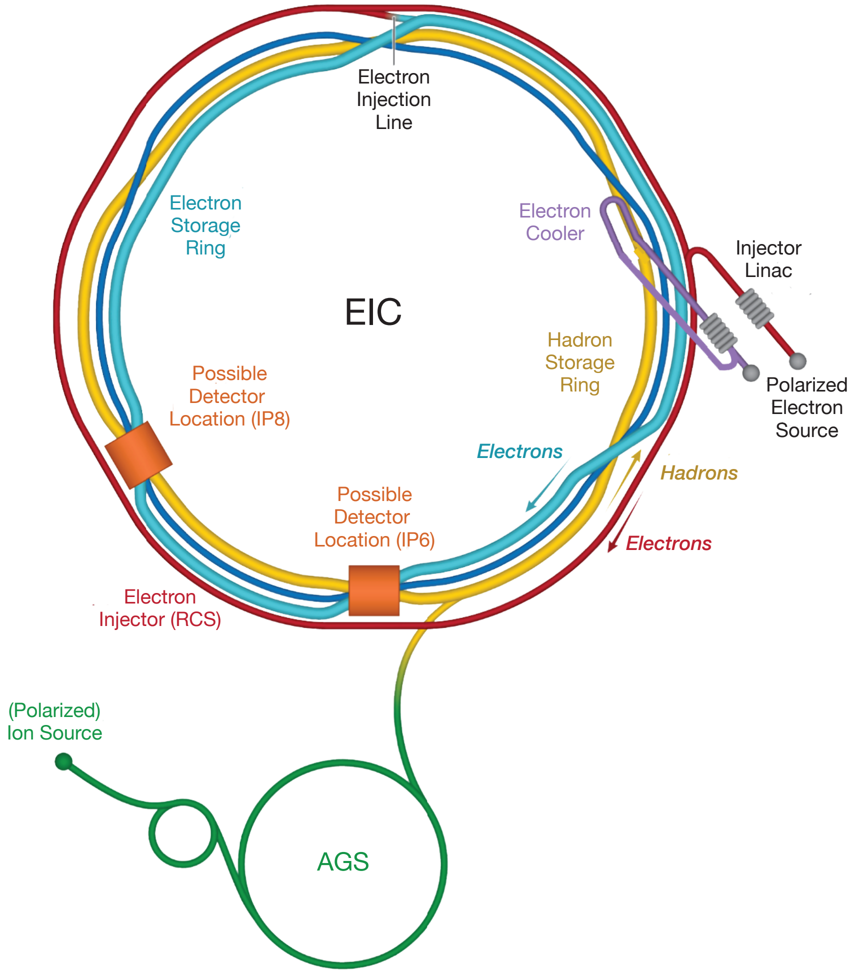
Markus Diefenthaler

The Electron-Ion Collider (EIC)



Frontier accelerator facility in the U.S.

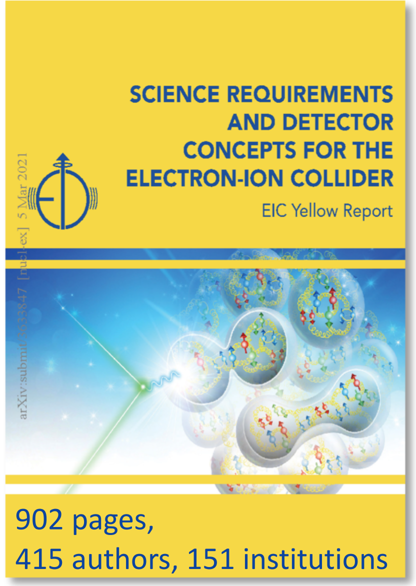
- **World's first collider of**:
  - Polarized electrons and polarized protons,
  - Polarized electrons and light ions (d, ^{3}He),
  - Electrons and heavy ions (up to Uranium).

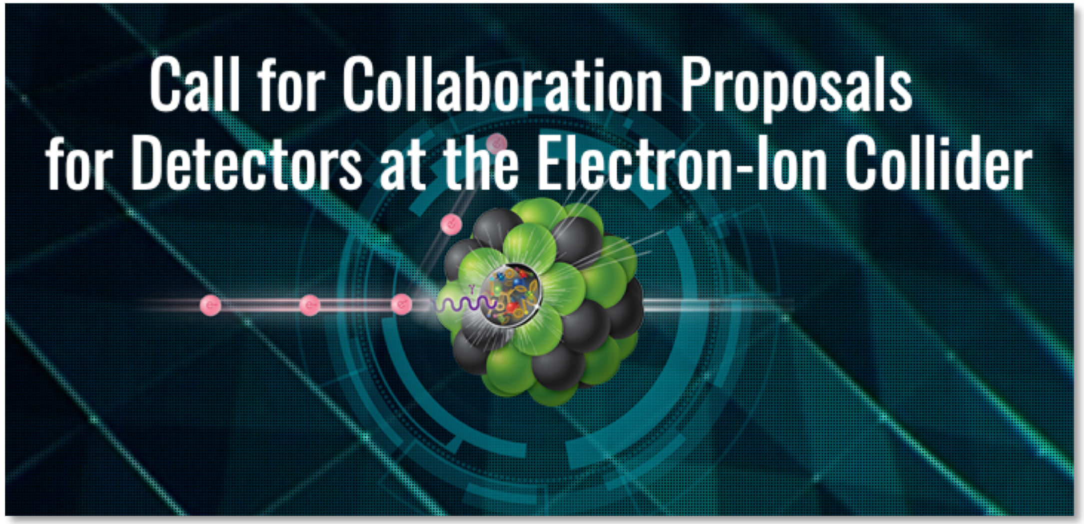
- The EIC will enable us to embark on a **precision study of the nucleon and the nucleus at the scale of sea quarks and gluons**, over all of the kinematic range that are relevant.

- Jefferson Lab and BNL will be host laboratories for the EIC Experimental Program. Leadership roles in the EIC project are shared.

- For Software & Computing at the EIC, Jefferson Lab will bring expertise from the 12 GeV CEBAF science program, leadership, and a vision.

Support of the EIC Initiatives in 2020–2021

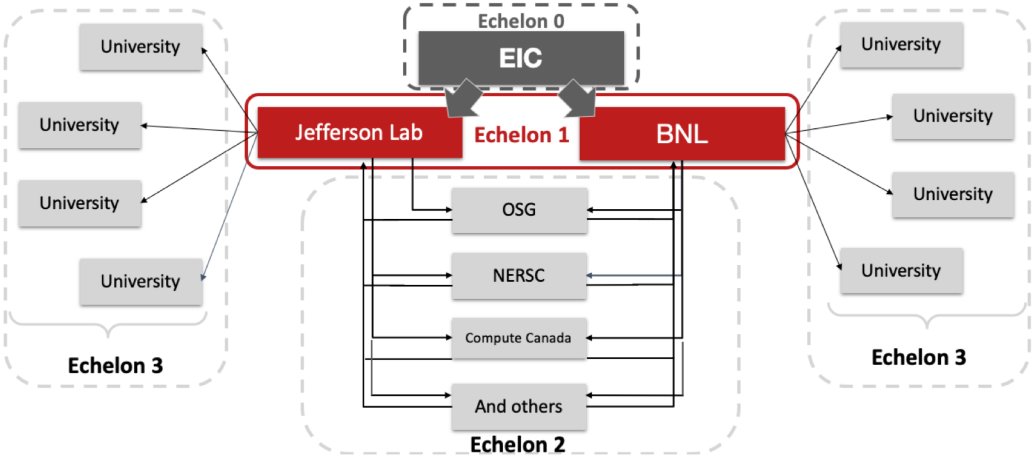




- The EIC Yellow Report describes the physics case, the resulting detector requirements, and the evolving detector concepts for the experimental program at the EIC.
- The studies leading to the EIC Yellow Report were commissioned and organized by the **EIC User Group** with strong involvement from Jefferson Lab.
- Jefferson Lab provided software & computing resources for the studies, mainly fast simulations and Geant4 simulations of detector components.

- Jefferson Lab and BNL issued a Call for Collaboration Proposals for Detectors, with three proposals being submitted on December 1.
- A scientific-technical committee of renowned and independent subject matter experts will advise Jefferson Lab and BNL on how to realize an optimal set of experimental equipment at the EIC.
- Jefferson Lab supported computational needs related to the proposals, full simulations of the large-scale detector systems being proposed for the EIC.

Federated Approach and Data Management



XRootD

RUCIO

Federated approach

- Both Jefferson Lab and Brookhaven are host laboratories of the EIC.
- The cooperation on Scientific Computing between the labs is managed by weekly meetings at the division/facility level.
- **Developments towards integrated research infrastructure**:
  - Containerization and cvmfs share allowed to run software on various sites.
  - Progress towards federated ID for access to shared computing resources.

Data management

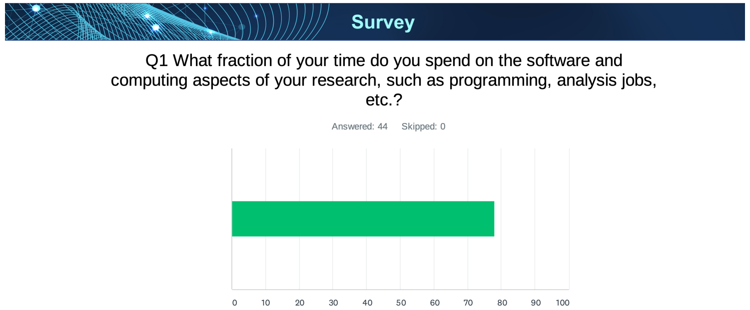
- We are building up **XRootD** as high performance data system for the EIC and our experimental program:
  - XRootD development is driven by the physics community. This will allow us to work with the developers in case of issues or feature requests.

- **We have been advised to**
  - *"systematically explore available Scientific Data Management Systems"*, e.g., Rucio,
  - develop *"a **data strategy** at JLab **across experiments and theory**"*, including a **data and analysis preservation strategy** for our experimental program.

Our Vision for Software & Computing at the EIC

"The purpose of computing is insight, not numbers." Richard Hamming (1962)

Software & computing are an integral part of our research:



Survey among NP Ph.D. students and postdocs in preparation of "Future Trends in NP Computing"

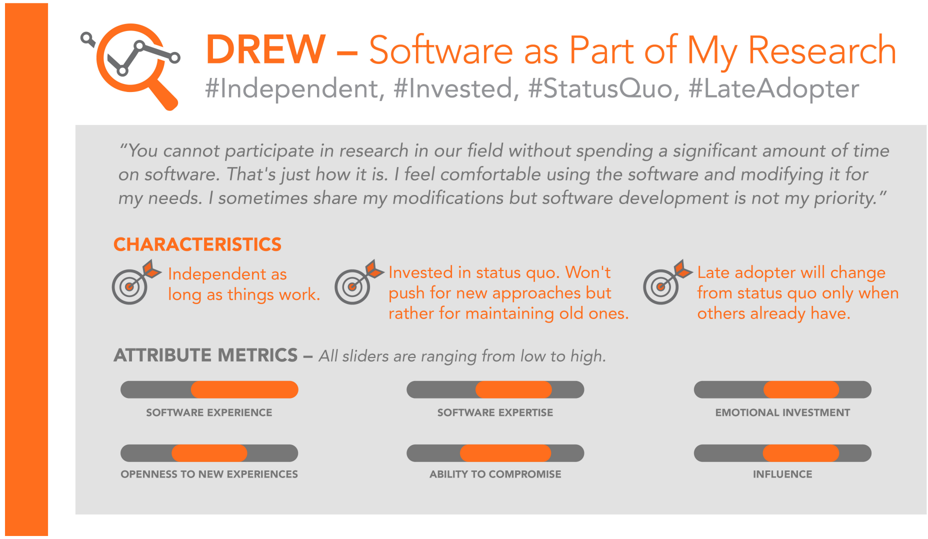
- **Goal** We work with our large Users Organization (over 1600 scientists from over 275 institutions) on data-intensive challenges and AI/ML and would like to ensure that also for the EIC scientists of all levels worldwide can participate in EIC analysis actively.
- **User-Centered Design**: To achieve this goal, we must develop simulation and analysis software using modern and advanced technologies while hiding that complexity and engage the wider community in the development.

Rapid turnaround of data for the physics analysis and to start the work on publications:

- **Goal**: Analysis-ready data from the DAQ system.
- **Compute-detector integration** with AI at the DAQ and analysis level.

User-Centered Design: Listen to Users, and/then Develop Software

- <u>State of Software Survey</u>: Collected information on software tools and practices during the Yellow Report Initiative.
- As part of the State of Software Survey, we asked for volunteers for focus-group discussions:
  - Students (2f, 2m), Junior Postdocs (2f, 3m), Senior Postdocs (2f, 3m), Professors (5m), Staff Scientists (2f, 3m), Industry (2f, 2m)
- **Results from the six focus-group discussions**:
  - Extremely valuable feedback, documented many suggestions and ideas.
  - Developed user archetypes with Communication Office at Jefferson Lab and UX Design Consultant:



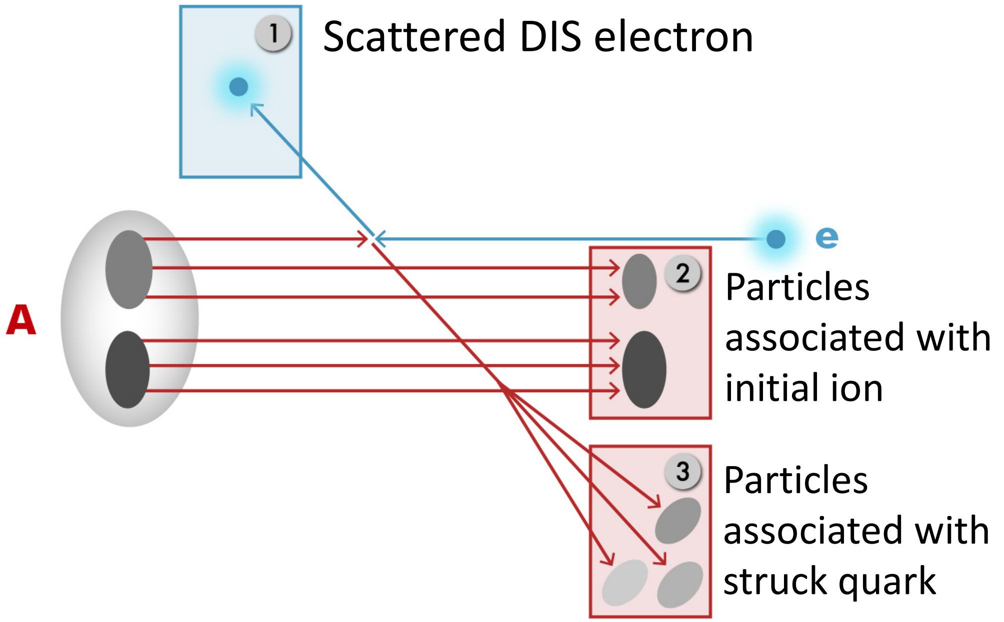
User Archetypes: Input to software developers as to which users they are writing software for:

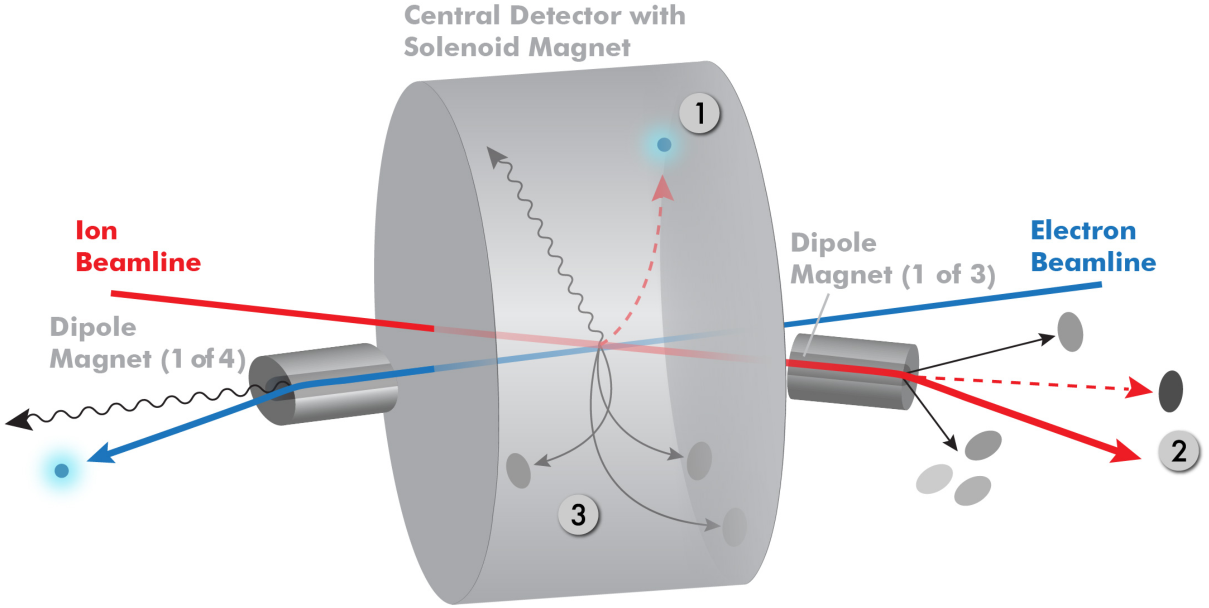
- Software is not my strong suit.
- Software as a necessary tool.
- **Software as part of my research.**
- Software is a social activity.
- Software emperors.

- We are repeating State of Software Survey now after detector collaboration proposals:
  - The regular software census will be essential to better understand and quantify software usage throughout the EIC community. During the next survey, we will also ask on feedback on the user archetypes.

Integrated interaction region and detector design to optimize physics reach

The aim is to get **~100% acceptance** for all final state particles, and measure them with good resolution.

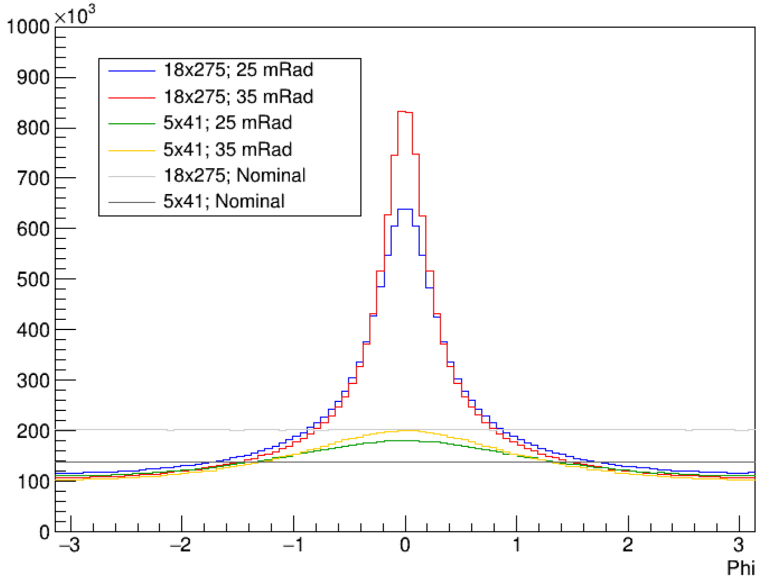


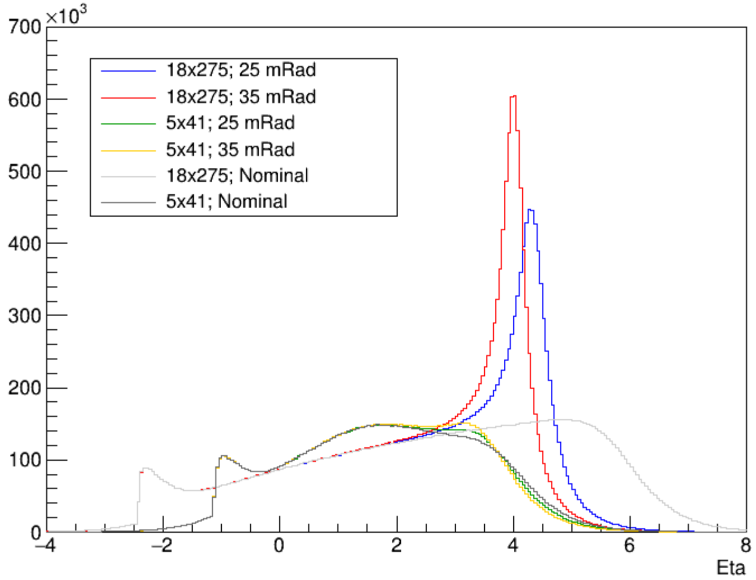


Experimental challenges:
- Beam elements limit forward acceptance.
- Central Solenoid not effective for forward.

Possible to get ~100% acceptance for the whole event:
- Beam crossing angle creates room for forward dipoles.
- Dipoles analyze the forward particles and create space for detectors in the forward ion and electron direction.

Accelerator and Beam Conditions Critical for EIC Simulations





- **Accelerator and beam effects that influence EIC measurements**
  - Beam crossing angle,
  - Crabbing rotation,
  - Beam energy spread,
  - Angular beam divergence,
  - Beam vertex spread.

- **Note for EIC Community** https://eic.github.io/resources/simulations.html
  - Profound consequences on measurement capabilities of the EIC and layout of the detectors,
  - How to integrate these effects in EIC simulations.
  - **Authors** J. Adam, E.-C.Aschenauer, M. Diefenthaler, Y. Furletova, J. Huang, A. Jentsch, B. Page.

Beyond that Include beam background estimates in simulations.

Extend our Vision beyond Machine-Detector Interface

Integration of DAQ, analysis and theory to optimize physics reach

Integration of DAQ, analysis and theory

- Research model with seamless data processing from DAQ to data analysis:
  - Not about building the best detector,
  - But the best detector that fully supports streaming readout and fast algorithms for alignment, calibration, and reconstruction in near real time.
  - For rapid turnaround of data for the physics analysis and to start the work on publications.

Streaming Readout for the EIC

Advantages of Streaming Readout

- **Simplification of readout**: No custom trigger hardware and firmware.
- Trigger-less readout idea for the **general-purpose detectors** of the EIC.
- **Opportunity to streamline workflows**: Merging of online and offline computing with combined software stack.
- Take advantage of other emerging technologies:
  - **AI**: Intelligent decisions in all aspects of data processing from detector readout and control to analysis.
  - **Heterogenous computing**.

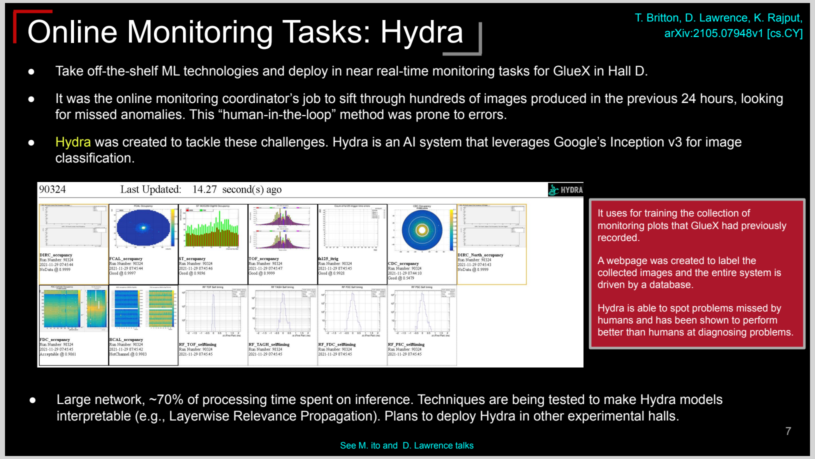
Development of Streaming Readout at Jefferson Lab

- EIC community fully determined to use streaming readout.
- Halls B and D have demonstrated the first operational test streaming readout DAQ system in on-beam tests (arXiv:2202.03085).
- Efforts and plans of some of the halls/experiments to evolve towards streaming readout and/or higher-level triggers will contribute towards EIC.
- INDRA-ASTRA project develops prototype for a fully automated, responsive detector system.

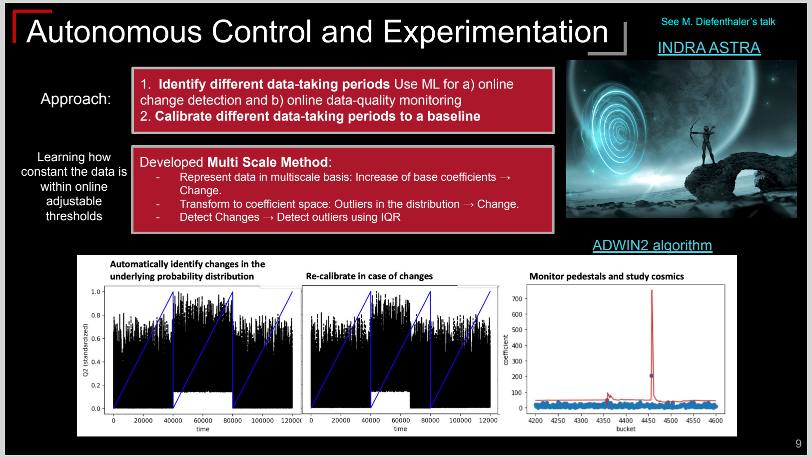
AI/ML for Streaming Readout at the EIC

Using Examples From Cristiano Fanelli's Presentation
Critical Path for Compute-Detector Model for the EIC

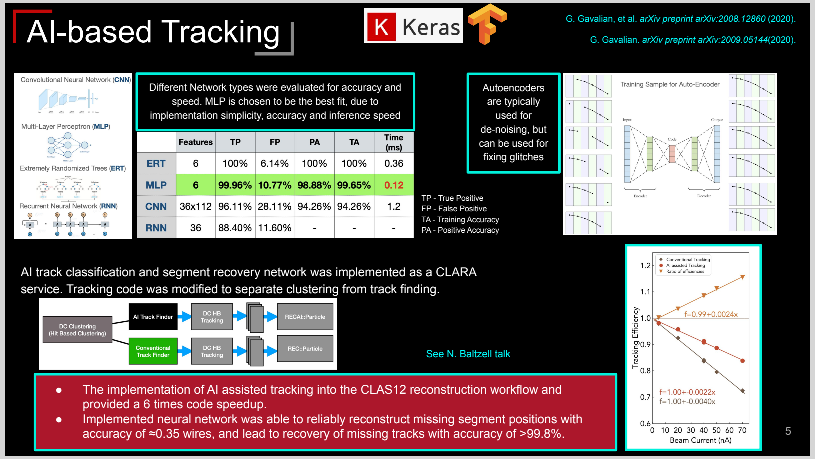
Automated Monitoring (arXiv:2105.07948)



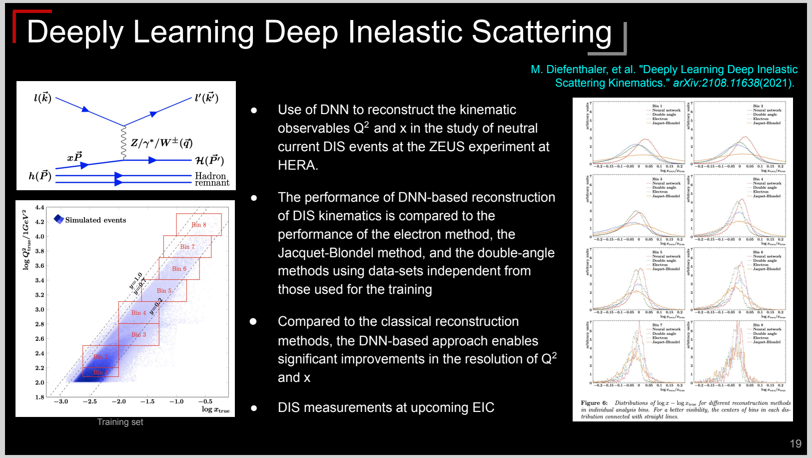
Automated Alignment and Calibrations



Tracking (arXiv:2009.05144, arXiv:2008.12860)



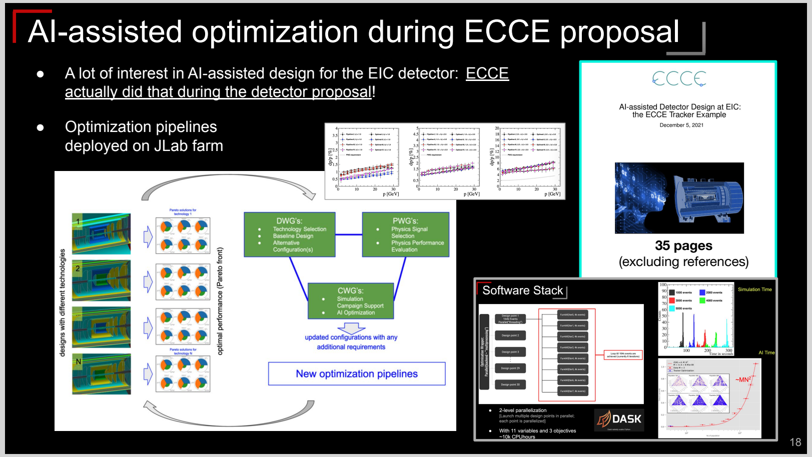
DIS Reconstruction (arXiv:2108.11638)

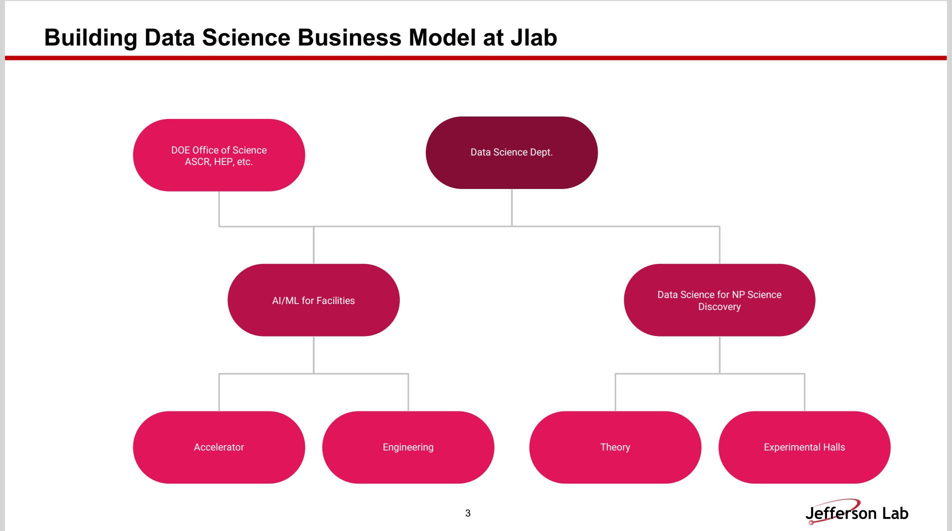


Jefferson Lab

AI/ML for the EIC

Detector Design Optimization





AI/ML will be an integral part of the EIC software using our expertise and experience from AI/ML applications from the 12 GeV CEBAF science program. We are incorporating AI/ML from the very beginning into the EIC, e.g., for detector design, and will systematically leverage on it during all phases of the project.

Interdisciplinary Approach: Data Science Department works with Accelerator Division, Experimental Physics Division, and the Theory Center

- Scientific, systematic approach to applying AI/ML approaches to EIC.
- Specific activation functions, network design for EIC applications.
- **Need to trust AI/ML**: Drive for uncertainty quantification.

Software Developments

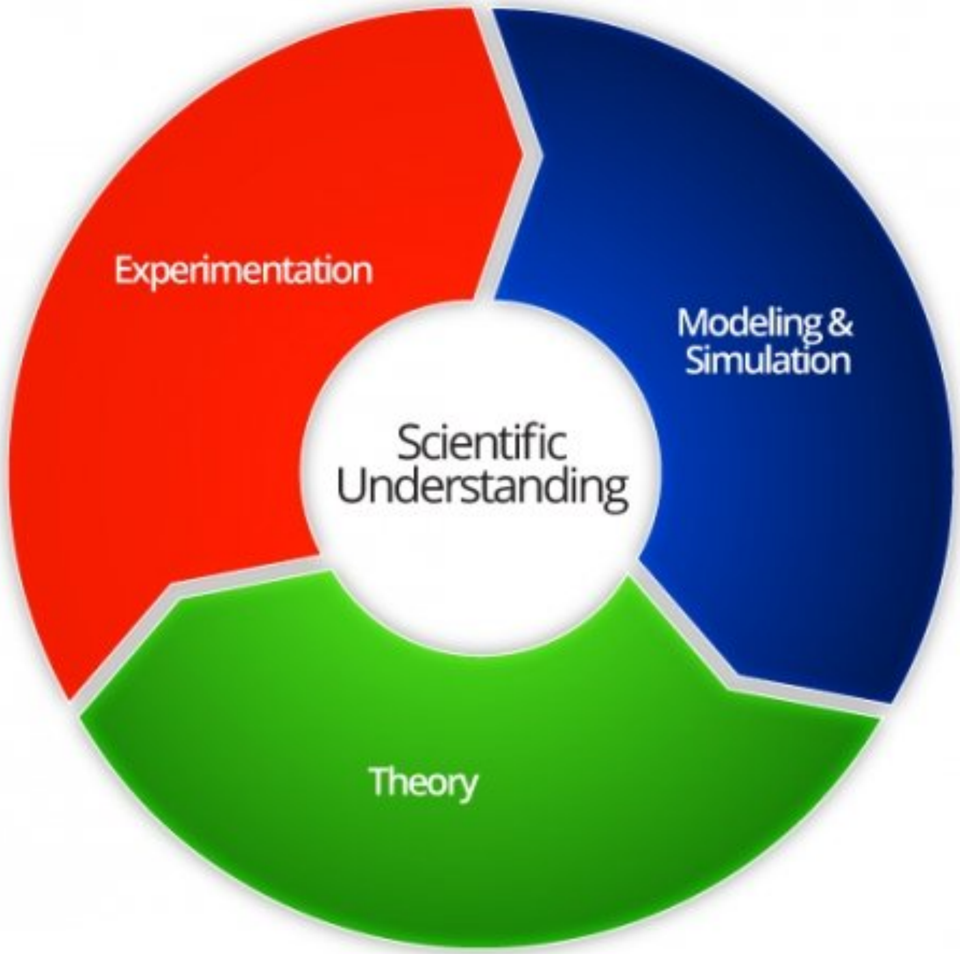
Software is in very early life stage

- Our focus on common software tools.
- We work with HEP standards, e.g., HepMC3 for MC event simulation,
- and engage with the wider community.

Community engagement

- **General-purpose MC event generators**: Herwig, Pythia, Sherpa
  - We lead community project on MC event simulation for ep and eA, current focus on validation with HERA data.
- **Geant4 collaboration**
  - Geant4 Technical Forum on NP experiments at Jefferson Lab in 2019
- **HEP Software Foundation**
  - Started to collaborate on Software & Computing Round Table and software tutorials. Discussion of NP Software Forum as part of HSF.
- **Software Working Group** within the EIC User Group.
  - We co-chair the SWG and work with the community to address software needs and evolving R&D.

Event Generators for the EIC



Monte Carlo Simulation of

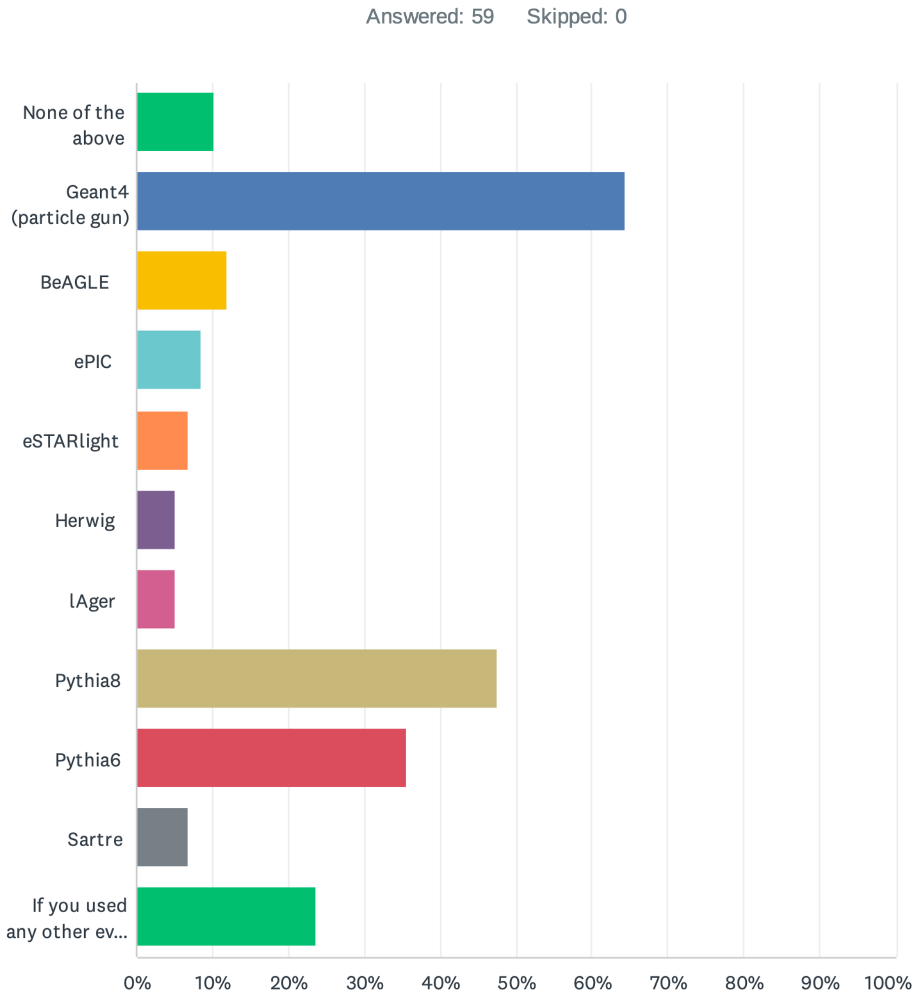
- electron-proton (ep) collisions,
- electron-ion (eA) collisions, both light and heavy ions,
- including higher order QED and QCD effects,
- including a plethora of spin-dependent effects.

Common challenges, e.g. with HL-LHC: **High-precision QCD measurements require high-precision simulations**.

Unique challenges MCEGs for electron-**ion** collisions and **spin-dependent** measurements, including novel QCD phenomena (e.g., GPDs or TMDs).

State of Software Survey in 2022

Q2 Over the past year, which event generation tool(s) did you use for EIC simulations? Check all that apply.

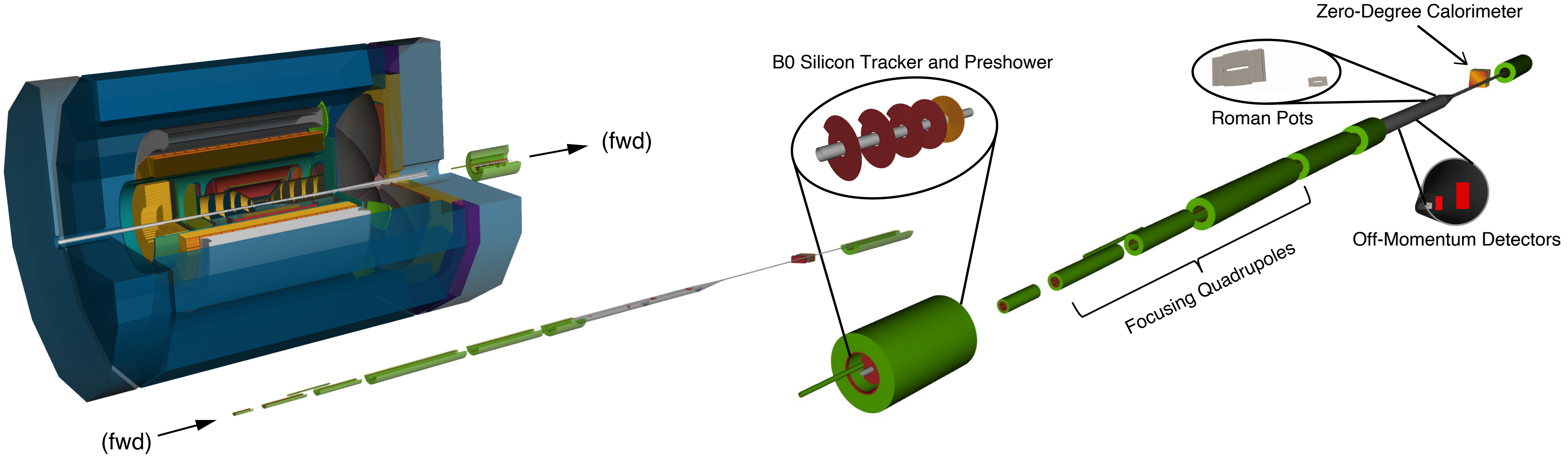


The EICUG SWG collected information on the **community's specific software tools and practices** during the call for detector collaboration proposals.

Other DEMPGen, Djangoh, elSpectro, TopHEG

Support Detector Design

- **Accelerate Simulations**
  - Detector design optimization using AI/ML as a service.
  - AI/ML for the simulation of calorimeters, Cherenkov detectors, etc.
    - Fast simulations fully integrated into Geant4.
- **Reconstruction**
  - AI/ML to accelerate reconstruction.
  - Reconstruction with far-forward detectors fully integrated (ATHENA example for IP6).



Our R&D Towards Next-Generation Detector Simulations: eA)ST

Detector Simulation

- **Turn-key application**
- **Built on top of Geant4 for full and fast simulations**
- **With library of potential detector options**

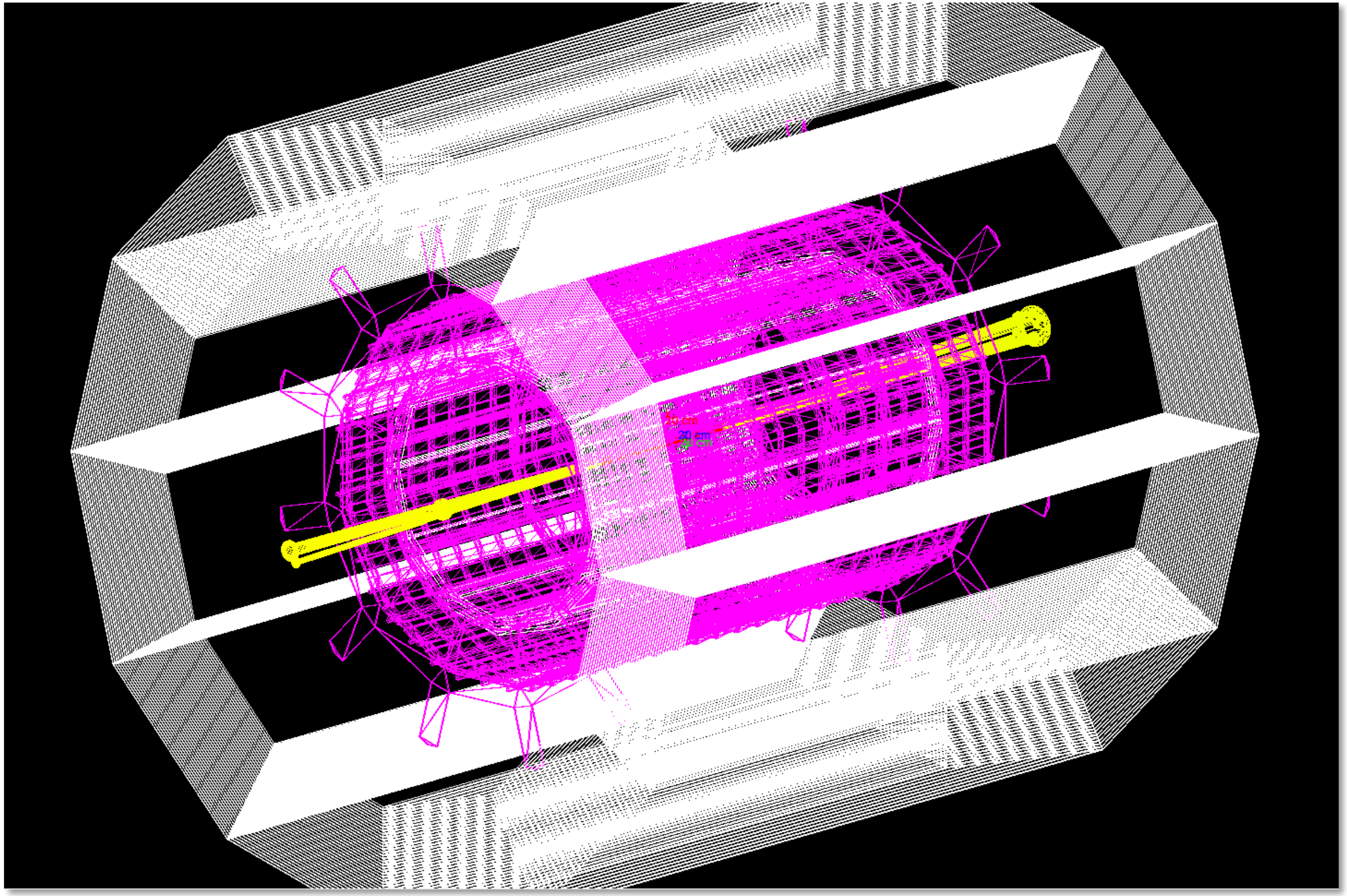
Requirements

- Ease of **leveraging new and rapidly evolving** technologies:
  - AI/ML to accelerate simulations
  - Heterogeneous architectures:
    - AI/ML is the best near term prospect for using LCF/Exascale effectively.
- Ability to reuse existing simulation work
- Ease of switching detector options
- Ease of switching between detailed and coarse detector descriptions

Project

- Support for high concurrency heterogeneous architectures and fast simulations integrated with full detector simulations allows to leverage AI/ML in Geant4.
- Makoto Asai, who led Geant4's multi-threaded reengineering to support high concurrency heterogeneous architectures, is now at Jefferson Lab and leading the next phase in concurrent Geant4, sub-event parallelism.
- We are building up team at Jefferson Lab on next-generation detector simulations with strong support from wider EIC community, in particular from BNL.

eAST: Reusing existing simulation works



Yellow volume is a beampipe imported from a CAD system (inside the pipe is filled by vacuum).

Red volumes are EM calorimeter support structure imported from another CAD system.

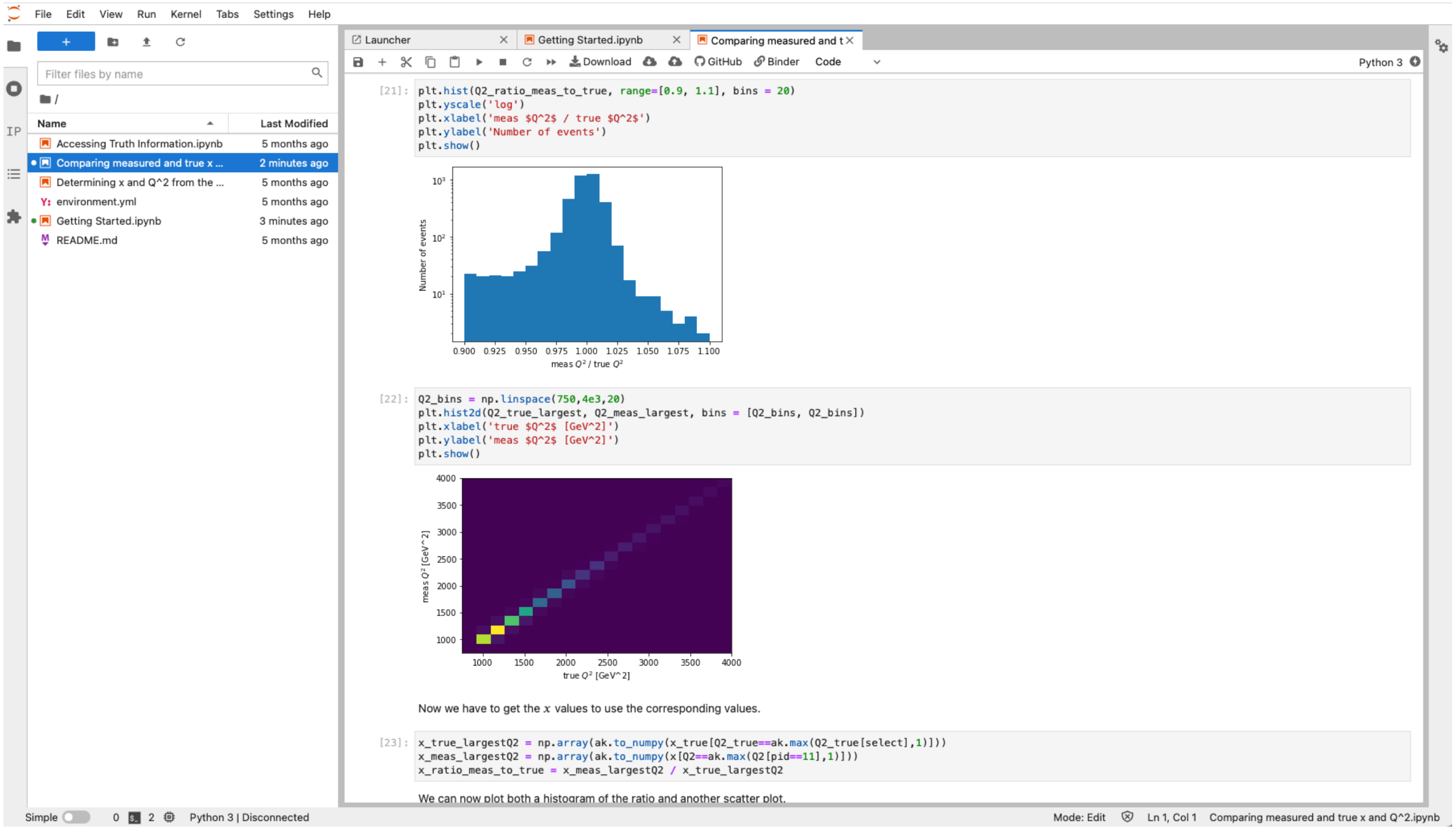
White volumes are hadron calorimeters imported from DD4HEP.

How to get started: https://mybinder.org/v2/gh/eic/python-analysis-bootcamp/main

Examples for DIS for a collider

File access via:

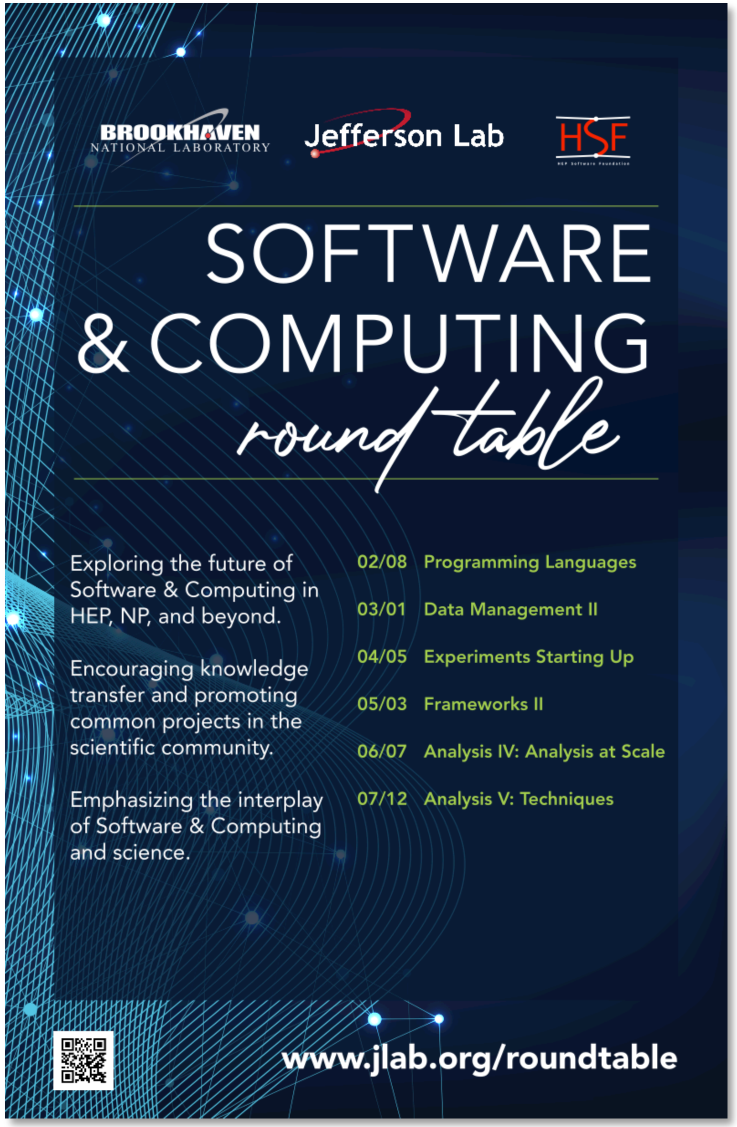
- **XRootD**
- **uproot**



How to get involved

- **You would like to develop the science program for the EIC**:
  - The main process at the EIC is: DIS. Get familiar with DIS measurements at a collider with multi-purpose detectors with full acceptance.
  - Start with our binder tutorials and prototype the analysis of your interest. You have open access to data from full detector simulations.
- **You would like to develop detectors for the EIC**:
  - Integrate your detector component into the detector concepts that are available in **eAST**.
- **You have developed a MCEG and would like to use it for the EIC**:
  - Add HepMC3 output to your MCEG and you can use it with all the detector simulation tools.
  - Do not worry about the integration of beam effects. The detector simulation tools handle that.
  - To get involved in MCEG R&D for the EIC: Engage with **MC4EIC**.
- **You would like to develop reconstruction algorithms for the EIC**:
  - Stay tuned. We are currently discussing the tool kit for the whole community.
- **You are interested in AI/ML for the EIC**:
  - Reach out to our Data Science Department (talk by Malachi Schram on Friday) or **AI4EIC** .
- **You would like to get involved in Software R&D**:
  - Join the EIC Software discussions in the EICUG: http://www.eicug.org/web/content/eic-software

Software & Computing Round Table



- Seminar series on the **interplay of computing and science**
- With O(50) participants per month
- Initiated at Jefferson Lab after the first "Future Trends in NP Computing" workshop in 2016 with two main goals:
  - **Knowledge transfer**
  - **Encourage common projects**
- Since 2020 **jointly organized with BNL and the HSF** with software & computing topics from the wider NP and HEP community.
- Recordings available on YouTube:

Markus Diefenthaler

mdiefent@jlab.org

Scientific Computing for the EIC

- The EIC will enable us to embark on a **precision study of the nucleon and the nucleus at the scale of sea quarks and gluons**. Software & Computing will be an integral part of EIC science.
- We focus on a **federated approach** for distributed computing and **common software projects** with the NHEP community.
- In synergy with the computing for the 12 GeV CEBAF science program, we are working to **accelerate science**:
  - AI/ML and heterogenous computing for next-generation simulations.
  - Seamless data processing from DAQ to analysis using streaming readout and AI/ML.