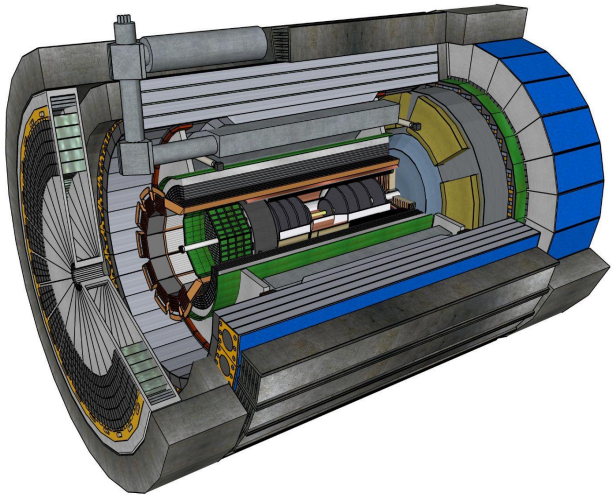
The ePIC Simulation Campaign Workflow on the Open Science Grid



Sakib Rahman
Nuclear and Particle Physics Software (NPPS)
Brookhaven National Laboratory
On behalf of the ePIC Production Working Group

21 October 2024
Computing In High Energy and Nuclear Physics (CHEP)
Venue: Jagiellonian University, Krakow, Poland

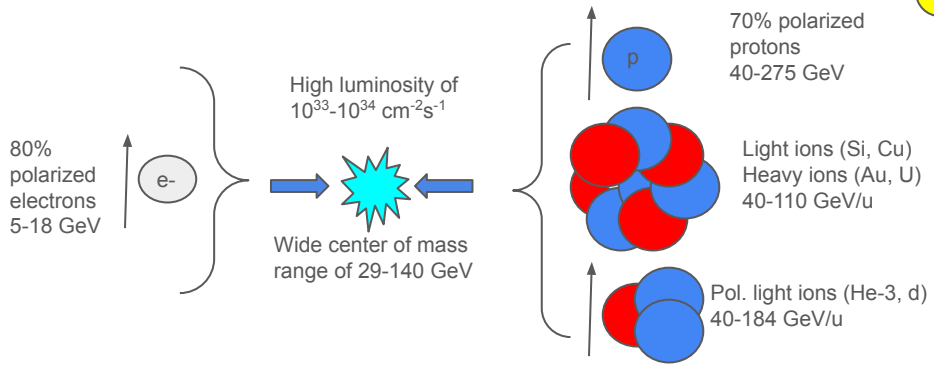
Introduction

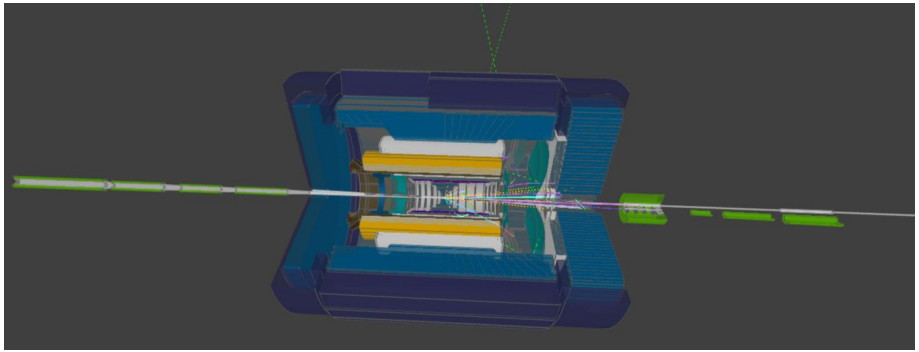
ePIC, the first experiment at the future Electron-Ion Collider (EIC), will be realized in partnership of host labs - Brookhaven National Laboratory (BNL) and Jefferson Lab (JLab).

International collaboration with 173 institutions worldwide will provide insight into the nucleon and nucleus down to the scale of sea quarks and gluons leveraging unique and versatile EIC beam specs.

Large-scale monthly simulation production campaigns on the Open Science Grid (OSG) form the basis for detector and physics studies in preparation for Technical Design Report. **Since May 2023, ePIC simulation campaigns have cumulatively consumed ~2000 core-years of OSG compute resources to simulate ~500 TB of data. Current compute and disk usage per monthly campaign: ~200 core-years and ~30 TB.**

Develop distributed computing capabilities in advance of operations phase.





Detector

Highly integrated, multi-purpose detector to measure all final-state particles of interest with close to full acceptance and good resolution.

Uniquely designed for asymmetric beam energies.

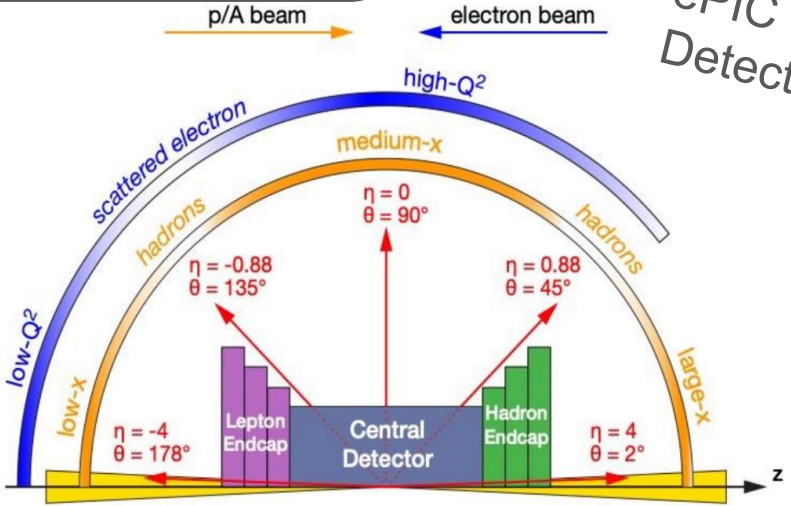
High-precision Si detectors for tracking particle trajectories in 1.7 T magnetic field.

EM and hadronic calorimeters for particle energy and jet measurements.

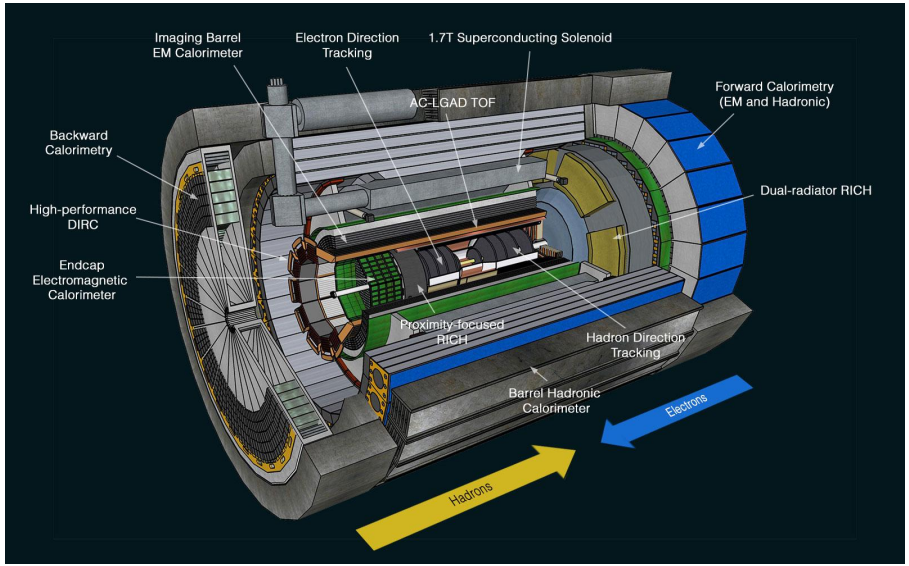
Ring Imaging Cherenkov, AC-LGAD Time of Flight, hpDIRC, etc. for particle ID.

Far forward and far backward systems allow measurement of particles close to beamline.

ePIC Central Detector



Wide kinematic coverage provided by central detector + far forward/far backward systems.



Simulation Campaigns

Some of the Datasets and Monte Carlo Generators In Production:

DIS (pythia8), SIDIS (pythia6-eic for minimum bias)

Exclusive
- DVMP (EpIC)
- DEMP (DEMPgen)
- DIFFRACTIVE_PHI (SARTRE)

Backgrounds
- Proton Beamgas (pythia8)
- Electron Beamgas (GETaLM)
- Merged = Beamgas+SynRad

List is continuously growing as per requests from Physics Working Groups.

Monthly Campaign Strategy

Tagged Train Release (s)

Currently require ~200 core-years of compute resources and ~30 TB of storage per monthly simulation campaign.

Common Interface Template

Can use the HTCondor submit template without the need to adapt it to specific sites.

Excellent Support System

Weekly meetings with OSG experts, support ticketing system, broad user base and experiences to learn from.

Leveraging the Open Science Grid (OSG)

Parallelism and Ease of Scaling

Able to run up to 10k jobs simultaneously on OSG resources.

Ease of access management

Collaborators from diverse institutional affiliations can authenticate through CILogon and access OSG resources.

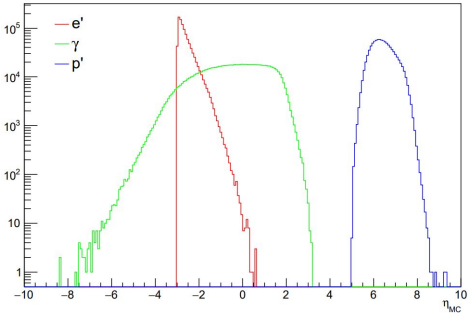
Integration of International resources

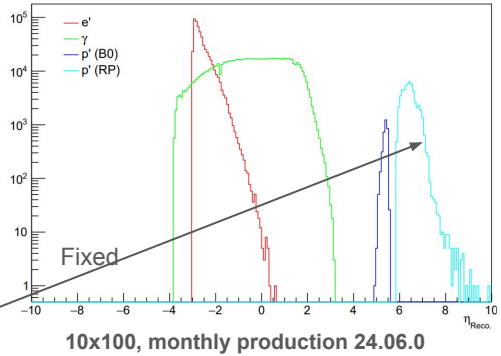
Demonstrated capability to access allocated international resources, such as, Digital Research Alliance of Canada, INFN-CNAF, etc.

Electron-Proton Deeply Virtual Compton Scattering

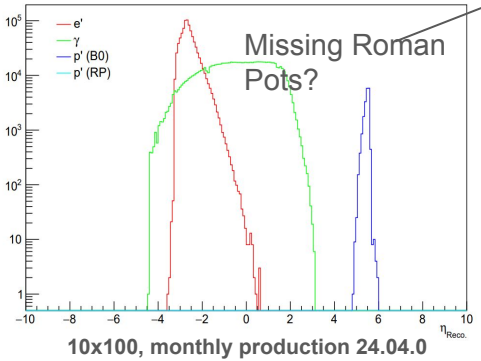
Credit: O. Jevons (Glasgow)

Relevant for nucleon tomography, origin of mass and spin.
Electron PID crucial and FF region critical for final state proton.





10x100, monthly production 24.06.0



10x100, monthly production 24.04.0

Evaluation of detector/physics performance and prompt identification of software issues enabled by regular campaign outputs.

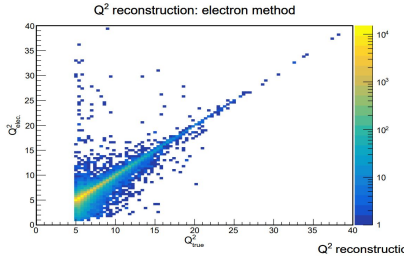
Leveraging OSG: Examples of Physics Analysis with Simulation Campaigns

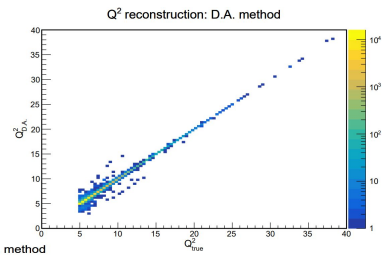
Elastic Electron-Proton Scattering

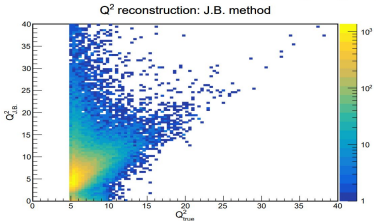
Credit: B. Schmookler (UC Riverside)

Significance: Highest Q^2 ever measured for elastic e-p scattering, and it would be the first time the elastic cross-section is measured at a high-energy collider.

Evaluation of different Q^2 reconstruction methods using monthly production 24.04.0.





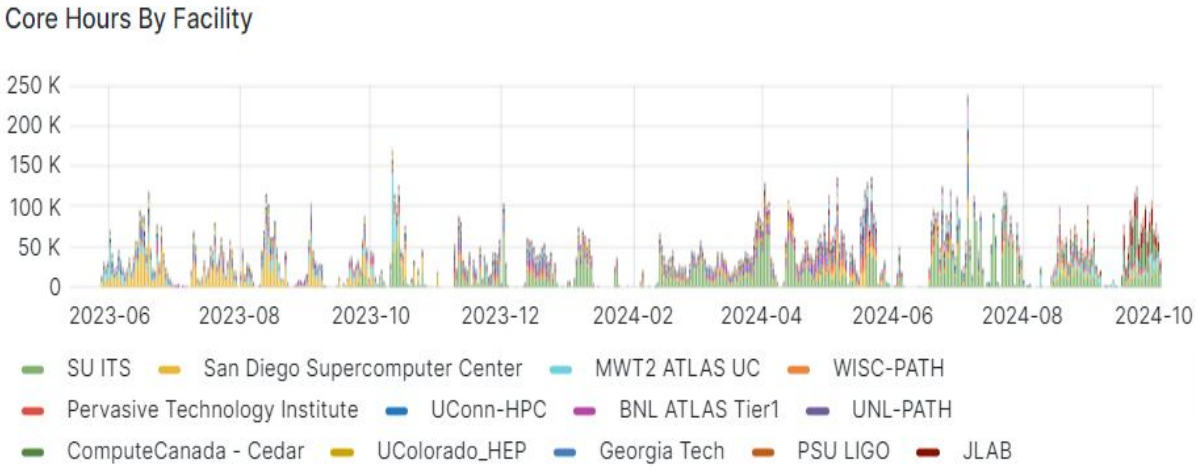


Leveraging OSG: Synergy Among Institutions Contributing Resources

3 operational access points (AP) maintained by JLab, BNL, and OSG.

Participation in regular meetings with Subatomic Physics National Team-Canada and OSG experts..

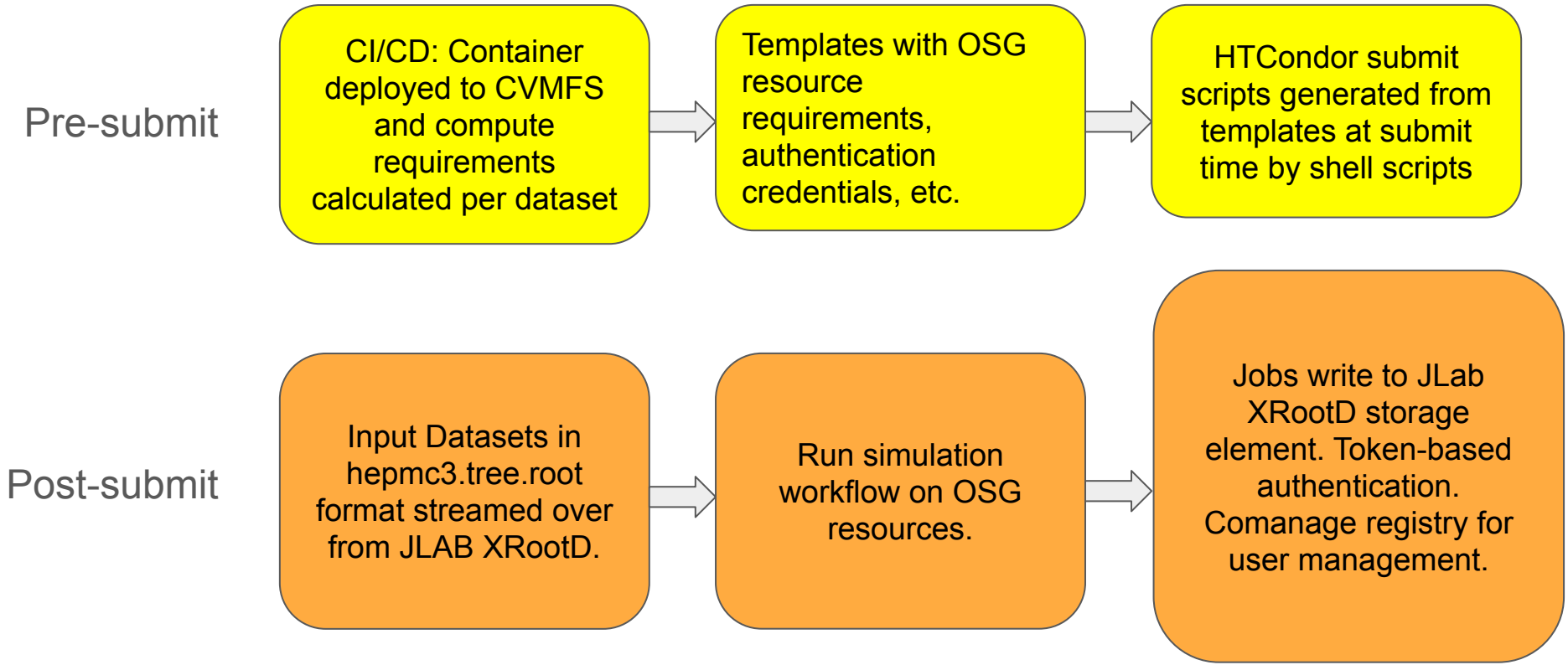
1PB XRootD storage element for ePIC campaigns + tape backup at JLab. Similar storage capacity being developed at BNL.



Production Rucio server at Jefferson Lab: temporary policy agreed and one-to-one Rucio data identifier (DID) name to storage path tested, bi-directional transfer between BNL and JLab tested, etc.

Since May 2023, ~2000 core-years consumed to produce ~500 TB campaign data

Current Production Workflow



Executed On Submit Node

https://github.com/eic/job_submission_condor

Environment variables

Shell script

Template variables

EXAMPLE

```
EBEAM=18 PBEAM=275
DETECTOR_VERSION=24.07.0
DETECTOR_CONFIG=epic_brycecanyon
JUG_XL_TAG=24.07.0-stable
./scripts/submit_csv.sh
osg_csv
hepmc3
DIS/CC/18x275/minQ2=100/DIS_CC_18x275_minQ2=100.csv
2
```

Executed On Job Nodes

https://github.com/eic/simulation_campaign_hepmc3

scripts/run.sh
integrated into cvmfs container
/cvmfs/singularity.opensciencegrid.org/eicweb/eic_xl:24.07.0

Lightweight Production Scripts and Templates

HTCondor Submit Script

Requirements

```
Requirements = HAS_SINGULARITY == TRUE &&
HAS_CVMFS_singularity_opensciencegrid_org == TRUE
&& OSG_HOST_KERNEL_VERSION >= 31000
request_cpus = 1
request_memory = 3.5 GB
request_disk = 5 GB
max_idle = 500
```

eic-shell singularity container regularly deployed to cvmfs by CI/CD pipeline. Spack is used for package management.

Project and Singularity Image

```
+ProjectName="ePIC"
+SingularityImage="/cvmfs/singularity.opensciencegrid.or
g/eicweb/eic_xl:24.07.0"
```

Job Exit Policy

```
on_exit_hold = (ExitBySignal == True) || (ExitCode != 0)
```

Dealing with Failed Jobs

If failure rate is high across sites, review ePIC software and container release.

If OSG site-specific, update blacklist and release:
+UNDESIRED_Sites = "Comma-separated list of sites"

Validation

Snakemake workflows run a range of detector and physics benchmarks for validation on the eicweb server at the Argonne National Laboratory

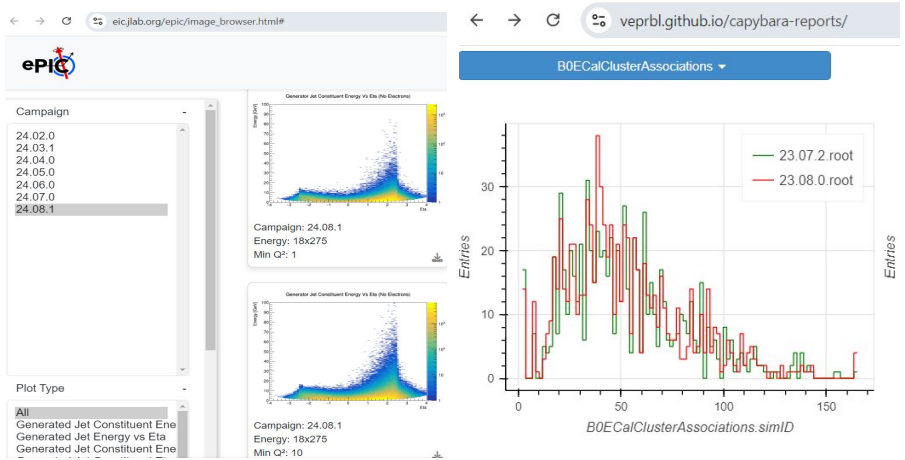
Image artifacts are transferred from eicweb to JLab and accessible via the Image Browser

Explore using compute capacity on OSG to run long running benchmarks- testing snakemake executor for HTCondor as well as feasibility of having runners on OSG ready to accept jobs from eicweb.

Important Utilities

epic-capybara: Compares and presents differences between campaigns

Image_browser: Searchable repository of image artifacts from different campaigns



How Can We Improve Our Workflow?

Complete integration of Rucio into production.

Upgrade to workload management system: PanDA or similar, under evaluation.

Integrate storage sites with Open Science Data Federation (OSDF). Abstract away issues related to token authentication and auto-renewals.

Event generation within jobs: working on version control for generators and datasets.

Summary

ePIC, the first experiment at the Electron-Ion Collider (EIC), involves 173 institutions exploring nucleon and nucleus structure with advanced EIC beam capabilities.

OSG provides standard interface, streamlined access management, parallelism and ease of scaling alongside excellent support system.

Aim to complete integration of Rucio into production, explore workload management systems, and on-the-fly event generation.

Large scale simulation campaigns for physics/detector studies and to develop distributed computing capability ahead of operations.

Production campaigns currently use tagged cvmfs-hosted container and lightweight shell scripts/HTCondor submit templates. A suite of validation tools compare results across campaigns.

Brookhaven
National Laboratory

Jefferson Lab
Exploring the Nature of Matter

Checkout other ePIC contributions at CHEP 2024

- Reconstruction Framework Advancements for Streaming Readout for the ePIC Experiment at the EIC
- Collaborative software and maintainability for ePIC experiment at EIC
- Cache Rules Everything Around Me: Building ePIC Containers With Spack
- Collaborative Tools for the ePIC Experiment