`epic-analysis`

Common Analysis Framework for (SI)DIS and More

Christopher Dilks

ePIC Collaboration Meeting

January 2023

Research supported by the

Duke UNIVERSITY

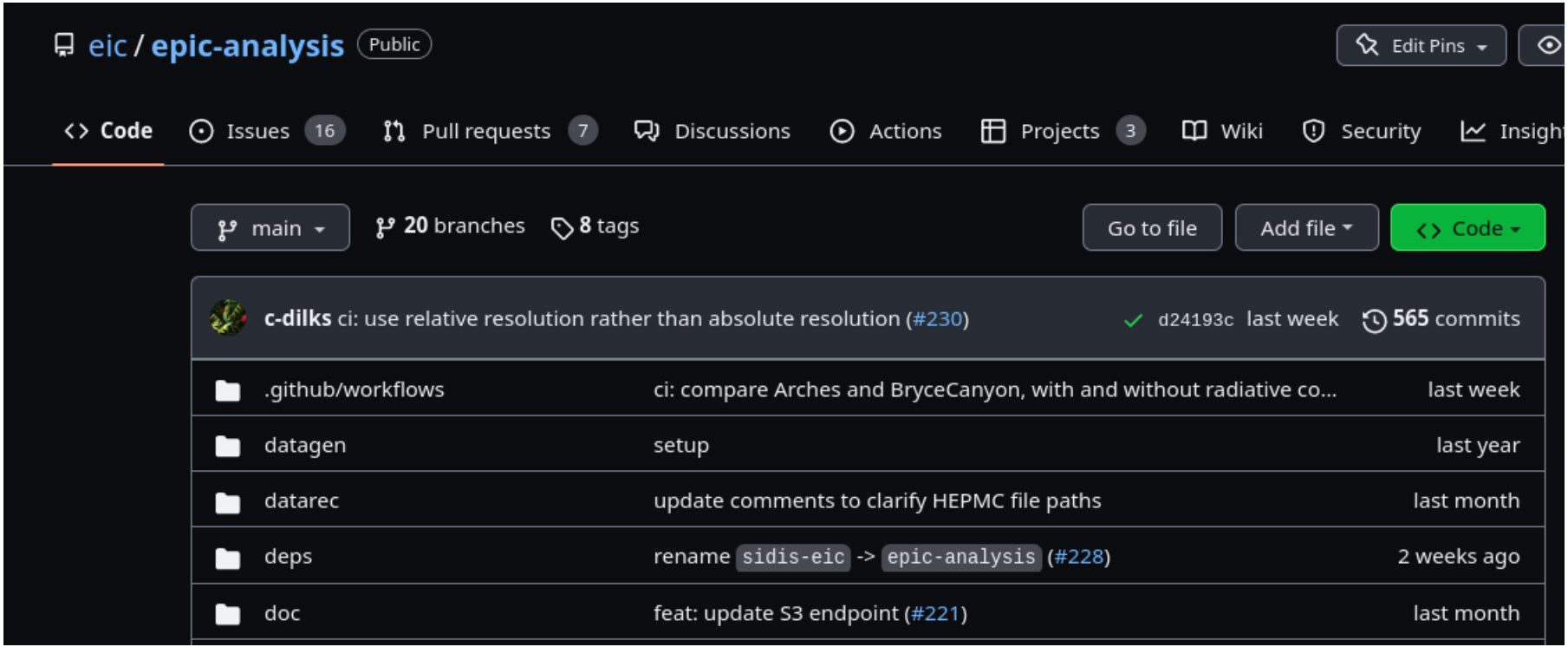
U.S. DEPARTMENT OF ENERGY | Office of Science

Outline

- **epic-analysis**
  - Support ePIC, ECCE, ATHENA, Delphes
  - Data retrieval automation
  - Kinematics Reconstruction
  - Q^2 weighting
- **Continuous Integration**
  - As a “slow” benchmark
  - Comparisons
- **Ongoing Plans**

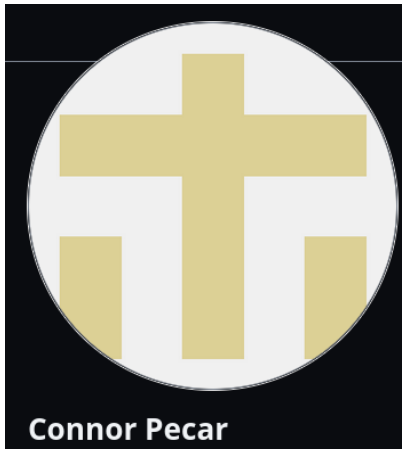
GitHub Repository

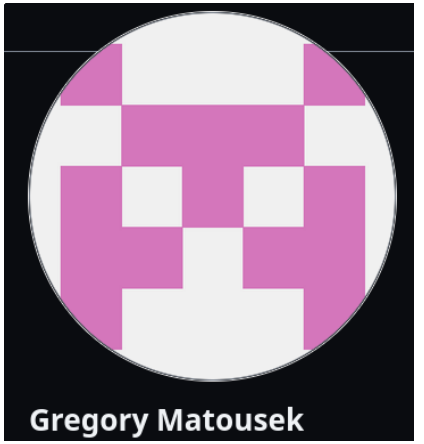
https://github.com/eic/epic-analysis

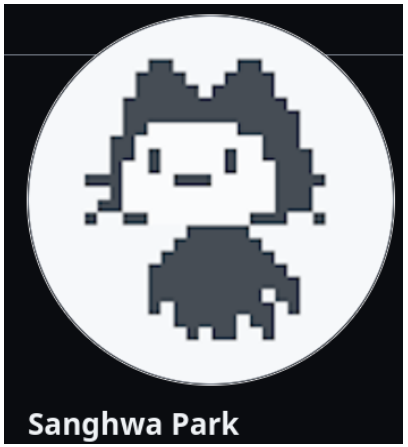


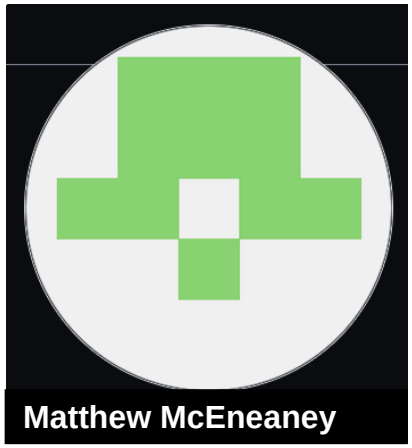
Thanks to our Contributors

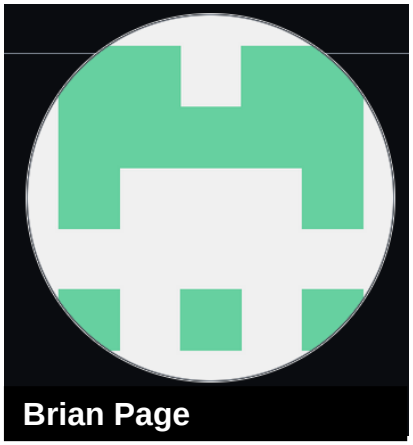
And many more for guidance and help!











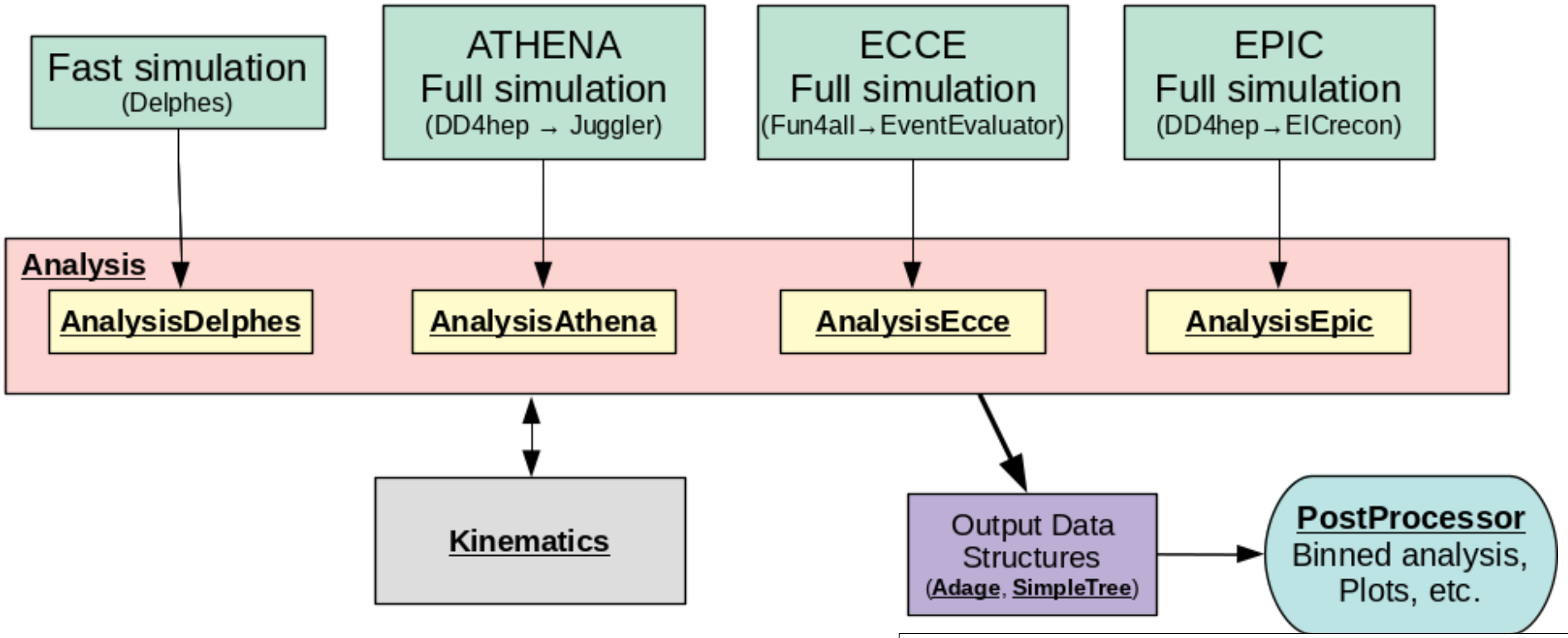
History

- **largex-eic**
  - Original focus on low-y region, large x and low Q2
  - Generalized to support SIDIS studies in ATHENA

- **sidis-eic**
  - Renamed for ePIC
  - Migrated to the eic organization on GitHub

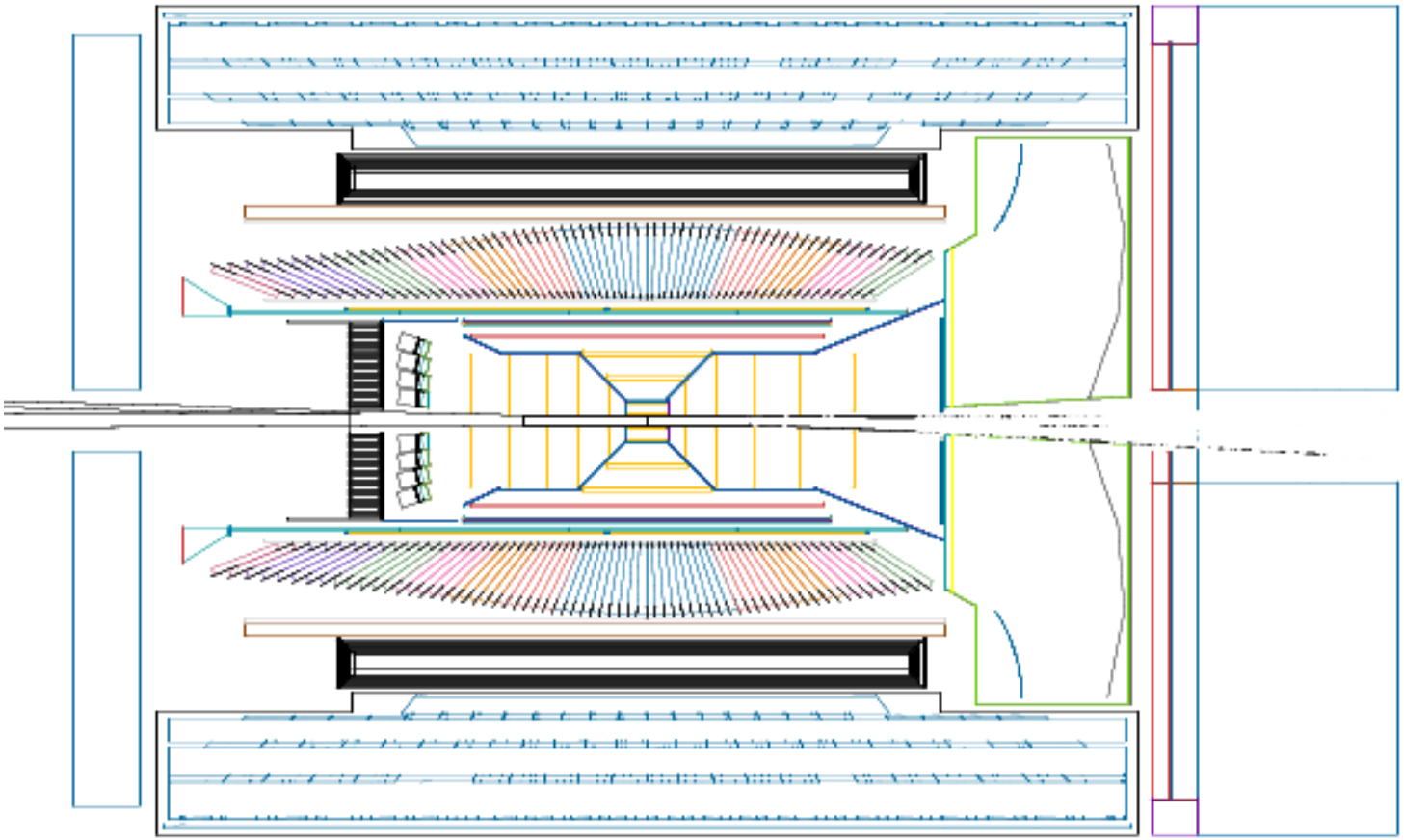
- **epic-analysis**
  - Interest from Jets, Heavy Flavor, and Inclusive working groups → more general name

epic-analysis



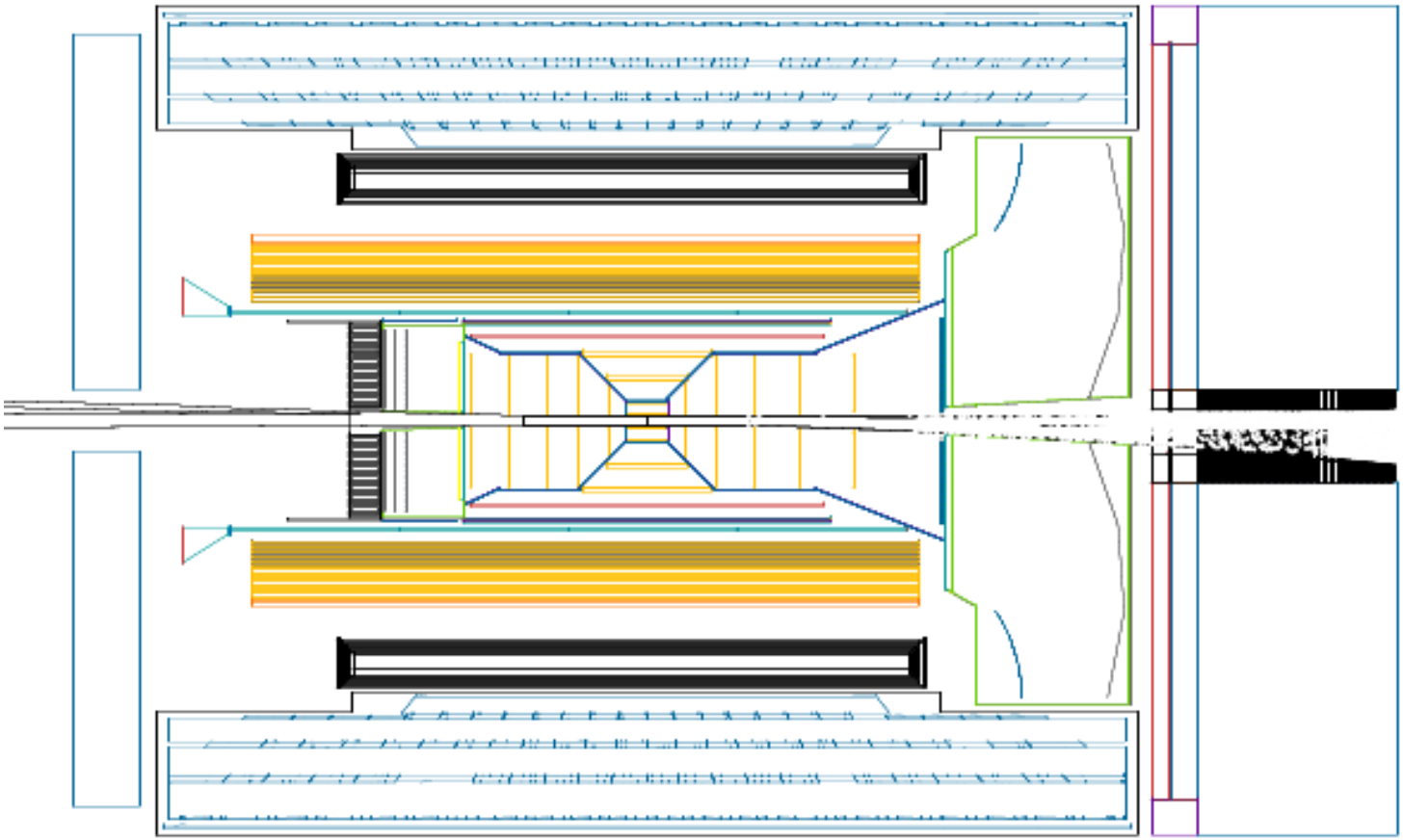
- TTrees
- Multidimensionally binned objects (Histograms, …)

ePIC Detector Configurations



Arches
SciGlass bEcal
mRICH
2nd MPGD behind DIRC

ePIC Detector Configurations



BryceCanyon
Imaging bEcal
pfRICH
Calorimeter Insert

s3tool

- Adapts to the varying file tree and naming conventions of past productions
- Downloads files from S3, or generates lists of URLs for streaming
- Organizes data by energy and Q2 bin
- Obtains cross section for Q2 weighting
- Generates “config files” for usage in epic-analysis
- Supports ePIC, ATHENA, ECCE, and can run HEPMC files through Delphes


```
            epic.22.11.3  ::  Pythia 6, with & without radiative corrections
            epic.22.11.2  ::  Pythia 8
               ecce.22.1  ::  Last ECCE Production, August 2022
athena.deathvalley-v1.0   ::  ATHENA Proposal production
          hepmc.pythia6   ::  HEPMC files from Pythia 6 for EPIC, with & without radiative corrections
          hepmc.pythia8   ::  HEPMC files from Pythia 8, used for ATHENA proposal
```

```
S3/eictest/EPIC/RECO/22.11.2/epic_arches/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
```

```
S3/eictest/EPIC/RECO/22.11.3/epic_arches/SIDIS/
├─ Lambda_ABCONV
└─ pythia6
   ├─ ep_18x275
   │  └─ hepmc_ip6
   │     └─ radcor
   └─ ep_5x41
      └─ hepmc_ip6
         ├─ noradcor
         └─ radcor
```

Ask us to support other productions, or open a PR

Kinematics Reconstruction Methods

- **Kinematics calculations performed in dedicated class(es)**
  - Used for both reconstructed and MC generated particles
  - Inputs: beams, scattered electron, hadronic final state, and observed particles (single hadrons for SIDIS, jets, etc.)
- **Calculations**
  - Inclusive variables (x, Q2, W, y, …)
    - 6 methods: electron, J.B., double angle, mixed, sigma, eSigma
  - SIDIS variables (p, p_T, z, ϕ_h, …)
  - Jet variables (z, p_T, $j_\perp$, …)
  - In general uses Lorentz invariant calculations; boost to specific frames when needed
- **Future Plan**
  - Cross check with upstream calculations from the reconstruction framework and/or upstream our methods

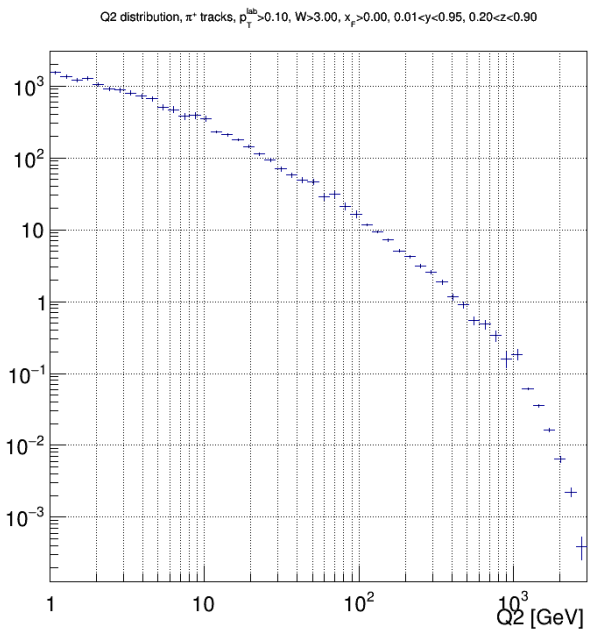
Kinematics Reconstruction with Machine Learning

- Reconstruction with pre-trained tensorflow model available (soon)
- Utilizing full hadronic final state and electron information with Particle Flow networks (http://energyflow.network, arXiv:1810.05165) to reconstruct virtual photon four momentum
  - Our application to ATHENA full simulation presented at DIS2022 (arXiv:2209.14489)
- Currently implemented with event-by-event predictions done in python script, bound using pybind11
- Current plan is to store trained model for each COM energy, specified simulation versions in epic-analysis repository
  - Requires tensorflow and energyflow python packages
  - Then callable with “ML” as reconstruction method in epic-analysis
- Validating results on current ePIC simulation, but release soon!

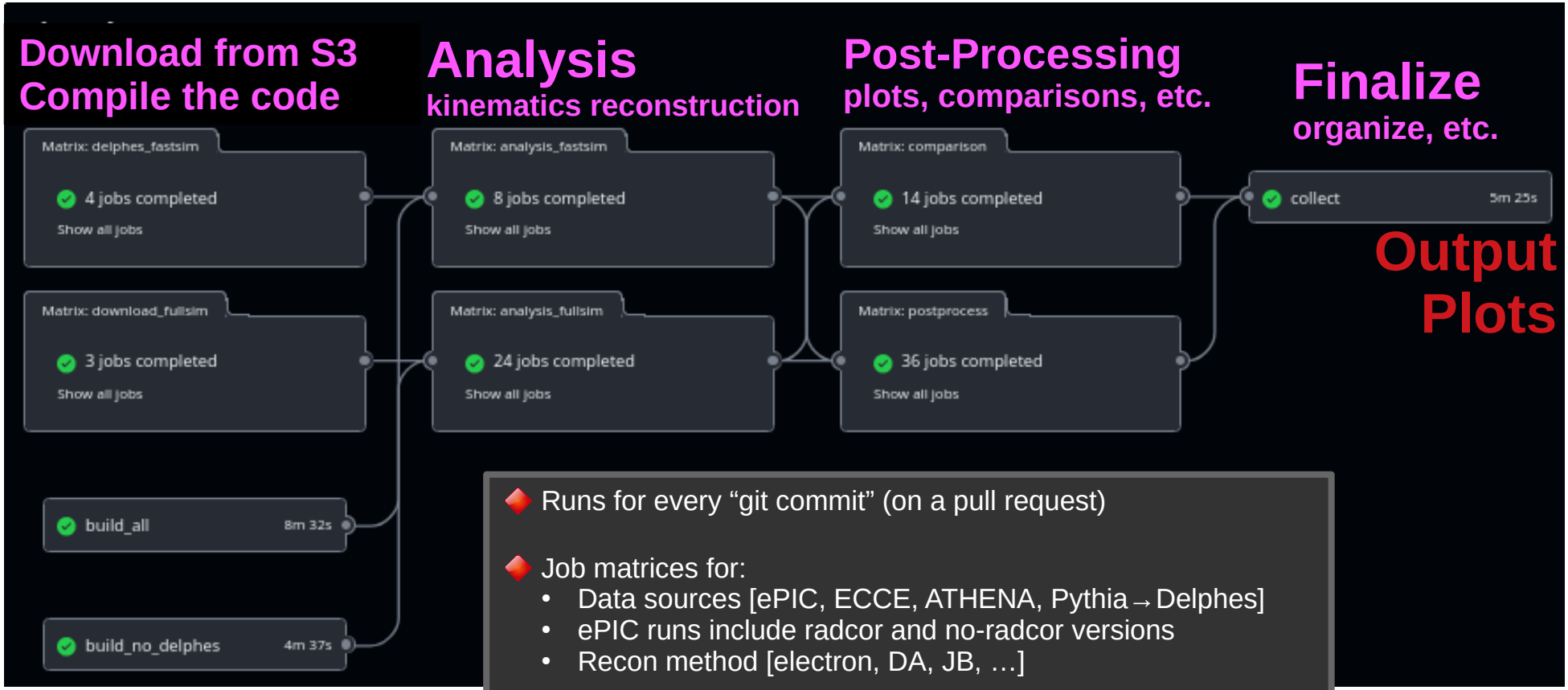
Slide from Connor Pecar

Q^2 Weighting

- **Data are produced in varying Q^2 ranges:**
  - 1 – 10 GeV^2
  - 10 – 100 GeV^2
  - 100 – 1000 GeV^2
  - 1000 GeV^2 and above
- **Use weights to combine them**
  - Uses Cross sections from Pythia and Number of Events analyzed
  - Automated by epic-analysis
    - Maintain a table of cross sections for varying beam energies
  - Allows for evaluation in a *broad* Q^2 range without waiting for rare high-Q^2 events



Continuous Integration (CI)

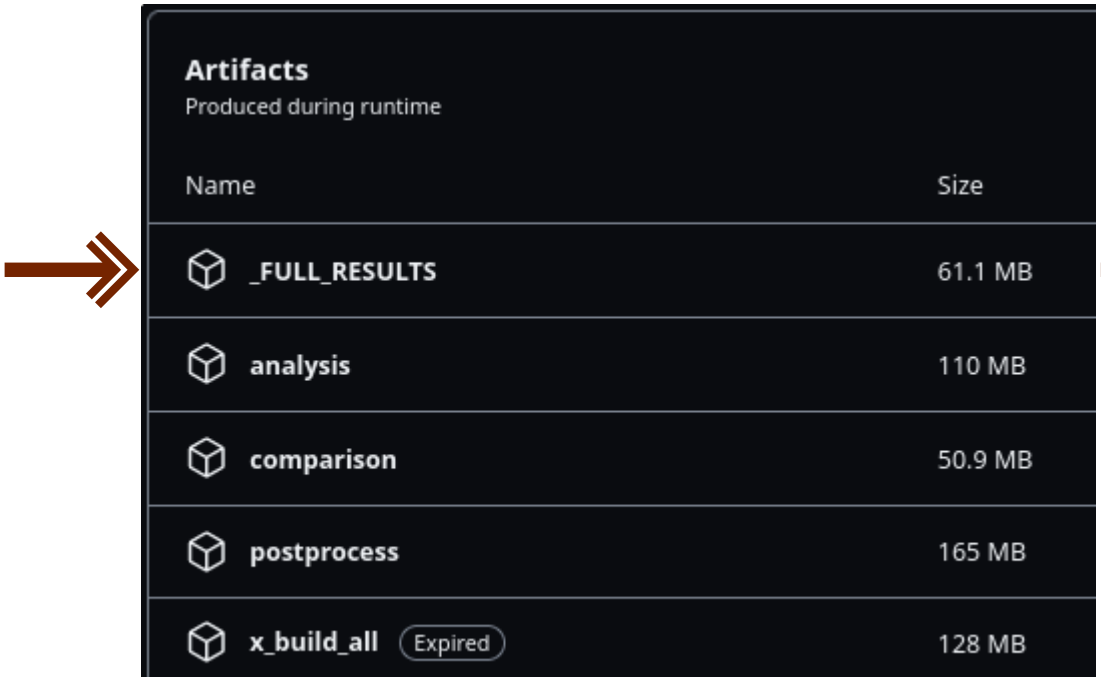


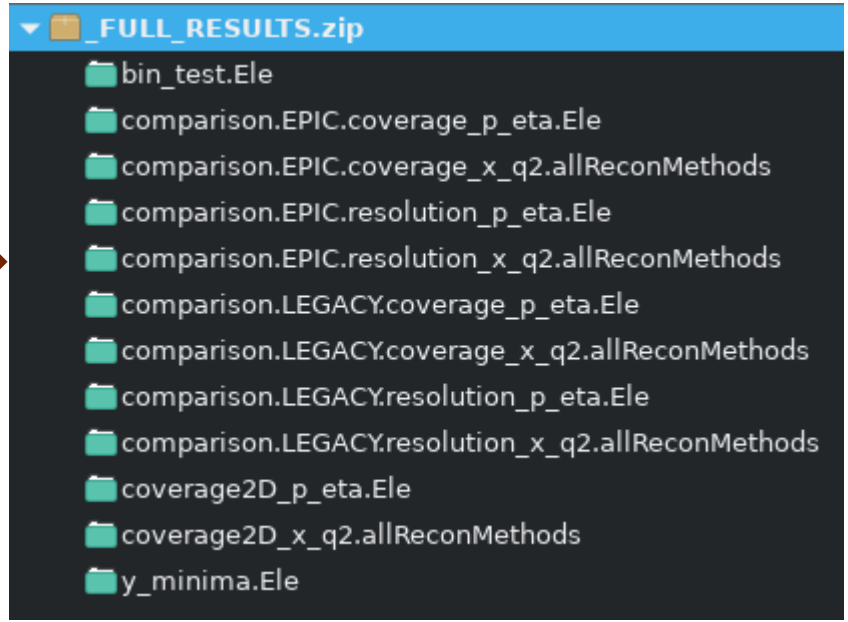
- Runs for every “git commit” (on a pull request)
- Job matrices for:
  - Data sources [ePIC, ECCE, ATHENA, Pythia→Delphes]
  - ePIC runs include radcor and no-radcor versions
  - Recon method [electron, DA, JB, …]
- Typically takes ~50 min

Continuous Integration (CI)

- Artifacts = CI output plots (and more)
- Final set of plots in '_FULL_RESULTS'
- Must be logged into GitHub to access

Example artifacts (high statistics):
https://github.com/eic/epic-analysis/actions/runs/3853010345





Productions used for CI

- **ePIC 22.11.3**
  - S3/eictest/EPIC/RECO/22.11.3/epic_{arches,brycecanyon}/SIDIS/pythia6
  - With and without radiative corrections
- **ECCE 22.1**
  - S3/eictest/EPIC/Campaigns/22.1/SIDIS/pythia6
- **ATHENA DeathValley 1.0**
  - S3/eictest/ATHENA/RECO/deathvalley-v1.0/DIS/NC
- **Delphes – Fast Simulations**
  - HEPMC files from S3/eictest/EPIC/EVGEN/SIDIS/pythia6
  - Run through Delphes (CI job)
  - Using the sample with radiative corrections
  - **Uses legacy ATHENA settings – needs to be updated!!**

Event Selection for CI

$$W > 3 \text{ GeV}$$
$$0.01 < y < 0.95$$
$$0.2 < z < 0.9$$
$$x_F > 0$$
$$p_T(\text{lab}) > 0.1 \text{ GeV}$$

Focusing on 18x275 (for now)

- Testing all available reconstruction methods
- Histograms in bins of
  - (x, Q^2)
  - (η, p)

In the interest of time, for these slides:

- Distributions and resolutions of: x, p_T, z, ϕ_h
- Also focusing on the Electron reconstruction method

Comparison Plots from CI

ePIC vs. ePIC

	Arches	Bryce Canyon
With radiative corrections	•	•
Without radiative corrections	•	•

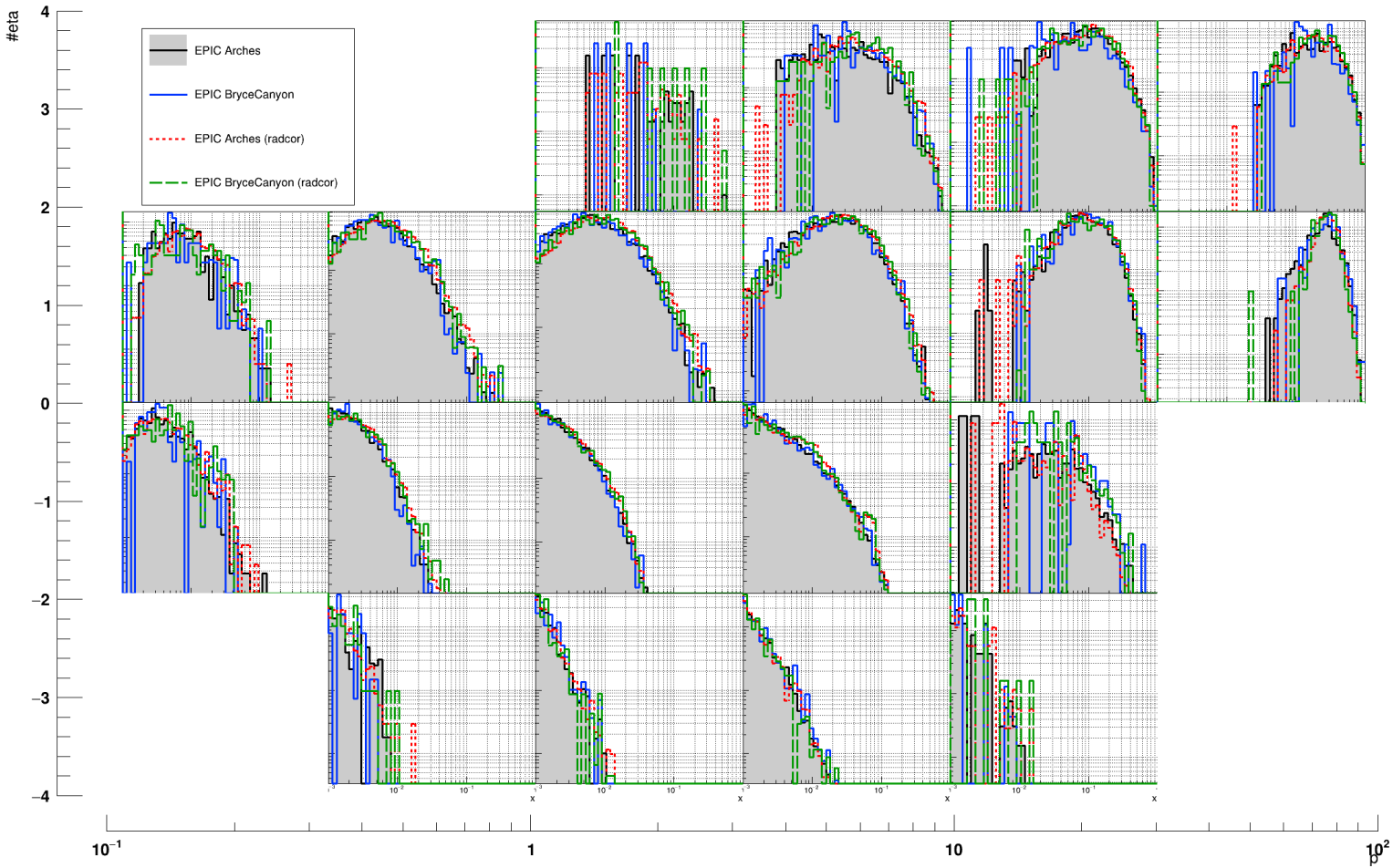
ePIC vs. Legacy

	ePIC Arches, with radiative corrections
Delphes (ATHENA card)	•
ATHENA	•
ECCE	•

CI allows us to make regular comparisons of ePIC configurations, effects from radiative corrections, and ePIC with legacy designs (ATHENA and ECCE)

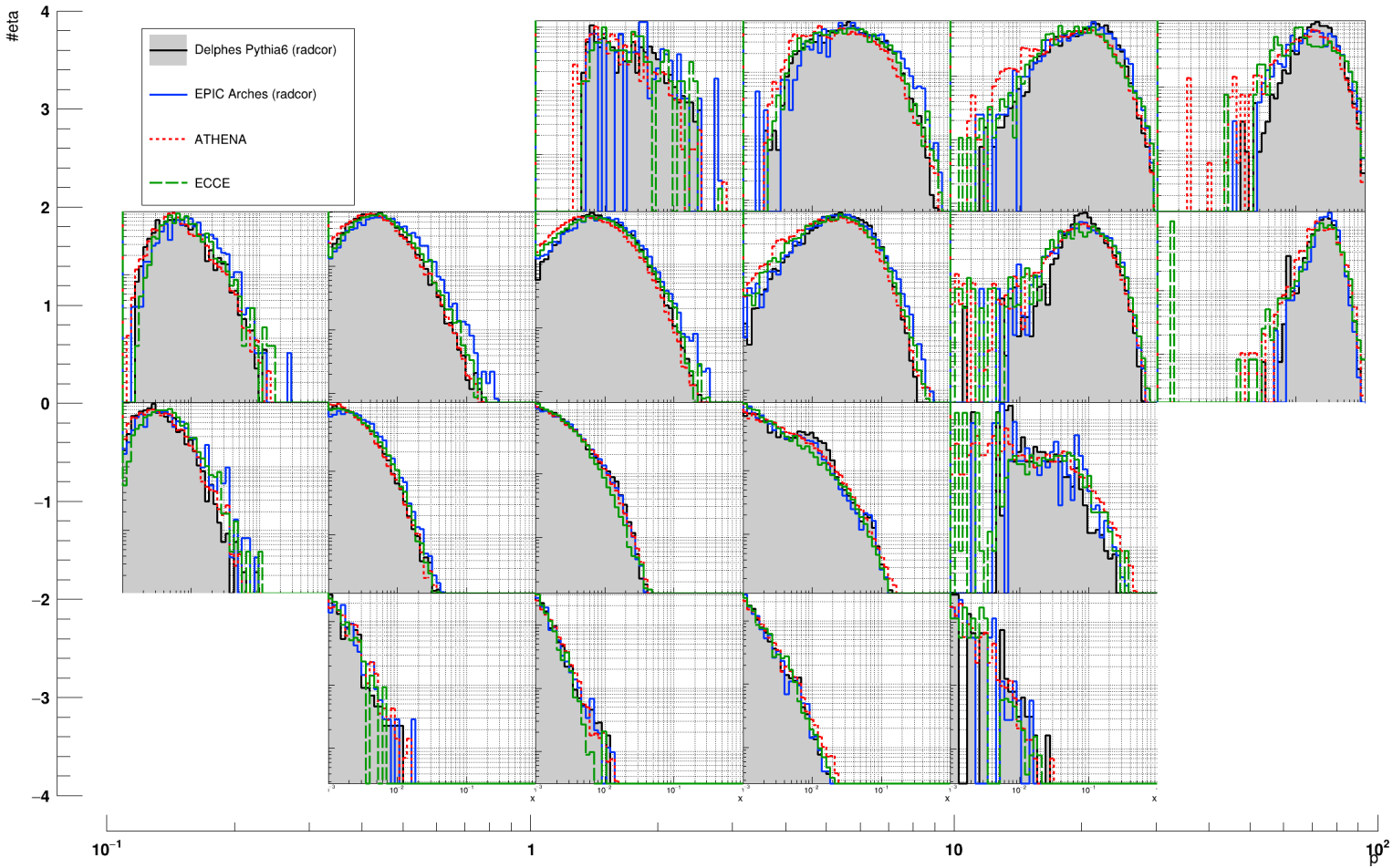
CI Comparisons – ePIC vs. ePIC

x distributions



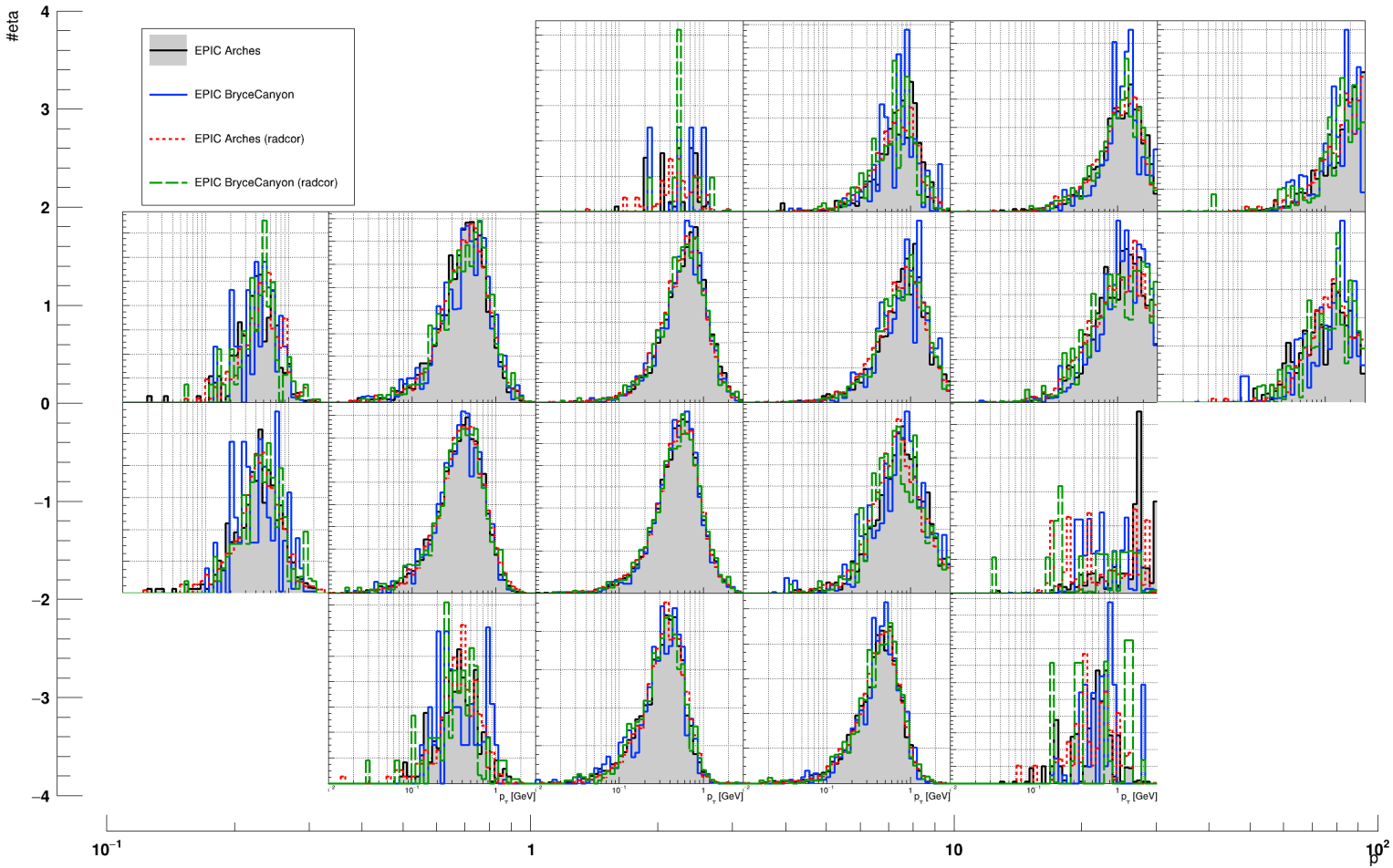
CI Comparisons – ePIC vs. Legacy

x distributions



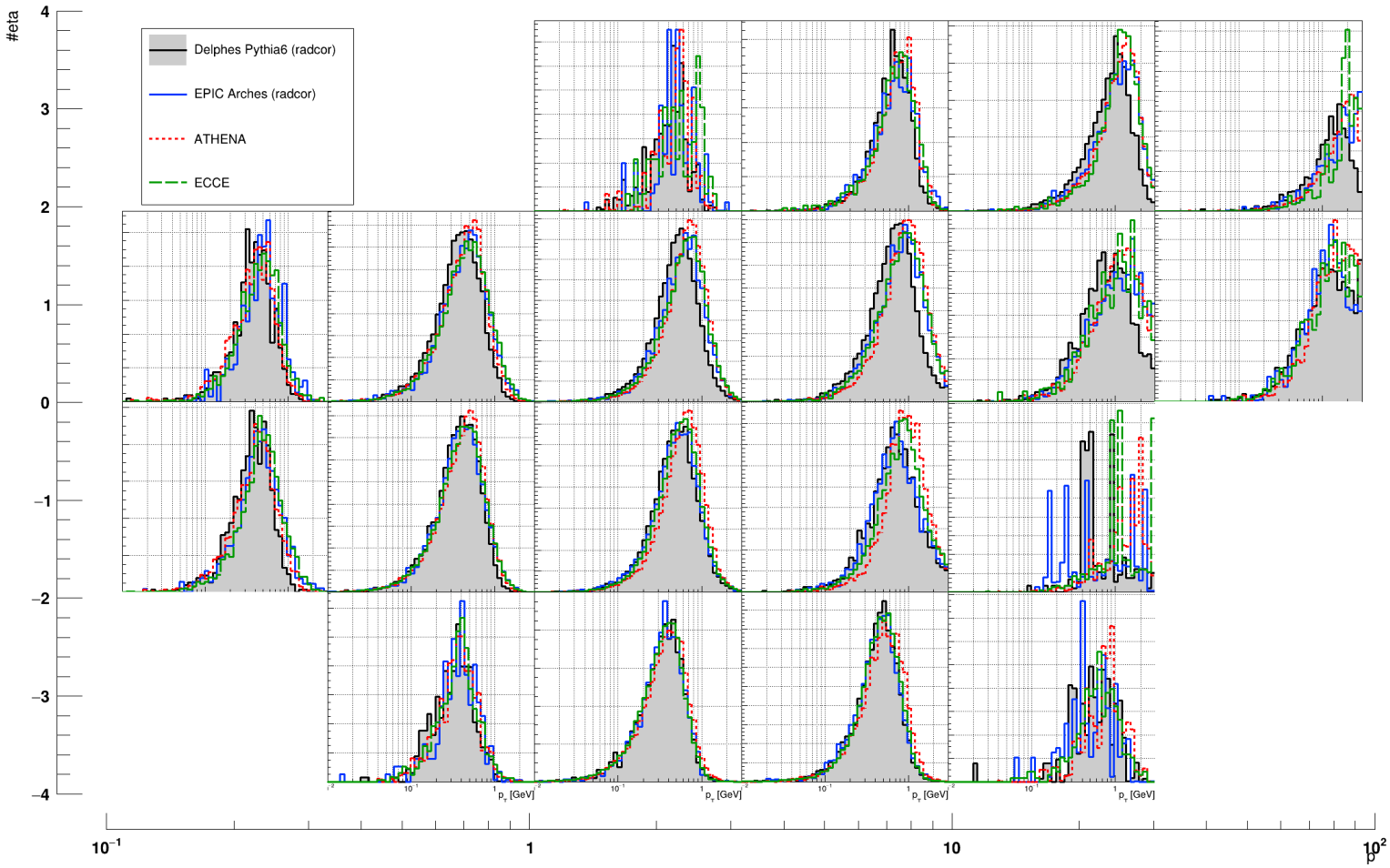
CI Comparisons – ePIC vs. ePIC

p_T distributions



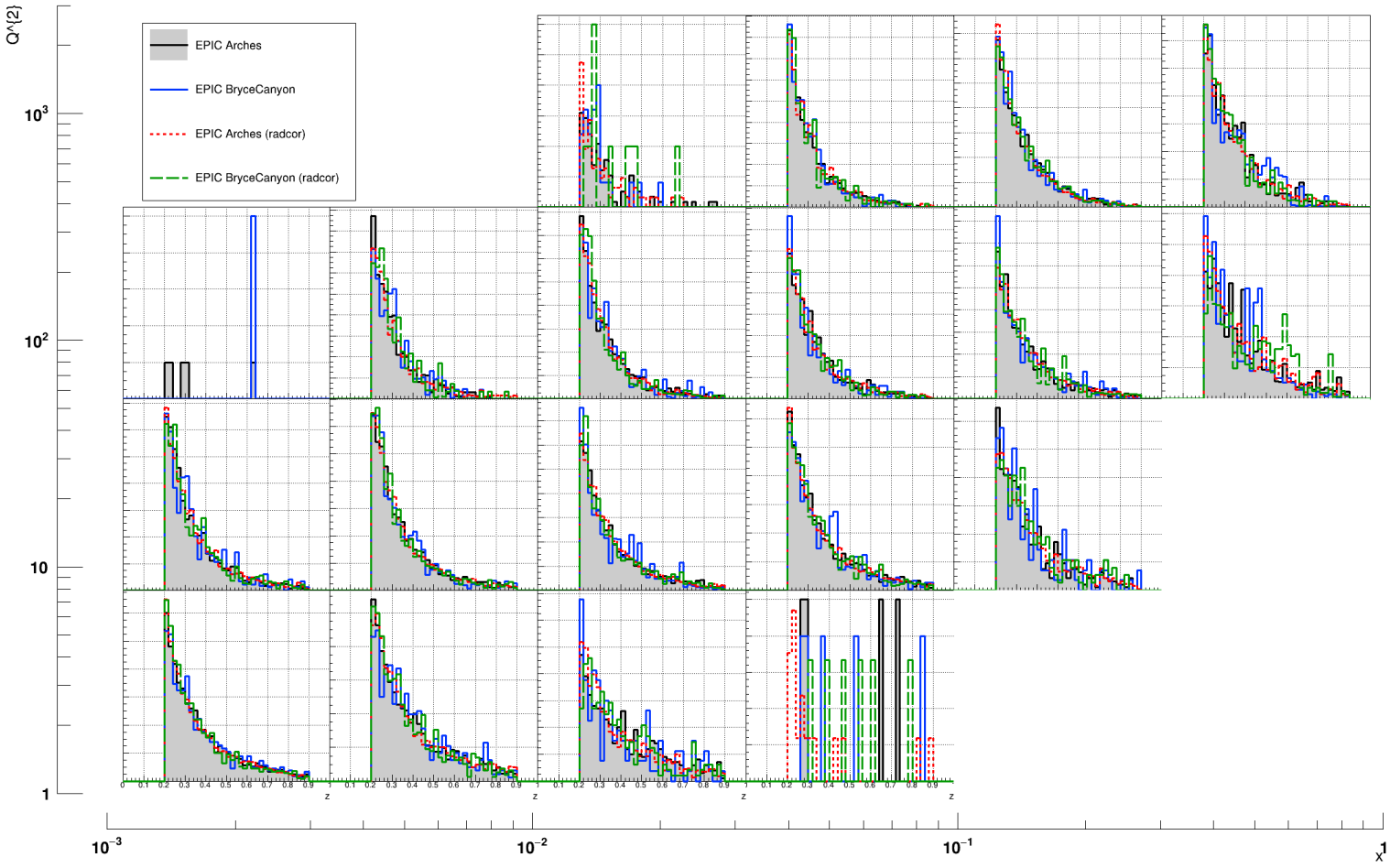
CI Comparisons – ePIC vs. Legacy

p_T distributions



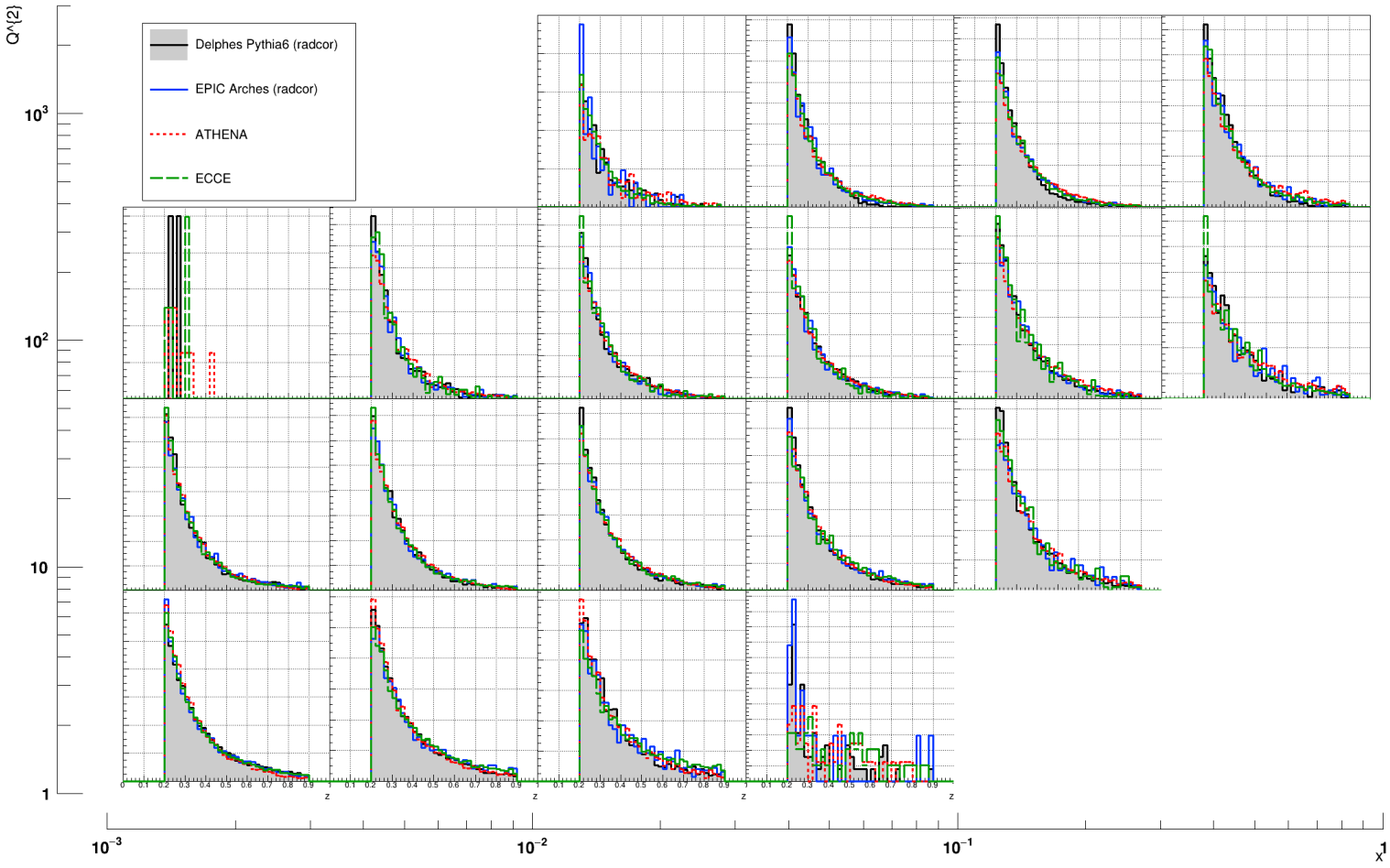
CI Comparisons – ePIC vs. ePIC

z distributions



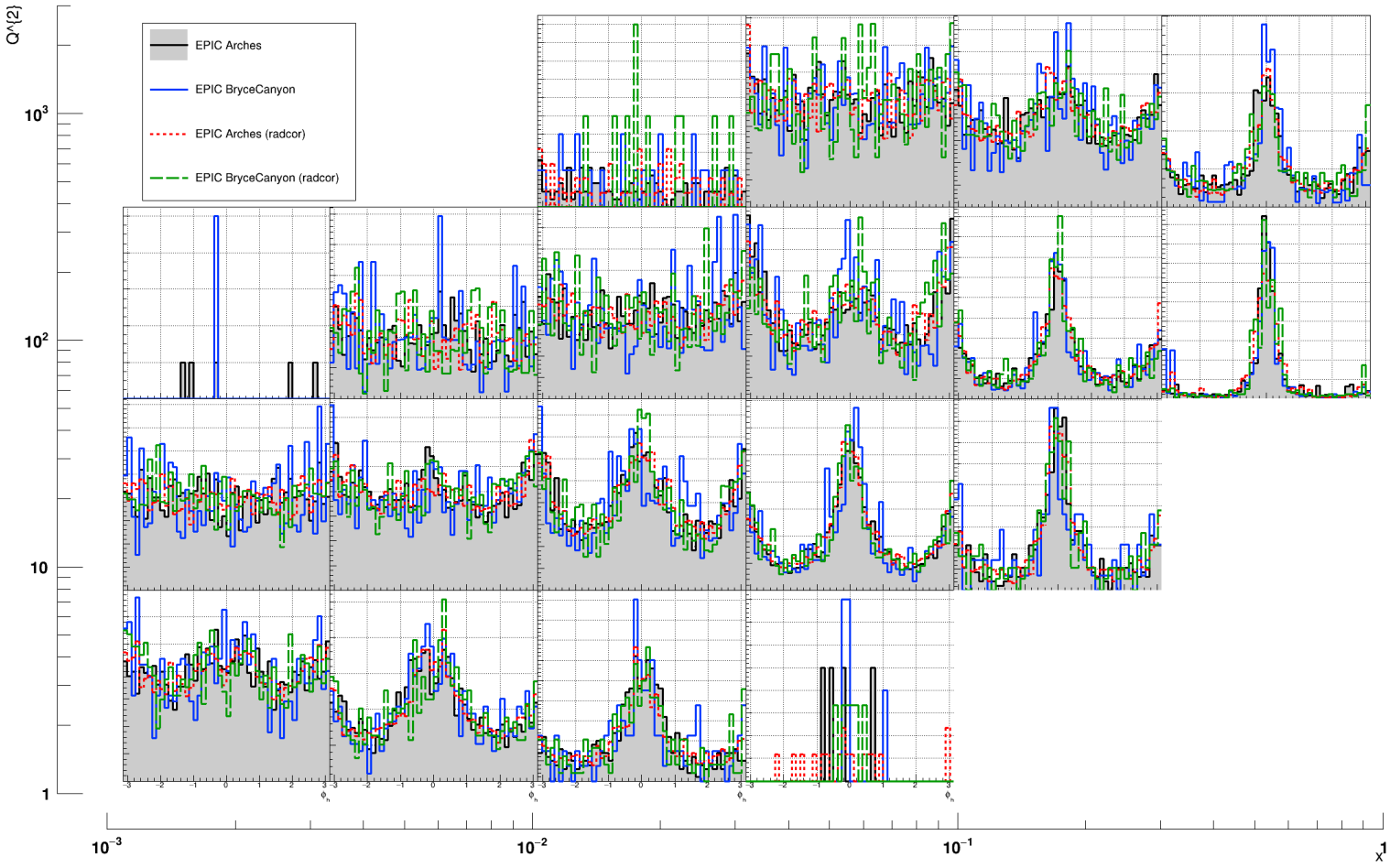
CI Comparisons – ePIC vs. Legacy

z distributions



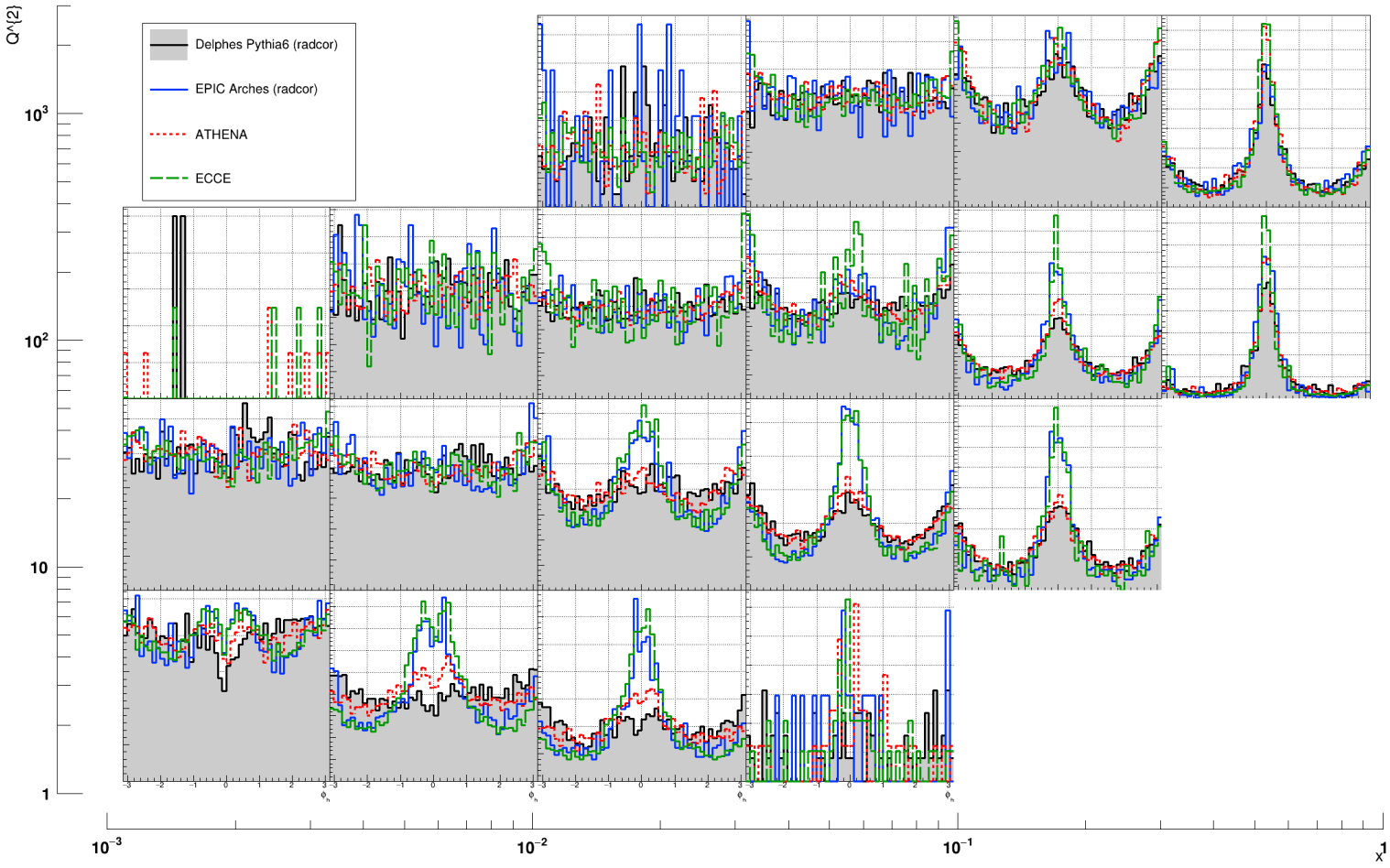
CI Comparisons – ePIC vs. ePIC

ϕ_h distributions



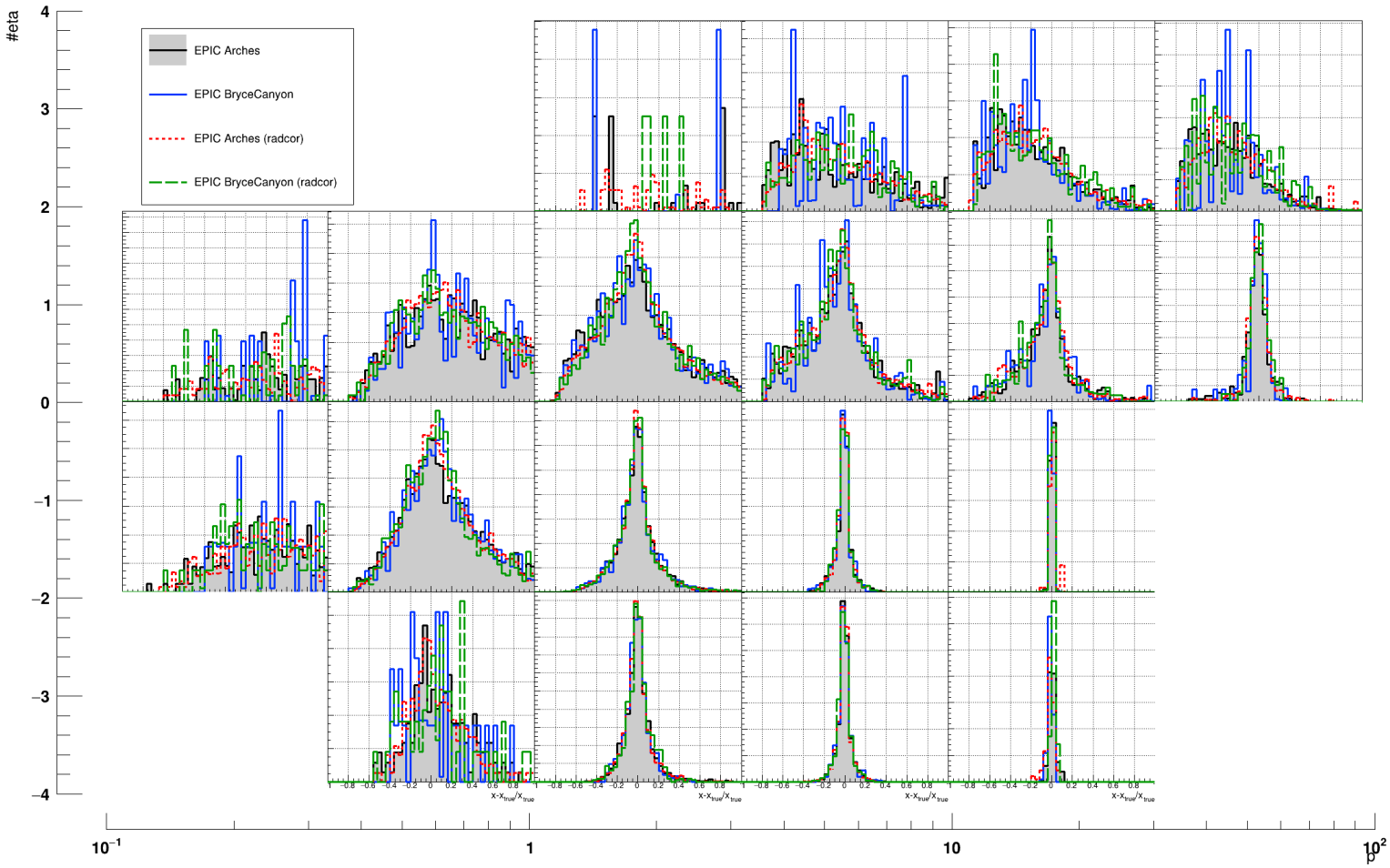
CI Comparisons – ePIC vs. Legacy

ϕ_h distributions



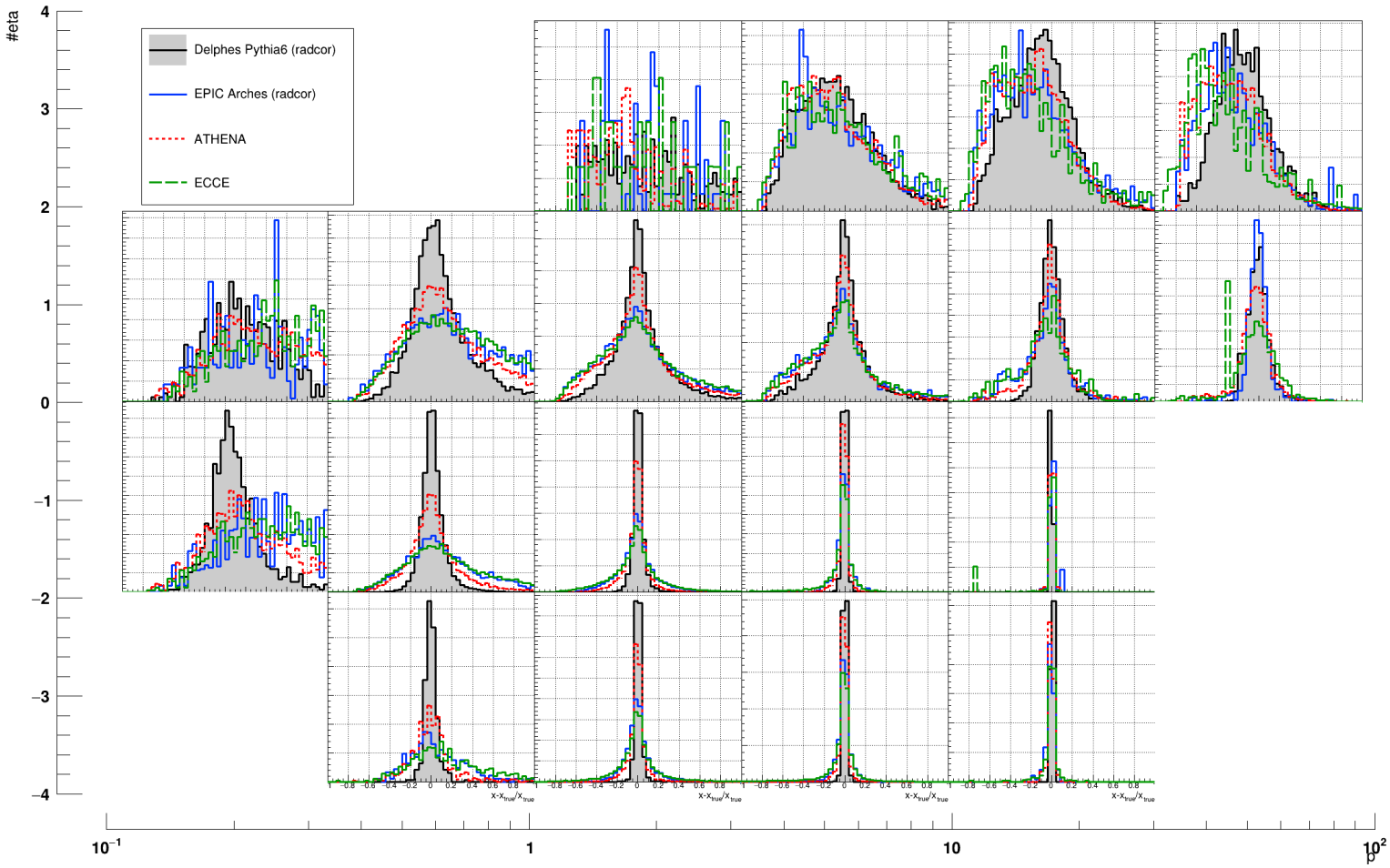
CI Comparisons – ePIC vs. ePIC

x resolutions



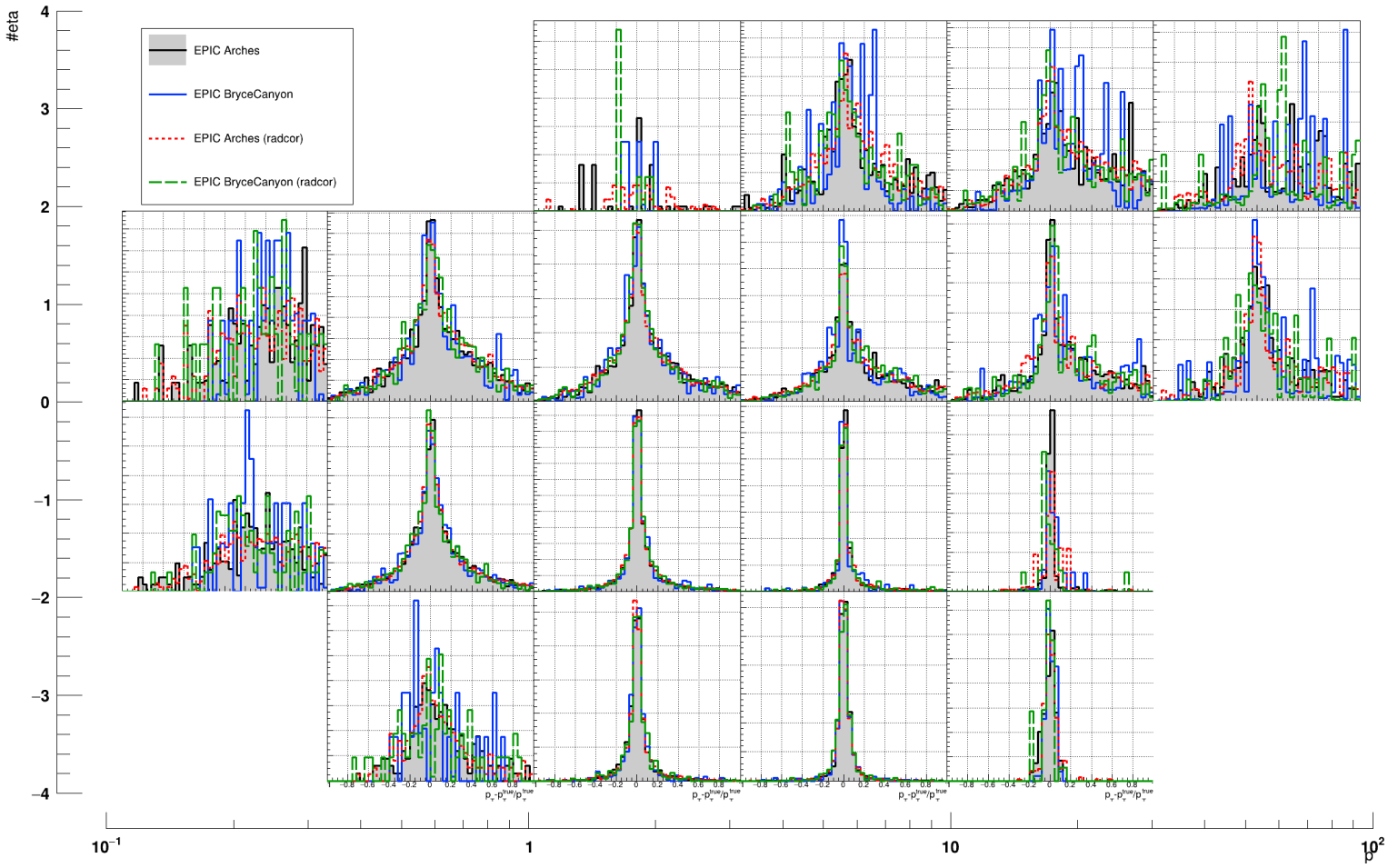
CI Comparisons – ePIC vs. Legacy

x resolutions



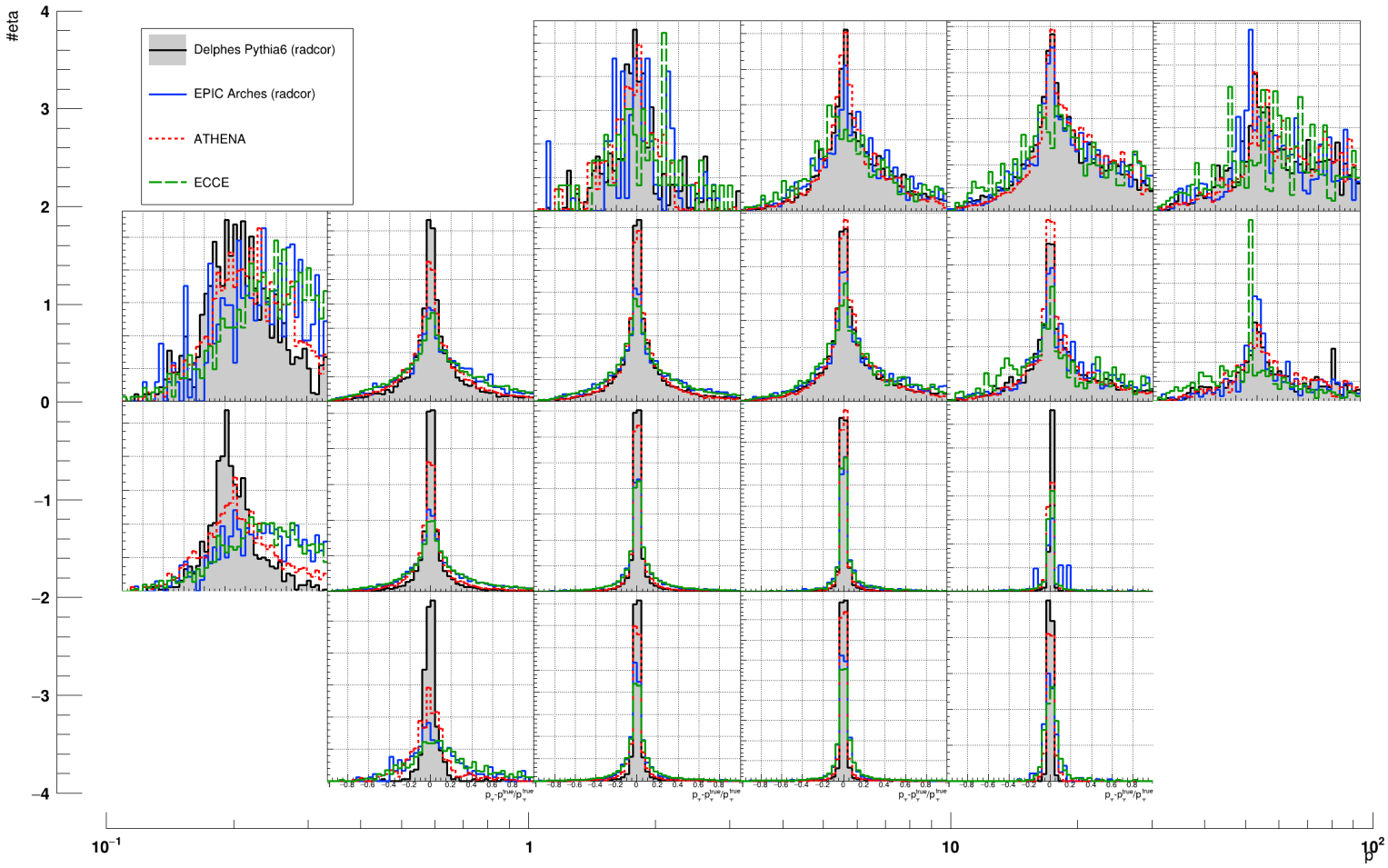
CI Comparisons – ePIC vs. ePIC

p_T resolutions



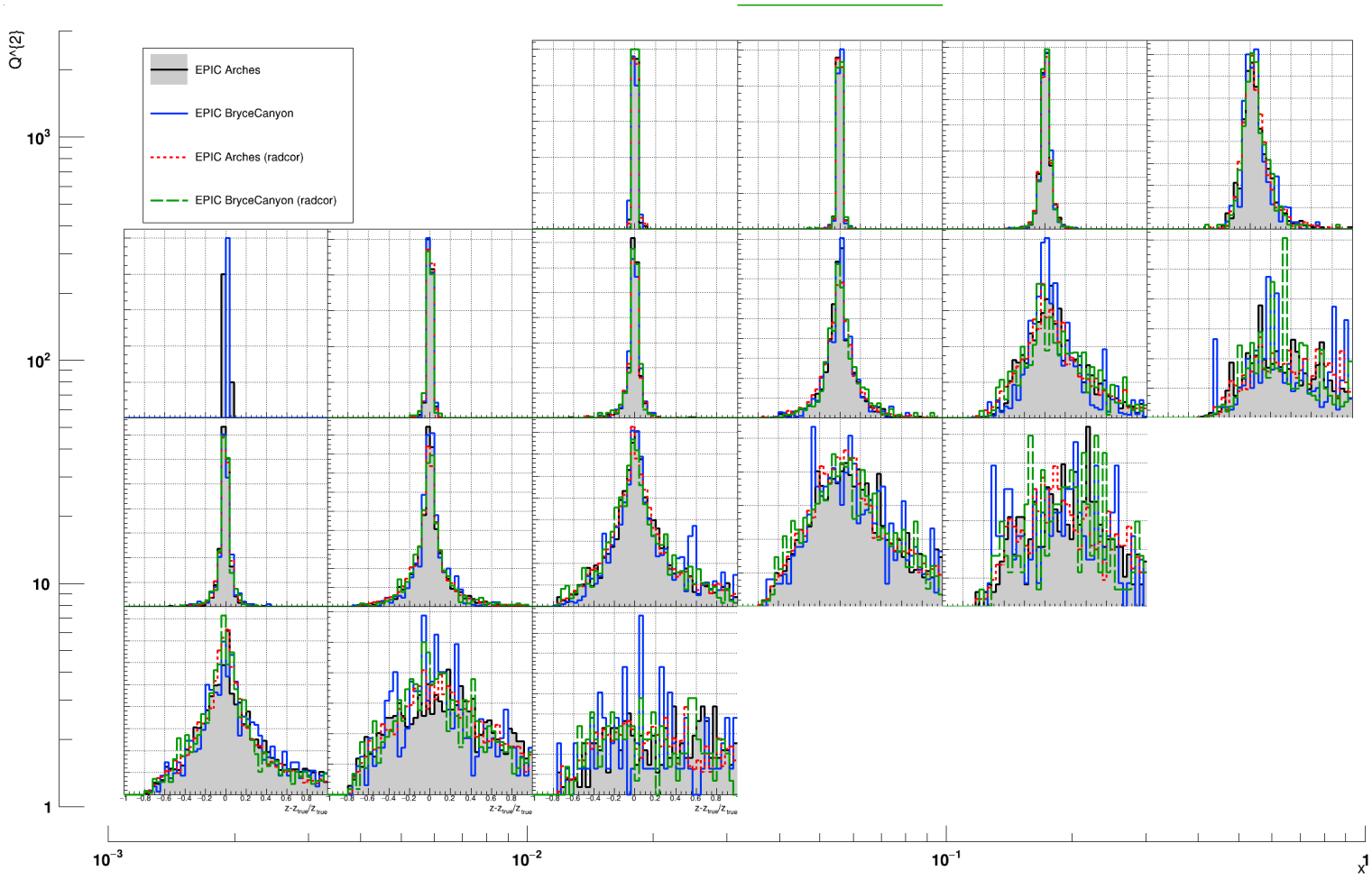
CI Comparisons – ePIC vs. Legacy

p_T resolutions



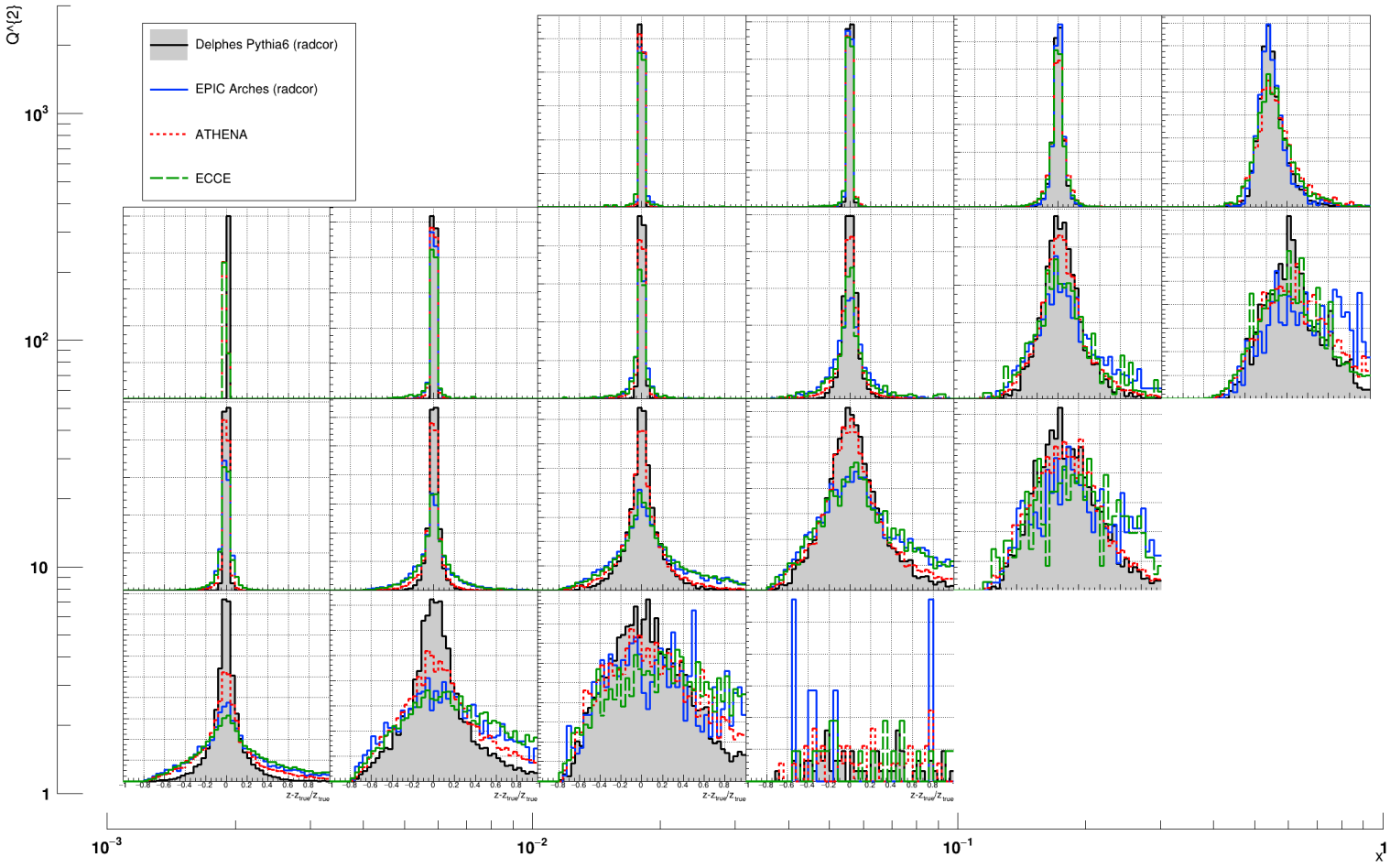
CI Comparisons – ePIC vs. ePIC

z resolutions



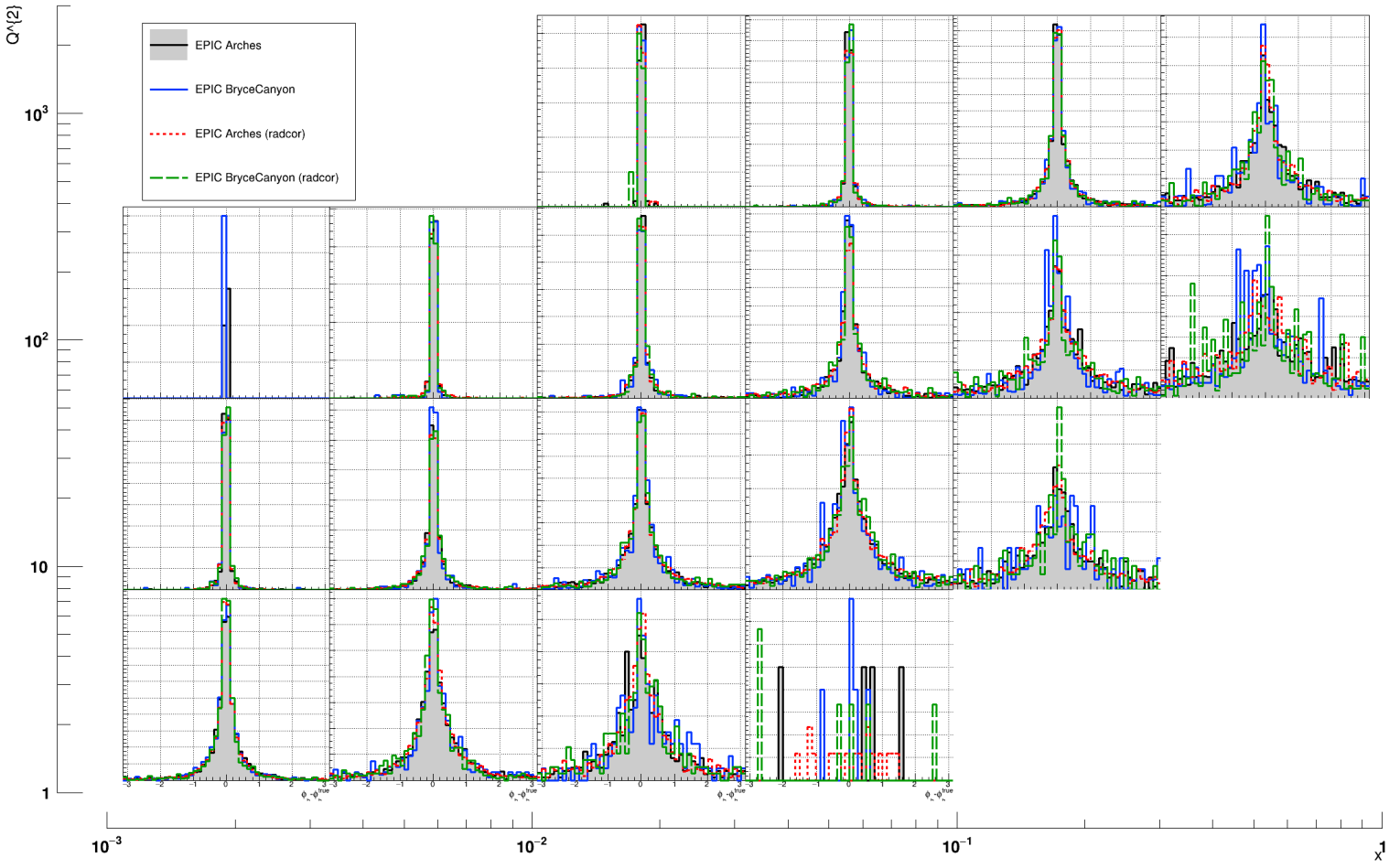
CI Comparisons – ePIC vs. Legacy

z resolutions



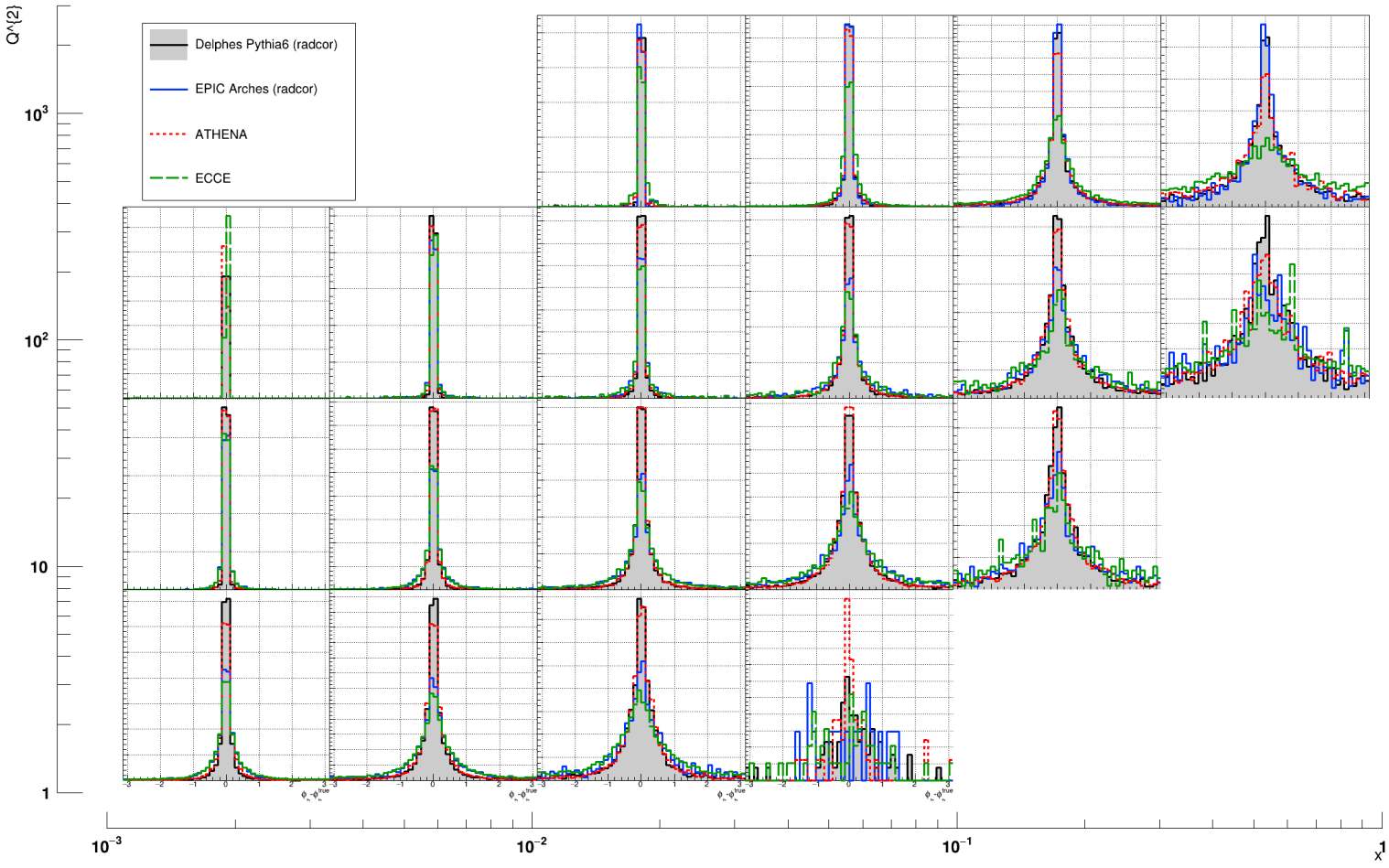
CI Comparisons – ePIC vs. ePIC

ϕ_h resolutions



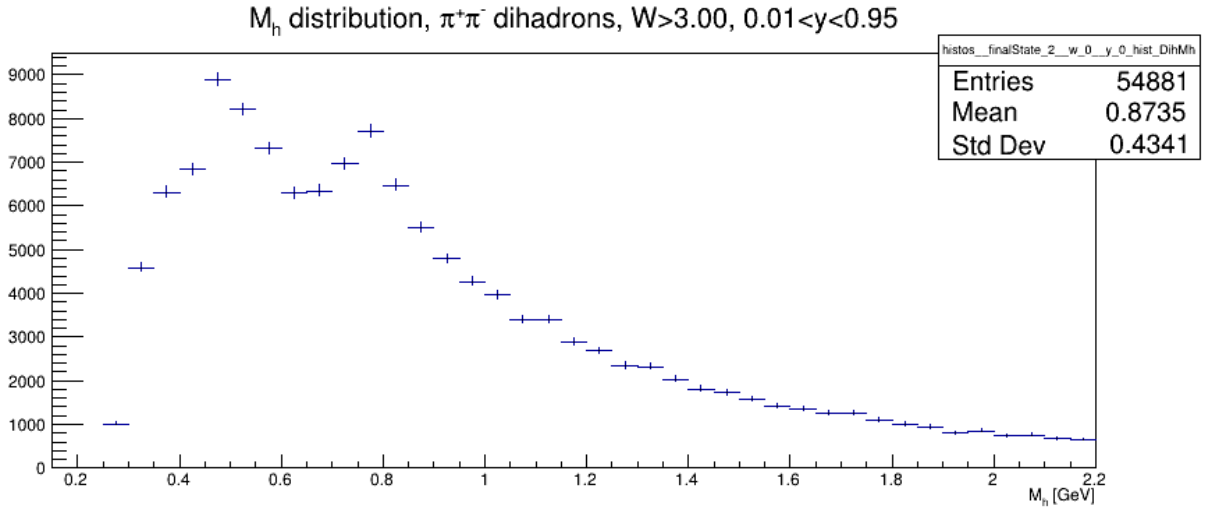
CI Comparisons – ePIC vs. Legacy

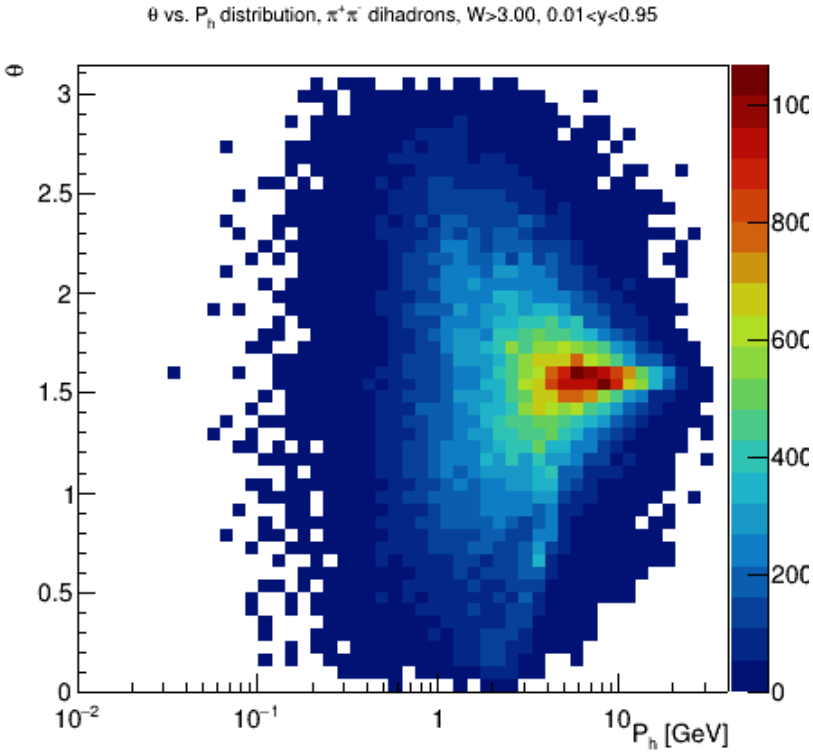
ϕ_h resolutions



Future Support: Dihadrons

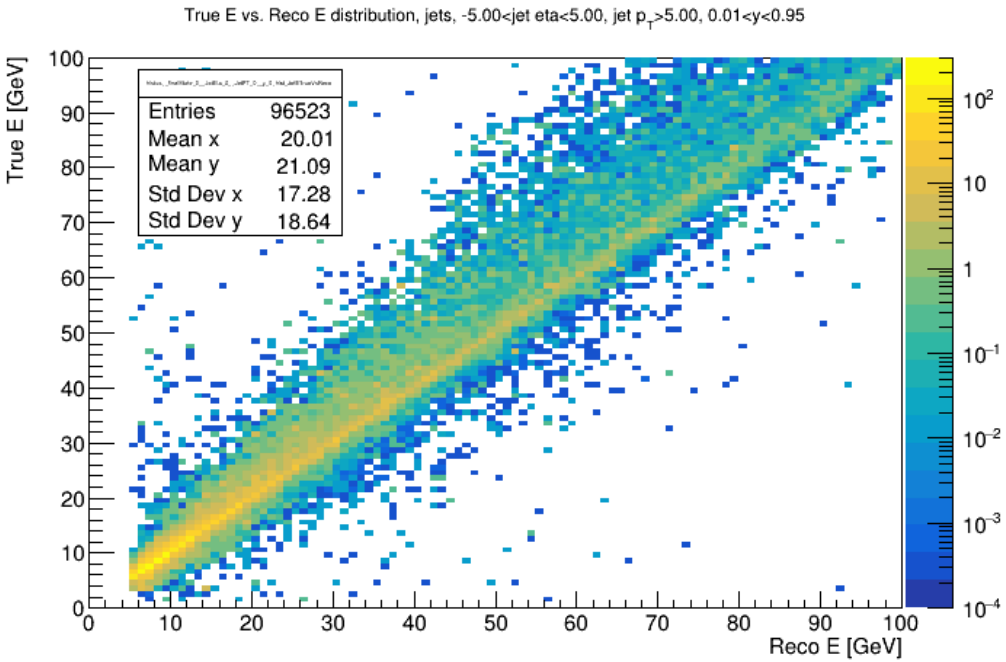
- Work in Progress: https://github.com/eic/sidis-eic/pull/192
  - Kinematics – done, but needs validation / cross check
  - Inclusive Pairing – done, but needs validation

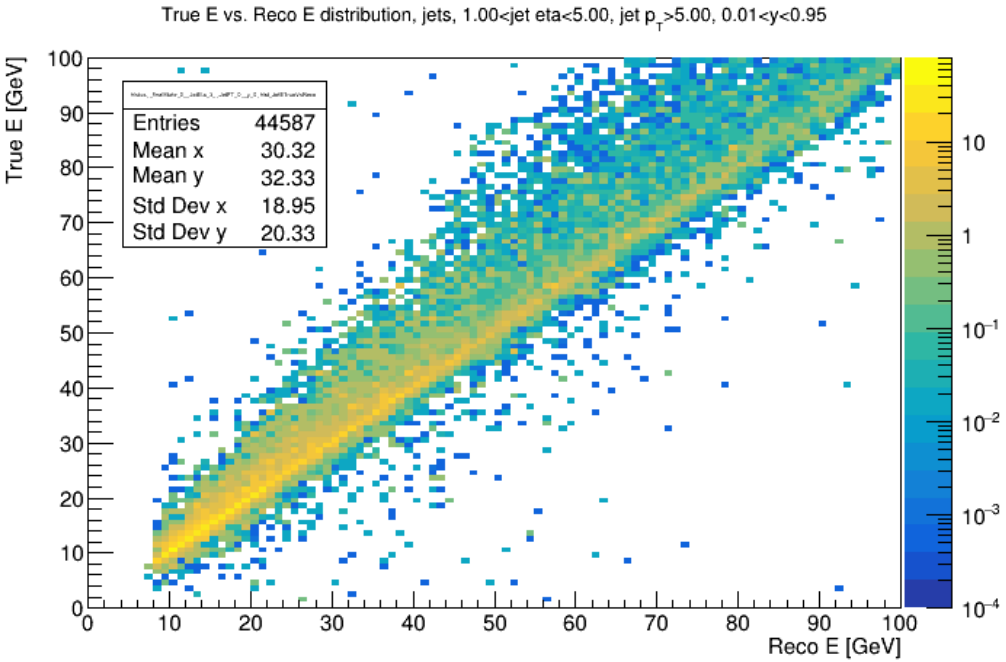




Future Support: Jets

- **Already supported in Delphes analysis chain, using fastjet anti-k_T**





$\eta > 1$

Summary

- **epic-analysis**
  - Supports ePIC full simulation, Delphes fast simulation, and legacy ECCE and ATHENA productions
  - Automates retrieval of data from S3 and Q^2 weighting
  - Kinematics Reconstruction via various methods
- **Continuous Integration**
  - As a “slow” benchmark
  - Comparisons of ePIC vs. ePIC and ePIC vs. Legacy
- **Short term plans**
  - Dihadrons
  - Jets
  - Scaling (support higher statistics)

backup

Handling Multidimensional Binning

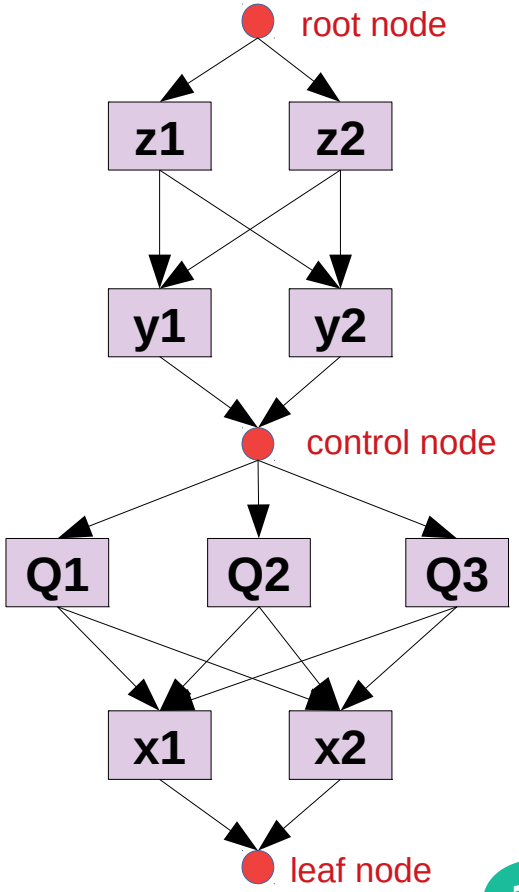
- Problem: The need for multidimensional analysis caused deeply nested for loops to spread throughout epic-analysis
  - Not maintainable and not generalized
  - Very susceptible to bugs

```
for (auto z_bin : z_bins) {
  for (auto y_bin : y_bins) {

    action_before_x_Q2_subloop( z_bin, y_bin );

    for (auto Q2_bin : Q2_bins) {
      for (auto x_bin : x_bins) {

        action_for_each_bin( z_bin, y_bin, Q2_bin, x_bin );

      }
    }

    action_after_x_Q2_subloop( z_bin, y_bin );

  }
}
```

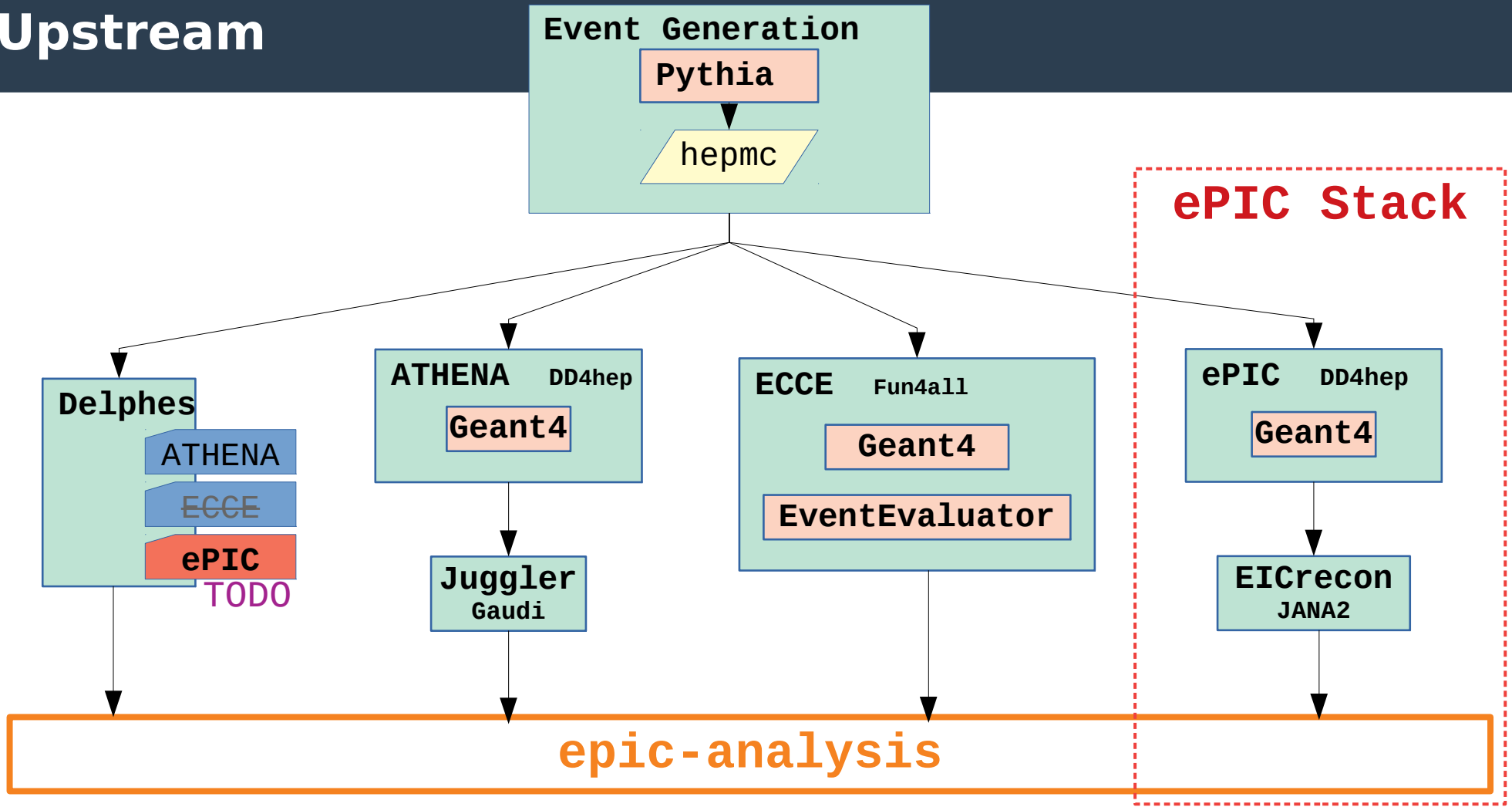
- Solution: use a Directed Acyclic Graph (DAG)
  - Fully connected layers of 1D bins
  - One path from root node to leaf node == 1 multidimensional bin
  - “Control nodes” store lambdas, executable during depth-first traversal

```
// define lambdas
action_before_x_Q2_subloop = ... ;
action_after_x_Q2_subloop  = ... ;
action_for_each_bin        = ... ;

// attach lambdas to the DAG
P->Op()->BeforeSubloop( {"x","q2"}, action_before_x_Q2_subloop );
P->Op()->AfterSubloop(  {"x","q2"}, action_after_x_Q2_subloop  );
P->Op()->Payload( action_for_each_bin );

// run
P->Execute();
```

4D Binning in (z,y,Q^2,x)



Upstream


Event Generation
Pythia
hepmc
ePIC Stack
Delphes
ATHENA
ECCE
ePIC
TODO
ATHENA DD4hep
Geant4
Juggler
Gaudi
ECCE Fun4all
Geant4
EventEvaluator
ePIC DD4hep
Geant4
EICrecon
JANA2
epic-analysis