Unbinned unfolding at the EIC

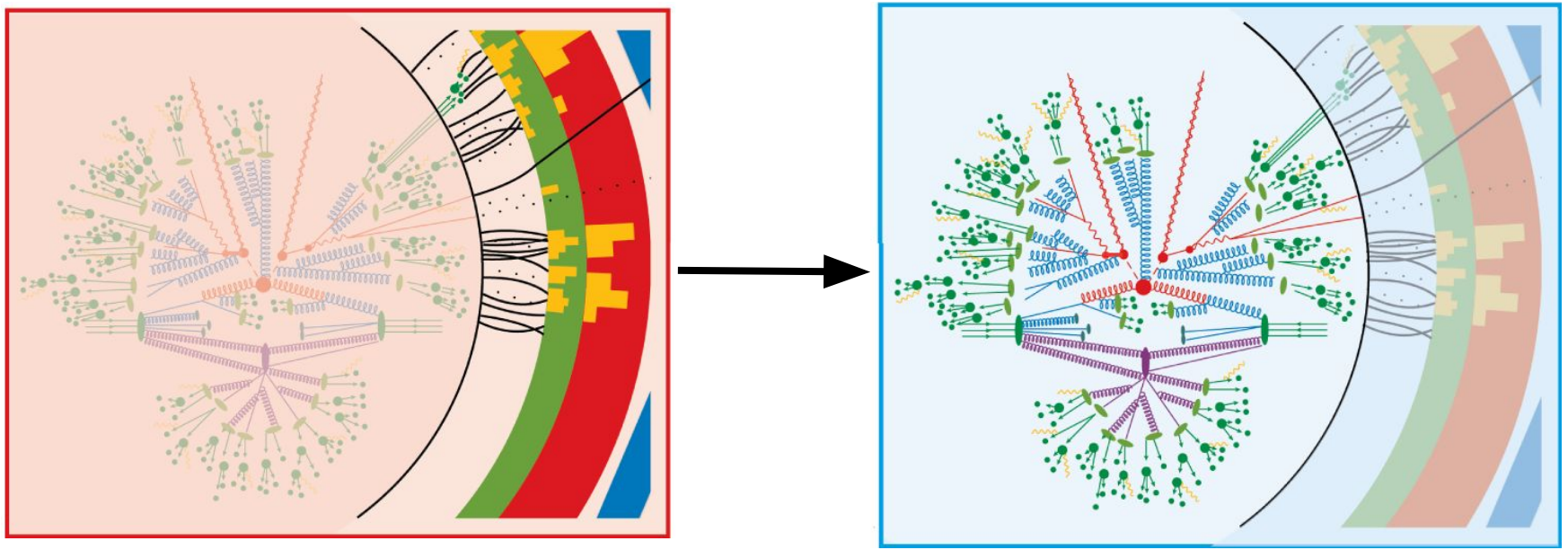
Ryan Milton

In collaboration with M. Arratia & B. Nachman's LBNL group

California EIC Consortium Meeting, UCLA
01/09/2025

Overview of unfolding

- Objective: Remove detector distortions from experimental data



Experimental data

Physics information

Binned unfolding

$$m = Rt$$

Binned unfolding

$$m = Rt$$

Measured data

Binned unfolding

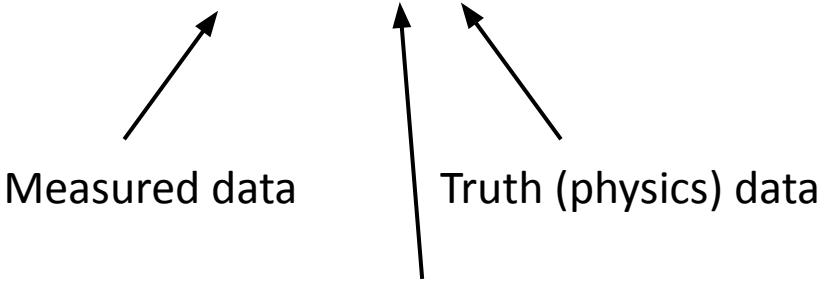
$$m = Rt$$

Measured data

Truth (physics) data

Binned unfolding

$$m = Rt$$



Response matrix

R_{ij} = Pr(measure in bin i | truth in bin j)

Binned unfolding

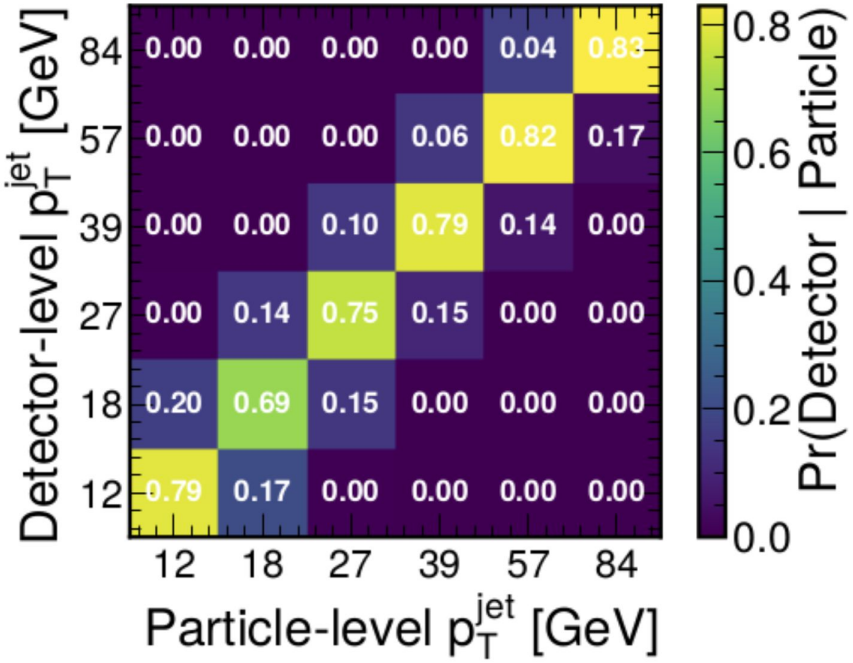
$$m = Rt$$

Measured data

Truth (physics) data

Response matrix

R_{ij} = Pr(measure in bin i | truth in bin j)

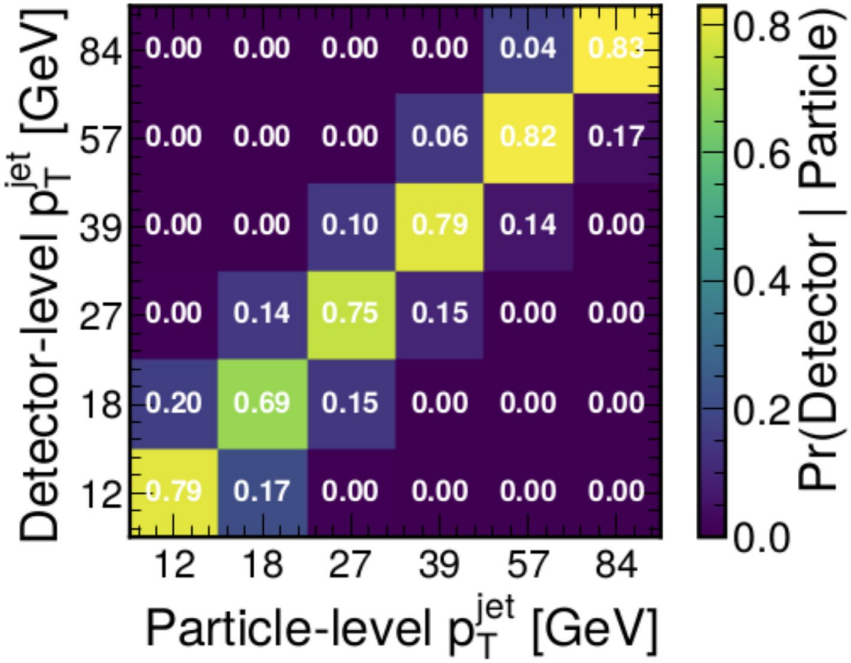


Example of response matrix

Binned unfolding

$$m = Rt$$

$$t = R^{-1}m$$



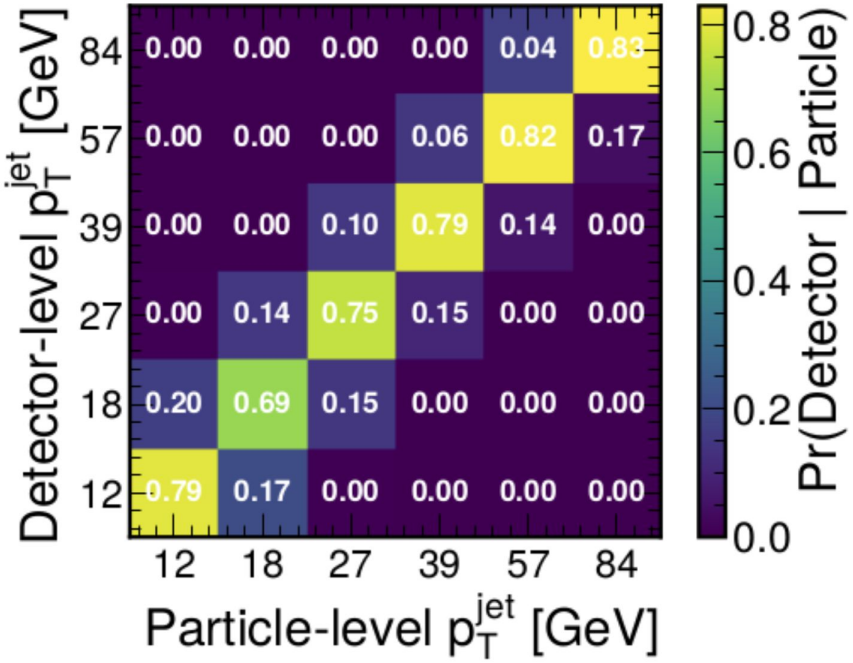
Example of response matrix

Binned unfolding

$$m = Rt$$

$$t = R^{-1}m$$

Main idea of binned unfolding is to invert the response matrix!



Example of response matrix

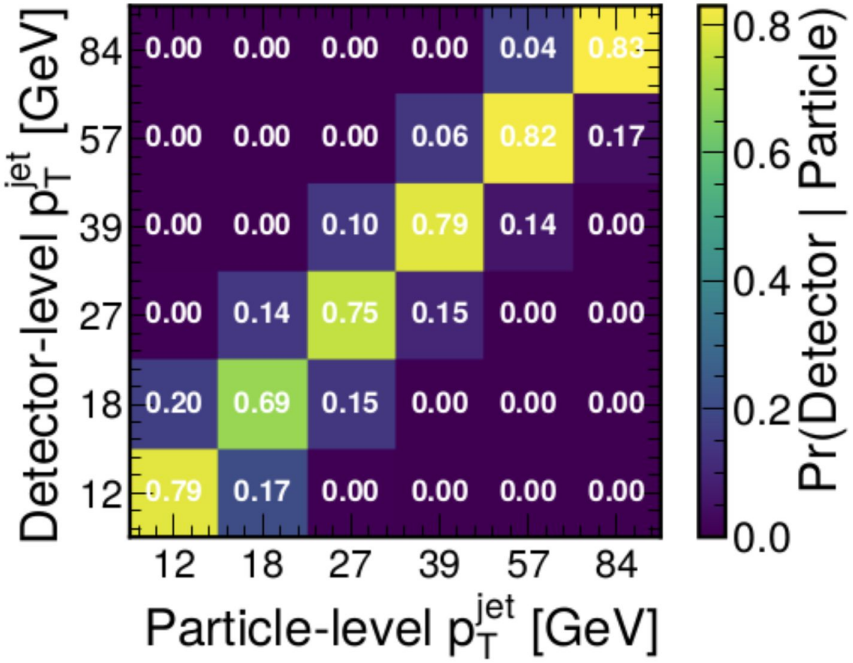
Binned unfolding

$$m = Rt$$

Examples:
Singular value decomposition,
Iterative Bayesian unfolding

$$t = R^{-1}m$$

Main idea of binned unfolding is to invert the response matrix!



Example of response matrix

Drawbacks of binned unfolding

- How to define optimal binning?
  - Must be chosen before unfolding procedure
  - Binning choices makes it difficult to compare between experiments and to publish data

Drawbacks of binned unfolding

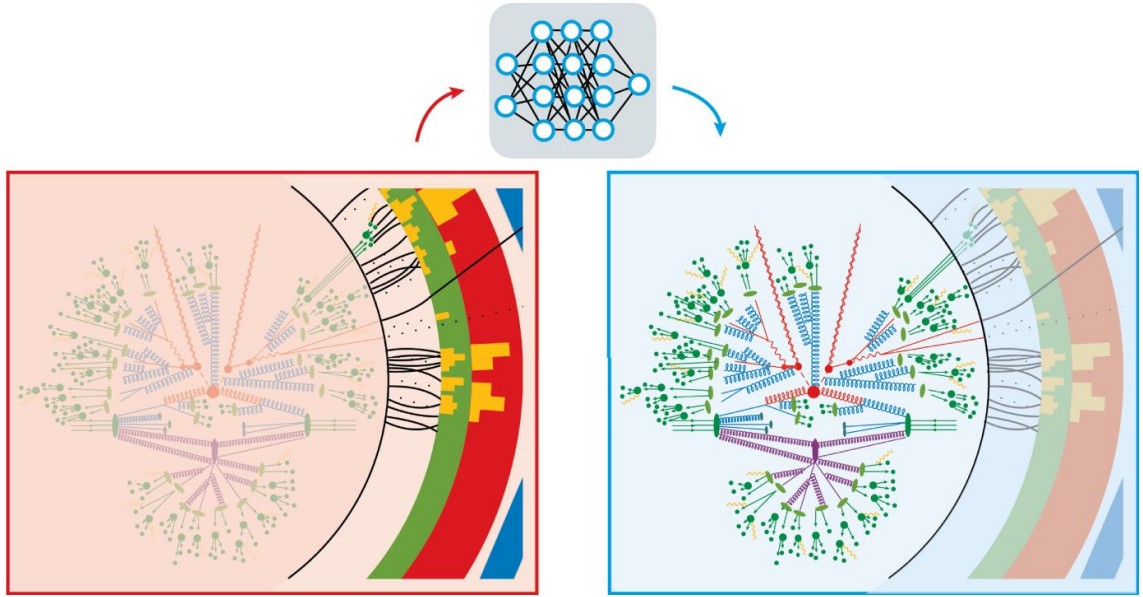
- How to define optimal binning?
  - Must be chosen before unfolding procedure
  - Binning choices makes it difficult to compare between experiments and to publish data
- Difficult to scale histograms to multiple dimensions by including multiple distributions

Drawbacks of binned unfolding

- How to define optimal binning?
  - Must be chosen before unfolding procedure
  - Binning choices makes it difficult to compare between experiments and to publish data
- Difficult to scale histograms to multiple dimensions by including multiple distributions
- Lose correlations between different observables during 1D unfolding

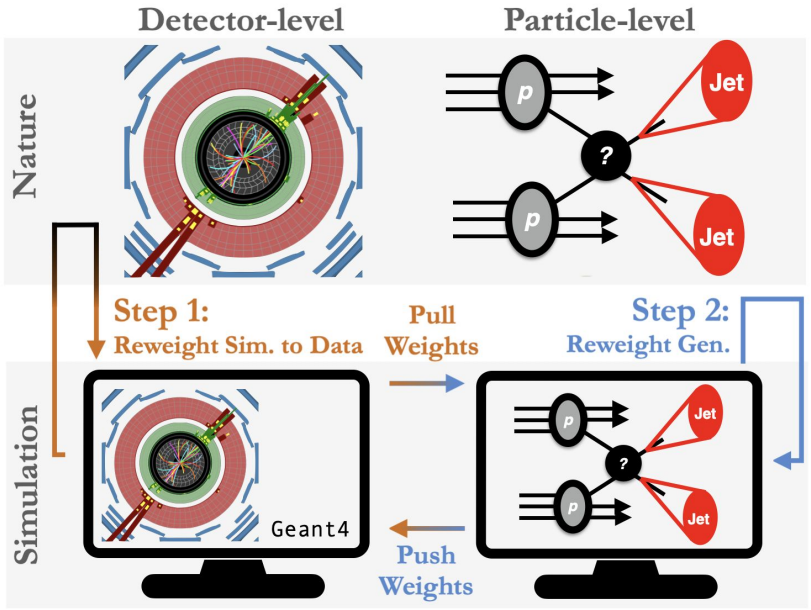
Drawbacks of binned unfolding

- Motivates an unbinned unfolding method using machine learning
- Naturally unbinned and can handle high dimensions



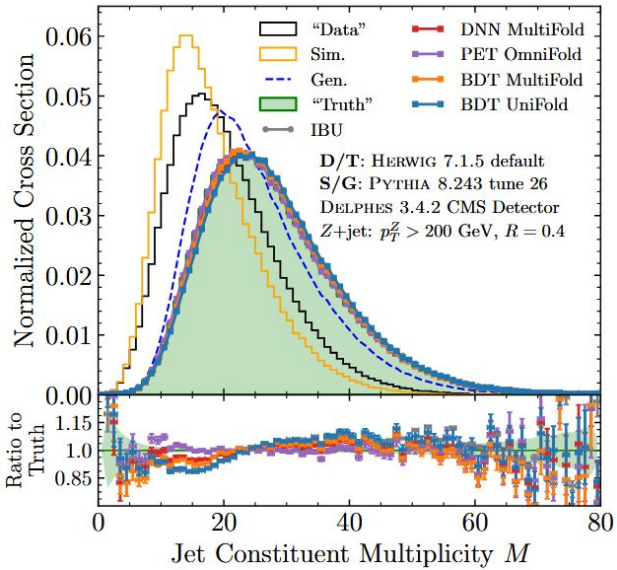
Unbinned unfolding with OmniFold

- Iteratively weights simulation with likelihood ratios
- Step 1: Classifier outputs likelihood ratios of data and reconstructed simulation
- Step 2: Classifier produces likelihood ratios of MC truth and MC truth weighted by step 1 likelihood ratios
- Final step 2 weights applied to MC truth for unfolded data



Unbinned unfolding in RooUnfold

- Lack of common tools for OmniFold
- Created a version of OmniFold in RooUnfold and wrote a paper about unbinned unfolding
- Make ML unbinned unfolding available for everyone!
- Funded by an NSF project with LBNL
  - NSF CSSI 2311666



RooUnfold / RooUnfold / Merge requests / !75

Draft: Omnifold

Open **Ryan Dale Milton** requested to merge `rmilton/RooUnfold:omnifo…` into `master`

Tools for Unbinned Unfolding

Ryan Milton,[a] Trevin Lee,[a] Vinicius Mikuni,[c] Miguel Arratia,[a] Jingjing Pan,[b] Fernando Torales Acosta,[b] Tanvi Wamorkar,[b,d] and Benjamin Nachman[b,c]

Unbinned unfolding in RooUnfold

- Accessible unbinned unfolding tool that is simple to use
- Implemented using boosted decision trees
- Don’t need powerful computing to use

```
omnifold = ROOT.RooUnfoldOmnifold()
omnifold.SetMCDataFrame(df_sim_truth)
omnifold.SetSimDataFrame(df_sim_reco)
omnifold.SetMeasuredDataFrame(df_measured)
omnifold.SetMCPassReco(sim_pass_reco)
omnifold.SetMCPassTruth(sim_pass_truth)
omnifold.SetMeasuredPassReco(measured_pass_reco)
omnifold.SetNumIterations(4)
unbinned_results = omnifold.UnbinnedOmnifold()
step2_weights = ROOT.std.get[1](unbinned_results)
```

Usage of OmniFold

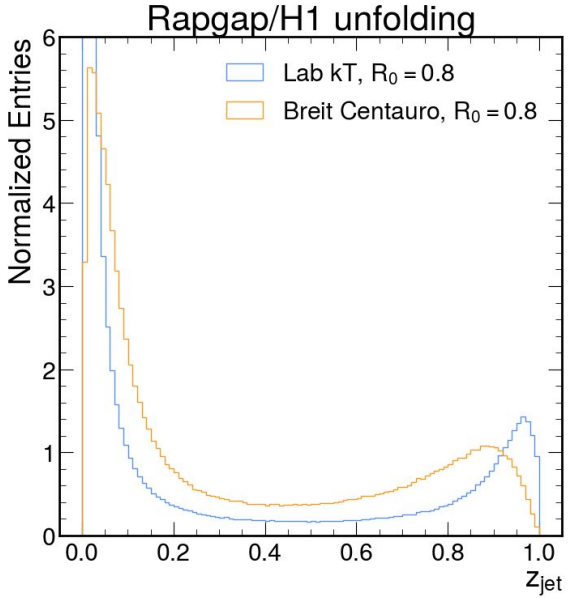
- Used in H1 for simultaneously unfolding multiple jet observables
- Can do new types of measurements by simultaneously unfolding **all particles**
  - Jets clustered in different reference frames with different clustering algorithms
  - Jet clustering with different recombination schemes (E-scheme, winner-takes-all)
  - Impact of clustering before and after boosting
  - Clustering with different radius values
- Unfolding in high dimensions!

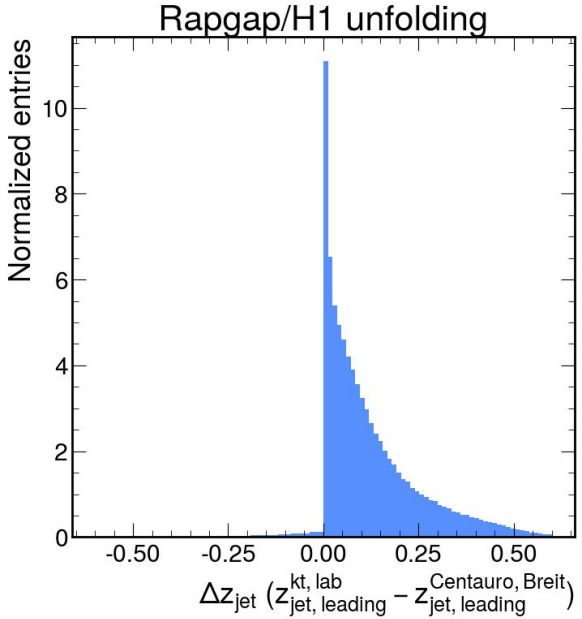
Usage of OmniFold

- Used in H1 for simultaneously unfolding multiple jet observables
- Can do new types of measurements by simultaneously unfolding **all particles**
  - Jets clustered in different reference frames with different clustering algorithms
  - Jet clustering with different recombination schemes (E-scheme, winner-takes-all)
  - Impact of clustering before and after boosting
  - Clustering with different radius values
- Unfolding in high dimensions!

z_{jet} measurements

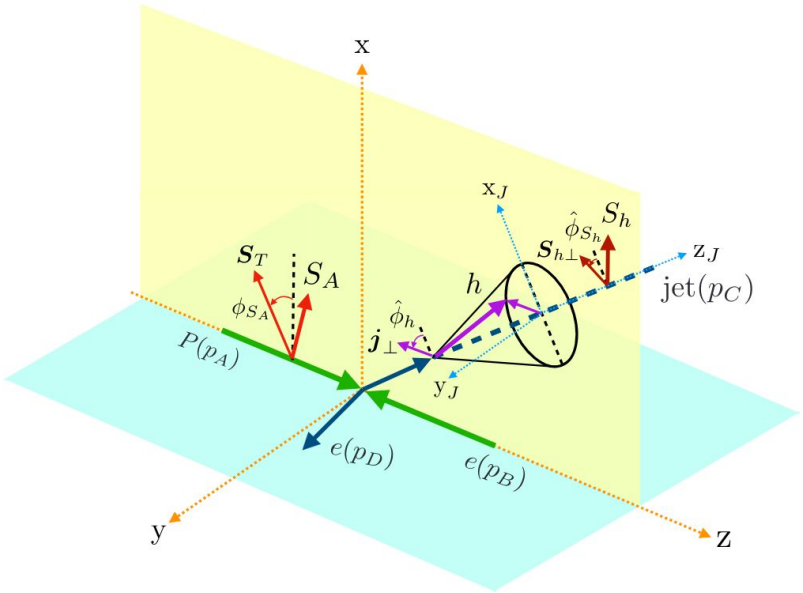
- Idea : Jets clustered in different reference frames (lab/Breit) with different clustering algorithms (kt/Centauro – M. Arratia et. al., PhysRevD.104.034005 (2021))
- Unfolding using all the hadronic final state particles
- z_{jet}: fraction of the struck-quark momentum carried by the jet





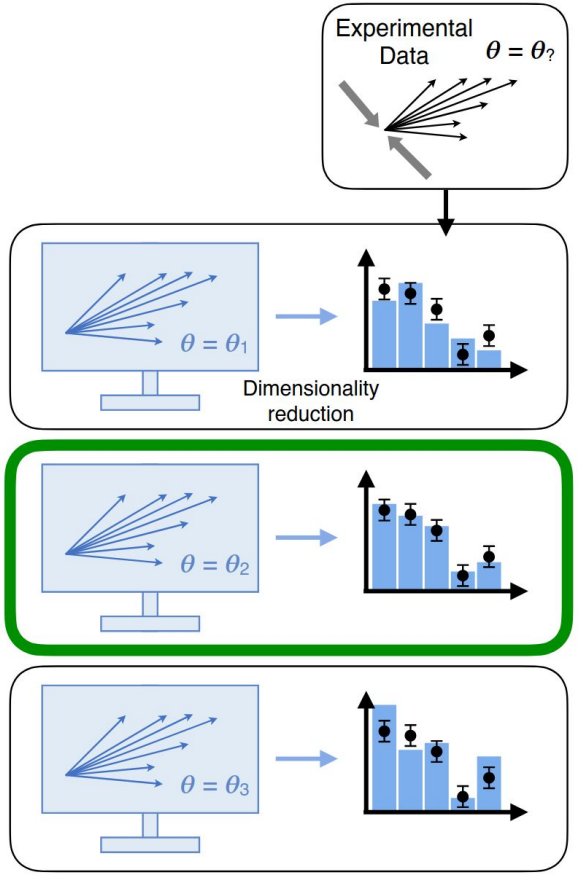
Application to the EIC

- TMDs are key for studying the 3D properties of the nucleon
- Can be probed via jets
- This uses 9 variables – 5 jet variables + 4 hadron variables
  - p_T^{jet}, η^{jet}, q_T^{jet}, $\Delta\phi^{jet}$, ϕ_s (proton spin)
  - Hadron longitudinal-momentum fraction
  - Hadron p_T w.r.t. jet axis, j_T
  - Azimuthal angle ϕ_h
  - Transverse spin of hadron ϕ_s^h
- Can unfold in multiple dimensions



Binned inference

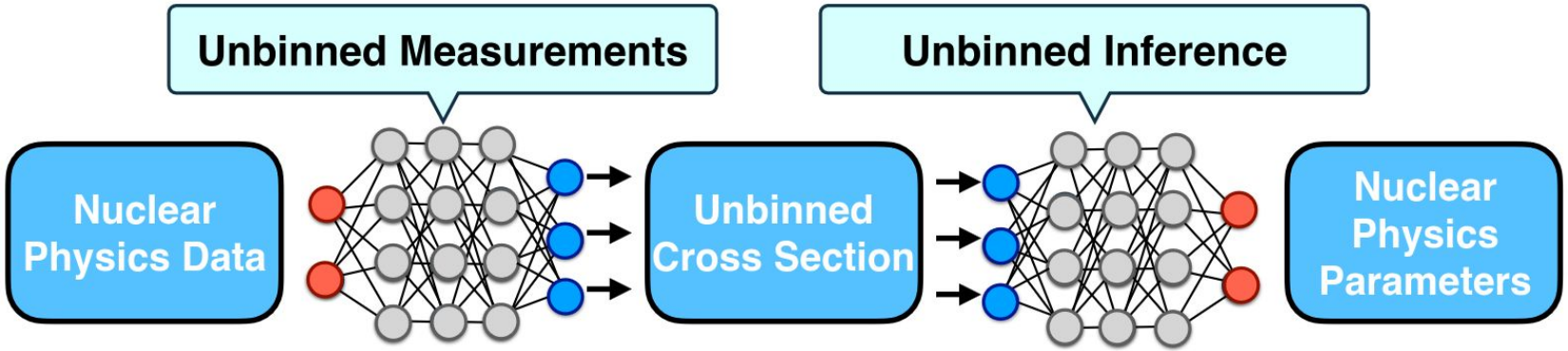
- To go from measurements to matching with theory, need to do inference
- Typically, a functional form is assumed and parameters are fit
  - Computationally expensive and depends on binning



Conventional inference

ML-based inference

- New approach: Train a model to output probability of data given parameters
- Use this model with unbinned unfolded data!



Summary

- Unfolding removes detector distortions from data
- Binned unfolding is limited, motivating an ML unbinned unfolding approach
- Promising applications to H1 and EIC, among other experiments
- Can create a fully unbinned pipeline with unbinned inference

Thanks!