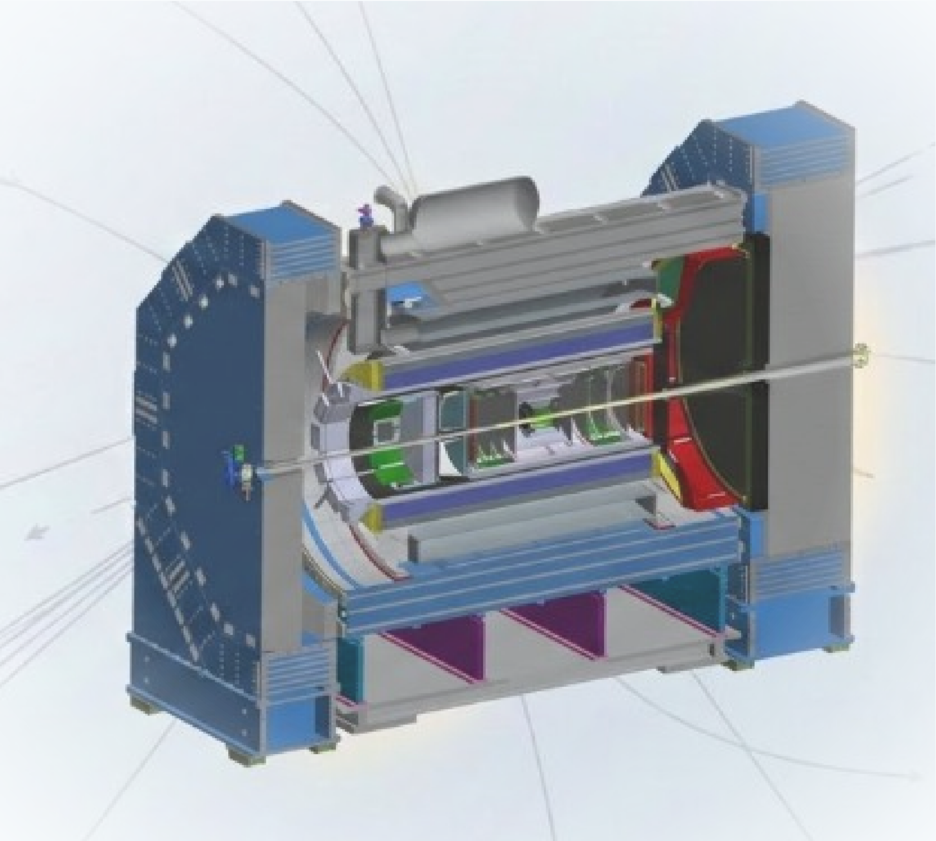
Exploring Potential Pathways to Accelerate ePIC Detector Simulation

Sakib Rahman

on behalf of the ePIC Collaboration

Brookhaven National Laboratory



Computing in High Energy and Nuclear Physics (CHEP)
Chulalongkorn University, Bangkok, Thailand
26 May 2026

The ePIC Experiment at the Electron-Ion Collider

1157

Members

183

Institutions

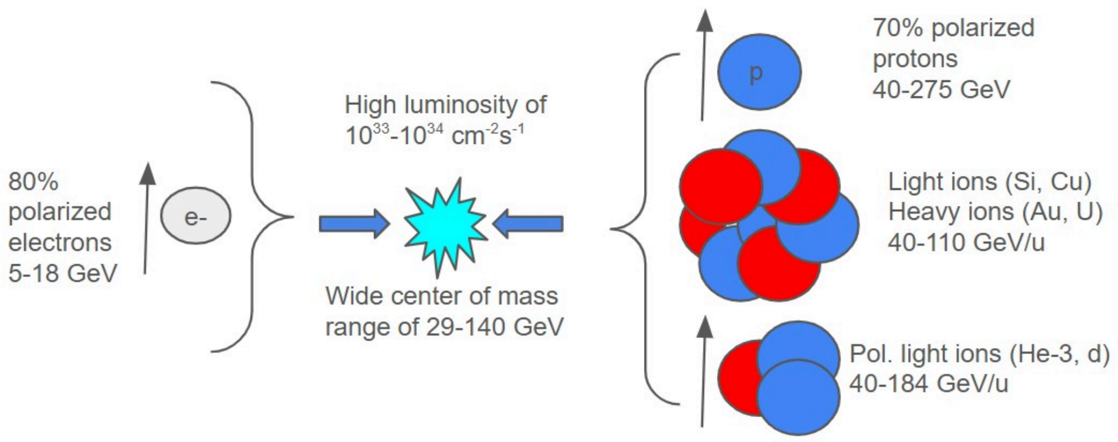
26

Countries

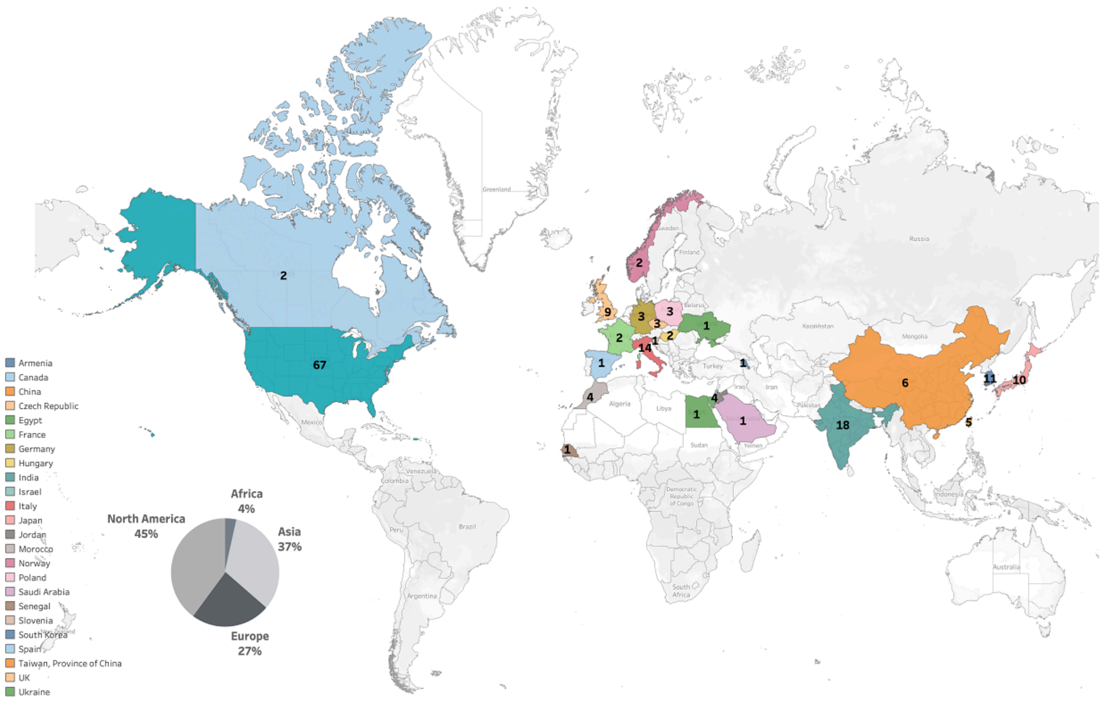
ePIC will be the first detector at the future Electron-Ion Collider. Data taking planned for the **early 2030s**

Being realized through a partnership between host labs: **Brookhaven National Laboratory (BNL)** and **Jefferson Lab (JLab)**

Will enable precision studies of nucleons and nuclei at the scale of sea quarks and gluons



EIC Beam Specifications



ePIC: A Global Collaboration

The ePIC Physics & Detector Simulations Working Group

- Sits within the **Software & Computing** branch of the ePIC collaboration
- Embrace **streaming readout** as data format (See Takuya's talk)
- Enable seamless and efficient detector design and integration in simulation, including validation with engineering design
- Implement background timing structure
- Deliver accurate MC sim with **Geant4** and **DD4hep**
- Coordinate with User Learning to inform the collaboration about our detector simulation tools (See Alexandr's talk)

SPOKESPERSON'S OFFICE		
TECHNICAL COORDINATION	SOFTWARE & COMPUTING	ANALYSIS COORDINATORS
Tracking	⭐ **Physics & Detector Simulations**	BSM & Precision EW
Electronics, Readout & DAQ	Reconstruction	Exclusive, Diffraction & Tagging
AC-LGAD	Streaming Computing	Jets & Heavy Flavor
Calorimetry	User Learning	Inclusive Physics
PID	Production	Semi-Inclusive Physics
	Validation	

Explore Acceleration Pathways

Multithreading, sub-event parallelism, and **GPU offload** to scale simulations across heterogeneous resources in the **ePIC Computing Model** (See Holly's talk)

ML-based Fast Simulation

Rapid iteration across detector configurations and reconstruction algorithms

Tractable Background Simulations

Acceleration reduces per-event cost to make large-scale background production and background-inclusive simulations tractable (See my other talk)

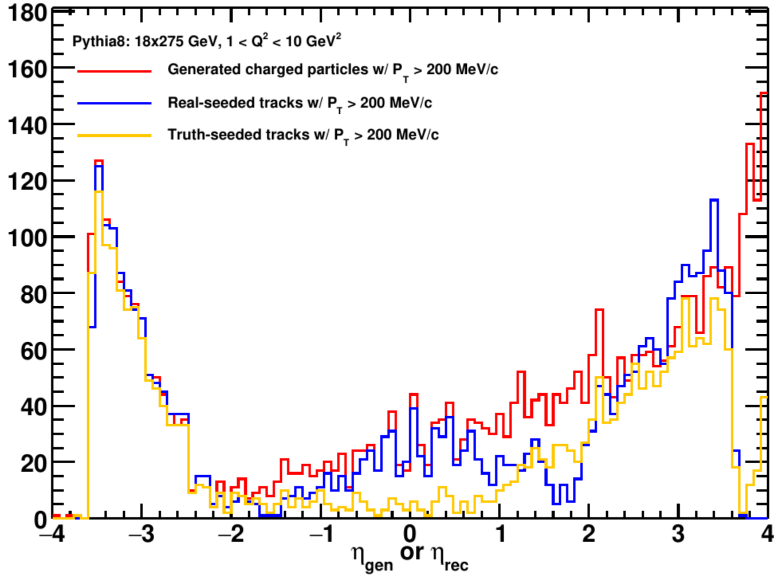
Accelerating ePIC with Multithreading: Physics Validation

⚠️ PRELIMINARY

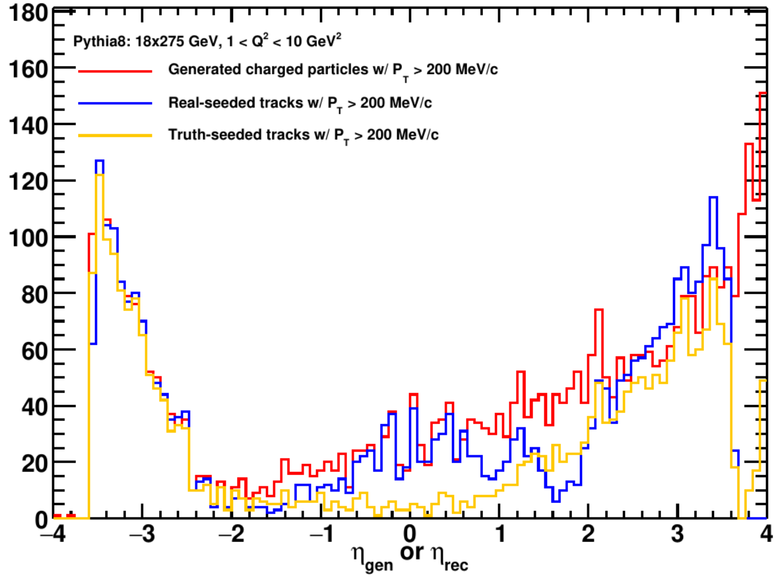
Tracking η distributions compared across thread counts using 1000 realistic deep inelastic scattering events. Generated charged particles, real-seeded and truth-seeded tracks show consistent agreement.

Enabled by DD4hep#1240 — multithreading support added to DD4hep.

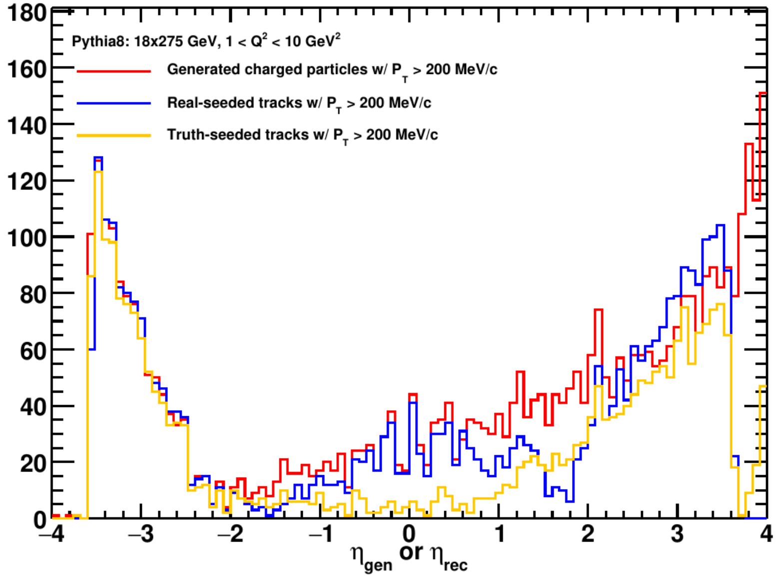
8 Threads



64 Threads



128 Threads



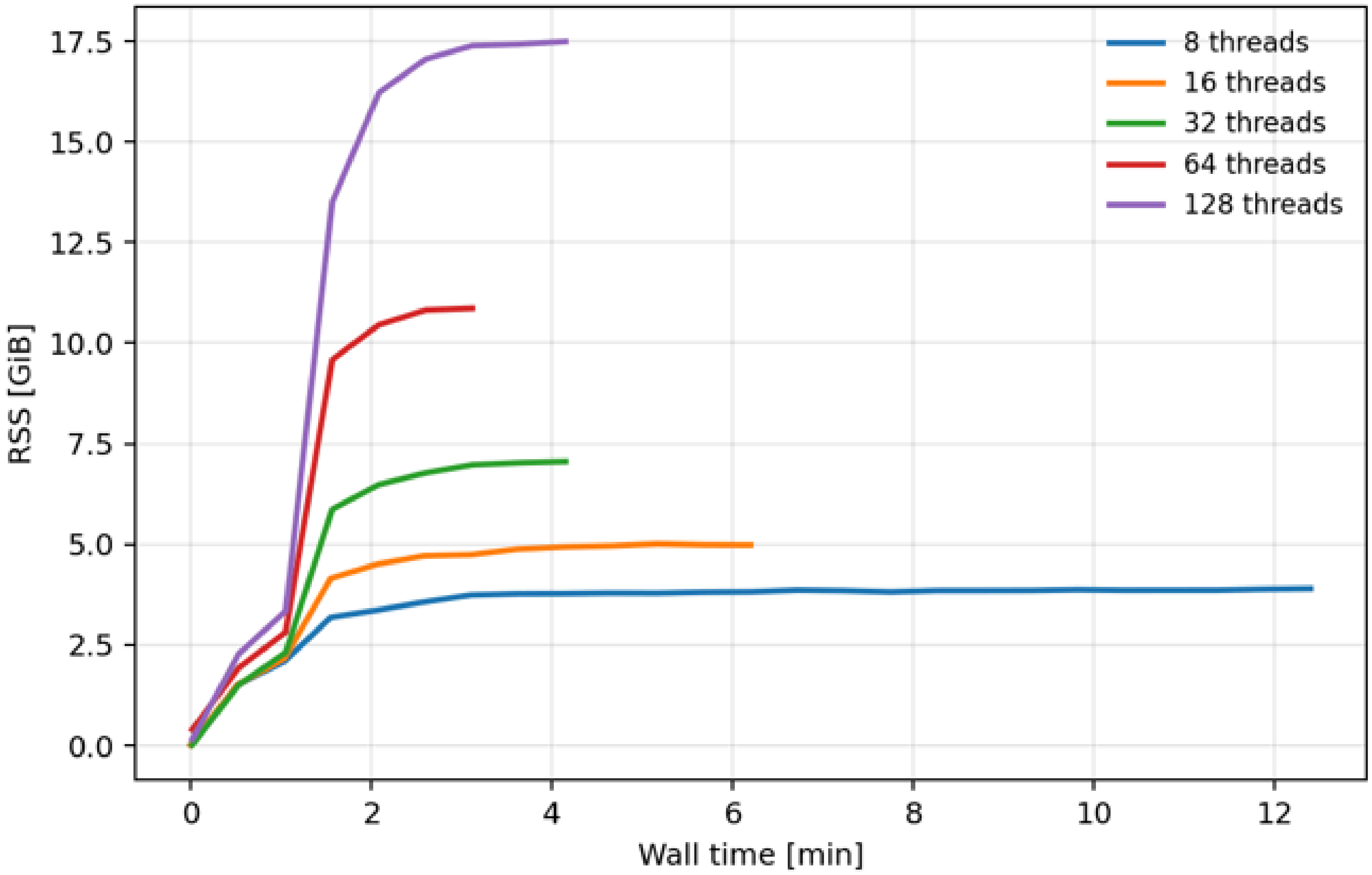
Geometry	epic_craterlake 26.04.1
Container	eic_dev_cuda:unstable-mr-249
CPU	1 × AMD EPYC 7763, 64 physical cores
Site	Perlmutter
Input	`root://dtn-eic.jlab.org//volatile/eic/EPIC/EVGEN/DIS/pythia8.316-1.0/NC/noRad/ep/18x275/q2_1to10/pythia8.316-1.0_NC_noRad_ep_18x275_q2_1to10_run000.hepmc3.tree.root`

Accelerating ePIC with Multithreading: Performance

⚠️ PRELIMINARY

npsim wall time and memory scaling across thread counts for 1000 realistic deep inelastic scattering events. Either 32 or 64 threads could be optimal running condition.

Geometry	epic_craterlake 26.04.1
Container	eic_dev_cuda:unstable-mr-249
CPU	1 × AMD EPYC 7763, 64 physical cores
Site	Perlmutter
Input	`root://dtn-eic.jlab.org//volatile/eic/EPIC/EVGEN/DIS/pythia8.316-1.0/NC/noRad/ep/18x275/q2_1to10/pythia8.316-1.0_NC_noRad_ep_18x275_q2_1to10_run000.hepmc3.tree.root`



Understanding the Life of ePIC Events

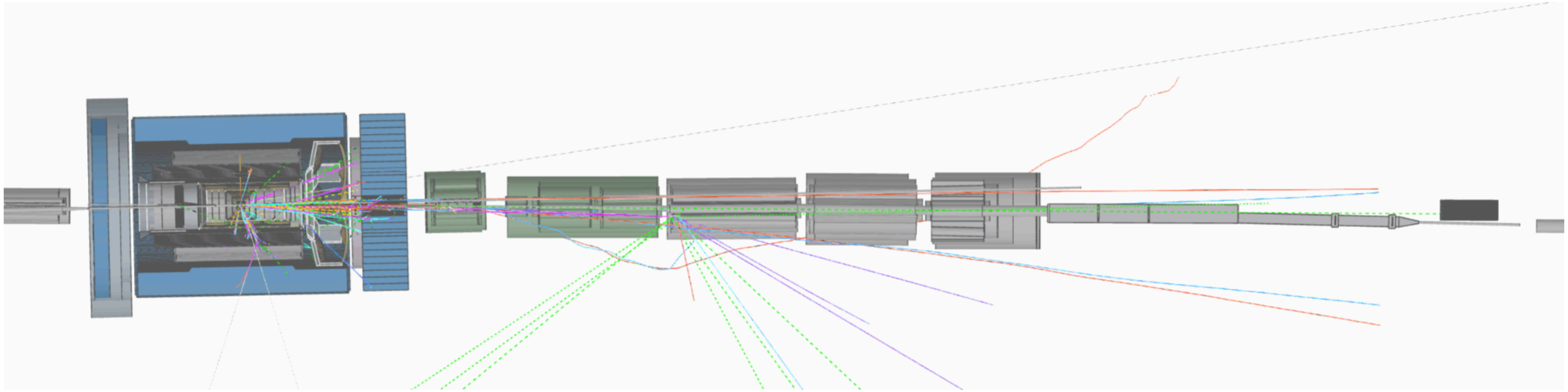
Central, far-forward, and far-backward subsystems provide broad pseudorapidity coverage, enabling the study of many different physics processes. We profile the lifetime of ePIC events to improve our computational efficiency.

Current Validation

Pythia8 NC DIS, 18×275 GeV, $Q^2 = 1$–$10\ \mathrm{GeV}^2$, no radiative corrections

Future Work

Extend to all relevant physics processes. Time budget and acceleration targets vary by dataset



ePIC Detector in the Firebird Event Display

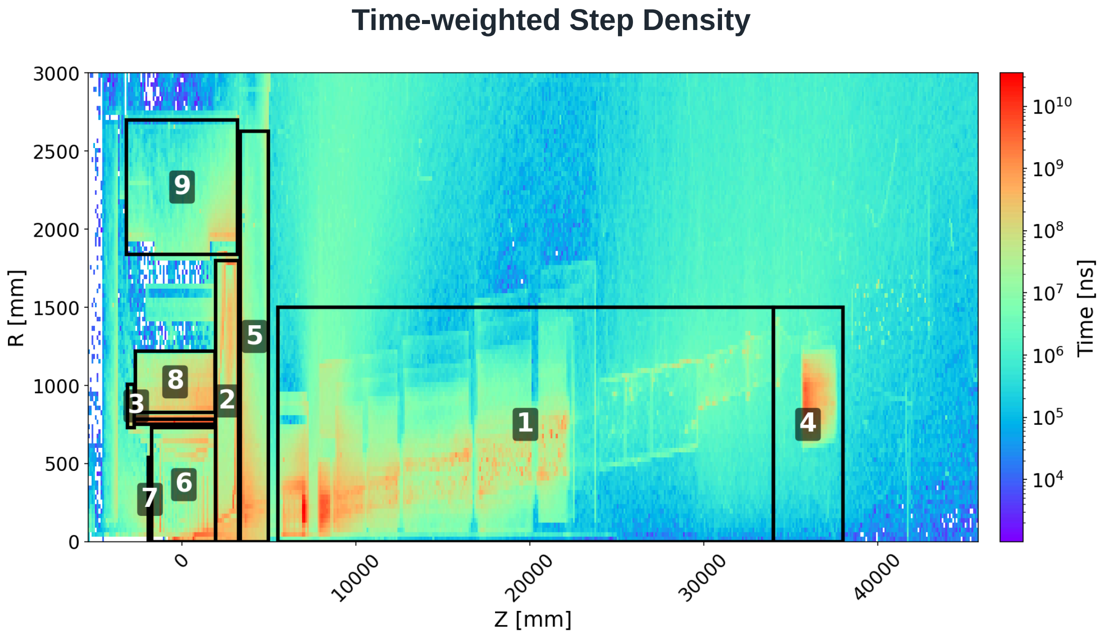
Acceleration Potential: Identify Simulation Time Sinks

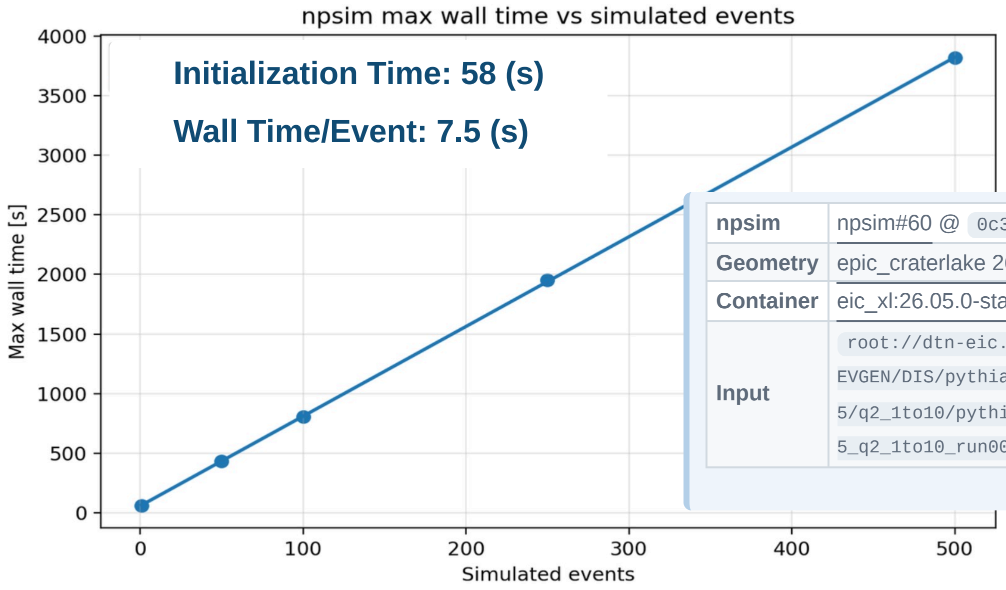
⚠️ **PRELIMINARY**

Time profiling of single-threaded run with realistic deep inelastic scattering events. The step density plot shows where stepping activity and simulation time is concentrated.

#	Subsystem	Step %
1	Far Forward (Other)	43.4%
2	dRICH	9.5%
3	hpDIRC	9.0%
4	Far Forward (ZDC)	8.1%
5	Forward EM+Hadron Cal	7.7%
6	Tracker / Beampipe	6.3%
7	EEEMCal	6.1%
8	Barrel Imaging+SciFi Cal	3.6%
9	Barrel HCal	1.5%

Focuses on stepping time only — track-level actions and other overhead not captured



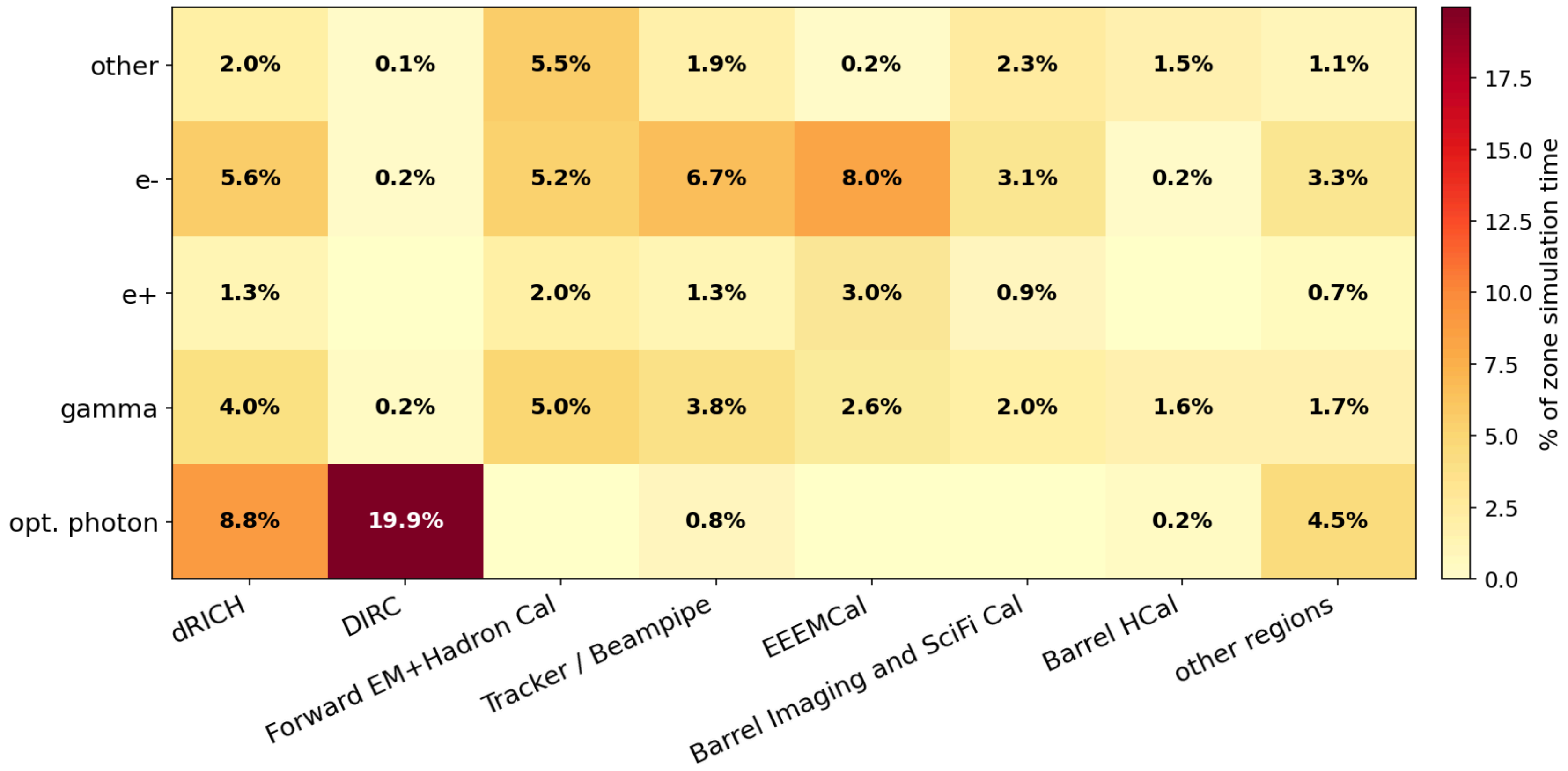


npsim	npsim#60 @ `0c316b6`
Geometry	epic_craterlake 26.05.0
Container	eic_xl:26.05.0-stable
Input	`root://dtn-eic.jlab.org//volatile/eic/EPIC/EVGEN/DIS/pythia8.316-1.0/NC/noRad/ep/18x275/q2_1to10/pythia8.316-1.0_NC_noRad_ep_18x275_q2_1to10_run000.hepmc3.tree.root`

Acceleration Potential: Central Detectors

⚠️ PRELIMINARY

Central Detectors — 43.6% of stepping time



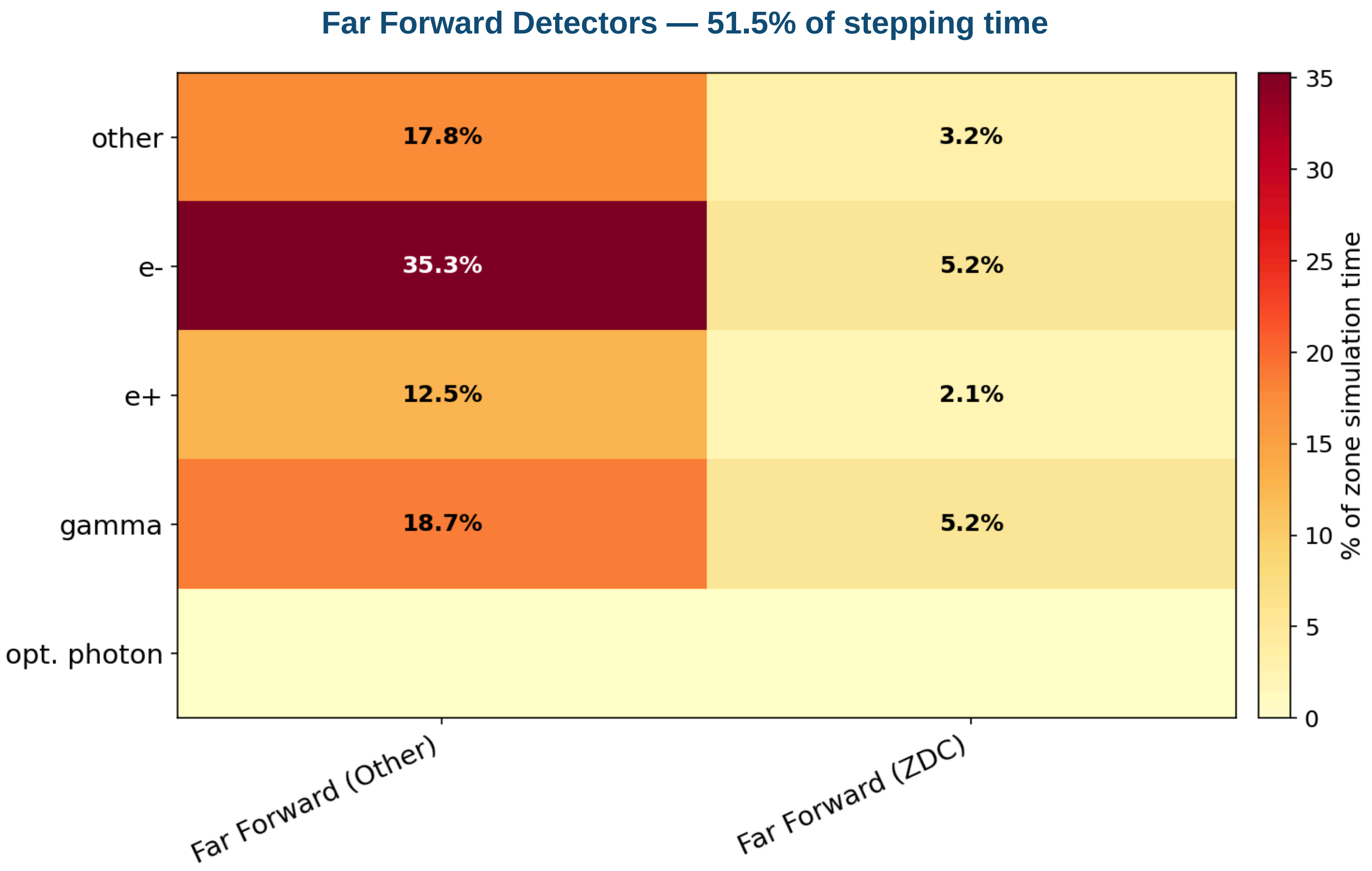
Subsystem / PID tuning

- Optical-only acceleration sufficient for hpDIRC and dRICH
- Central-only production: theoretical max **~1.4×** speedup

npsim	npsim#60 @ `0c316b6`
Geometry	epic_craterlake 26.05.0
Container	eic_xl:26.05.0-stable
Input	`root://dtn-eic.jlab.org//volatile/eic/EPIC/EVGEN/DIS/pythia8.316-1.0/NC/noRad/ep/18x275/q2_1to10/pythia8.316-1.0_NC_noRad_ep_18x275_q2_1to10_run000.hepmc3.tree.root`

⚠️ PRELIMINARY

Acceleration Potential: Far Forward Detectors



Full ePIC production

- Far Forward dominates stepping time
- Must accelerate both EM and optical transport for meaningful gains

npsim	npsim#60 @ `0c316b6`
Geometry	epic_craterlake 26.05.0
Container	eic_xl:26.05.0-stable
Input	`root://dtn-eic.jlab.org//volatile/eic/EPIC/EVGEN/DIS/pythia8.316-1.0/NC/noRad/ep/18x275/q2_1to10/pythia8.316-1.0_NC_noRad_ep_18x275_q2_1to10_run000.hepmc3.tree.root`

Simphony: GPU Acceleration for Optical Photons

Simphony (prev. EIC-Opticks) is a fork developed at BNL of Opticks (Simon Blyth). During full Geant4 Monte Carlo simulation, optical photon transport is accelerated using the GPU-based NVIDIA OptiX framework.

EXPERIMENT-AGNOSTIC IMPROVEMENTS

Ease of Deployment

- Spack package
- Tagged containers

Additional Core Functionality

- Wavelength shifting
- External JSON configuration (in place of environment variables)

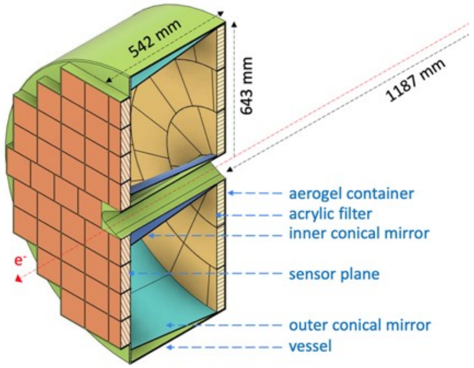
Examples of New Use Cases

- Event batching — improves GPU efficiency for ePIC-like low-to-moderate photon yields per event
- Monte Carlo truth propagation
- Maximize CPU and GPU utilization through overlapping execution
- DD4hep integration

Geometry-specific Parameter Tuning

- Max photon steps per kernel launch — prevents long-lived stragglers stalling threads
- Ray offset tuning for better fidelity

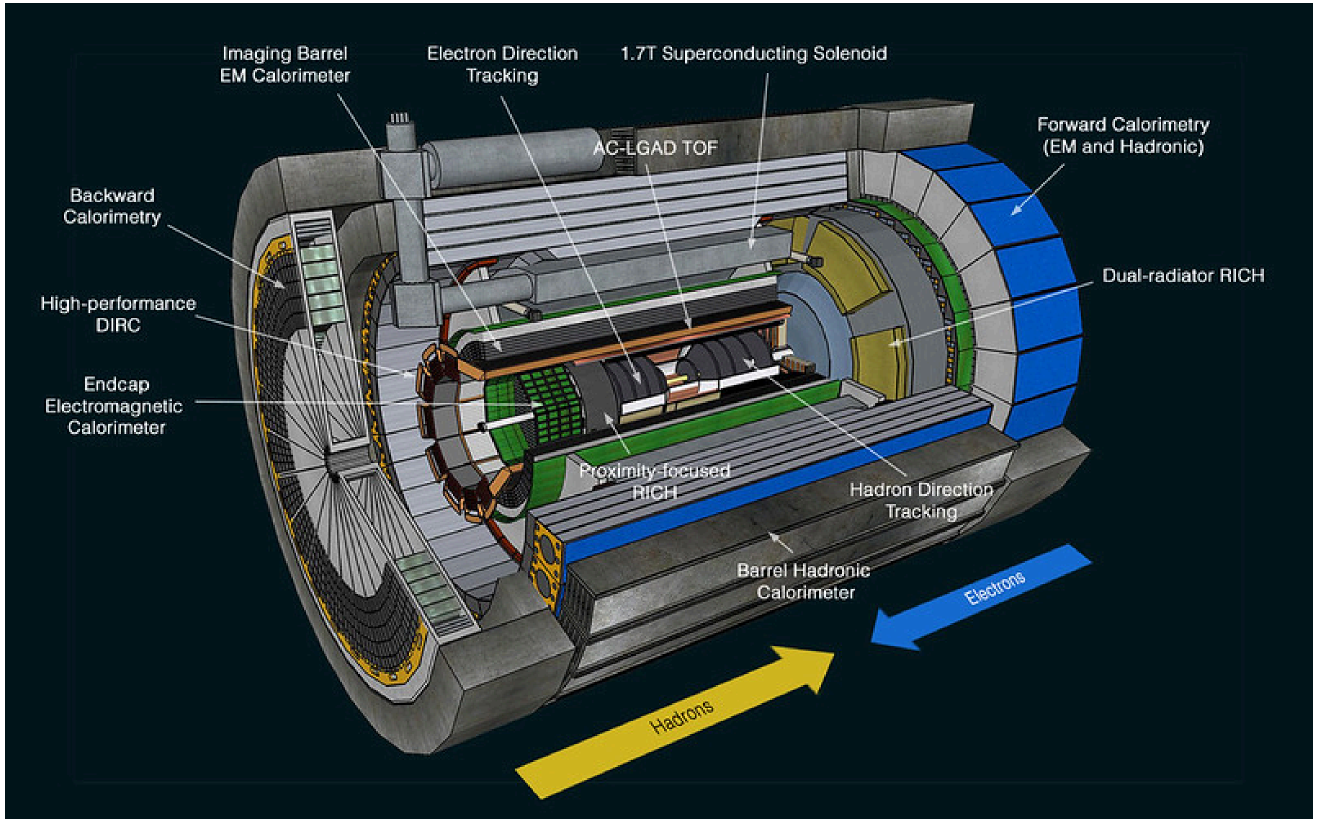
Simphony Use Case: Standalone ePIC Subsystem Design Simulations

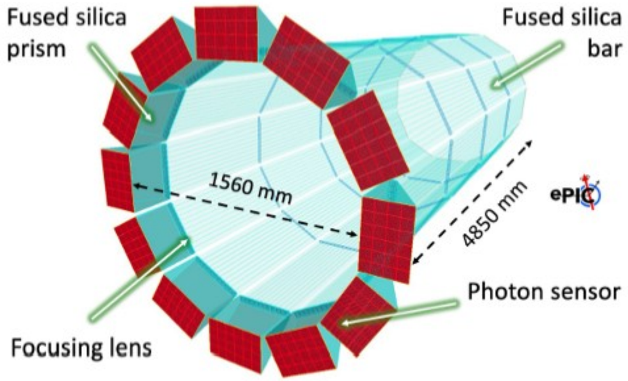


pfRICH · Dongwi et al. (2025)

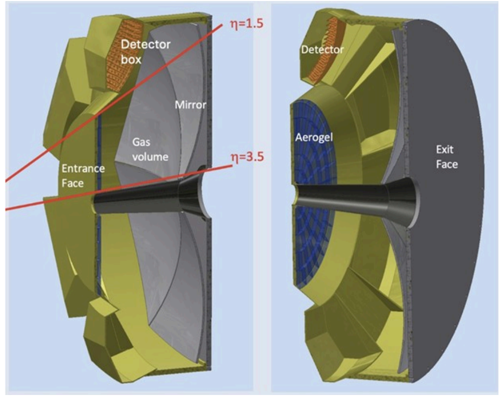
Optical photon transport dominates simulation time for PID detectors: hpDIRC, pfRICH and dRICH.

C. Chatterjee et al. (2024)





hpDIRC · G. Kalicy (2025)

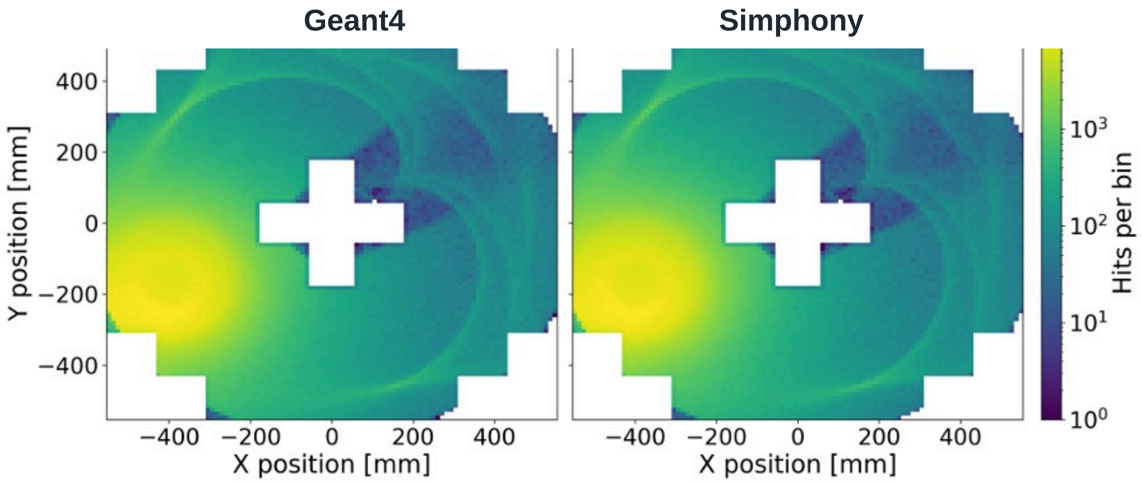


dRICH · Contalbrigo et al. (2026)

Simphony: Validation and Performance for ePIC Use Cases

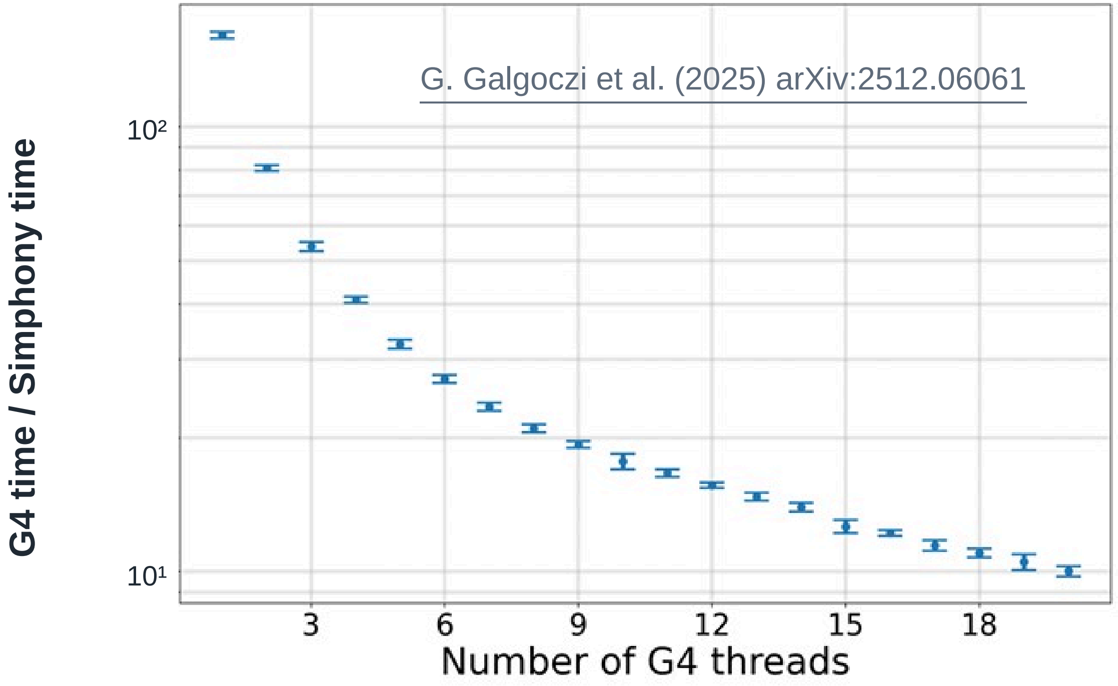
Physics Validation — pfRICH

- 50,000 electrons · p = 5 MeV/c · ~44M optical photons
- Recorded hits agree within statistical uncertainties:
  Geant4: 8.693 ± 0.003 M
  Simphony: 8.694 ± 0.003 M



Performance — pfRICH

GPU: NVIDIA GeForce RTX 4090 · CPU: Intel Xeon w7-3445



- Validation and performance studies for dRICH and hpDIRC ongoing
- Preliminary hpDIRC result: **~260× speedup** vs single-thread Geant4 with 10M optical photons
- We hope to reproduce these studies across a range of heterogeneous resources

AdePT and Celeritas: EM GPU Acceleration

AdePT and **Celeritas** both offload EM shower transport to GPU by leveraging the Geant4 Tracking Manager, and can potentially enable GPU-accelerated ePIC simulation.

Exploring Celeritas in ePIC

- Added DD4hep plugin
- Added support for realistic ePIC 3D RZ field maps
- Partial MC truth propagation
- *Work in progress:*
  - Optical photon support in DD4hep plugin for comparisons with Simphony
  - Benchmarks with realistic DIS events on simple pre-shower geometry
  - Benchmarks on standalone calorimeter subsystem

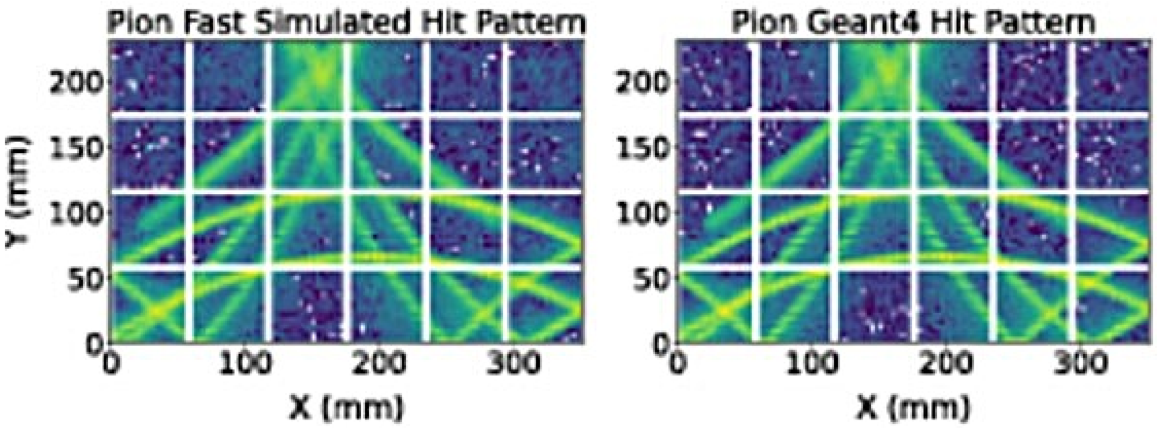
Exploring AdePT in ePIC

- Added DD4hep plugin
- End-to-end pipeline with MC truth propagation
- Example SiD geometry: fully working, initial benchmarking done
- *Work in progress:*
  - Implementing solution for tessellated regions in ePIC geometry
  - Detailed ePIC simulation benchmarking

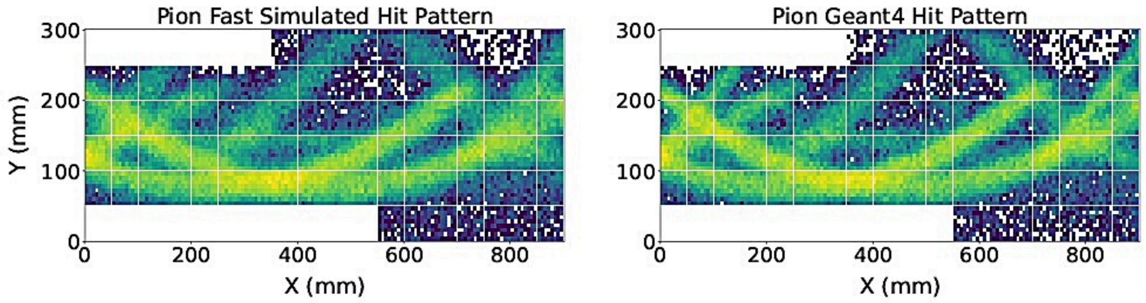
Accelerated Simulations: hpDIRC Example and Beyond

Foundation Model for DIRC Detectors

- Transformer-based Foundation Model with a Mixture-of-Experts (MoE) architecture developed for DIRC detectors [1,4]
- Single pre-trained FM supports multiple downstream applications via fine-tuning:
  - High-fidelity, fast simulations (topology + photon yield)
  - Near real-time PID / Classification
  - Noise Filtering
- Demonstrated on the ePIC hpDIRC and the GlueX DIRC at Jefferson Lab [1,2,3,4], and now being extended to calorimetry [5]



Foundation Model for ePIC



Foundation Model for GlueX

[1] J. Giroux, C. Fanelli, *MLST* 7.1 (2026): 015031 [2] C. Fanelli et al., *MLST* 6.1 (2025): 015028 [3] J. Giroux et al., *MLST* 6 (2025): 040501 [4] C. Fanelli et al., arXiv:2604.24775 (2026) [5] C. Cardona et al., arXiv:2603.28804 (2026)

Acceleration Challenges

CUDA Architecture Portability

GPU payloads and all dependencies must be built against the CUDA arch of the destination hardware; mismatched builds cause runtime crashes

GPU Parameter Tuning

GPU-specific parameters must be tuned to event topology per dataset and available hardware resources

Workflow Integration

PanDA WMS adaptive brokerage will match containers to GPU resources, but the container build matrix grows with each new arch target

Physics Validation Fidelity

GPU-offloaded transport must reproduce Geant4 reference within analysis-level tolerances across all relevant observables

ML Model Generalization

- Models must be trained to take into account different beam conditions, physics processes and kinematic ranges
- Retraining is needed when geometry changes as well

Event Batching and Sub-event Level Parallelism

- Address GPU under-utilization due to simplistic ePIC event topology
- Integration challenges yet to be carefully studied

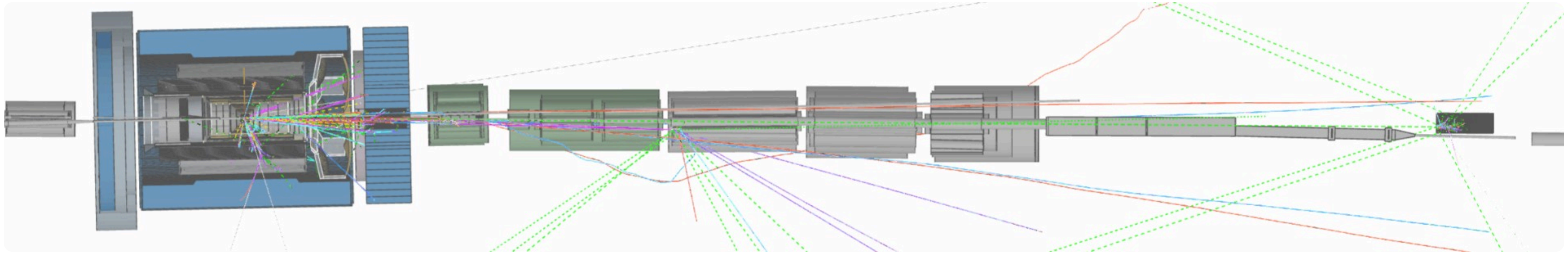
Summary and Near-Term Work

What We Showed

- Multithreading enabled in DD4hep
  - Physics-validated, optimal at 32–64 threads
- Preliminary profiling shows that optical photons and EM showers dominate simulation cost
- Simphony:
  - GPU acceleration for optical photons in ePIC PID detectors
  - Order-of-magnitude speedup vs 20 Geant4 threads, validated on pfRICH

What We Are Exploring

- AdePT and Celeritas:
  - GPU acceleration for EM tracks
  - Integration and benchmarking in progress
- ML-based fast simulation/Foundation models
  - hpDIRC example
  - Large-scale training data generation with Simphony
- Event batching and sub-event level parallelism



Related ePIC Talks at CHEP 2026

The ePIC Streaming Computing Model
Holly Szumila-Vance

Scaling ePIC Simulation Production: Distributed Workflow and Data Management
Sakib Rahman

Development of Streaming Data Reconstruction for ePIC Experiment at EIC
Takuya Kumaoka

ePIC User Learning Training and Documentation Strategies
Alexandr Prozorov