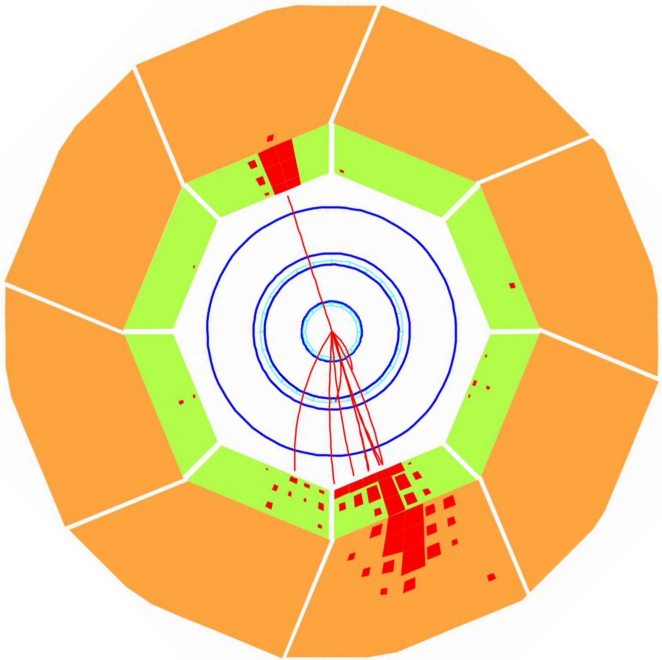


Machine Learning-Assisted Measurement of Lepton-Jet Azimuthal Angular Asymmetries in Deep-Inelastic Scattering at HERA

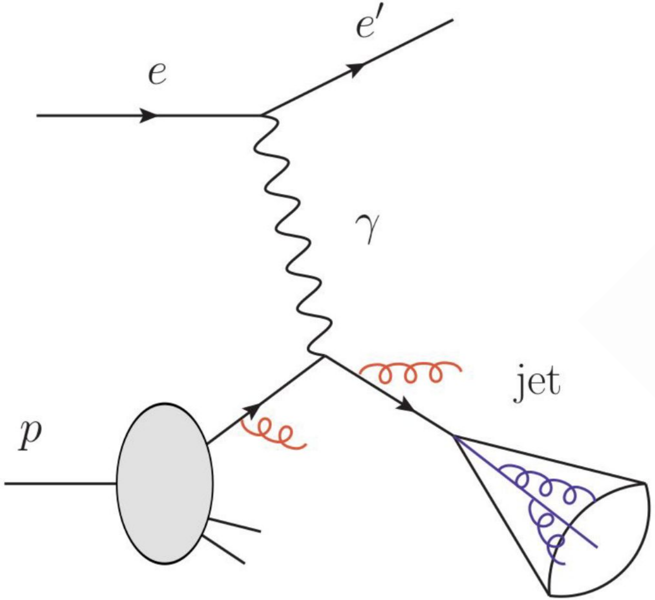
+ Deep learning full phase-space measurement of high Q2 ep collisions at HERA

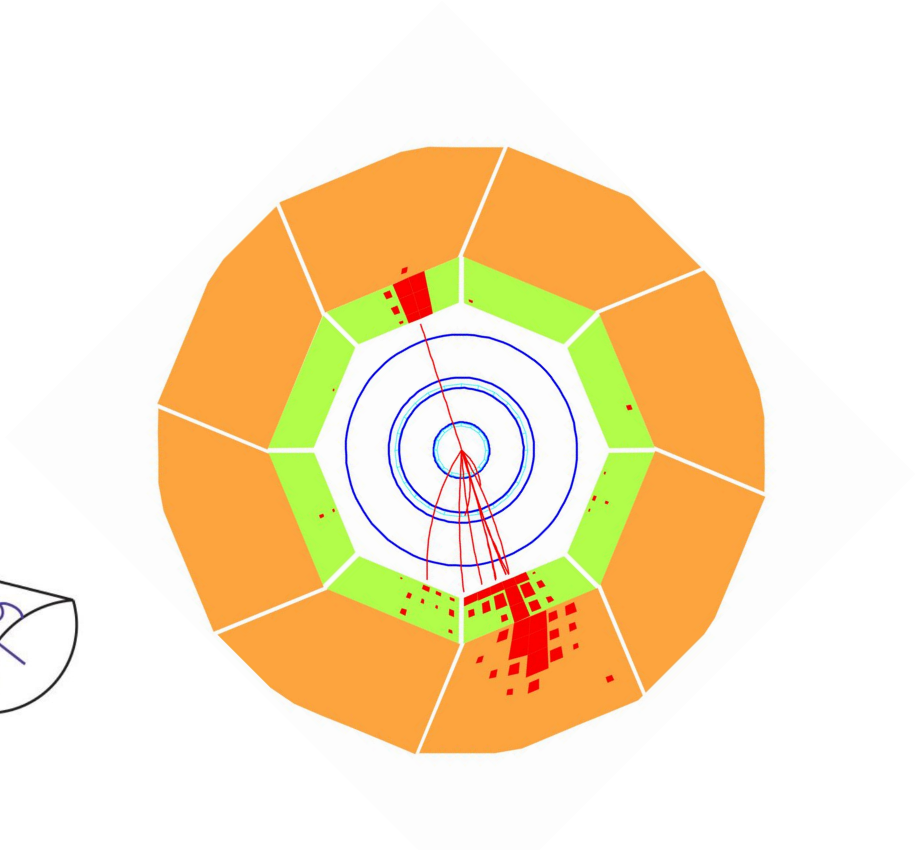
Miguel Arratia,
on behalf of the H1 Collaboration

UC RIVERSIDE

DIS 2025 at Cape Town, South Africa, March 25th 2025

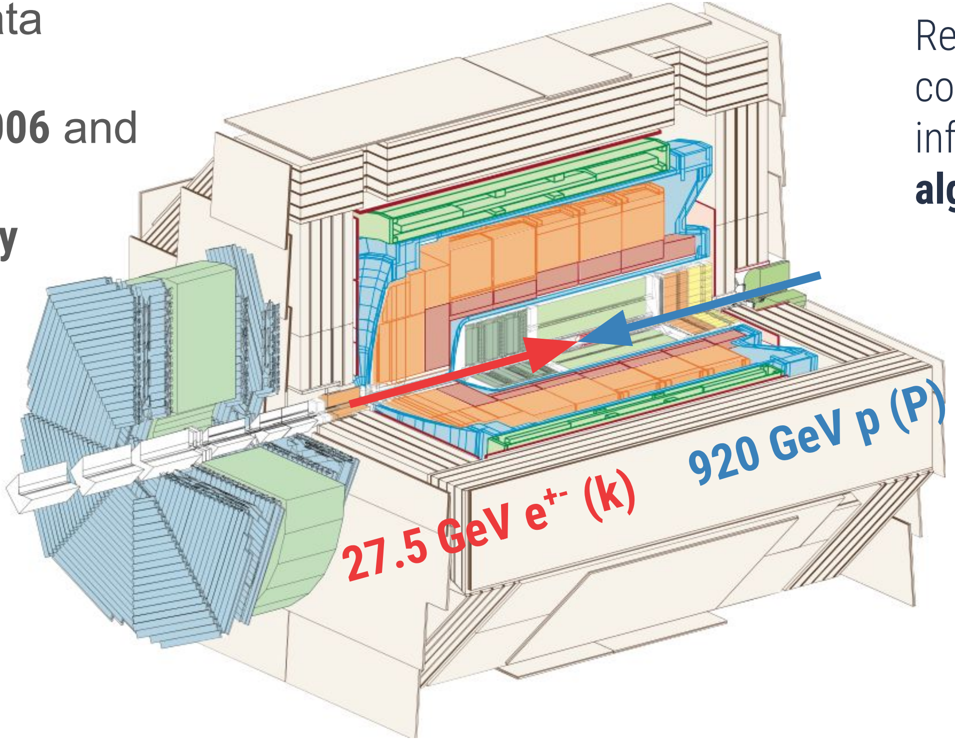
Goal: Measurements of jet in DIS for novel QCD studies, enabled with novel machine-learning approaches





Experimental setup

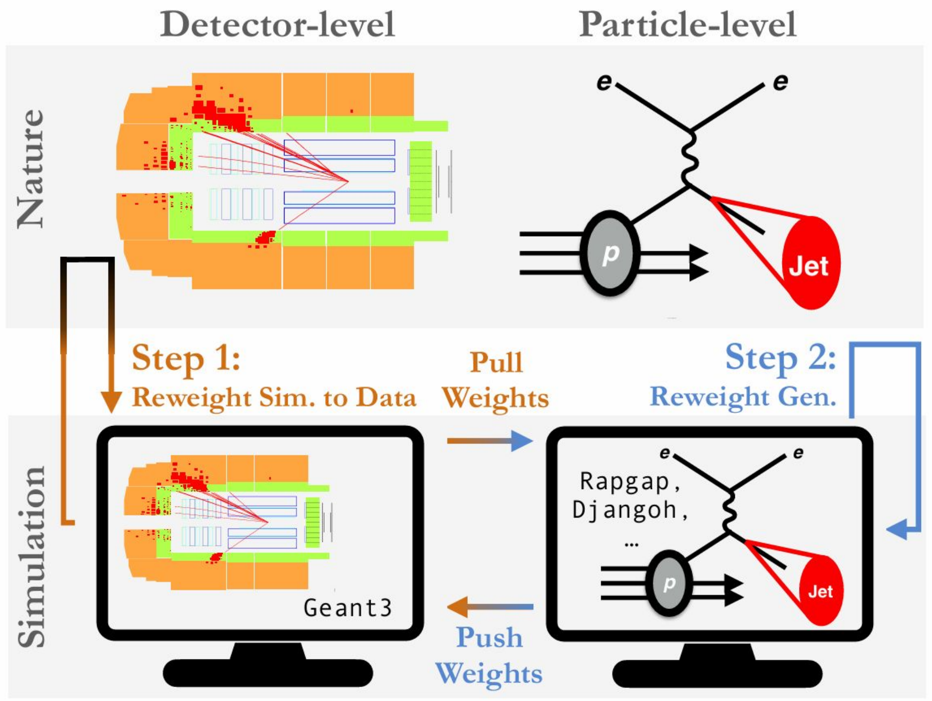
Using **228 pb^{-1}** of data collected by the **H1 Experiment** during **2006** and **2007** at **318 GeV center-of-mass energy**



Reconstructed hadrons using combined detector information: **energy flow algorithm**

OmniFold

Andreassen et al. PRL 124, 182001 (2020).
See also [2503.09720] Tools for Unbinned Unfolding



2-step iterative process

- **Step 1**: Reweight simulations to look like data
- **Step 2**: Convert learned weights into functions of particle level objects
- Use **classifiers** to learn the **reweighting** functions!

New manuscript:

Machine Learning-Assisted Measurement of Lepton-Jet Azimuthal Angular Asymmetries in Deep-Inelastic Scattering at HERA

H1 Collaboration • V. Andreev (Higher Sch. of Economics, Moscow) Show All(143)

Dec 18, 2024

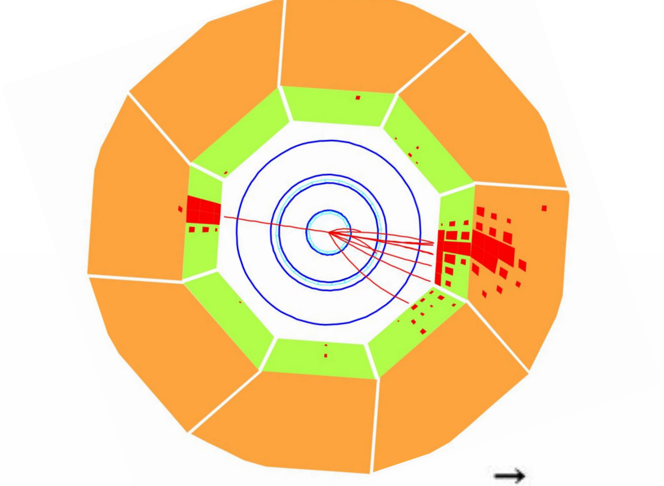
15 pages

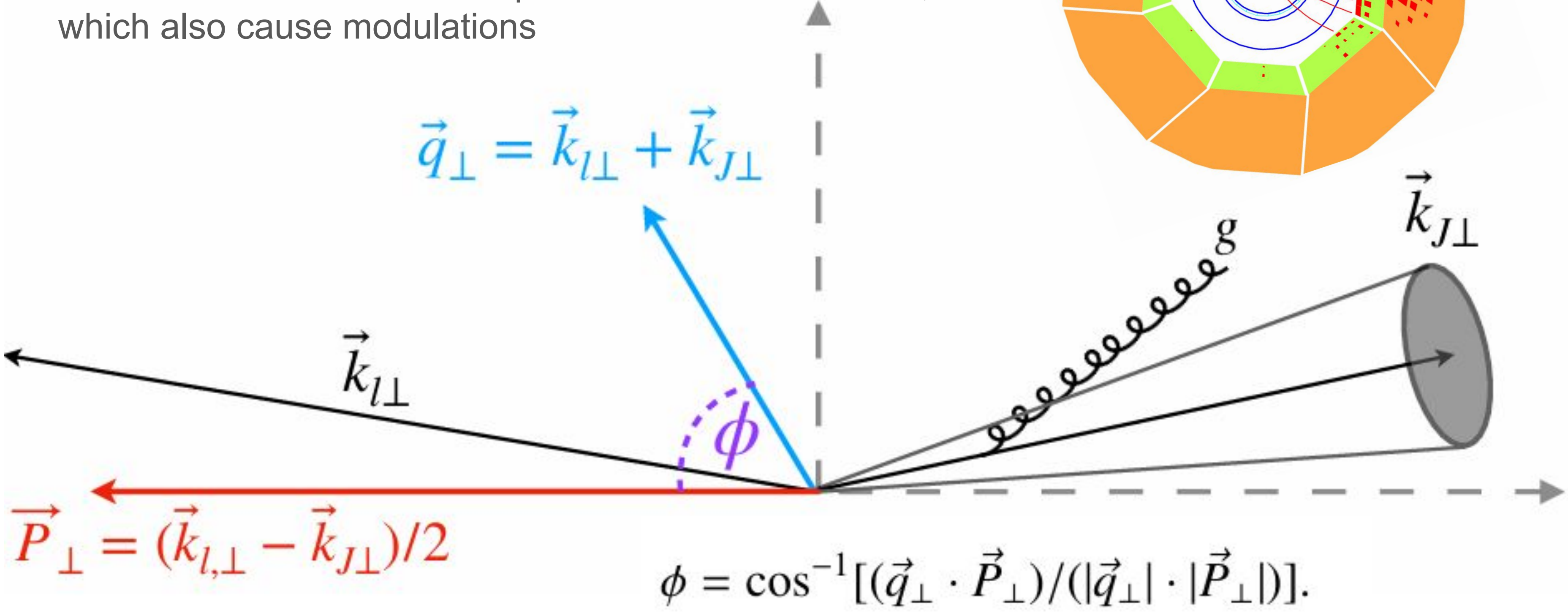
e-Print: 2412.14092 [hep-ex]

Report number: DESY24-200

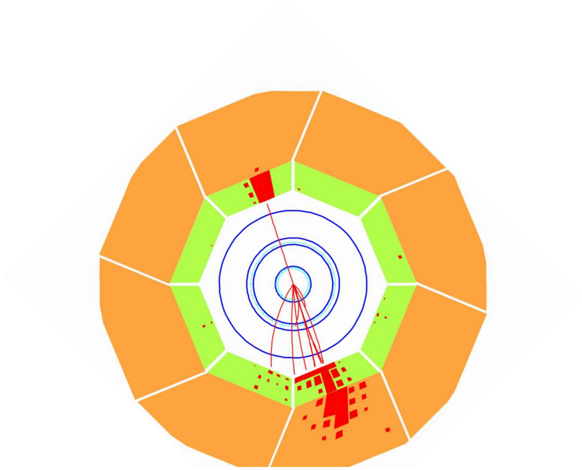
Motivation for Lepton-jet azimuthal modulation

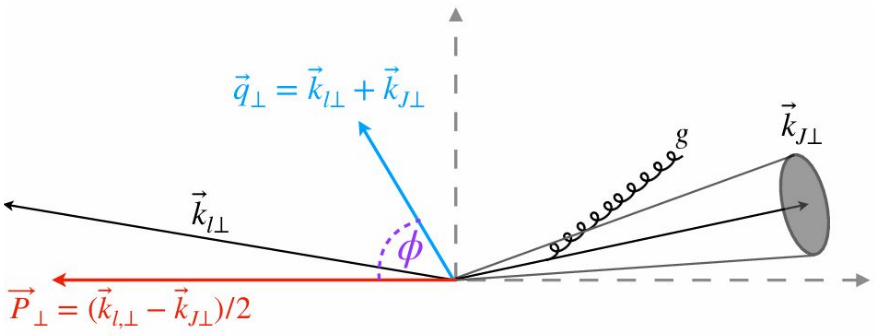
- Sensitive to out-of-cone soft-gluon radiation, which is calculable in pQCD.
- Crucial reference for future spin measurements at EIC, which also cause modulations





Reusing unfolding as in previous lepton-jet paper PRL 128 (2022) 132002)





8D input features use as input to unfolding.

Unbinned unfolding, so new observable can be computed (in this case, azimuthal angle of qT vector).
→ Showcases versatility of Omnifold method.

Fiducial cuts:

$Q^2 > 150\ GeV^2$

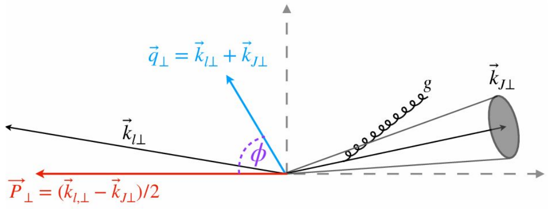
$0.2 < y < 0.7$

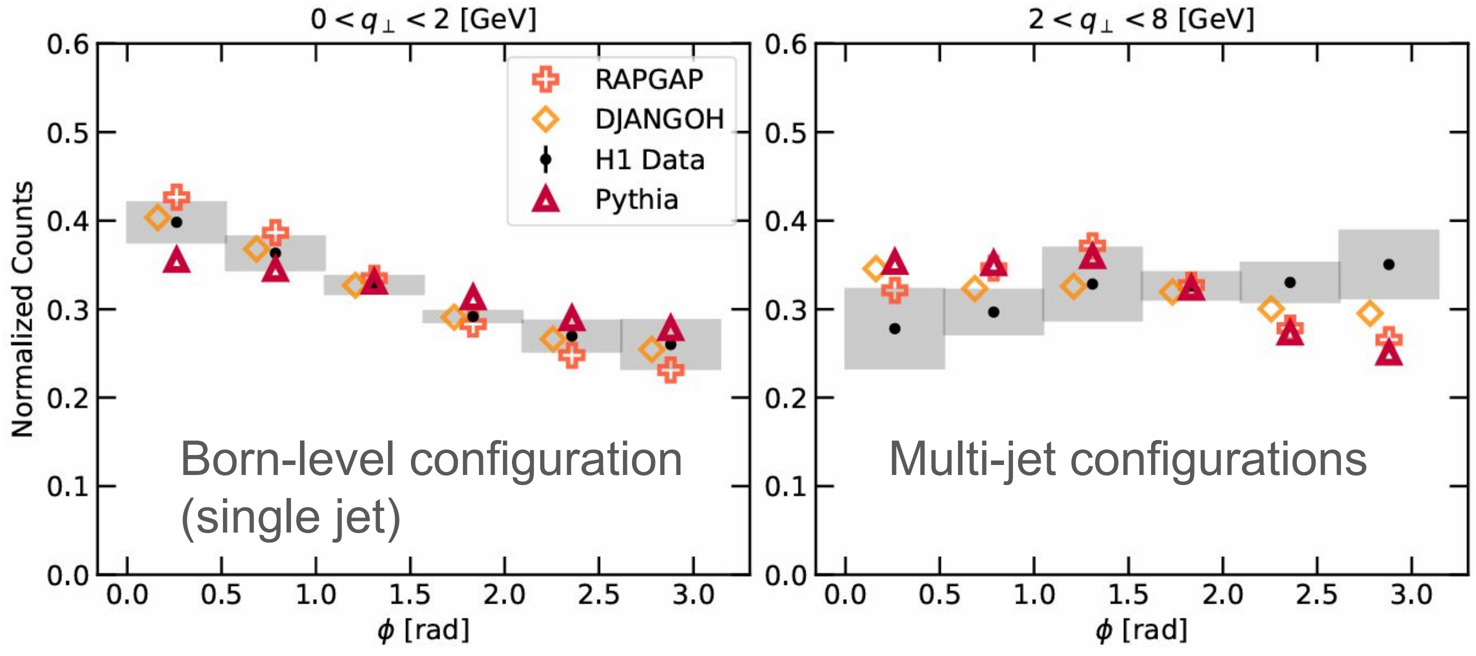
$p_T^{jet} > 10\ GeV$

$-1 < \eta^{jet} < 2.5$

$k_T, R = 1.0$

Azimuthal modulation

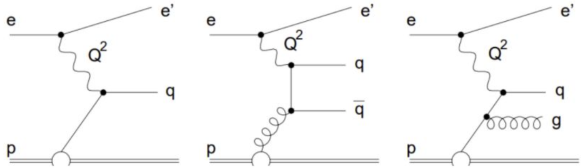




- Simple cosine at low qT, relatively well described by MC models
- Higher harmonics at higher qT, not well described by MC models.

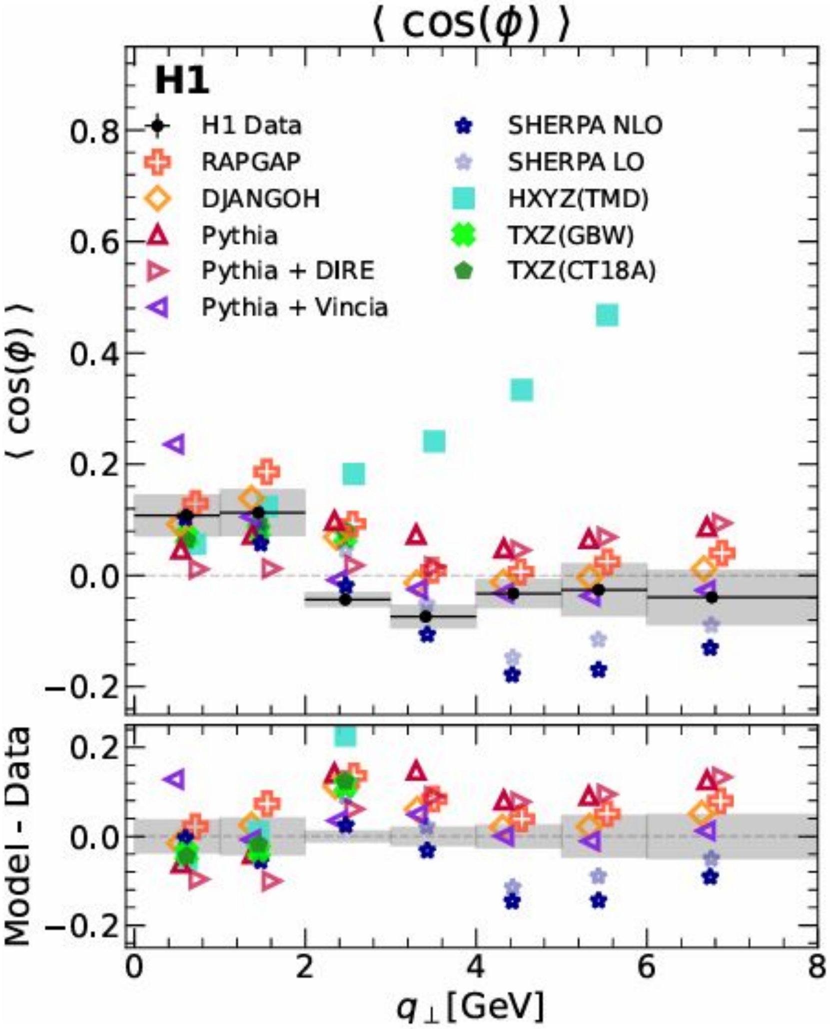
Fourier Analysis, 1st moment

First harmonic positive at low qT (single-jet), and drops to negative at higher qT (multijets)



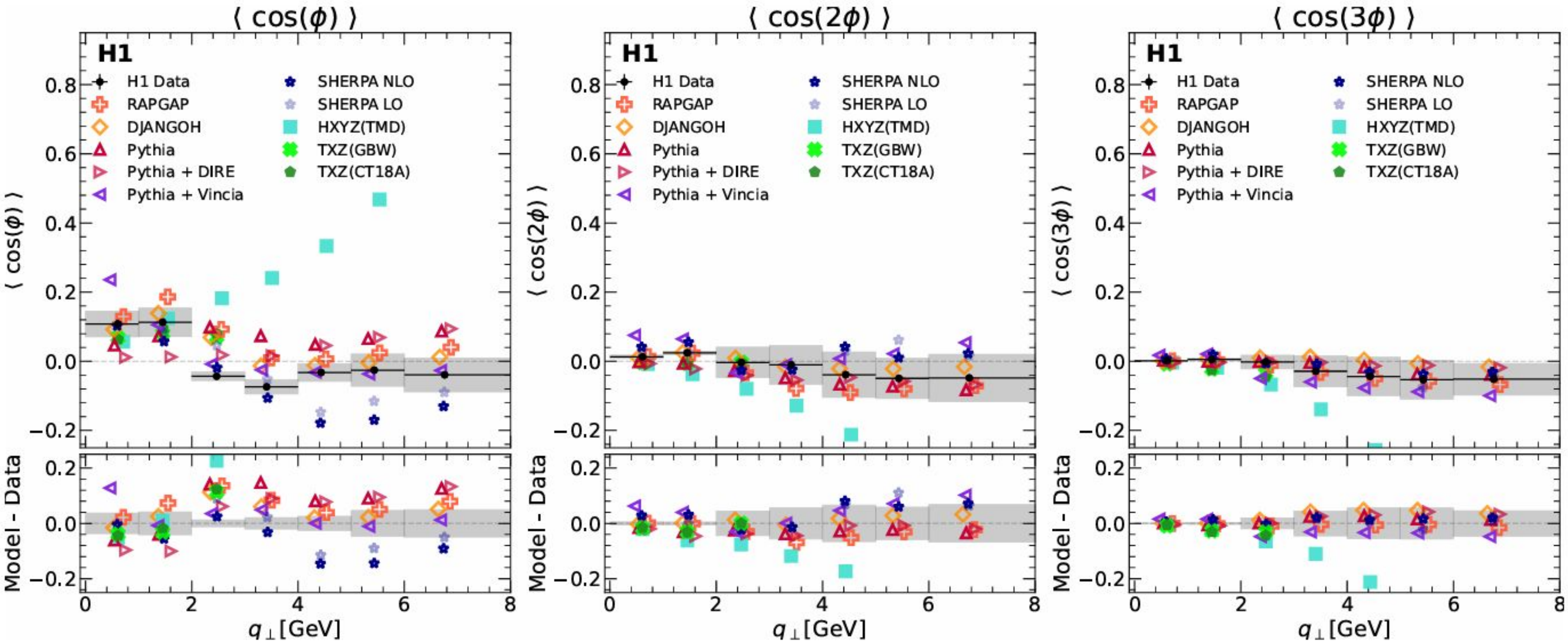
Given its accuracy, this measurement has large sensitivity to parton shower treatment and MC generator model.

Analytical calculations describe the data only at low qT, fail miserably at higher qT. Clearly missing higher-order treatment.



Fourier Analysis,

Strong discriminating power for theory calculations, MC parton showers, etc.



New Preliminary Result (for DIS2025!)

Towards Unfolding All Particles in High Q^2 DIS Events

H1 Collaboration

Abstract

The differential cross section with respect to all final state particle momenta and electric charge in high Q^2 events is measured in deep-inelastic positron-proton scattering using data collected with the H1 detector of HERA. The unbinned and full phase-space unfolding is implemented using the OmniFold machine learning algorithm, with the point-edge transformer neural network architecture. To illustrate the utility of this measurement, we present a number of projections from the full phase space, including observables previously measured by H1 and new observables that are challenging using conventional methods. For example, we show a simultaneous measurement of jets in the laboratory and Breit frames. The data are corrected for detector acceptance, efficiency, and resolution effects and uncertainties are estimated at the per-particle level and we are working towards releasing the result in an unbinned format.

Phase space definition, particle input to unfolding

The inputs to the unfolding are:

– Per-particles features: $[p_T, \eta, \phi, C]$

– Global DIS quantities: $[Q^2, y, p^e_x, p^e_y, p^e_z]$,

Fiducial Phase space definition:

- $0.2 < y < 0.7$
- $Q^2 > 150$ GeV^2

Particle selection:

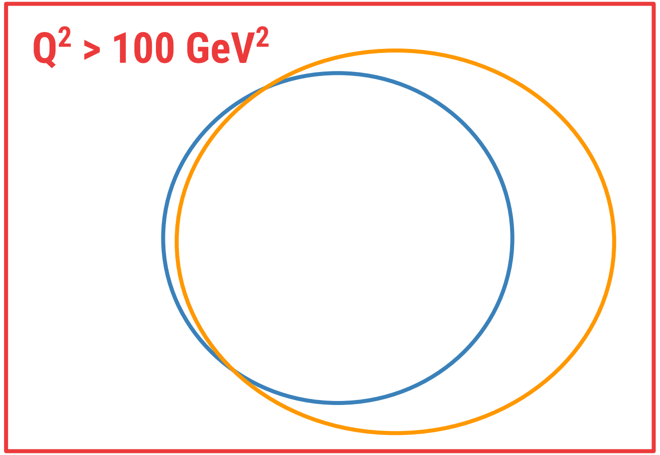
- $p_T > 0.1$ GeV
- $-1 < \eta_{lab} < 2.75$
- Charge information used if $\eta_{lab} < 2$

Reco Phase space definition:

- $0.08 < y < 0.7$
- $Q^2 > 150$ GeV^2
- p_T miss < 10 GeV,
- $45 < em/p_z < 65$

Particle selection:

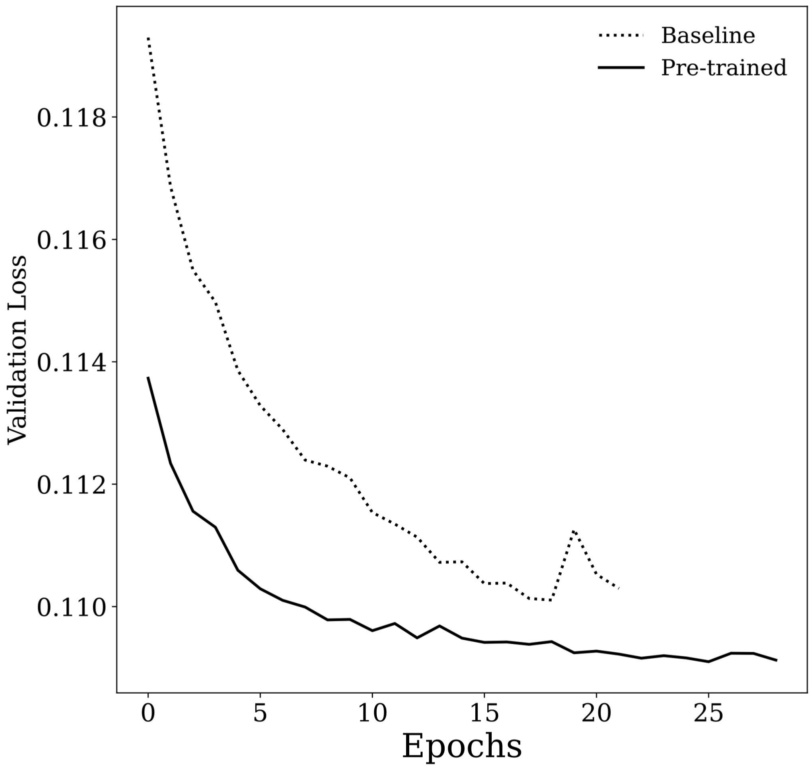
- $p_T > 0.1$ GeV
- $-1 < \eta_{lab} < 2.75$



Red box used during unfolding, but only fiducial results shown

- Pass reco selection: **Red -> Orange**: **77%**
- Pass fiducial selection: **Red -> Blue**: **58%**
- Pass fiducial and reco selection: **Blue -> Orange**: **96%**
- Don't pass fiducial but pass reco: **Red -> Orange** (without **blue**): **50%**

Pretraining for Unfolding



Pre-training:

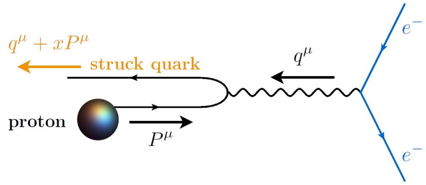
- We use a smaller version of the OmniLearn model[1]
- We have around 20M simulated samples in Djangoh and Rapgap, but only around 200k data events
- We can improve the reweighting quality by **pre-training** the model first
- Train the model to **classify Rapgap from Djangoh** and use that as the initialization for the rest of the unfolding

1: **V. Mikuni**, B. Nachman, arXiv:2404.16091

Results

	Lab k_T	Breit k_T	Centauro
$\Delta\phi^{\mathrm{jet}}$	Y	N	N
$\ln(\lambda_1^1)$	Y	N	N
p_T^{jet}	Y	Y	Y
z^{jet}	Y	Y	Y

Breit Frame



Jet Selections

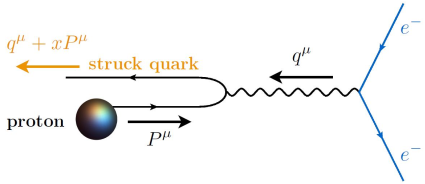
- **Jet algorithm:** kT and Centauro both with R = 1
- **Jet pT cuts:**
  - Lab frame kT jets: $p_T > 10$ GeV
  - Breit frame kT jets: $p_T > 5$ GeV
  - Centauro jets: $z > 0.2$ (no pT cut)

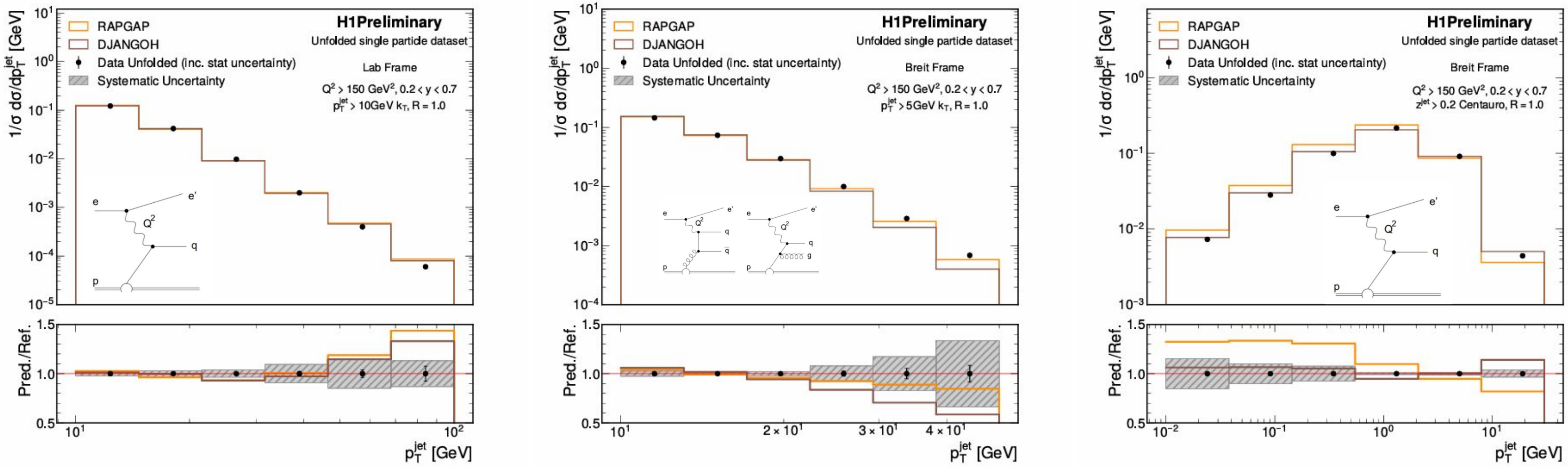
Centauro jet algorithm is a longitudinally invariant method designed for DIS studies

Energy-Energy Correlator is also measured in the Breit frame.

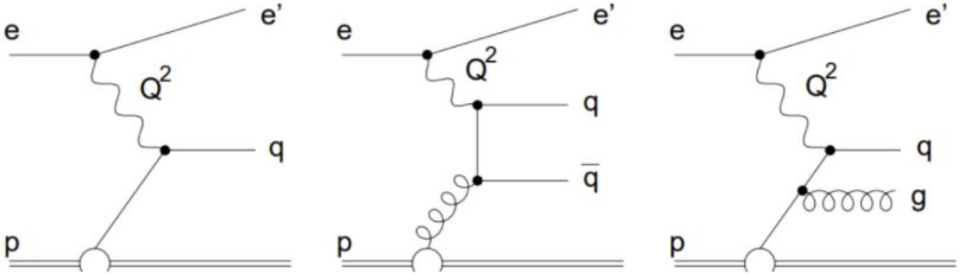
Asymmetric jet clustering in deep-inelastic scattering:
Arratia et al. Phys.Rev.D 104 (2021) 3, 034005

Simultaneous measurement of jet transverse momentum spectra in lab and Breit-frame with anti-kT and Centauro

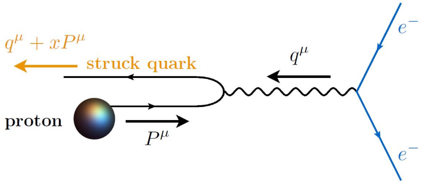


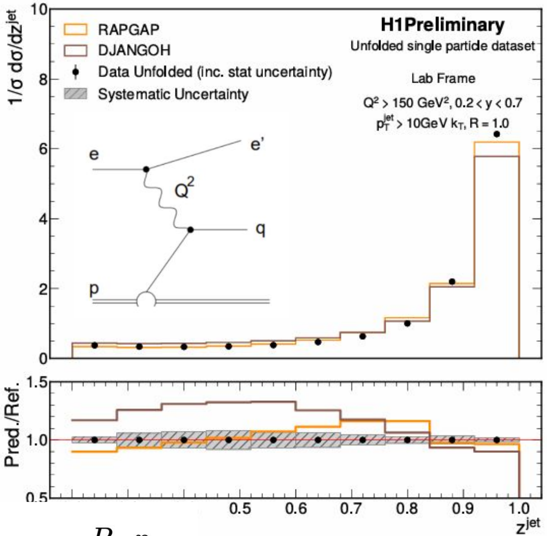


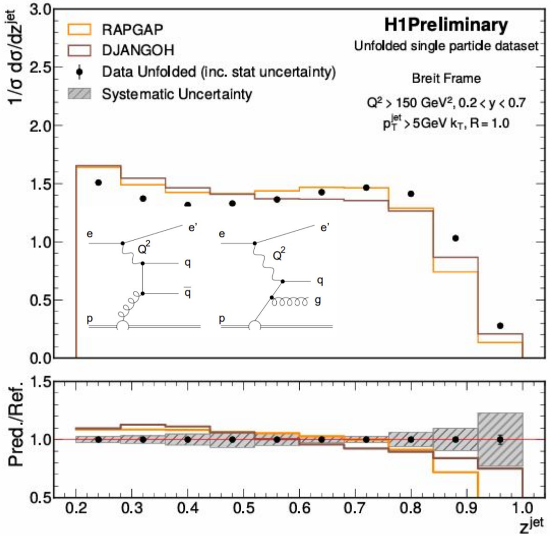
Lab frame and algorithm/pT threshold select different subprocesses

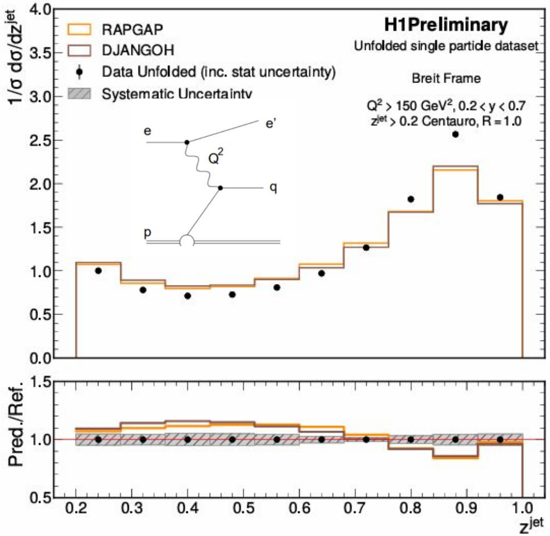


Simultaneous measurement of jet energy fraction in lab and Breit-frame with anti-kT and Centauro



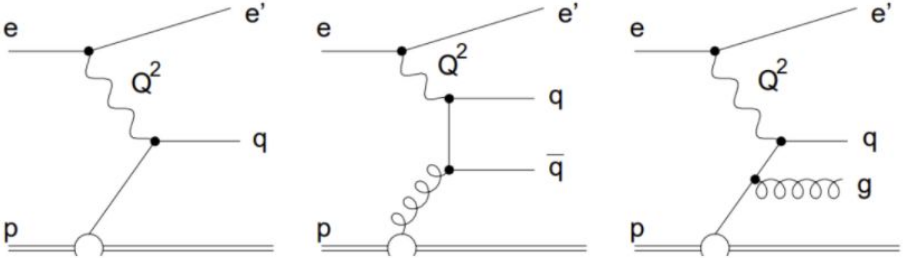






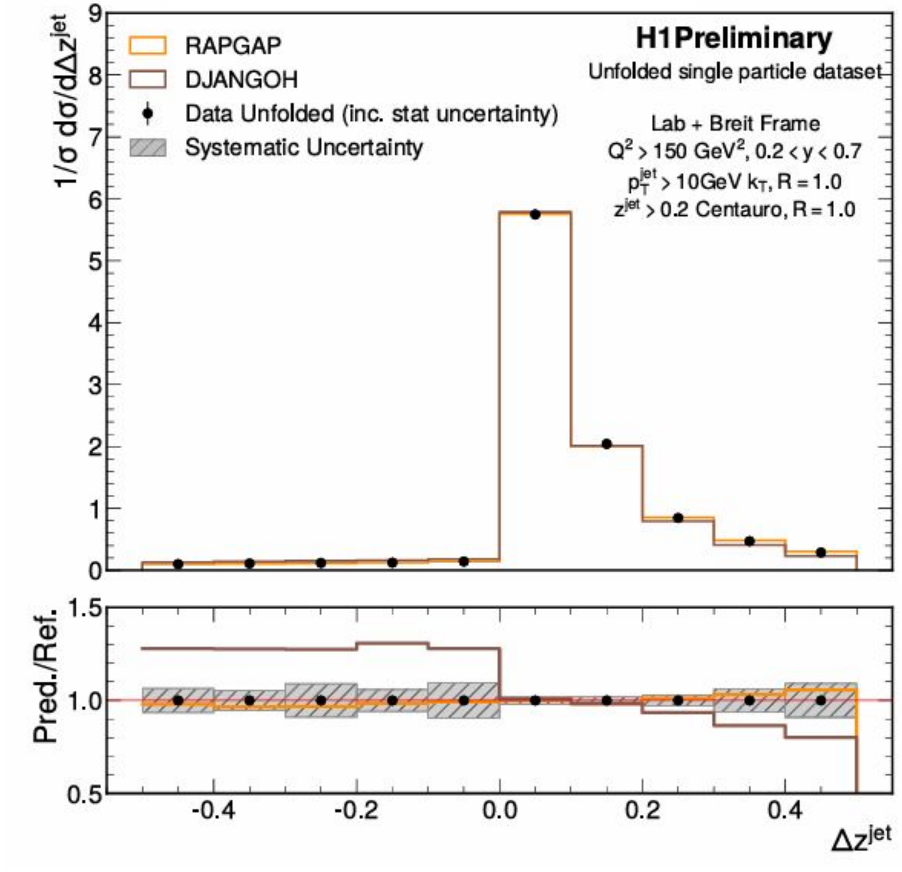
$$z_{\text{jet}} = \frac{P \cdot p_{\text{jet}}}{P \cdot q}$$

Lab frame and algorithm/pT threshold select different subprocesses

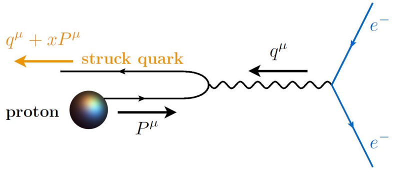


Difference between lab and breit frame z

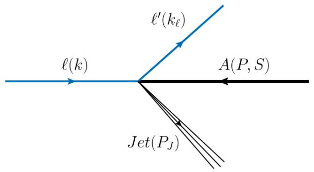
$$\Delta z^{\text{jet}} = z^{\text{leading jet}}_{\text{lab},\, k_T} - z^{\text{leading jet}}_{\text{Breit, Centauro}},$$



Breit Frame,
Centauro algorithm



Laboratory frame,
anti-kT algorithm



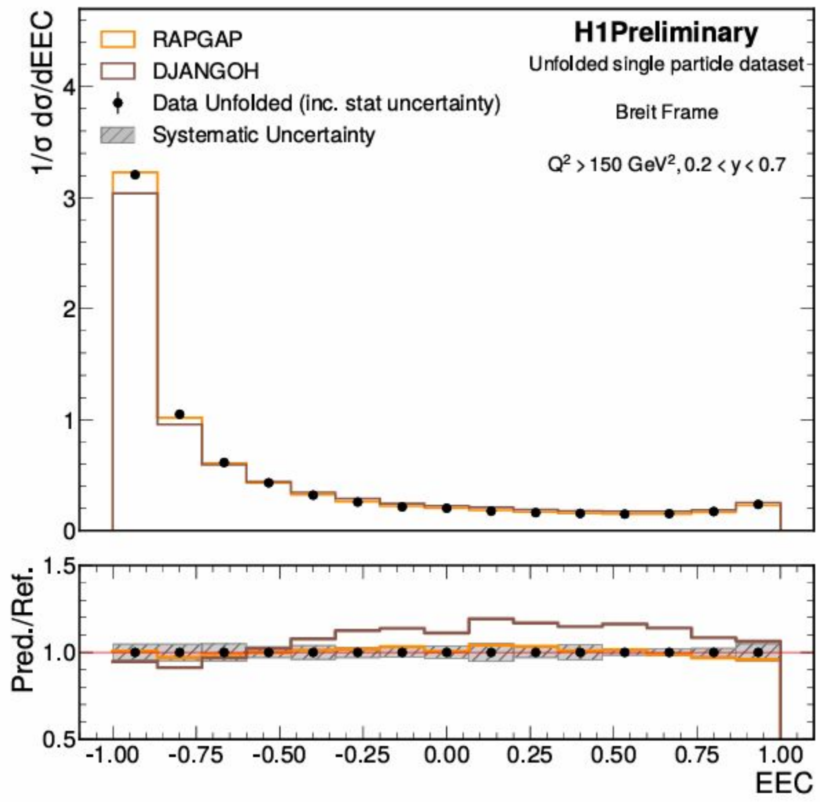
- Either case should be selecting Born-level configuration. Naively (“at leading log”) the z distribution should be the similar. **It is not. Reasons unclear.**
- At least one factor is QED radiation that smears boost to Breit frame (QED is not yet accounted for in this preliminary result)

Energy-energy correlator

$$\mathrm{EEC_{DIS}} = \sum_a \int \frac{d\sigma_{ep\to e+a+X}}{\sigma} z_a \, \delta(\cos\theta_{ap} - \cos\theta),$$

$$z_a \equiv \frac{P \cdot p_q}{P \cdot (\sum_i p_i)},$$

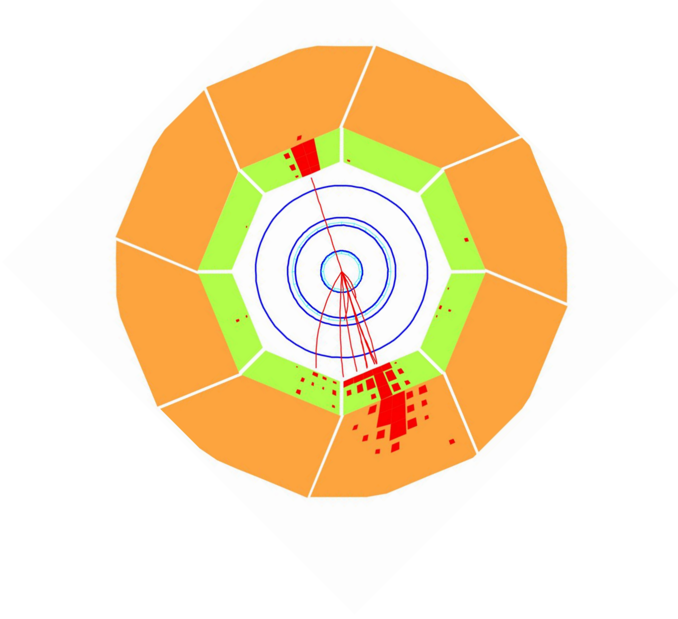
Precise and accurate measurement, complements other recent measurements in pp and AA.



Summary conclusions

- Measurement of azimuthal modulation of lepton-jet asymmetries reveal non-zero asymmetries. Fourier analysis shows data not fully described by any MC or theory calculation→ **Important reference to refine theory for future studies of hadron structure at HERA and EIC**

- Measurement of single-particle unfolding demonstrates versatility beyond traditional unfolding methods: simultaneous measurements in lab and Breit frame, and different jet algorithms, plus EEC and other observables.
→ **Novel studies reveal strong sub-process DIS selection. This handle can be exploited e.g. for improved quark/gluon tagging and TMD studies.**



Backup

Systematic uncertainties

Systematic uncertainties included in the results

- **HFS energy scale:** +- 1%
- **HFS azimuthal angle:** +- 20 mrad
- **Lepton energy:** +- 0.5%
- **Lepton azimuthal angle:** +- 1 mrad
- **Model uncertainty:** differences in unfolded results between Djangoh and Rapgap
- **Non-closure uncertainty:** Differences between the expected and obtained values of the closure test
- **Statistical uncertainty:** Standard deviation of 50 bootstrap samples with replacement (will increase to 100 when done running)

Uncertainties not yet added

- **QED uncertainty**: Use the variation of measured quantities when radiation is turned off in the simulation