Markus Diefenthaler (Jefferson Lab)

ePIC Computing & Software Review Report

ePIC Computing & Software Review, September 26–27, 2024

A committee of experts, the **EIC Computing and Software Advisory Committee**, has been formed to advise the host laboratories on the progress and status of computing and software for the ePIC collaboration. Reviews are expected to take place on a regular cadence, with a **charge reflective of** the **EIC schedule**, the **stage of the ePIC experiment**, and **impending deadlines**. This charge covers an **assessment of the ePIC computing model in preparation for the November 2024 RRB meeting**. The scope of this review also includes the organization of the newly formed ECSJI.

EIC Computing and Software Advisory Committee (ECSAC), present at the review

Mohammad Al-Turany (GSI)

Simone Campana (CERN)

Christoph Pauss (MIT)

Verena Martinez Outschoorn (UMass Amherst)

Frank Würthwein (UCSD)

Presentations at the Review

- **Welcome**
  - **Speaker:** David Dean, Deputy Director for Science at Jefferson Lab
- **The EIC and the ePIC Detector: Getting to the Heart of Matter**
  - **Speaker**: Holly Szumila-Vance (Florida International University)
- **ePIC Software & Computing Overview**
  - **Speaker**: Torre Wenaus (BNL)
- **ePIC Streaming Computing Model**
  - **Speaker**: Markus Diefenthaler (Jefferson Lab)
- **ePIC Software Components: Building with the Community**
  - **Speaker**: Dmitrii Kalinkin (University of Kentucky)
- **International Computing and ePIC**
  - **Speaker**: Wouter Deconinck (University of Manitoba)
- **ECSJI: Status Report**
  - **Speaker**: Alexei Klimentov (BNL)

Selected Highlight from Presentations: Giving Back to the Community

- Contributed fixes for:
  - **ACTS**: improvements in tracking and infrastructure, with plans for more development
  - **DD4hep**: general bugs, HepMC3 ROOT and gzip support, new readout segmentation types, …
  - **EDM4hep**: build infrastructure
  - **Geant4**: addressed bugs UI, performance optimizations
  - **HepMC3**: addressed bugs in ROOT/XRootD IO, build infrastructure
  - **PODIO**: API improvements, build infrastructure, memory access semantics, streaming over network
  - **ROOT**: addressed bugs in UI, regressions in ROOT file handling
  - **Snakemake**: adressed general bugs
  - **Spack**: maintenance work on the package set
  - **UpROOT**: S3 object storage support (for legacy test setup at BNL/SDCC)
- EICrecon provides an ever expanding set of reusable LGPL3-licensed algorithms
- For HEP-NP projects, we encourage our users to submit their questions and problems to ePIC Helpdesk first

Selected Highlight from Presentations: 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.

ePIC Computing & Software Review Closeout

The report is being finalized and is expected to be available for the EIC RRB meeting on November 12-13, 2024.

In this presentation, we provide comments on the draft report, which is not publicly available.

ECSAC: ***"Overall, we think the ePIC computing & software preparations are in excellent shape for this early in the process."***

Charge Question 1.1

Is there a comprehensive and cost-effective short and long-term plan for the software and computing of the experiment?

- **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?**

Findings:

- Currently available computing resources sufficient for immediate needs of producing a TDR in 2025.
- Several software developments made in the last year, including build events from timeframes.
- No dedicated DOE effort for software and computing: Available effort based on synergistic contributions and best effort.

Comments:

- We congratulate ePIC for the achievements made in the last year in terms of software development. We believe that our recommendations from last year's review in that respect were addressed.
- There has been significant progress towards the TDR, showing good readiness of the software and simulation.
- **ePIC together with host labs need to develop a long term staffing needs plan for software and computing**:
  - Timing for such long term dedicated funding must be commensurate with ePIC Software & Computing being in a position to successfully deliver towards the miniDAQ milestone in FY28.

Recommendations:

- Assign dedicated effort to ePIC Software & Computing before the next ECSAC meeting.

Charge Question 1.1: Further Context

ECSAC: "**It is our impression that the S&C team has a clear plan for where near and long term additional effort would be invested into.**"

Short/Medium-Term (next 3 years), from homework questions

- 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), from homework questions

- 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

Charge Question 1.2

Is there a comprehensive and cost-effective short and long-term plan for the software and computing of the experiment?

- **Is the design of the ePIC computing model and resource needs assessment adequate for this stage of the project?**

Findings:

- Role of Echelon 2 w.r.t. to streaming needs to be clarified.
- ePIC Software & Computing developed detailed resource assessment leading to predict the storage and compute needs for EIC Phase I.

Comments:

Recommendations:

- None.

Charge Question 1.3

Is there a comprehensive and cost-effective short and long-term plan for the software and computing of the experiment?

- **Is the ePIC computing and model flexible? Can it evolve and integrate new technologies in software and computing?**

Findings:

- ePIC Computing model different from the current NHEP experiments to allow for more flexibility.
- Flexibility aims to:
  - Foster collaboration in software development,
  - Leveraging opportunities at facilities for contributing computing services.

Comments:

- Flexibility of the computing model is an important mechanism for engaging partners and leveraging synergies.
- At the same time it comes with the cost of complexity and that has to be properly balanced.

Recommendations:

- None.

Charge Question 2

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?

Findings:

- ePIC Software & Computing plans well integrated with standard practices across NHEP.
- ePIC Software & Computing uses many common tools and are active contributors to several.
- Statement of Software Principles designed:
  - To take advantage of future developments, and
  - To be flexible to adapt to new technologies.

Comments:

- **ECSAC**: *"We congratulate the S&C team for a job well done."*

Recommendations:

- **None.**

Charge Question 3

Are the ECSJI plans to integrate into the software and computing plans of the experiment sufficient?

Findings:

- Presentation on the role of the ECSJI, the EIC Computing & Software Joint Institute between the host labs.

Comments:

- Upcoming long-term decision on ECSJI's role in funding for EIC Computing & Software.

Recommendations:

- None.

Charge Question 4

Are the plans for the integrating international partners' contributions flexible and adequate at this stage of the project?

Findings:

- **Globally distributed computing model** being developed that is similar to existing global computing efforts.
- **Increasing number of international partners** in the process of joining the computing and software infrastructure.

Comments:

- Streaming readout makes the computing model different from other such models in NHEP.
- International partners appear to be at a healthy and commensurate level with the size of the project at this stage.
- **ECSAC surprised that we did not present potential Echelon 2 contributions from within the U.S.**:
  - **ECSAC:** *"This is a gap where potential resources and effort could be tapped into."*

Recommendations:

- 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.

Review Summary

ECSAC excited about ePIC Software & Computing status and plans:

- A lot of engagement. Intense, productive discussions.
- Many comments on how to present the compute-detector integration by ePIC and our computing plan to the NHEP community and how to improve it.

Two Recommendations from ECSAC Review:

- **To host labs**: Assign dedicated effort to ePIC Software & Computing before the next ECSAC meeting.
- 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.

Next Steps

- Share revised report on ePIC Computing model with collaboration (by end of next week).
- Incorporate ECSAC feedback in ePIC Software & Computing in EIC RRB meeting (November 12-13).
- Follow up on review comments and recommendations (stay tuned).

Backup

Highlight: Storage Resource Estimates

Actual needs in 2034.

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**

Echelon 1 sites arrive data, two copies
One copy (can and may be more) across Echelon 2 sites for alignment, calibration, and reconstruction use cases.

Highlight: Computing Resource Estimates

Actual needs in 2034,

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**

See prompt reconstruction.

Roughly 10% of data stream.

Must keep up with data taking; assume 2x headroom.

Reprocessing includes simulation as well as data.

Simply adding together the core counts is an overestimate. Reconstruction core hours used only part time.