Brookhaven
National Laboratory

ePIC Simulation Production WG

Sakib Rahman (BNL) and Thomas Britton (JLab)
15 January 2025

Charge

Prepare for and run simulation campaigns based on priorities from the Technical and Analysis coordinators.

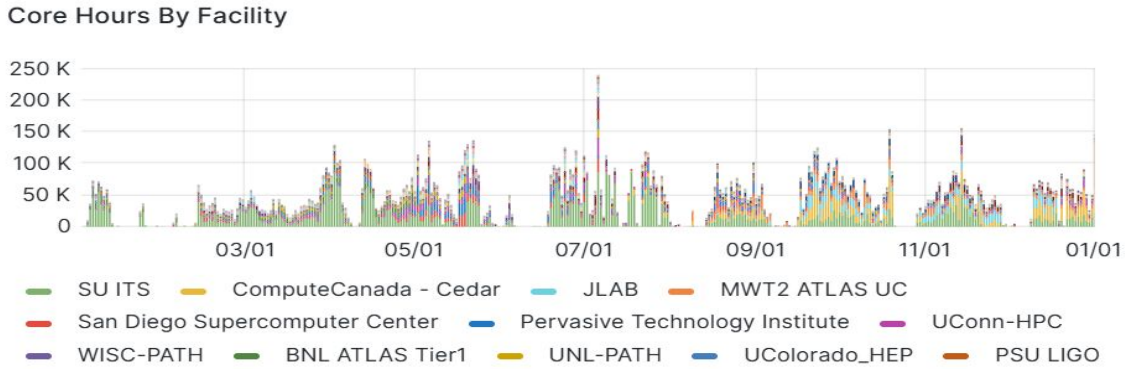
Develop automated production workflows that scale with the needs of the collaboration.

Highlights from Year 2

2024 Simulation Production On The OSG At a Glance:
CPU Usage: ~15 M Core Hours
Disk Usage: ~328 TB : RECO: 317 TB, EVGEN: 11 TB.
Centrally submitted through JLab submit host

Infrastructure improvements

- Initiated and started testing new BNL submit host (late 2024)
- Successful integration of international resources (Digital Research Alliance of Canada and INFN)
- Improved throughput from JLab
- Added 200TB of XrD storage at BNL with token-based read-write access
- Lustre disks made available at JLab to replace part of current storage
- Rucio instance stood up



Rucio in Production

- Successful test jobs running on OSG and writing back to JLAB XrD storage.
- Set up a cloud RSE for logging
- January, 2025 campaign expected to be the first campaign with Rucio

Simulation Production Links

- Live campaign updates
- Input Preprocessing Policy
- Default List of Datasets
- File access instructions
- NEW Dataset Request Form

Current Priorities

- Stress testing new Rucio-based workflow starting with January, 2025 campaign
- Start running background mixed samples
- UCONN showed interest in providing more compute resources for ePIC production. Working with OSG to work out the configuration details
- Optimize throughput from INFN
- Integrate BNL XrD as an RSE and include it in the production workflow in some capacity
- Integrate BNL submit host in production
- Improve and streamline process of requesting new datasets: new dataset request form and first production meeting of every month will be used to review what datasets would be run in that campaign
- Integrate more stakeholders in production workforce: seek out members from Physics and Detector working groups who are interested in being liaison for their respective group - attend production meetings or even taking responsibility for running certain datasets

Engagement

Actively contributing members

- Sakib Rahman
  - Monthly production workflow and updates
  - New dataset integration
  - International resource integration
  - Coordinate with BNL Scientific Data and Computing Center regarding BNL resources for ePIC
- Thomas Britton
  - JLab (OSG) Infrastructure
  - RUCIO
- Anil Panta
  - RUCIO
- Wouter Deconinck
  - Code review and software insights

~8-12 attendees in every meeting.

Required involvement from Collaboration Members

Consequences of insufficient workforce

- When things run fine little involvement is required
  - However, when problems arise and both Sakib and Thomas are away it can cause significant disruptions to the monthly campaigns
- This risk can be mitigated by training more individuals in managing the infrastructure and giving more ownership to other working groups

What more WG members could do:

- More members could help develop more automated systems for running and monitoring workflows
  - Or free up other members to do so
- More members engaged with the process could better manage the needs of other working groups by acting as a Liason.
- Those with other commitments, such as students, may not be able to fully take over mission critical tasks

Backup Slides

Charge

Prepare for and run simulation campaigns based on priorities from the Technical and Analysis coordinators.

Develop automated production workflows that scale with the needs of the collaboration.

Priorities from 2023

- Implement and document our Simulation Production Strategy, together with Validation WG. ✅
- Survey current production resources and identify potential future resources. ✅
- Inform when the Distributed Computing WG needs to start. (Around the time of adoption of PanDA or similar)

Current Priorities

Working with the PWGs to revise the **reference list of physics processes** and **related MC samples** to be included in the simulation campaigns for the TDR:

- Made good progress in meeting on Feb. 14: Preliminary list.
- Continued discussion on March 13 with additional feedback on exclusive, diffractive, and tagging processes.

Rucio

- Finish schema proposal
  - Obtain sign-off
- Migrate current data sets to the new naming scheme
- Fully integrate into the workflows of the production working group and beyond
  - Including users

International Collaboration

- Coordination with OSG PATh collaboration and Subatomic Physics National Team (SPNT) Canada.
- Testing job flow from JLab submit host to Alliance Canada resources (not part of the OSpool)
- International resource providers are required to accept OSG jobs

Improving Production Throughput and Monitoring

- Had low throughput on OSG during February and March. Issue with OSG central collector compounded by token authentication configuration.
- Aim is identify issues as soon as they happen and report the problem through OSG ticketing system if it's not within our control.
  - Working to mitigate having to go through a third party for diagnosis

Engagement

Production WG Meeting Attendance

8-12 attendees varying from week to week

- Sakib Rahman
- Thomas Britton
- Wouter Deconinck
- Markus Diefenthaler
- Dmitri Kalinkin
- Kolja Kauder
- Torre Wenaus
- Anil Panta
- John Lajoie
- .....

Apologies if any regular attendees left out :)

Actively Contributing Members

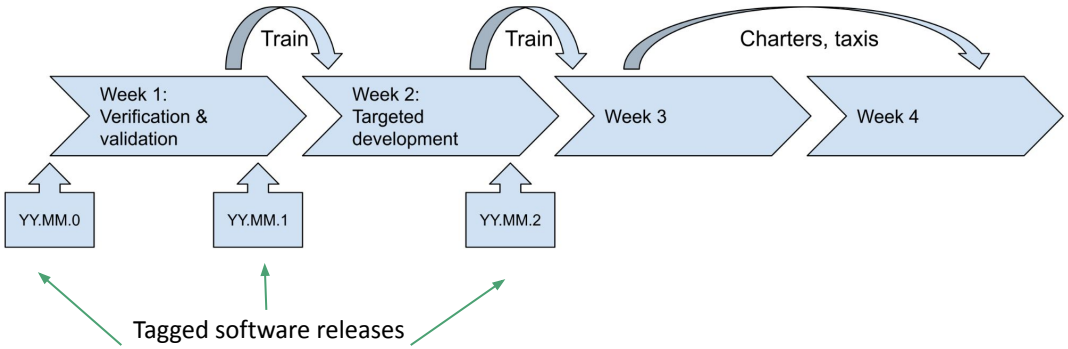
- Sakib Rahman
  - Monthly production workflow and updates
  - New dataset integration
  - International resource integration
- Thomas Britton
  - JLab (OSG) Infrastructure
  - RUCIO
- Anil Panta
  - RUCIO
- Akshaya Vijaya
  - Surveying and monitoring production outputs
- Wouter Deconinck
  - Code review and software insights

Monthly simulation productions since May 2022

Simulation Campaign Strategy

Objectives

1. Achieve **continuous deployment** of the software used for detector and physics simulations
2. Ensure **regular updates** of simulation productions for detector and physics studies, and for geometry and algorithm development
3. Implement **timely validation and quality control** for simulation productions on datasets that require substantial time and resources



Train: Major central campaign at a fixed (monthly) schedule

Charter: Special interest runs for working groups

Taxi: Bespoke runs for individual users

See “Simulation Production Strategy” document

Latest Completed Production Campaigns (24.02.0, 24.02.1, 24.03.1)

For live campaign updates, follow the firehose mattermost channel.

To see what was run in past campaigns, review our campaign history pages for reconstructed output files and full geant4 simulation output files.

Update is provided at the end of the campaign via email/web. To learn how to access files on xrootd, review our FAQ page.

To learn about our simulation and analysis framework, review the tutorials on the ePIC Collaboration Landing Page.

Example:
Listing and viewing Reconstructed Outputs for a particular dataset

xrdfs root://dtn-eic.jlab.org ls
/work/eic2/EPIC/RECO/24.03.1/epic_craterlake/DIS/NC/18x275/minQ2=10

where the different segments indicate server address, base address, detector config, physics process and beam properties respectively.

The corresponding event generator files will be at
/work/eic2/EPIC/EVGEN/DIS/NC/18x275/minQ2=10
and geant4 simulation output (for small subset) will be at
/work/eic2/EPIC/FULL/24.03.1/epic_craterlake/DIS/NC/18x275/minQ2=10

Copy files locally using xrdcp or open them in root directly

Current Campaign (24.04.0)

We have organized a group of datasets (Physics Processes and Backgrounds) that will be run every campaign (moving gradually towards version control for everything). Besides regular production campaign datasets, we will accommodate charter requests from Physics Working Groups depending on availability of resources after approval by Analysis and Software coordinators (AC/SCs).

Reminder: Criteria for MCEGs to be Included in Production

1) Must not duplicate effort. Need to have reference generator for each process.
2) Must be in hepmc3.tree.root format.
3) Must be version-tracked in a publicly accessible repository: Source code, steering files, run cards, etc. Follow the input preprocessing guidelines.

File Nomenclature and Organization

Organization of files	Example
<physics processes>/<generator repository release tag>/<electron momentum>x<proton momentum>/q2_<minimum q2>to<maximum q2>/<generator repository release tag>_<physics processes>_<electron momentum>x<proton momentum>_q2_<minimum q2>to<maximum q2>_run<index>.hepmc3.tree.root	DIS/NC/pythia6.428-1.0/10x100/q2_10to100/pythia6.428-1.0_DIS-NC_10x100_q2_10to100_run001.hepmc3.tree.root

Datasets in Production

Physics in Production (Unversioned)

Dataset	Expected Corehours	Generator
DIS NC 10x100 [minQ2=1, minQ2=10, minQ2=100, minQ2=1000] 18x275 [minQ2=1, minQ2=10, minQ2=100, minQ2=1000] 5x41 [minQ2=1, minQ2=10, minQ2=100]	 49k 98k 24k	Pythia8
DIS CC 10x100 [minQ2=100, minQ2=1000] 18x275 [minQ2=100, minQ2=1000] 5x41 [minQ2=100]	 21k 47k 7k	Pythia8
EXCLUSIVE TCS ABCONV [10x100 (hel minus), 18x275(hel minus/plus), 5x41(hel minus/plus)	 45k	?
EXCLUSIVE UCHANNEL PI0 UCHANNEL RHO	 1k 0.3k	?
EXCLUSIVE DVCS ABCONV	 14k	?

Physics in Production (Unversioned)

Dataset	Expected Corehours	Generator
SINGLES 3to50degrees(e-, e+, pi-, pi+, pi0, kaon-, kaon+, gamma, proton) 45to135degrees(e-, e+, pi-, pi+, pi0, kaon-, kaon+, gamma, proton) 130to177degrees(e-, e+, pi-, pi+, pi0, kaon-, kaon+, gamma, proton) etaScan(e-, mu-, gamma)	 29k 24k 23k 9k	

Physics in Production (Versioned)

Dataset	Expected Corehours	Generator
SIDIS 10x100 [q2_0to1] 18x275 [q2_0to1] 5x41 [q2_0to1] 10x275 [q2_0to1]	 75k 144k 46k ?	pythia6-eic 1.0.0
LAMBDA	?	LambdaGen pythia8.306-1.0
D0	?	D0Gen pythia8.306-1.0
DVMP	0.3k	DVMPdataset EpIC1.0.0-1.1
DVMP LAGER	?	LAGER
DEMP	22k	DEMPGen 1.0.0
EXCLUSIVE DIFFRACTIVE PHI ABCONV PHOTOPRODUCTION JPSI PHOTOPRODUCTION JPSI ABCONV	 51k ? ?	SARTREdataset sartre-1.39-1.0

Backgrounds In Production (Unversioned)

Dataset	Expected Corehours	Generator
MERGED (10x100)	0.06k	?

Backgrounds In Production (Versioned)

Dataset	Expected Corehours	Generator
PROTON BEAMGAS [275 GeV, 100 GeV]	33k	pythia8.306-1.0
ELECTRON BEAMGAS	?	?