ePIC-Analysis

Common Physics Analysis Software for the EIC

Christopher Dilks

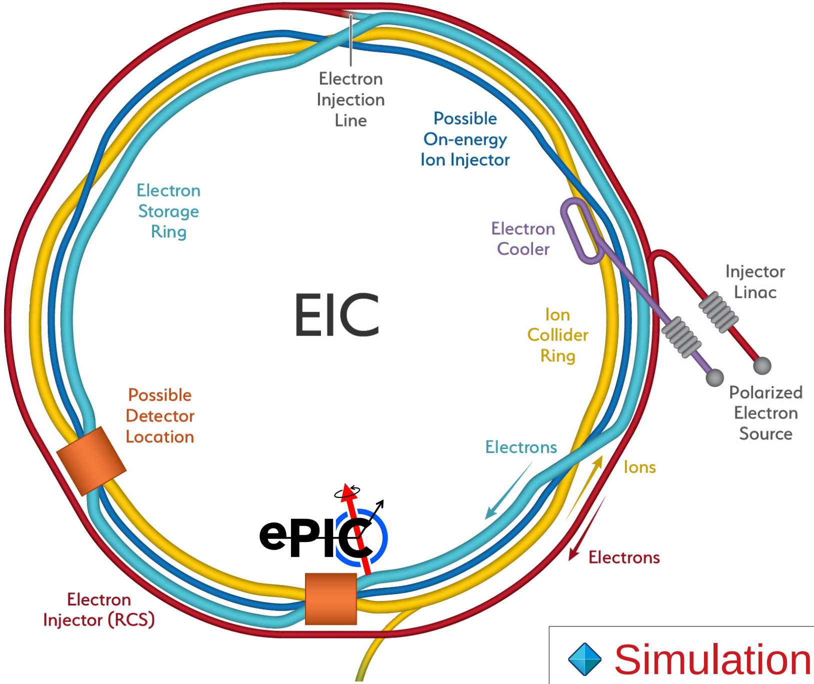
Duke University

for the ePIC Collaboration

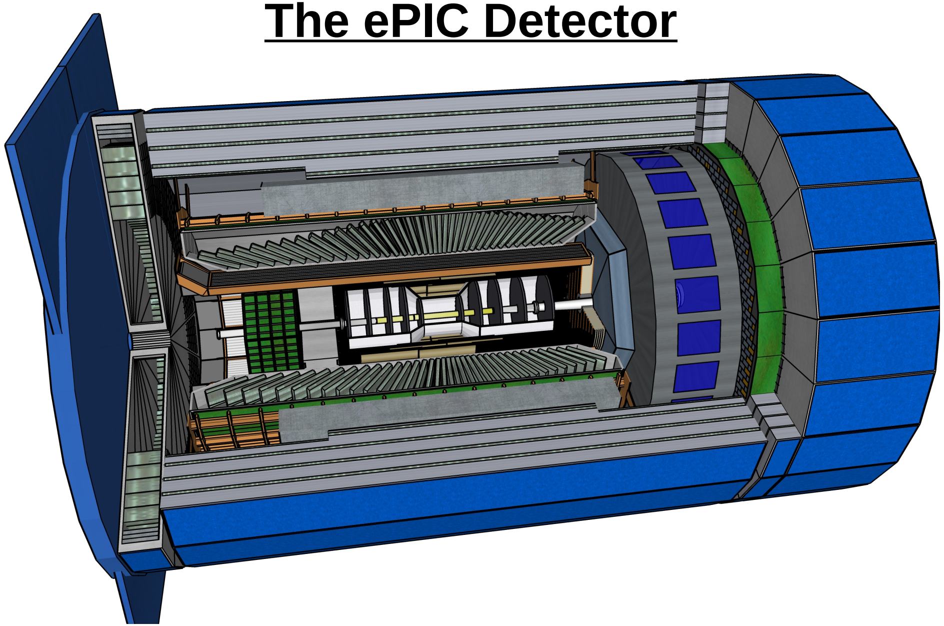
Research supported by the

U.S. DEPARTMENT OF ENERGY | Office of Science

ePIC and the Electron-Ion Collider



The ePIC Detector



- Simulation, reconstruction, and physics studies to help design an optimal detector for future experiments at the EIC
- Analysis of the physics is the primary goal

Some (recent) History

From conceptual designs…

SCIENCE REQUIREMENTS AND DETECTOR CONCEPTS FOR THE ELECTRON-ION COLLIDER

EIC Yellow Report

… to proposal designs …

2021

ATHENA

ECCE

CORE

2022

ePIC

… and to a future experiment

Different Designs… and Different Software*…

Event Generation

Full Simulation

Fun4all

DD4hep

GEANT4
A SIMULATION TOOLKIT

Fast Simulation

DELPHES
fast simulation

EIC-Smear

Reconstruction

EICrecon
JANA2

Juggler
(Gaudi)

* not a complete list

Different Designs… and Different Software*…

Event Generation

Full Simulation

Fun4all

DD4hep

GEANT4
A SIMULATION TOOLKIT

Fast Simulation

DELPHES
fast simulation

EIC-Smear

Reconstruction

EICrecon
JANA2

Juggler (Gaudi)

Requires adaptability of:

- Users and developers
- Code
  - Detector designs
  - Reconstruction algorithms
  - **Physics Analysis**

Duke UNIVERSITY

ePIC-Analysis: Common ahysics analysis framework

DELPHES
fast simulation

ATHENA

Fast simulation
(Delphes)

ATHENA
Full simulation
(DD4hep → Juggler)

ECCE
Full simulation
(Fun4all → EventEvaluator)

EPIC
Full simulation
(DD4hep → EICrecon)

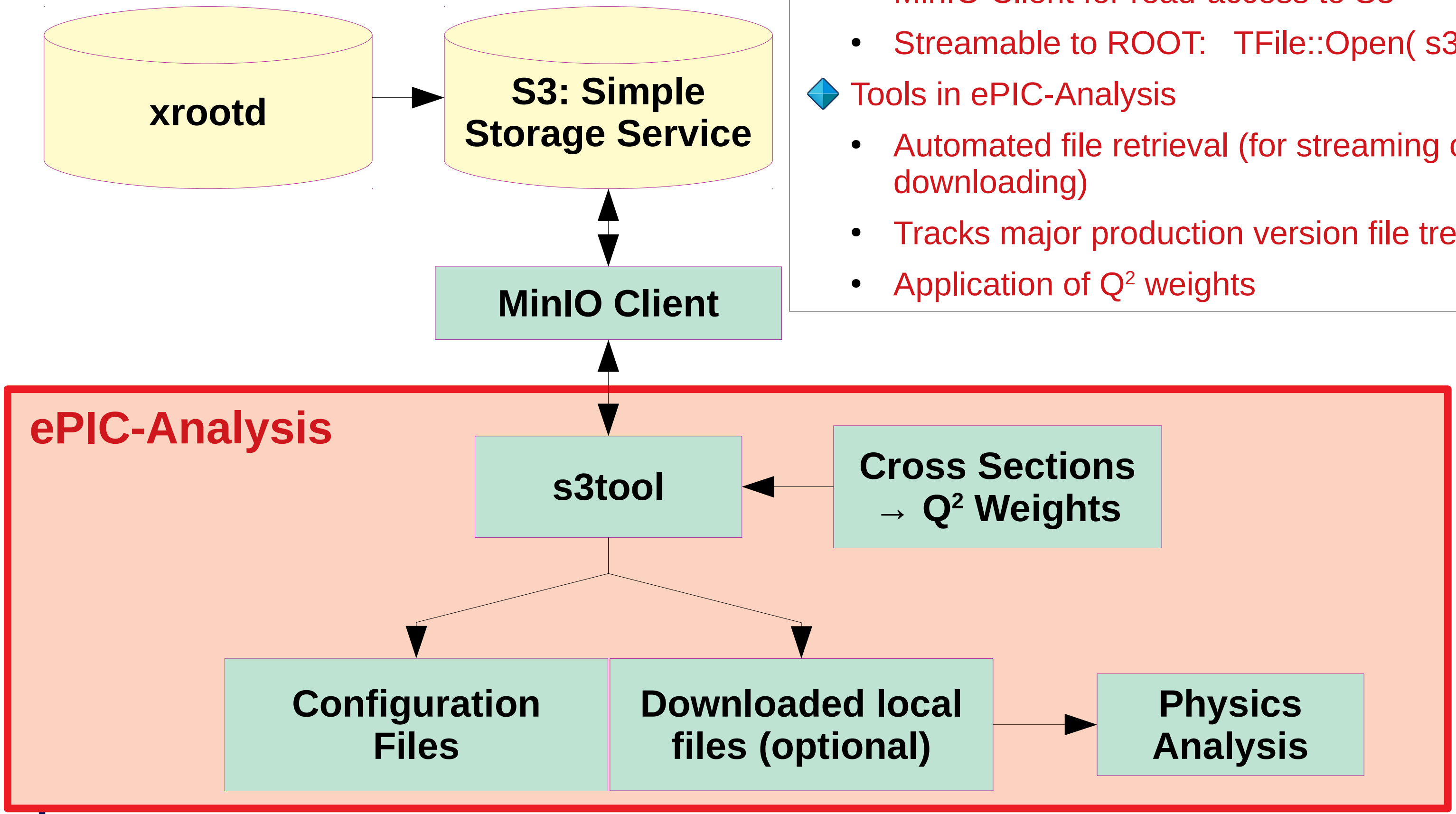
ePIC-Analysis

- Adapted to various upstream simulation sources
- Common physics reconstruction methods for DIS, SIDIS, and Jets
- Common physics analysis techniques
- Continuous Integration to benchmark detector design evolution

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

Data Storage and Retrieval

- Simulation/Reconstructed data are hosted on S3 and xrootd
  - MinIO Client for read-access to S3
  - Streamable to ROOT: TFile::Open(s3_URL)
- Tools in ePIC-Analysis
  - Automated file retrieval (for streaming or downloading)
  - Tracks major production version file trees
  - Application of Q^2 weights

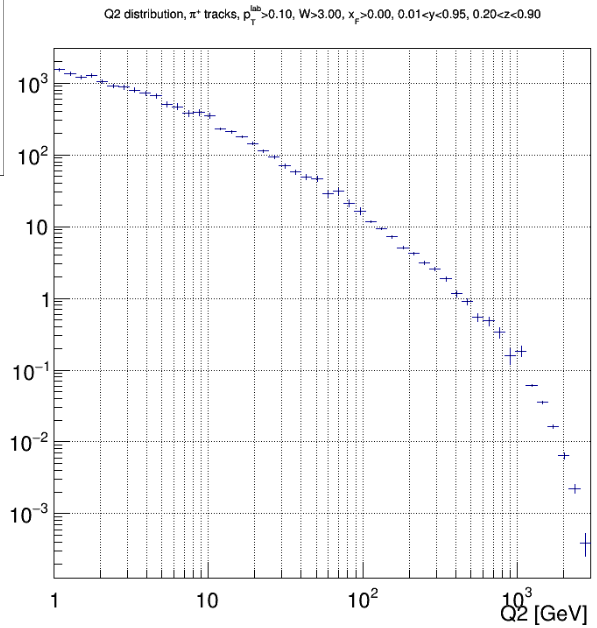


Q^2 Weighting

- Cross section falls rapidly with Q^2 → high Q^2 events are rare
  - Generate events in various bins of Q^2
  - Re-weight them using the cross sections to combine their data
  - Populates statistics even at very high Q^2
  - Allows for study of a broad range of Q^2, without having to wait for rare high Q^2 events
- ePIC-Analysis provides a common Q^2 weighting implementation

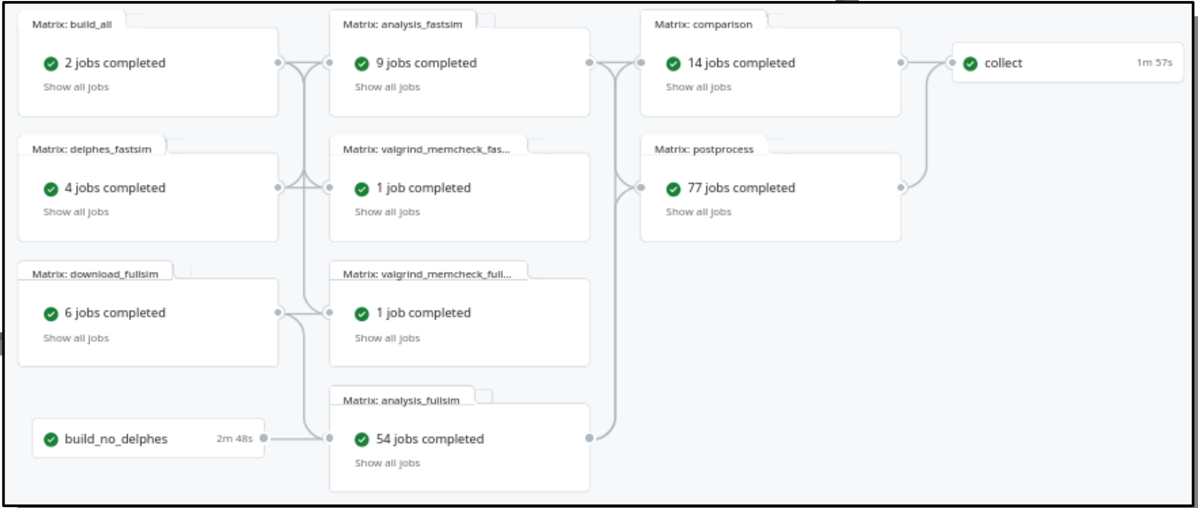
Q2 Bins

- 1 – 10 GeV^2
- 10 – 100 GeV^2
- 100 – 1000 GeV^2
- 1000 GeV^2 and above



Continuous Integration

- **Runs for every git commit (on a pull request)**
  - Could receive triggers from upstream simulation and reconstruction repositories
  - Make a change in geometry or reconstruction, check the impact on the physics
- **Job matrices:**
  - ePIC full simulation / Delphes fast simulation / previous designs
  - With / without radiative corrections
  - Kinematics reconstruction methods (electron / hadronic / mixed / …)
- **Build tests and Valgrind**
- **Artifacts: plots**
  - Coverage
  - Resolution
  - Multidimensional binning



Focusing on semi-inclusive pion production from electron and proton beam energies of 18 and 275 GeV

$$e + p \rightarrow e + \pi^{+} + X$$

Artifacts

- Histograms in bins of
  - (x, Q^2)
  - (η, p)

Semi-Inclusive Deep Inelastic Scattering (SIDIS) Cuts

$$W > 3 \text{ GeV}$$

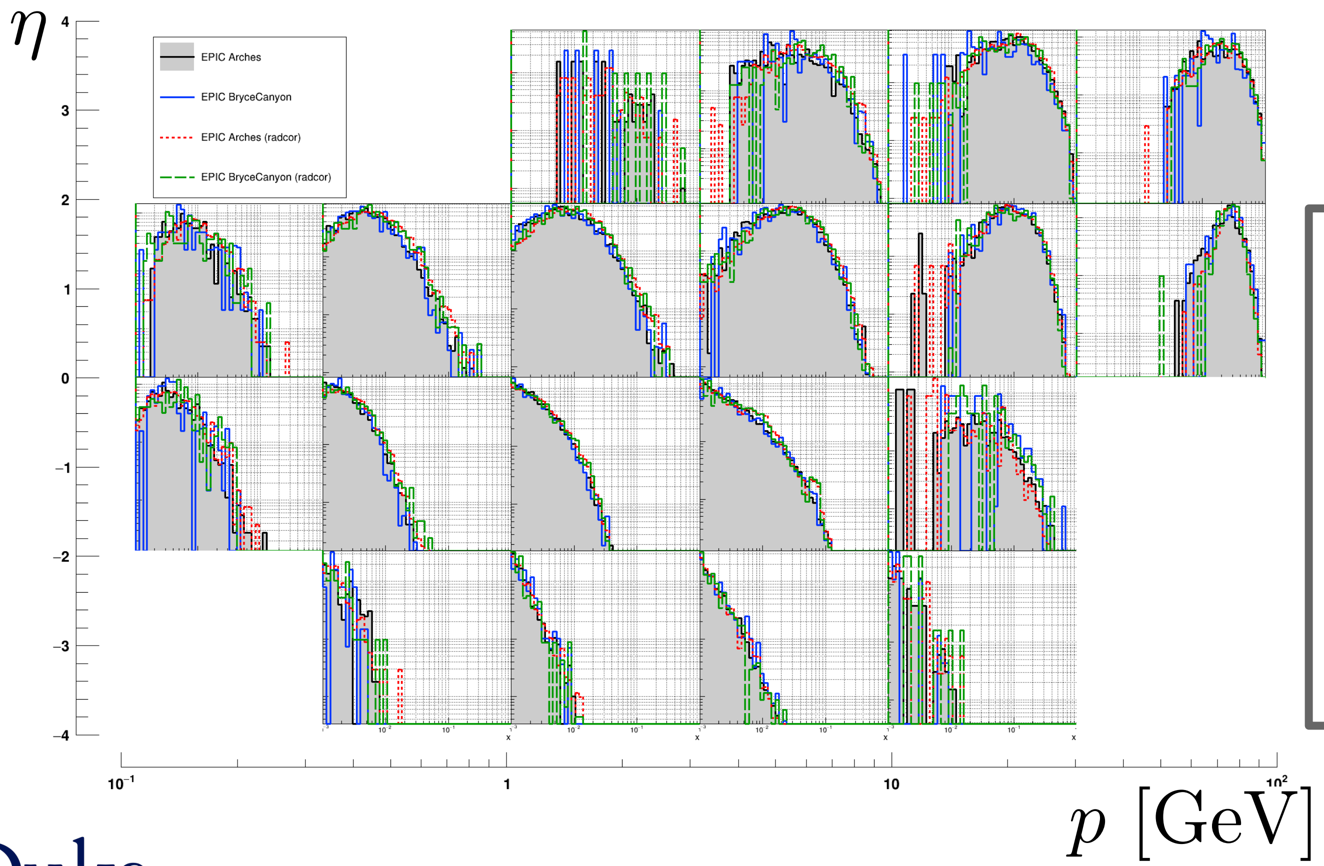
$$0.01 < y < 0.95$$

$$0.2 < z < 0.9$$

$$x_F > 0$$

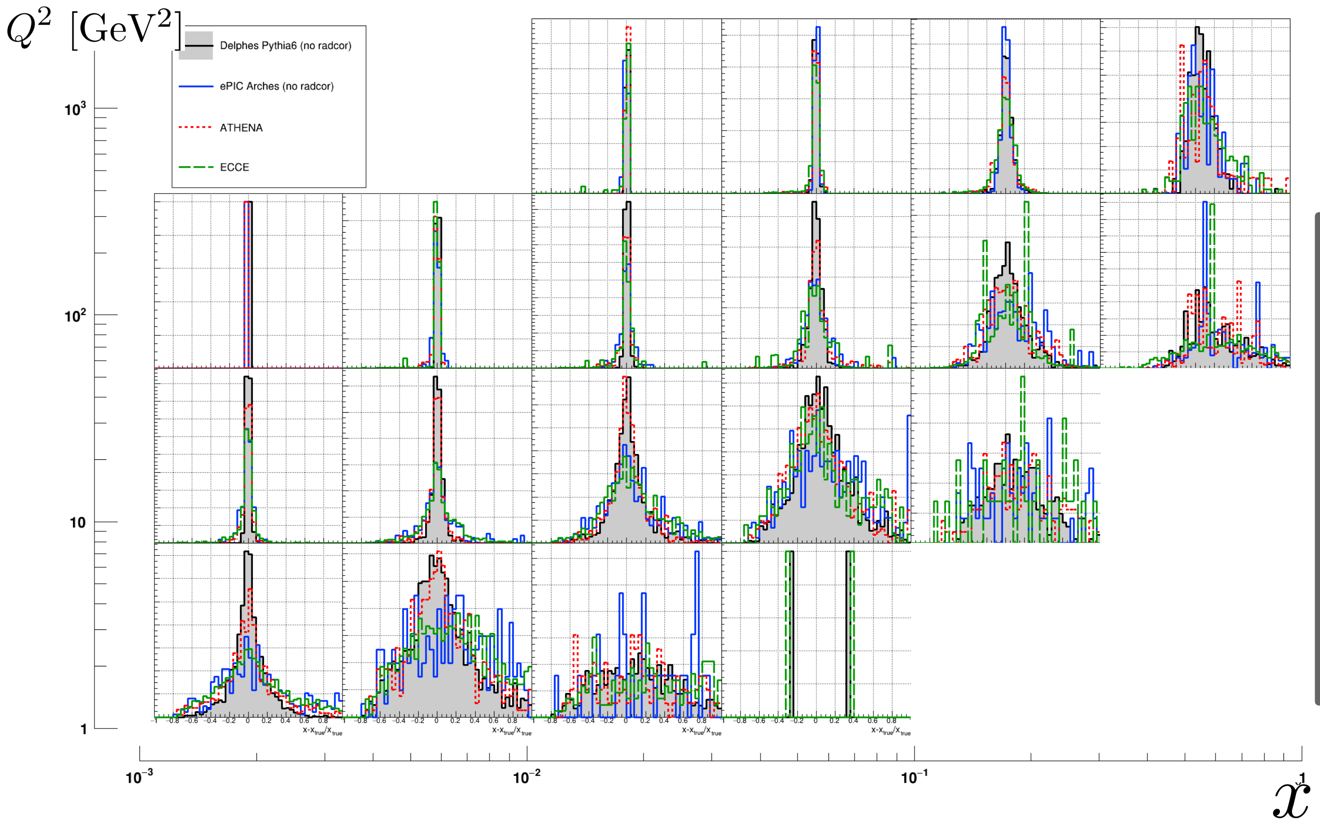
$$p_T(\text{lab}) > 0.1 \text{ GeV}$$

Comparison of two different ePIC design options



- x distributions in bins of (p,η)
- Comparisons:
  - With and without radiative corrections
  - ePIC designs “Arches” and “BryceCanyon”

Comparison of two different ePIC design options



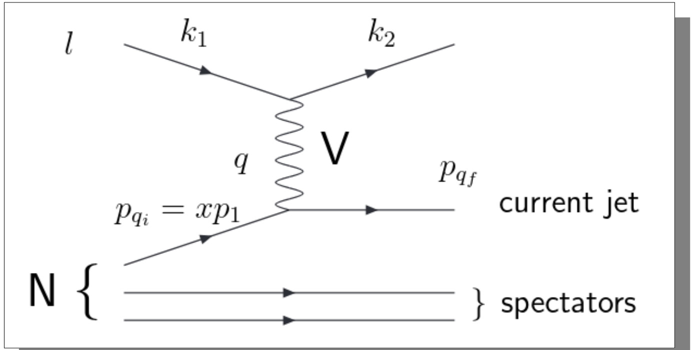
- x resolutions in bins of (x,Q^2)
- Comparisons with previous designs
  - ePIC / ECCE / ATHENA
- Kinematics reconstructed by electron
  - poor resolution in the small y region (low Q^2 and high x)

Kinematics Reconstruction

- Study SIDIS in a *particle collider* context
- Kinematics (x,Q^2,y) can be obtained from initial and final particle momenta
  - Need to develop tools for accurate reconstruction of these event kinematics

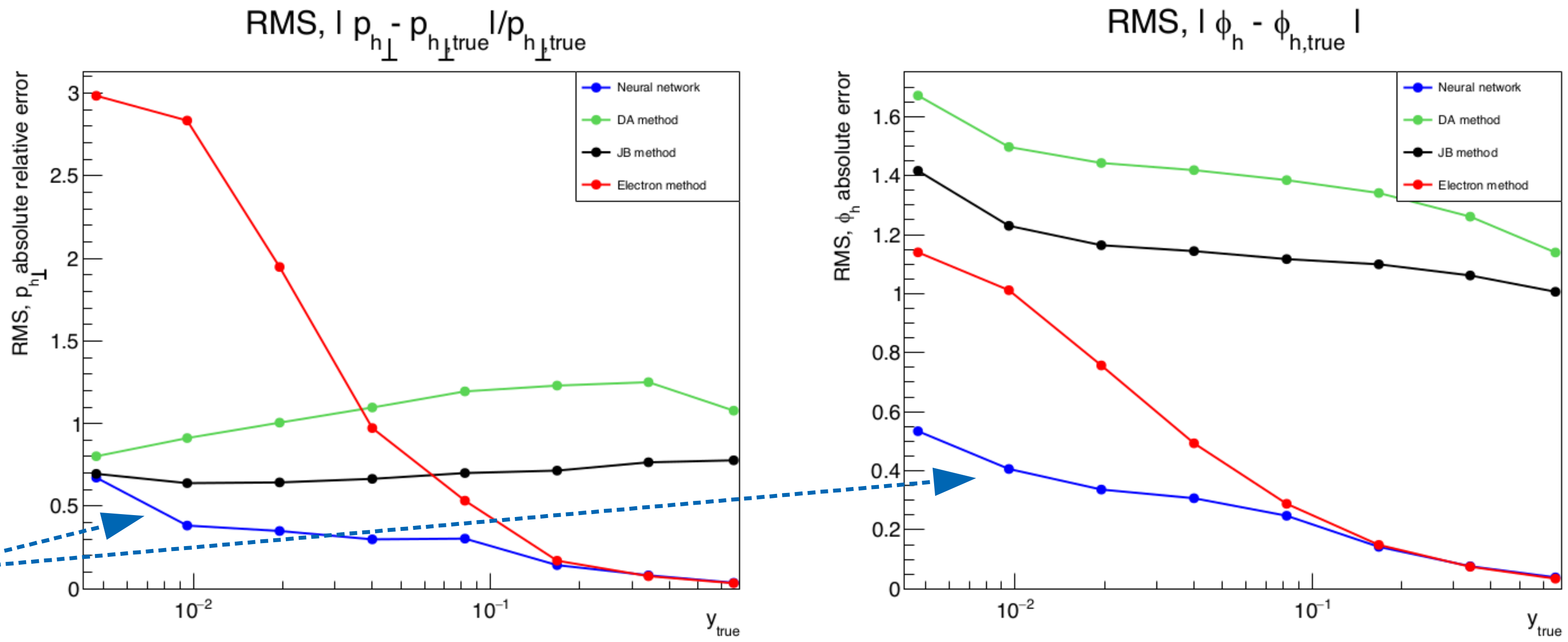
<u>Available methods in **ePIC-Analysis**</u>

i)	*Leptonic variables*	$q \equiv q_l = k_2 - k_1, \quad y_l = p_1.(k_1 - k_2)/p_1.k_1$
ii)	*Hadronic variables* [81]	$q \equiv q_h = p_2 - p_1, \quad y_l = p_1.(p_2 - p_1)/p_1.k_1$
iii)	*Jacquet-Blondel variables* [82]	$Q^2_{JB} = (\vec{p}_{2,\perp})^2/(1-y_{JB}), \quad y_{JB} = \Sigma/(2E(k_1))$ $\Sigma = \sum_h (E_h - p_{h,z})$
iv)	*Mixed variables* [81]	$q = q_l, y_m = y_{JB}$
v)	*Double angle method* [83]	$Q^2_{DA} = \frac{4E(k_2)^2\cos^2(\theta(k_2)/2)}{\sin^2(\theta(k_2)/2) + \sin(\theta(k_2)/2)\cos(\theta(k_2)/2)\tan(\theta(p_2)/2)},$ $y_{DA} = 1 - \frac{\sin(\theta(k_2)/2)}{\sin(\theta(k_2)/2) + \cos(\theta(k_2)/2)\tan(\theta(p_2)/2)},$
vi)	*θy method* [84]	$Q^2_{\theta y} = 4E(k_2)^2(1-y_{JB})\frac{1+\cos(\theta(k_2))}{1-\cos(\theta(k_2))}, \quad y_{\theta y} = y_{JB}$
vii)	*Σ method* [85]	$Q^2_{\Sigma} = \frac{(\vec{k}_{2,\perp})^2}{1-y_{\Sigma}}, \quad y_{\Sigma} = \frac{\Sigma}{\Sigma + E(k_2)[1-\cos(\theta(k_2))]}$
viii)	*$e\Sigma$ method* [85]	$Q^2_{e\Sigma} = Q^2_l, \quad y_{e\Sigma} = \frac{Q^2_l}{s x_{\Sigma}}$



Prog.Part.Nucl.Phys. 69 (2013) 28-84, 1208.6087 [hep-ph]

Kinematics Reconstruction With Machine Learning



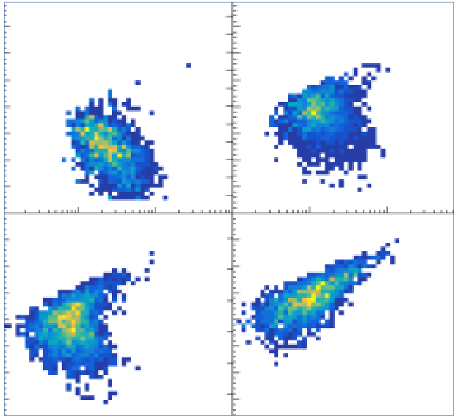
Particle Flow Network

AI for kinematics reconstruction shows promising results!

C. Pecar, 2nd Workshop on AI for the EIC (Oct. 2022)

See also M. Diefenthaler, et al., Eur.Phys.J.C 82 (2022) 11, 1064

Output Data Structures



- **ROOT objects**
  - Specific TTrees
    - SIDIS
    - Jets
    - and more
  - Histograms
- **Support for multidimensional binning of objects**
  - 1D Binning of observables is not enough!
  - The cross section is multidimensional, thus we need to perform analysis in multidimensional bins

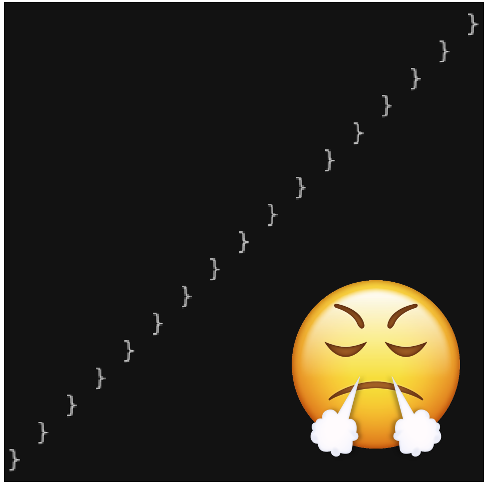
Duke UNIVERSITY

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

$$(z, y, x, Q^2)$$

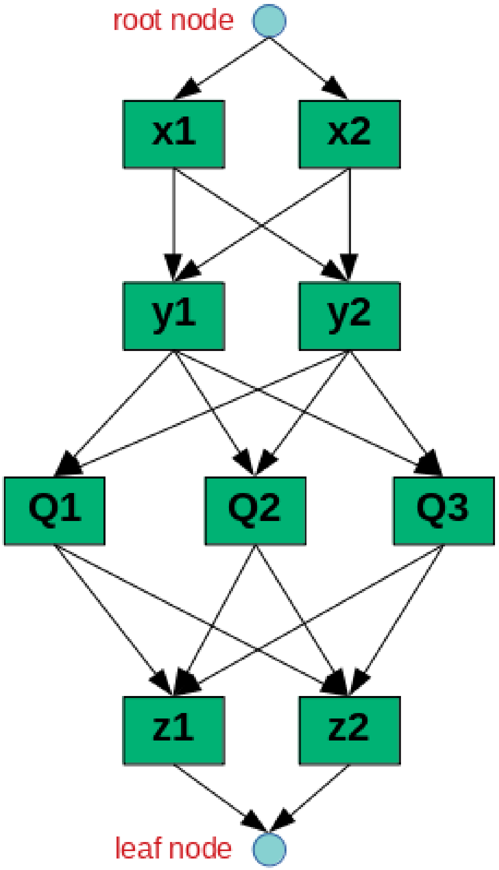
```
for (auto z_bin : z_bins) {
  for (auto y_bin : y_bins) {

    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 );

  }
}
```



Adage

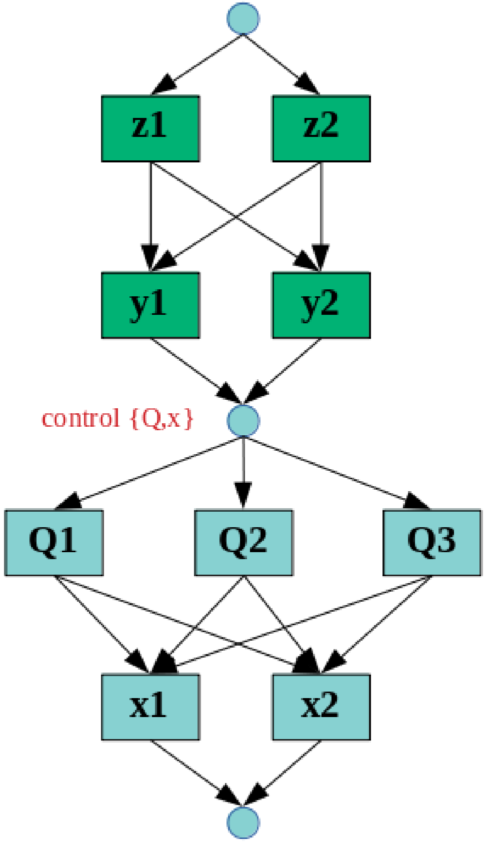
https://github.com/c-dilks/adage

- Generalize multidimensional binning implementation with a Directed Acyclic Graph (DAG)
  - Fully connected layers of 1D bins
  - One full path from root node to leaf node == 1 multidimensional bin
- Store 1st order functions as additional “control nodes”, between layers of 1D bin nodes
  - Executable during depth-first traversal
  - Attach your code to the data structure and run it!



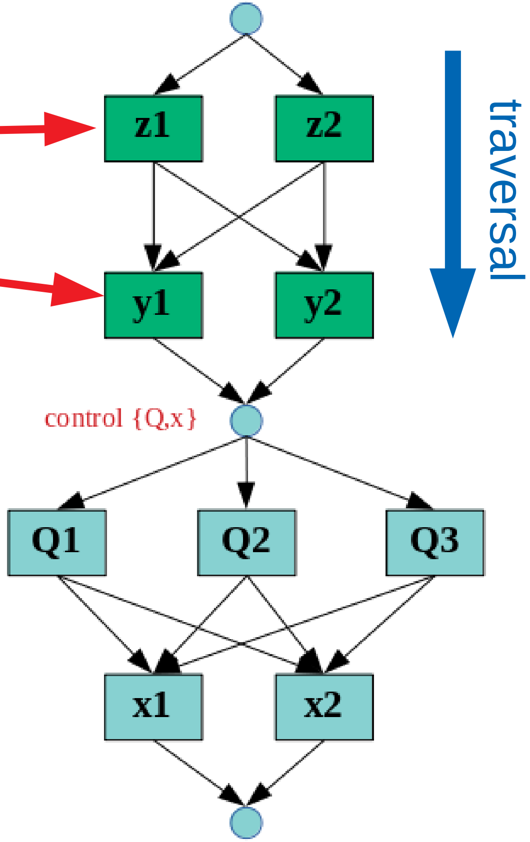
Adage

```
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 );

  }
}
```



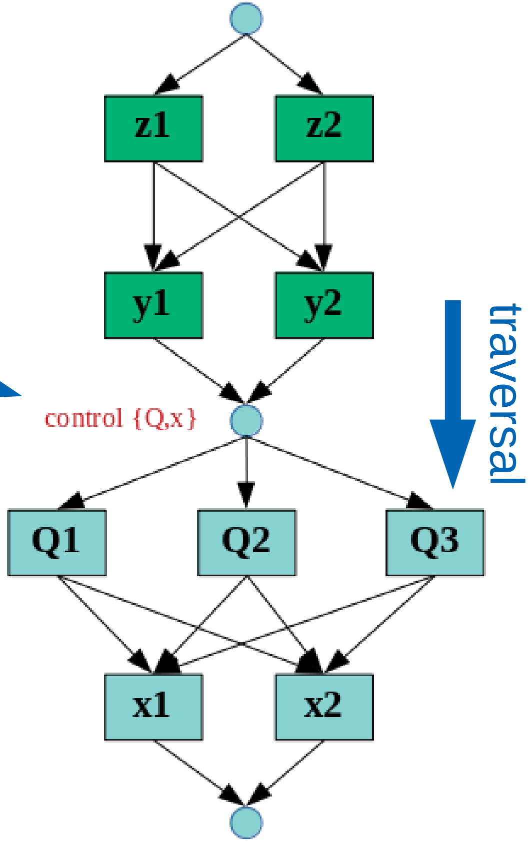
Adage

```
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 );

  }
}
```



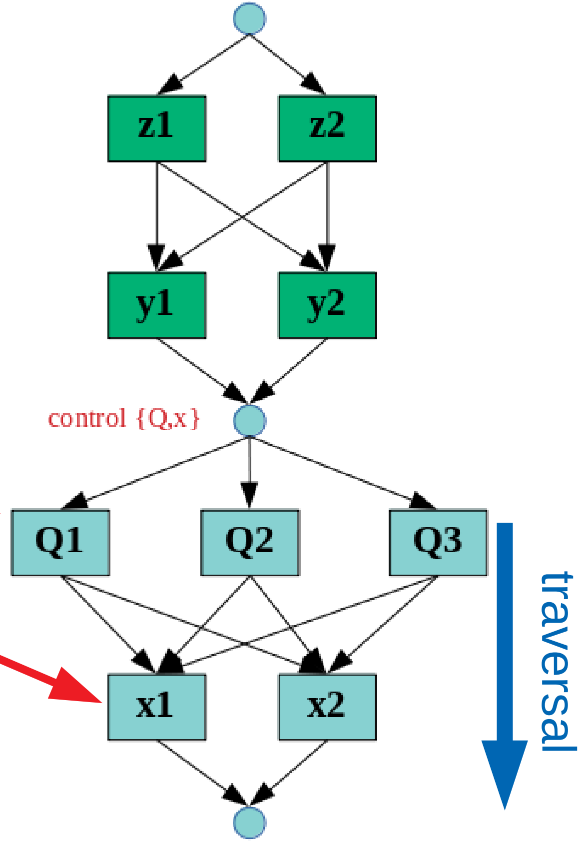
Adage

```
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 );

  }
}
```



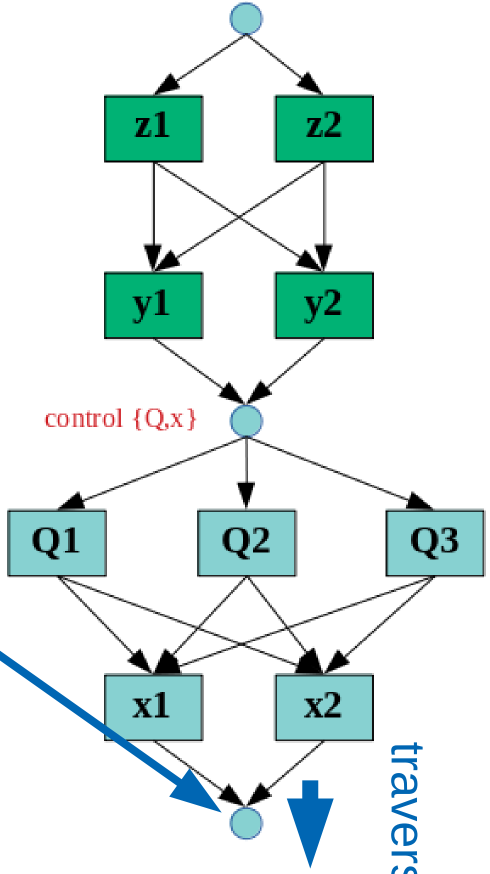
Adage

```
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 );

  }
}
```



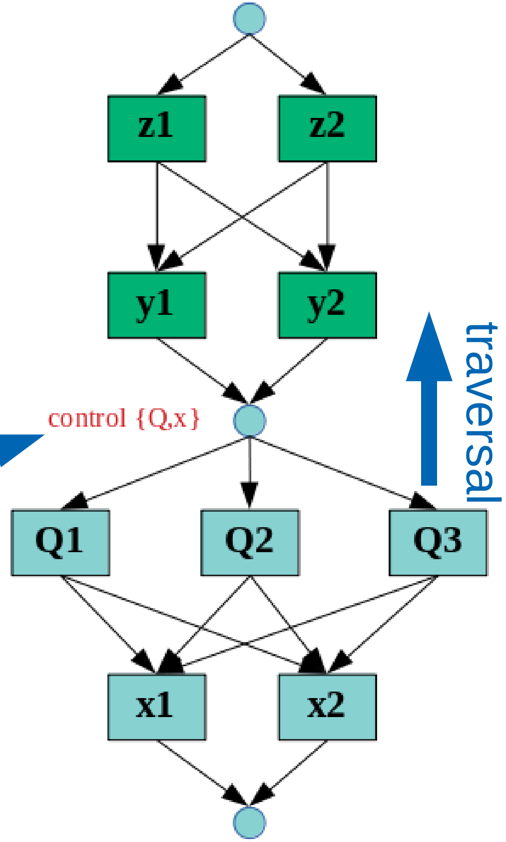
Adage

```
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 );

  }
}
```



Adage

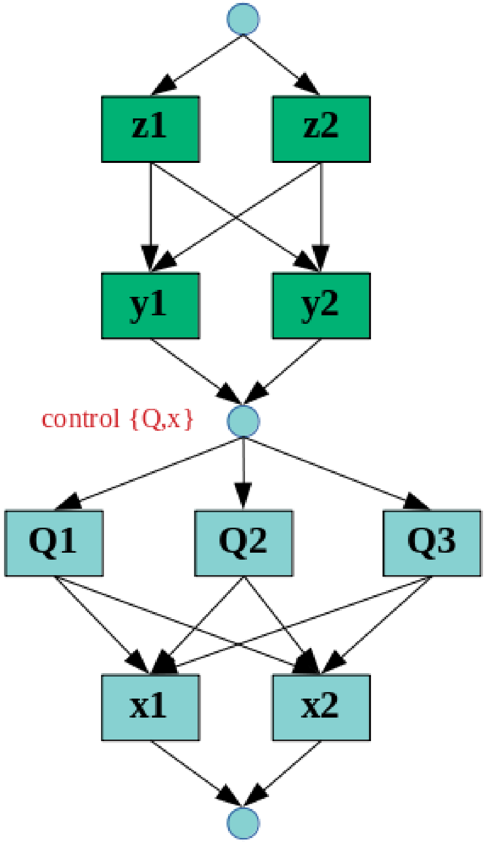
```
for (auto z_bin : z_bins) {
  for (auto y_bin : y_bins) {

    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 );

  }
}
```



Adage

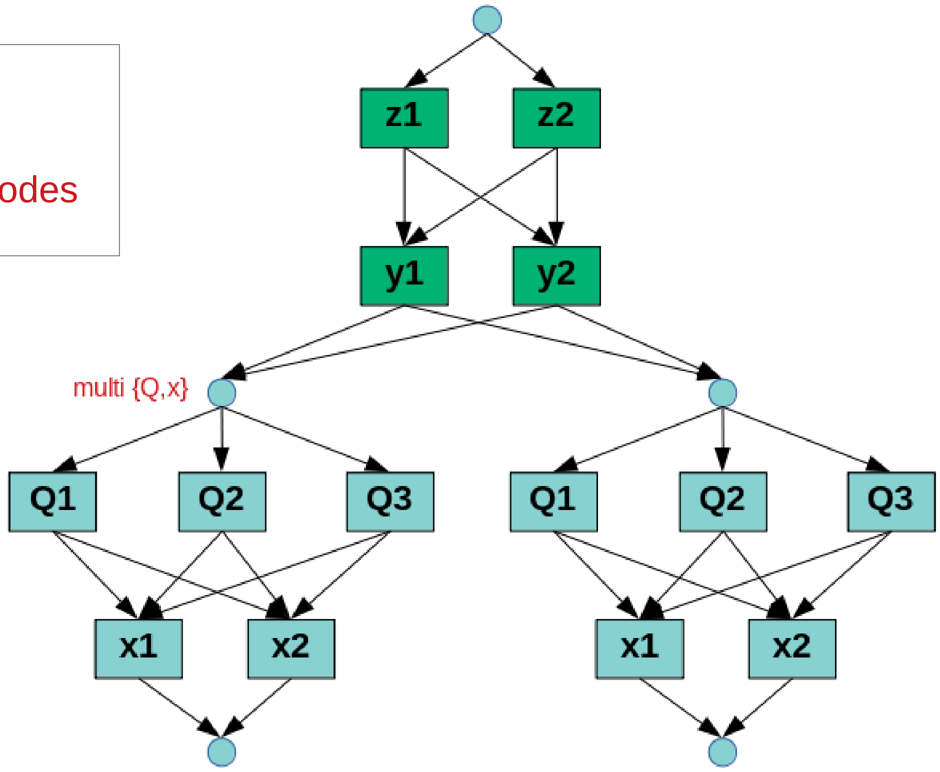
In Practice:

```
// define bins
...

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

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

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



Adage

- Additional Support
  - Conditional execution of subloops
  - Repeated subloops, with different control nodes



Duke UNIVERSITY

Summary

- ePIC-Analysis is a common framework for physics analysis
- Supports various upstream sources from the present as well as the past
- Was critical for the ATHENA proposal design
- Continues to support ePIC and will be integrated in the full software stack
  - See David Lawrence's talk: EIC Software Overview

… and before concluding …

Thanks to Our Contributors!

Christopher Dilks

Connor Pecar

Duane Byer

Sanghwa Park

Gregory Matousek

Matthew McEneaney

Ralf Seidl

Brian Page

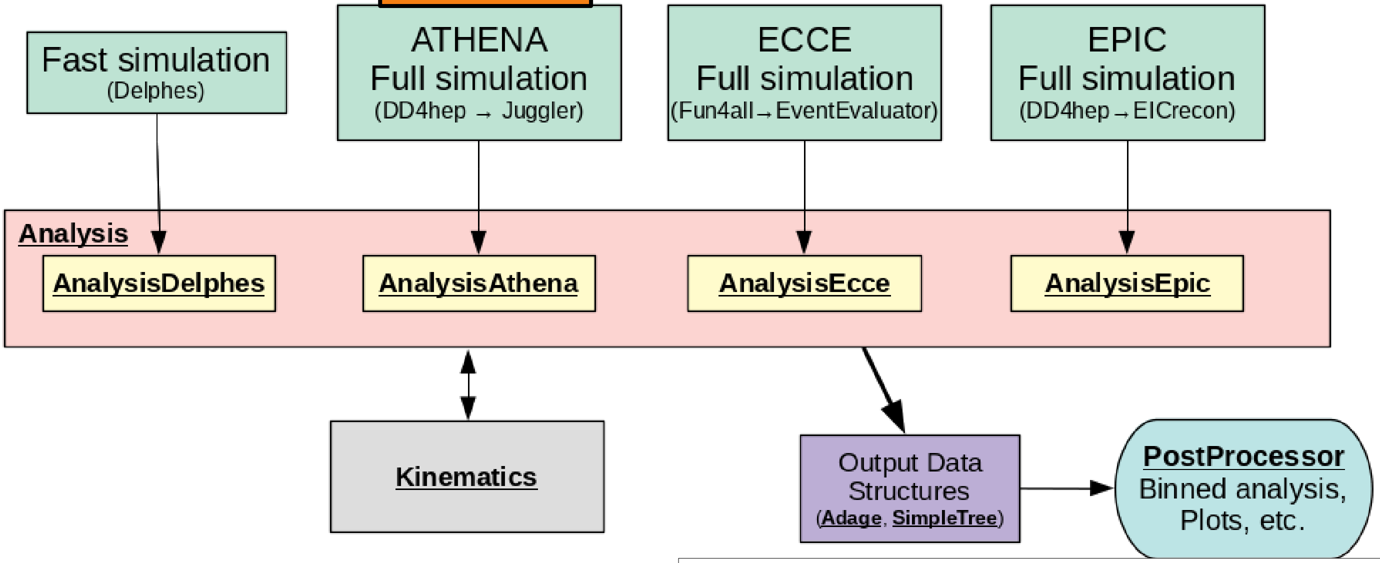
Kevin Adkins

Dmitry Kalinkin

And many more who have contributed advice and help!

BACKUP

ePIC-Analysis Structure



- ROOT: TTrees and Histograms
- Adage: multidimensional binning for anything

Kinematics Reconstruction

- **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