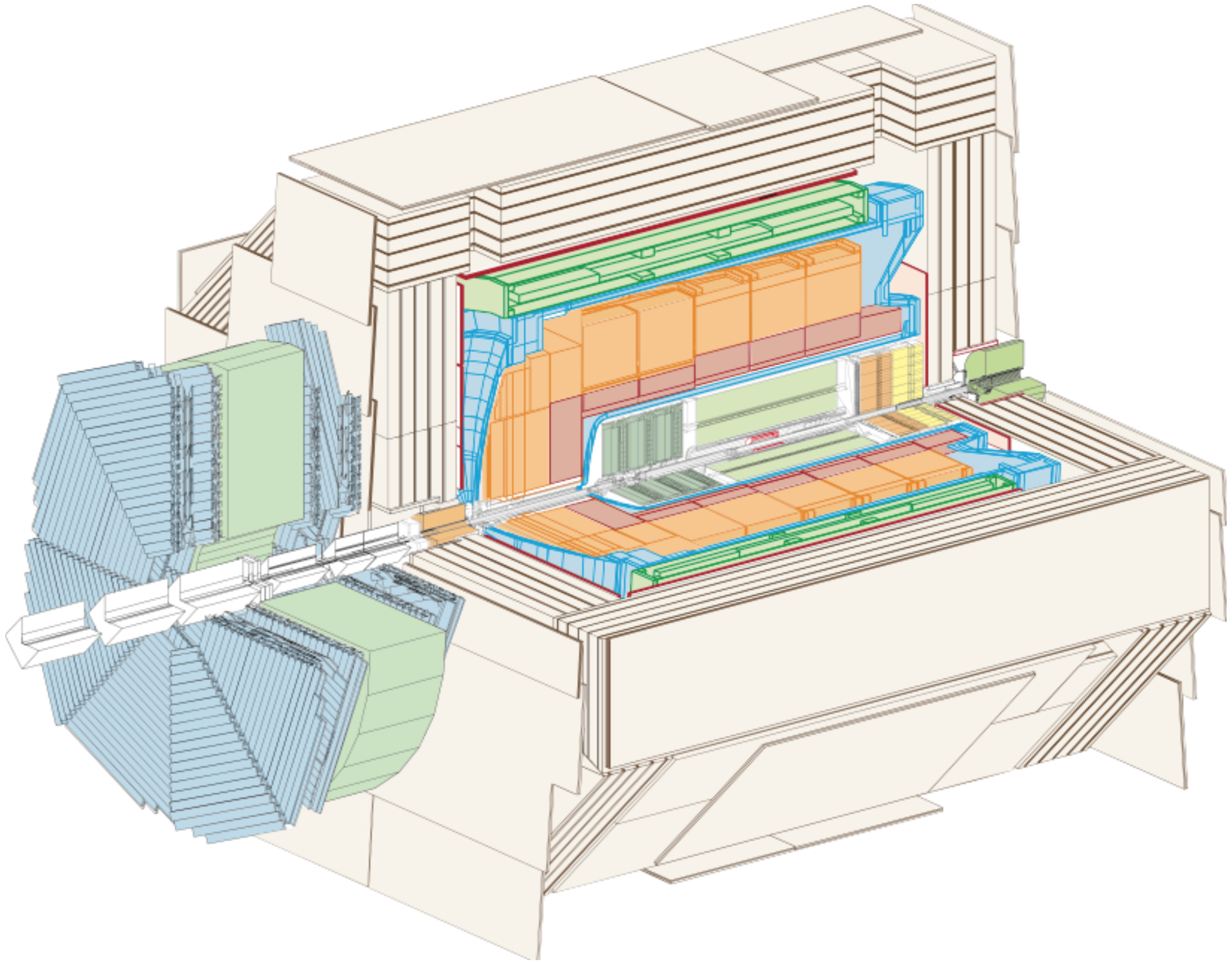
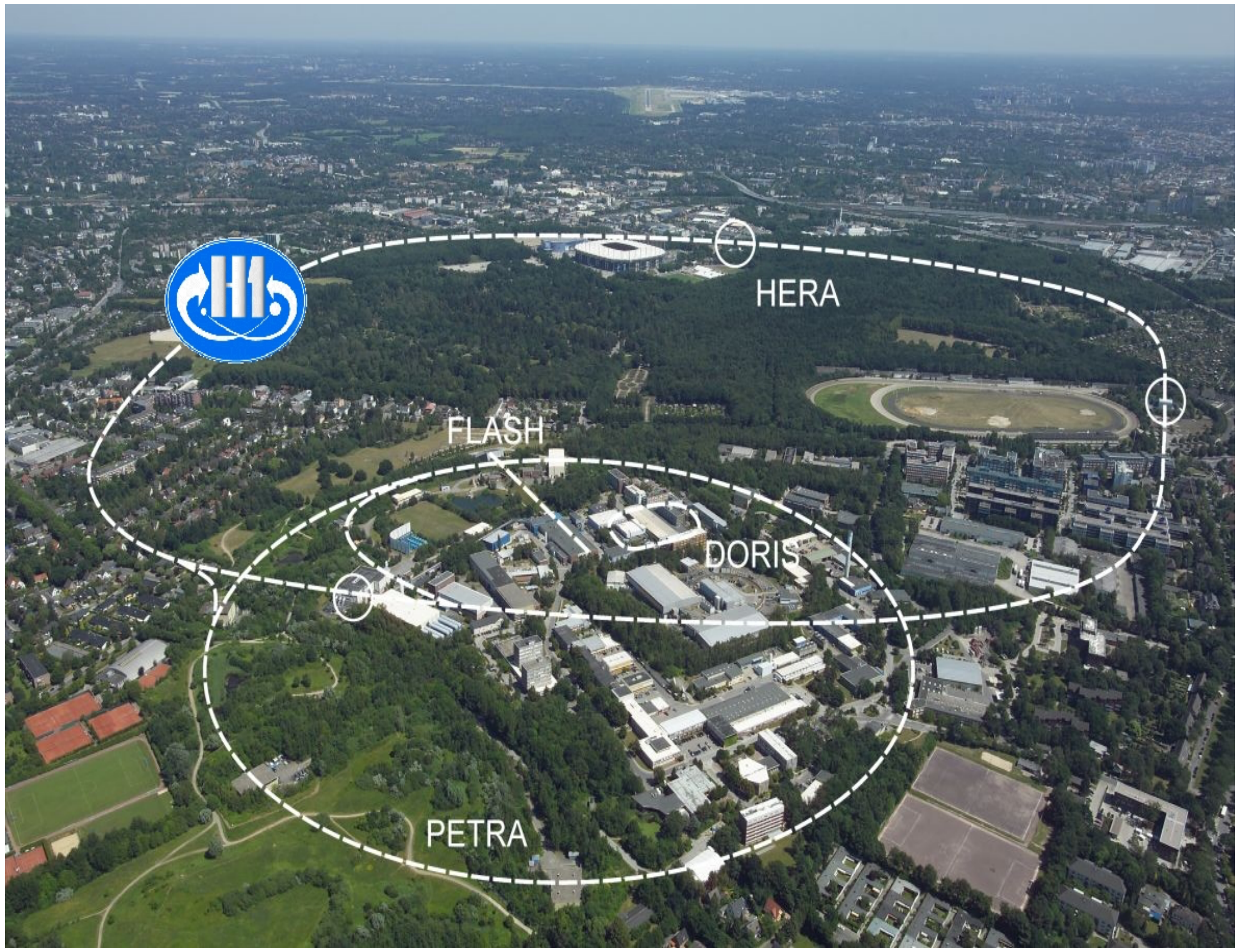
Unfolded Measurements With H1 using MultiFold

Fernando Torales Acosta
Benjamin Nachman

on behalf of the H1 Collaboration

H1 at HERA





- **H1 Detector at the positron-proton collider, HERA. Hosted in Hamburg Germany**
- **Major goal was to study internal structure of the proton through deep inelastic scattering**

$$e(k) + q(p_1) \rightarrow e'(k_\ell) + jet(k_J) + X$$

Quick Overview

- Lots of discussion on unfolding multi-differential cross sections
  - A bit more!
- Let's also look at a previously inaccessible observable
  - Moments requiring Un-binned unfolding

H1 Data

- Same data / selection / unfolding as arXiv:2108.12376
  - “Measurement of lepton-jet correlation in deep-inelastic scattering with the H1 detector using machine learning for unfolding”

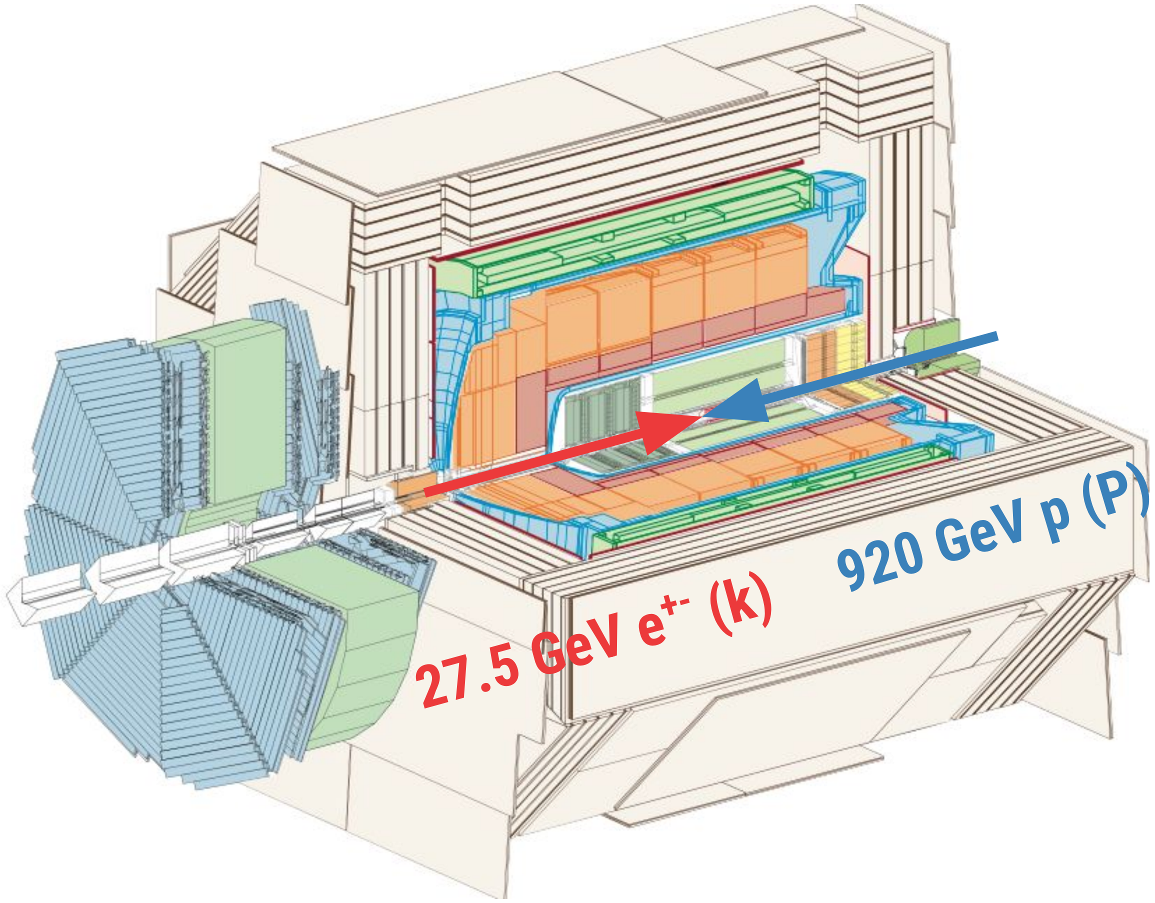
- H1 Data from 2006 and 2007 periods at 130 pb^{-1}
  - $\sqrt{s} = 320$ Gev, eP collisions

- Fiducial Cuts:
  - $0.2 < y < 0.7$
  - $Q^2 > 150\ \mathrm{GeV}^2$
  - $p_T^{\mathrm{jet}} > 10$ GeV
  - $-1 < \eta_{\mathrm{lab}} < 2.5$
  - $\mathrm{k_T}, R = 1.0$
  - $q_\perp / p_{\mathrm{T},jet} < 0.3$

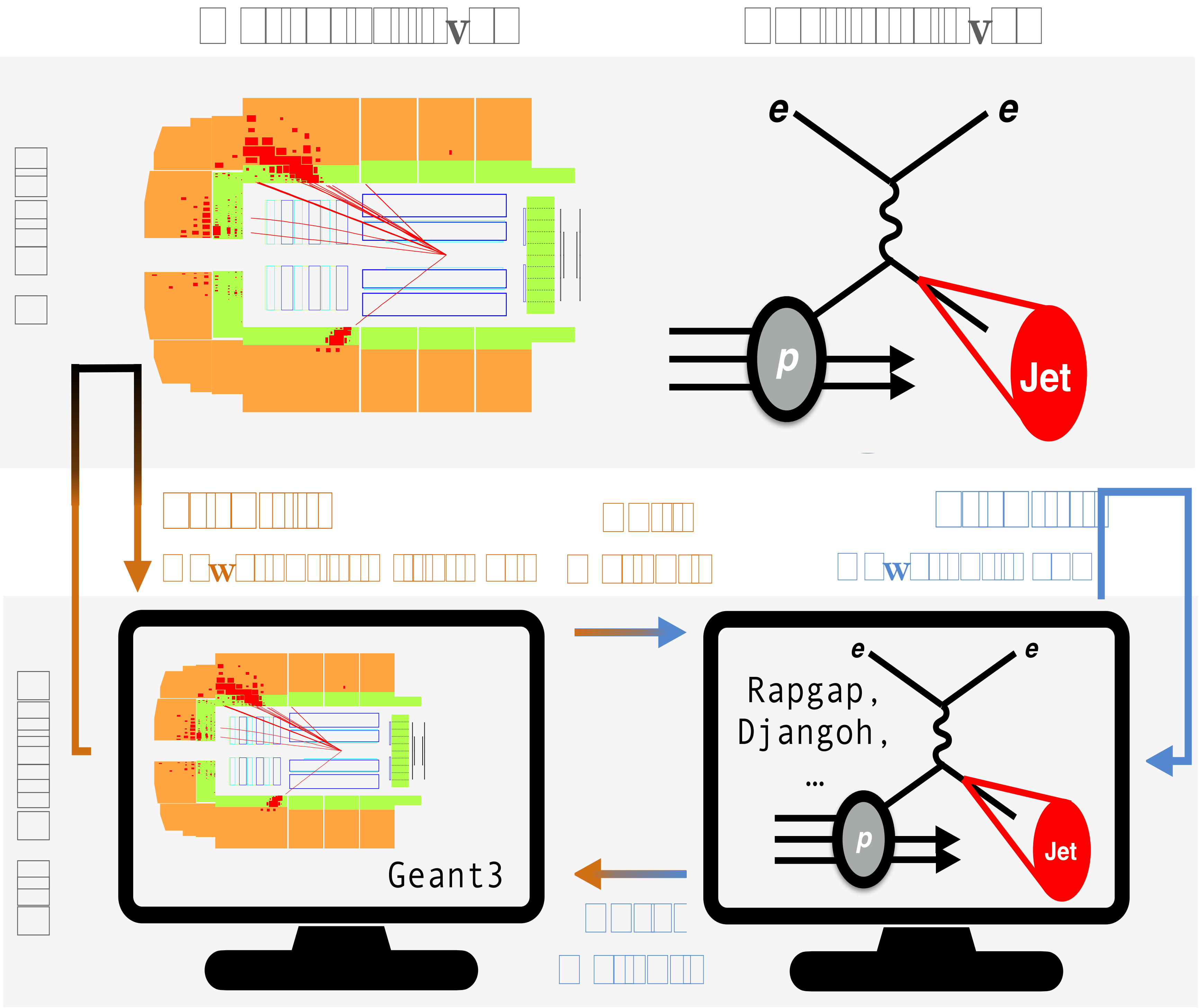
For asymmetry Measurement:

Cut on $q_\perp / p_{\mathrm{T},jet}$ to satisfy $P_\perp \gg q_\perp$:

$$p_{\mathrm{T,jet}} \approx P_\perp / 2$$



OmniFold



2 step iterative approach

1. Events from detector level sim. are reweighted to match the data
2. Create a "new simulation" by transforming weights to a proper function of the generated events

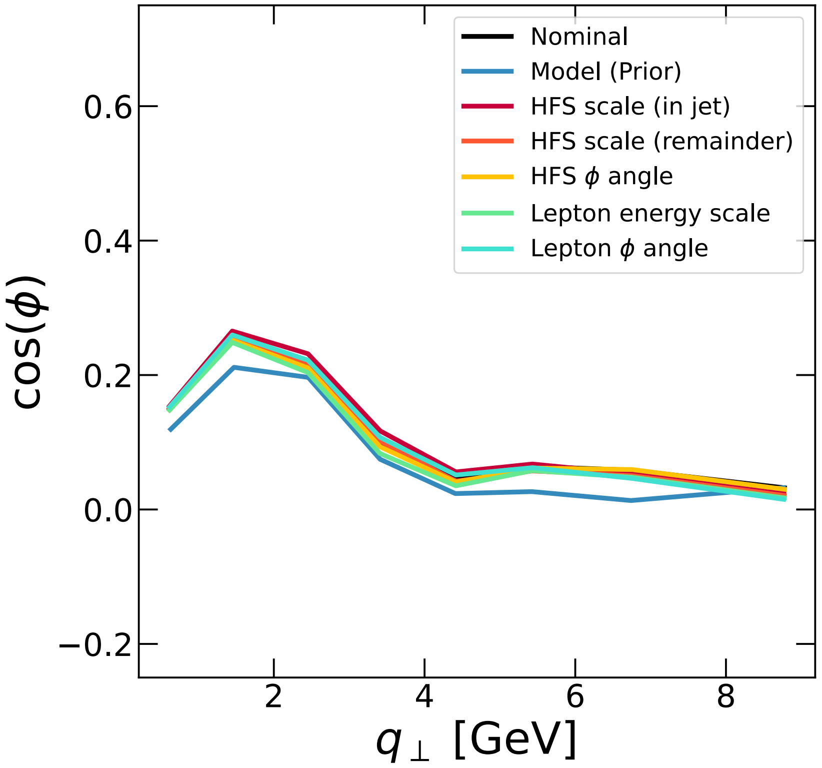
Classifiers used to approximate **2** likelihood functions:

1. reco MC to Data reweighting
2. **Previous** and **new Gen** reweighting

Uncertainties

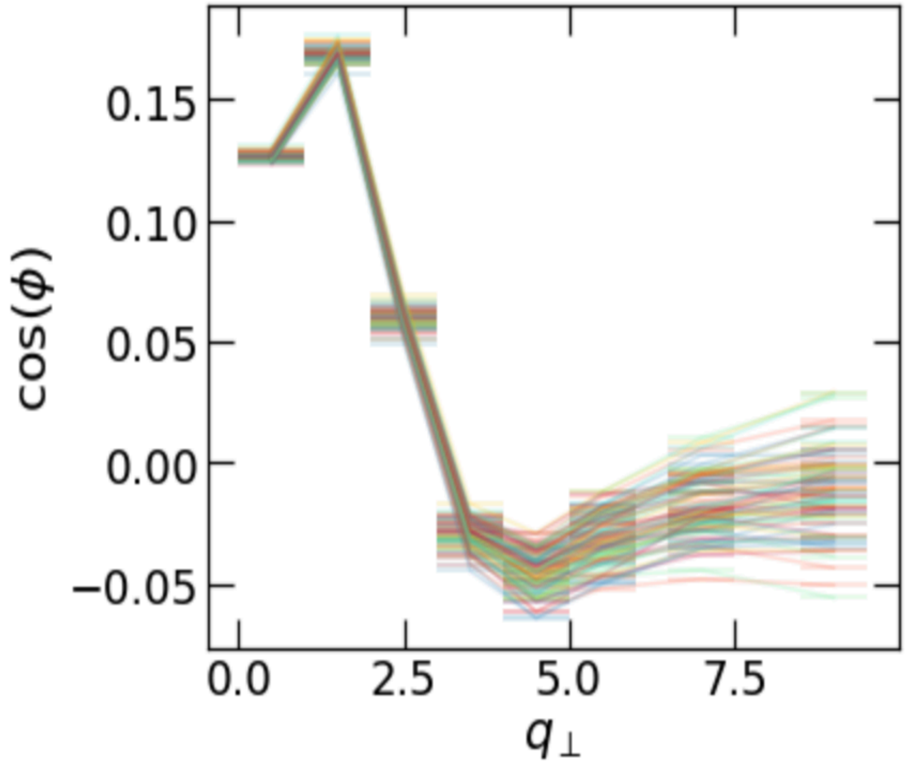
Systematic uncertainties

- **HFS energy scale:** +- 1%
- **HFS azimuthal angle:** +- 20 mrad
- **Lepton energy:** +- 0.5% (mainly affects Q^2)
- **Lepton azimuthal angle:** +- 1 mrad (mainly affects Q^2)
- **Model uncertainty:** differences in unfolded results between Djangoh and Rapgap
- **QED uncertainty**: Use the variation of measured quantities when radiation is turned off in the simulation

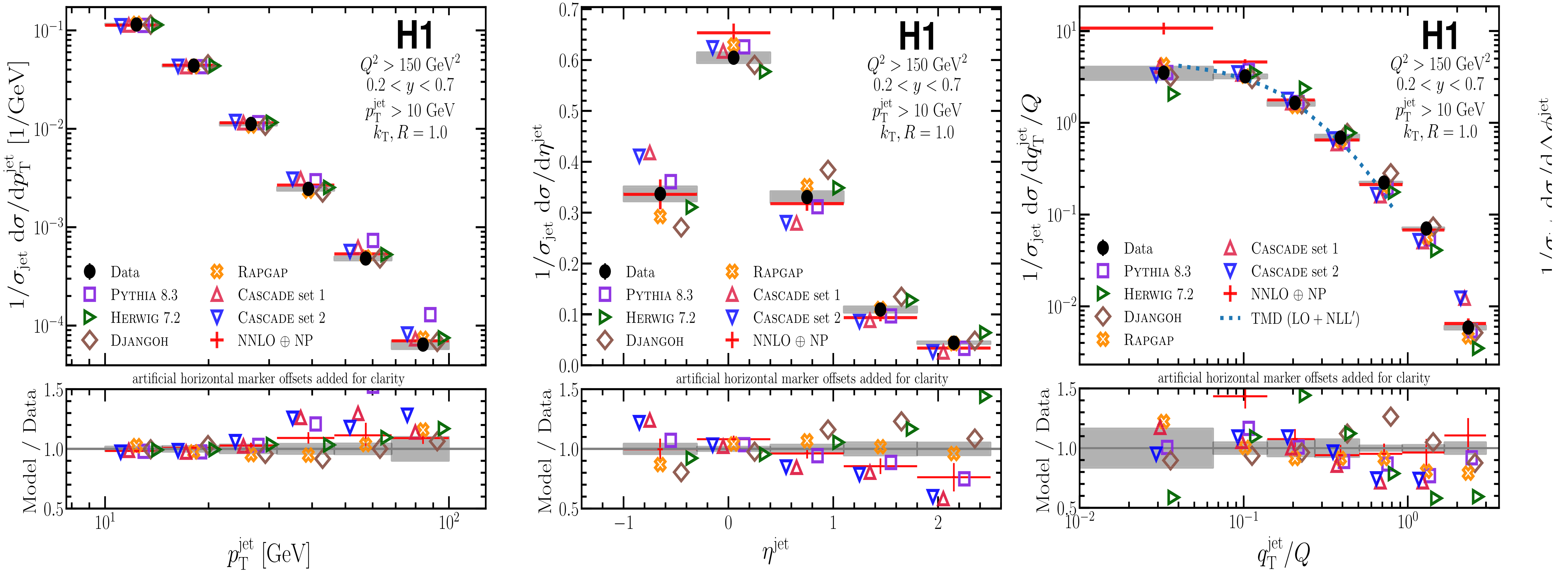


Statistical Uncertainty

- **Bootstrapping**
- Each event is given an new initial weight
- ~100 bootstraps
- Repeat entire unfolding process



H1 Differential Cross sections

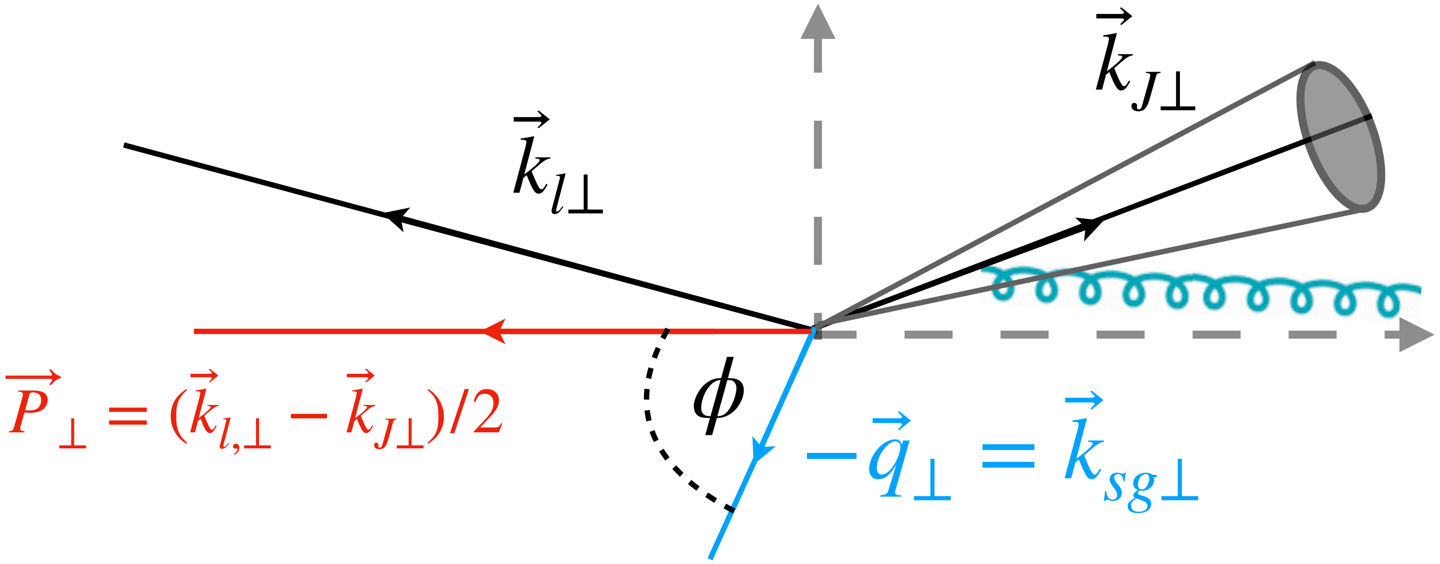


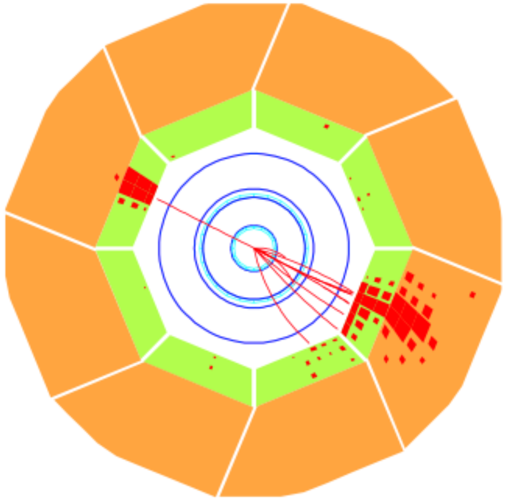
- Production cross sections as a function of p_T^{jet}, η^{jet}, and jet momentum balance q_T^{jet}/Q
- Simultaneous unfolding of eight observables:
    - $p_x^e, p_y^e, p_z^e, p_T^{jet}, \eta^{jet}, \phi^{jet}, \Delta\phi^{jet}, q_T^{jet}/Q$

Lepton Jet Asymmetry

- Total transverse momentum of the outgoing system $\vec{q}_\perp = \vec{k}_{\ell\perp} + \vec{k}_{J\perp}$, is typically *small* but *nonzero*
- Imbalance can come from perturbative initial and final state radiation
  - e.g. Emission of soft gluon with momentum $k_{\perp g}$
  - unrelated to TMDs or intrinsic transverse momentum of target gluons
- Depending on kinematics, soft gluon radiation can dominate
  - Radiative corrections enhanced approximately as $(\alpha_s \ln^2 P_\perp^2/q_\perp^2)^n$

$$P_\perp \gg q_\perp$$





$e(k) + q(p_1) \to e'(k_\ell) + jet(k_J) + X$

Lepton Jet Asymmetry

Key Ingredients:

- $q_\perp$ = ***Total*** **transverse momentum**

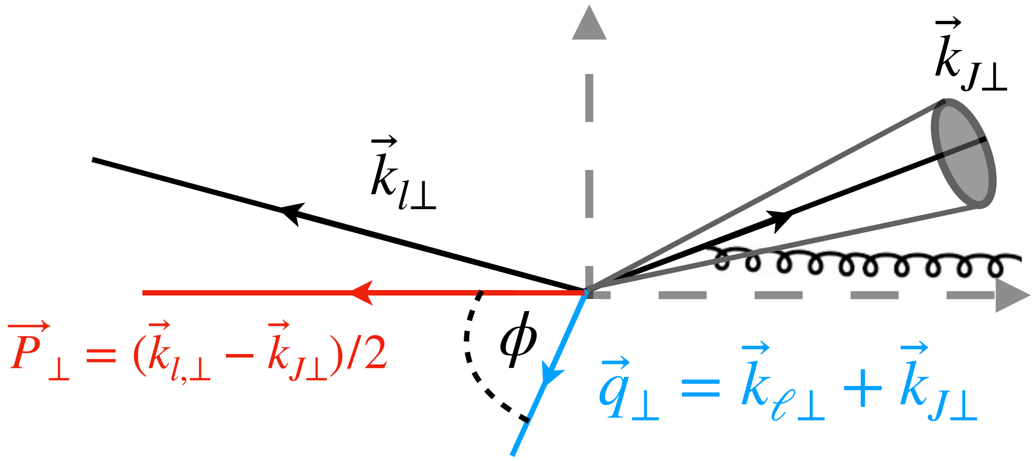
- $P_\perp$ = **Transverse momentum *difference***

- ϕ = Angle between $q_\perp$ and $P_\perp$

$$\vec{q}_\perp = \vec{k}_{\ell\perp} + \vec{k}_{J\perp}$$

$$\overrightarrow{P_\perp} = (\vec{k}_{\ell\perp} - \vec{k}_{J\perp}) \ / \ 2$$

$$\phi = \mathrm{acos}[(\vec{q}_\perp \cdot \overrightarrow{P_\perp}) \ / \quad \vec{q}_\perp \quad \overrightarrow{P_\perp} \quad]$$



Final Observable:
$\langle \cos(n\phi) \rangle$ **for n = 1, 2, 3**

Multifold used to unfold:

$p_x^e,\ p_y^e,\ p_z^e,\ p_\mathrm{T}^\mathrm{jet},\ \eta^\mathrm{jet},\ \phi^\mathrm{jet},\ \Delta\phi^\mathrm{jet},\ q_\mathrm{T}^\mathrm{jet}/Q$

Momentum conservation:

$$\vec{q}_\perp = -\sum_i^{soft} \vec{k}_{i\perp}$$

Motivation

1. Probes soft gluon radiation $S(g)$
   - Soft gluon radiation can be the primary contribution to asymmetry
   - 10.1103/PhysRevD.104.054037

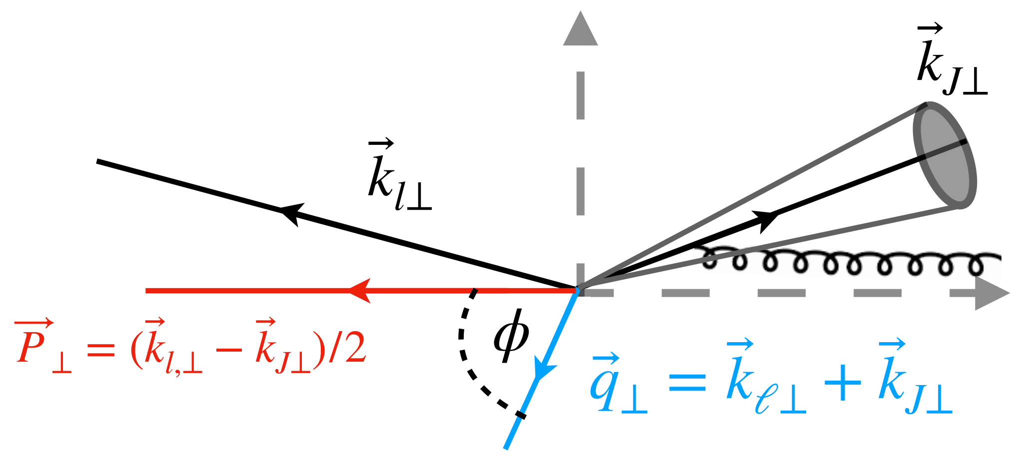
2. Asymmetry is perturbative
   - Opportunity to compare to unfolded H1 data

3. May represent a vital reference for other signals, in particular TMD PDF measurements
   - Factorize contributions TMD PDFs and Soft gluon radiation

4. Observable is sensitive to gluon saturation phenomena, possibly measurable at the EIC
   - 10.1103/PhysRevLett.130.151902

Putting it Together

$$\phi = \mathrm{acos}[(\vec{q}_\perp \cdot \overrightarrow{P_\perp}) \,/\, \vec{q}_\perp \quad \overrightarrow{P_\perp}\ ]$$



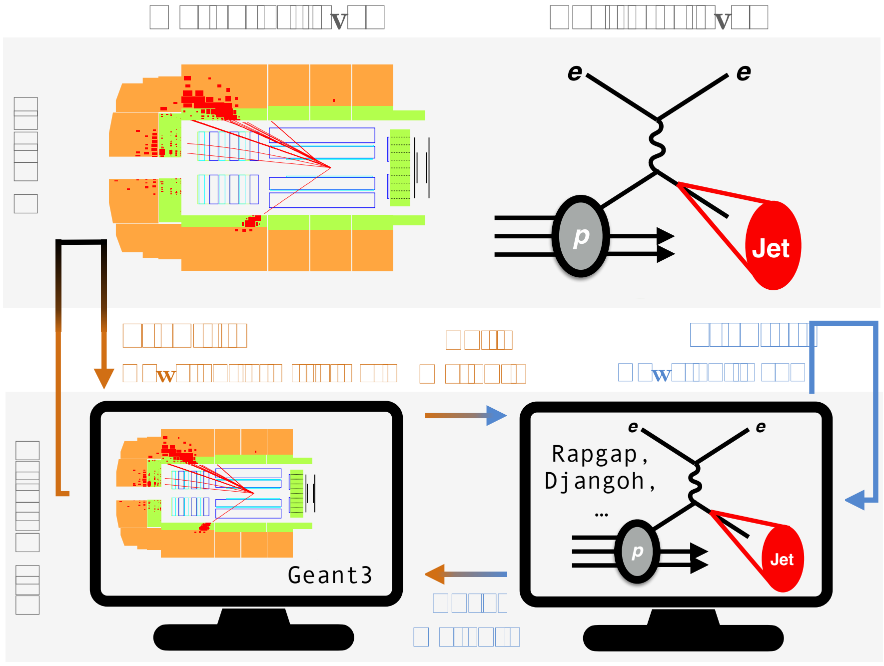
1. Obtain the azimuthal asymmetry angle, ϕ, in each event
2. Obtain unfolding event weight from MultiFold Step 2, ω_i, for each event, i

Calculate

$\langle\cos(n\phi)\rangle$ **for n = 1, 2, 3**

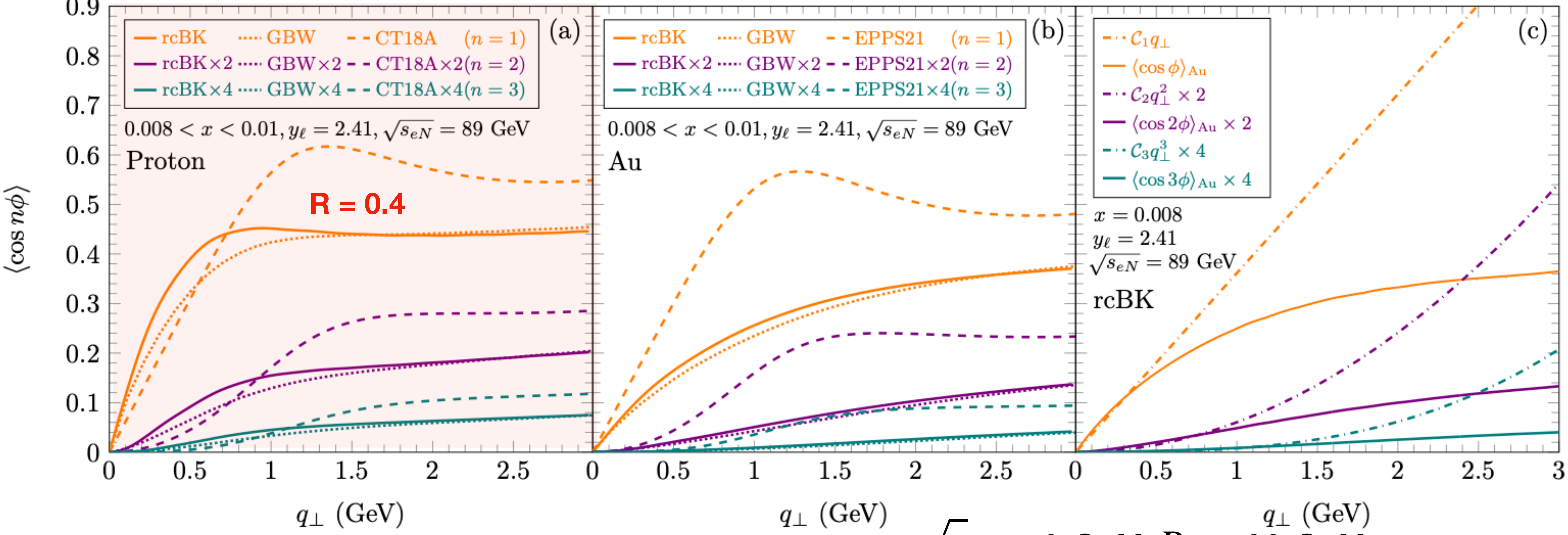
Report in bins of $\vec{q}_\perp$ GeV/c

$$\frac{\sum_i \omega_i \cos(n\phi_i)}{\sum_i \omega_i} \text{ for } n = 1,\ 2,\ 3$$



EIC Calculation @ HERA kinematics

arXiv: 2211.01647



$$\vec{q}_\perp = \vec{k}_{\ell\perp} + \vec{k}_{J\perp}$$

$$\overrightarrow{P_\perp} = (\vec{k}_{\ell\perp} - \vec{k}_{J\perp}) \;/\; 2$$

$\sqrt{s}$ =140 GeV, $P_\perp$ = 20 GeV,

y_l = 1.5, Q = 25 GeV

Radiative corrections

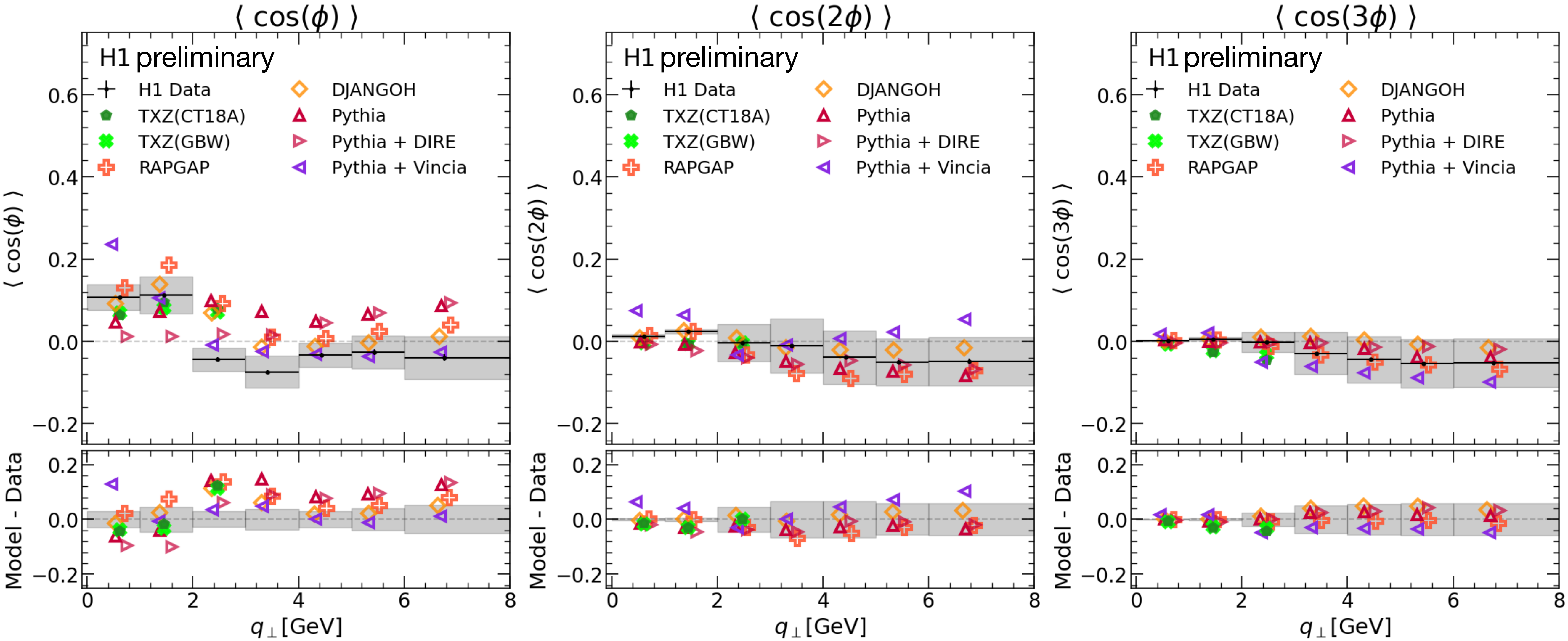
enhanced $\propto (\alpha_s \ln^2 P_\perp^2/q_\perp^2)^n$

Plots above are for R = 0.4. Calculation done for this measurement w/ R = 1.0,

Very good example of observable from ‘legacy’ dataset influencing future colliders

Harmonics of saturation with the inputs GBW model and a TMD calculation CT18A PDF

Moments of Asymmetry Results



- Three harmonics of the azimuthal angular asymmetry between the lepton and leading jet as a function of $q_\perp$.
- Predictions from multiple simulations as well as a pQCD calculation are shown for comparison.
- Measurement indicates calculation may break down near $q_\perp \approx 3$ GeV

Conclusions

- Promising measurement to probe soft gluon radiation
  - Important reference for lepton-jet DIS measurements!
  - Comparisons with pQCD calculations, and 3 generators
  - Theory has qualitatively similar shape, but *underestimates contribution*
  - May point to larger non-perturbative contributions to this observable

- MultiFold
  - This work presents a measurement of *moments*, requiring the *un-binned unfolding!*

- H1is a great example of exciting measurements using legacy datasets
  - First multidimensional un-binned unfolding using OmniFold
  - Novel observable with important implication for EIC
  - Simultaneous unfolding for Jet Substructure

https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.128.132002

https://doi.org/10.1016/j.physletb.2023.138101

Deutsches Elektronen-Synchrotron DESY

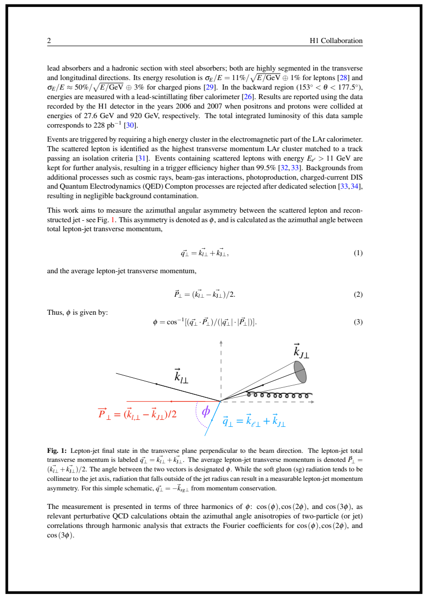
H1prelim-23-031, DIS 2023
March 14, 2024

Machine learning-assisted measurement of azimuthal angular asymmetries in deep-inelastic scattering with the H1 detector

H1 Collaboration

Abstract

Jet-lepton azimuthal asymmetry harmonics are measured in deep inelastic scattering data collected by the H1 detector using HERA Run II collisions. When the average transverse momentum of the lepton-jet system, $|\vec{P}_\perp|$, is much larger than the total transverse momentum of the system, $|\vec{q}_\perp|$, the asymmetry between $\vec{P}_\perp$ and $\vec{q}_\perp$ is expected to be generated by initial and final state soft gluon radiation and can be predicted using perturbation theory. Quantifying the angular properties of the asymmetry therefore provides a novel test of the strong force and is also an important background to constrain for future measurements of intrinsic asymmetries generated by the proton's constituents through Transverse Momentum Dependent (TMD) Parton Distribution Functions (PDF). Moments of the azimuthal asymmetries are measured using a machine learning technique that does not require binning and thus does not introduce discretization artifacts.



2 H1 Collaboration

lead absorbers and a hadronic section with steel absorbers; both are highly segmented in the transverse and longitudinal directions. Its energy resolution is $\sigma_E/E = 11\%/\sqrt{E/\mathrm{GeV}} \oplus 1\%$ for leptons [28] and $\sigma_E/E \approx 50\%/\sqrt{E/\mathrm{GeV}} \oplus 3\%$ for charged pions [29]. In the backward region ($153° < \theta < 177.5°$), energies are measured with a lead-scintillating fiber calorimeter [26]. Results are reported using the data recorded by the H1 detector in the years 2006 and 2007 when positrons and protons were collided at energies of 27.6 GeV and 920 GeV, respectively. The total integrated luminosity of this data sample corresponds to 228 pb^{-1} [30].

Events are triggered by requiring a high energy cluster in the electromagnetic part of the LAr calorimeter. The scattered lepton is identified as the highest transverse momentum LAr cluster matched to a track passing an isolation criteria [31]. Events containing scattered leptons with energy $E_{e'} > 11$ GeV are kept for further analysis, resulting in a trigger efficiency higher than 99.5% [32, 33]. Backgrounds from additional processes such as cosmic rays, beam-gas interactions, photoproduction, charged-current DIS and Quantum Electrodynamics (QED) Compton processes are rejected after dedicated selection [33, 34], resulting in negligible background contamination.

This work aims to measure the azimuthal angular asymmetry between the scattered lepton and reconstructed jet - see Fig. 1. This asymmetry is denoted as ϕ, and is calculated as the azimuthal angle between total lepton-jet transverse momentum,

$$\vec{q_\perp} = \vec{k_{l\perp}} + \vec{k_{J\perp}}, \tag{1}$$

and the average lepton-jet transverse momentum,

$$\vec{P}_\perp = (\vec{k_{l\perp}} - \vec{k_{J\perp}})/2. \tag{2}$$

Thus, ϕ is given by:

$$\phi = \cos^{-1}[(\vec{q_\perp} \cdot \vec{P_\perp})/(|\vec{q_\perp}| \cdot |\vec{P_\perp}|)]. \tag{3}$$

Fig. 1: Lepton-jet final state in the transverse plane perpendicular to the beam direction. The lepton-jet total transverse momentum is labeled $\vec{q_\perp} = \vec{k_{l\perp}} + \vec{k_{J\perp}}$. The average lepton-jet transverse momentum is denoted $\vec{P}_\perp = (\vec{k_{l\perp}} + \vec{k_{J\perp}})/2$. The angle between the two vectors is designated ϕ. While the soft gluon (sg) radiation tends to be collinear to the jet axis, radiation that falls outside of the jet radius can result in a measurable lepton-jet momentum asymmetry. For this simple schematic, $\vec{q_\perp} = -\vec{k}_{sg\perp}$ from momentum conservation.

The measurement is presented in terms of three harmonics of ϕ: $\cos(\phi), \cos(2\phi)$, and $\cos(3\phi)$, as relevant perturbative QCD calculations obtain the azimuthal angle anisotropies of two-particle (or jet) correlations through harmonic analysis that extracts the Fourier coefficients for $\cos(\phi), \cos(2\phi)$, and $\cos(3\phi)$.

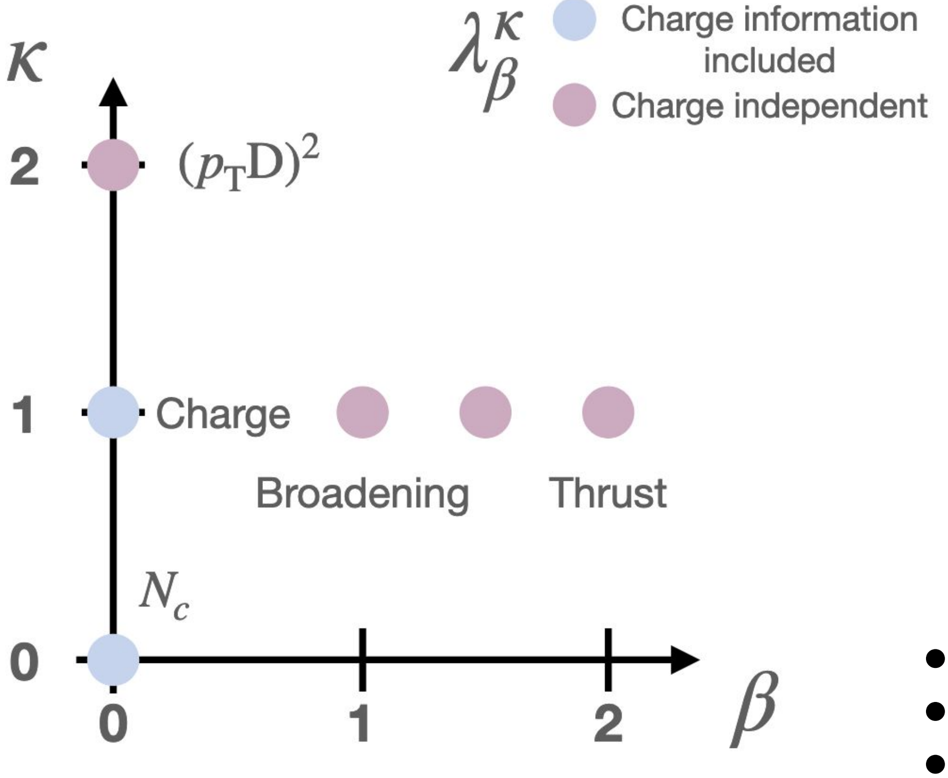
END

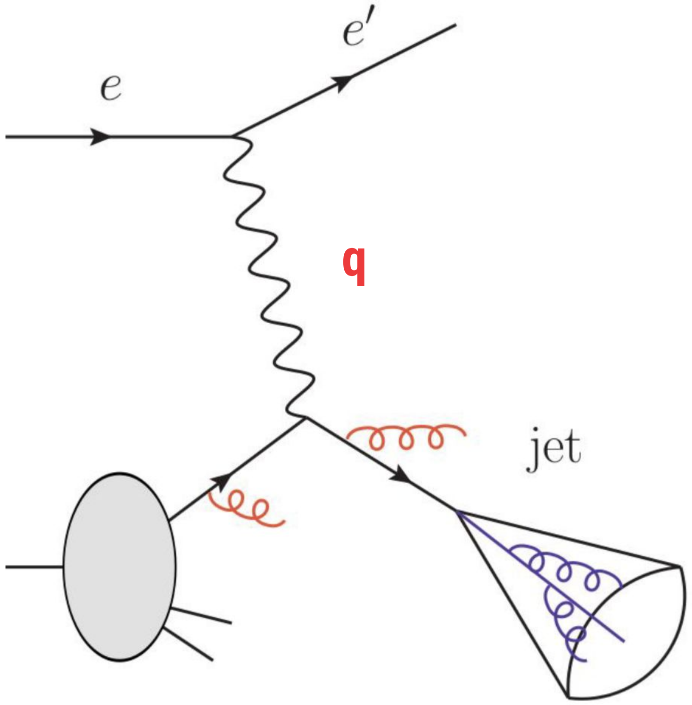
Backup

Jet Angularities

Use jet observables to study different properties of QCD physics:

- Infrared and collinear (IRC) safe λ^1_a, a = [0,0.5,1] and unsafe $\mathbf{p_TD}$ angularities
- Charge dependent observables: $\mathbf{Q_j}$ and $\mathbf{N_c}$
- Study the evolution of the observables with energy scale $\mathbf{Q^2 = -q^2}$



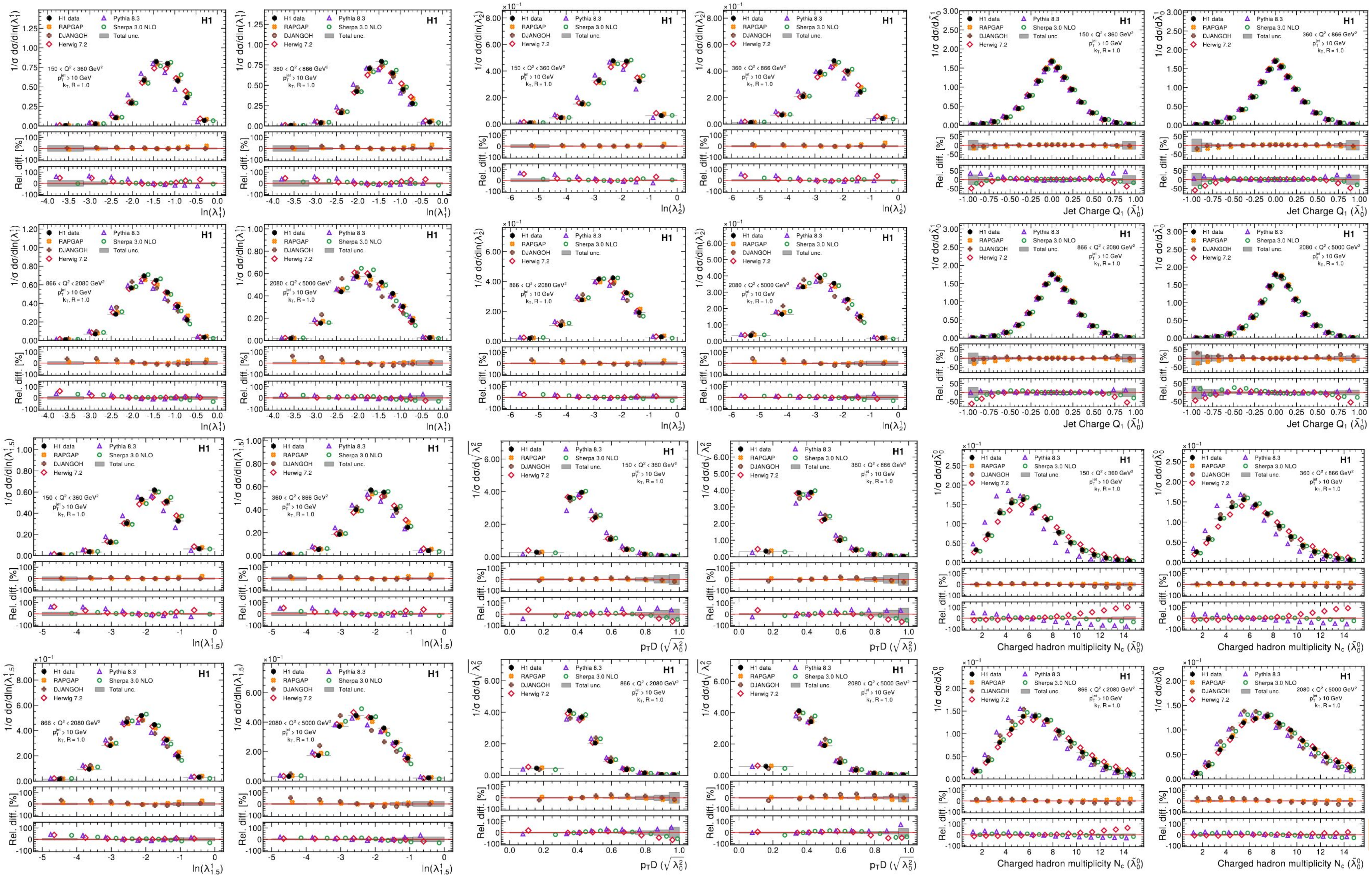


- z_i: longitudinal momentum fraction
- q_i: charge
- R_i distance from jet axis in (eta,phi)

$$\lambda^\kappa_\beta = \sum_{i \in \text{jet}} z_i^\kappa \left(\frac{R_i}{R_0}\right)^\beta$$

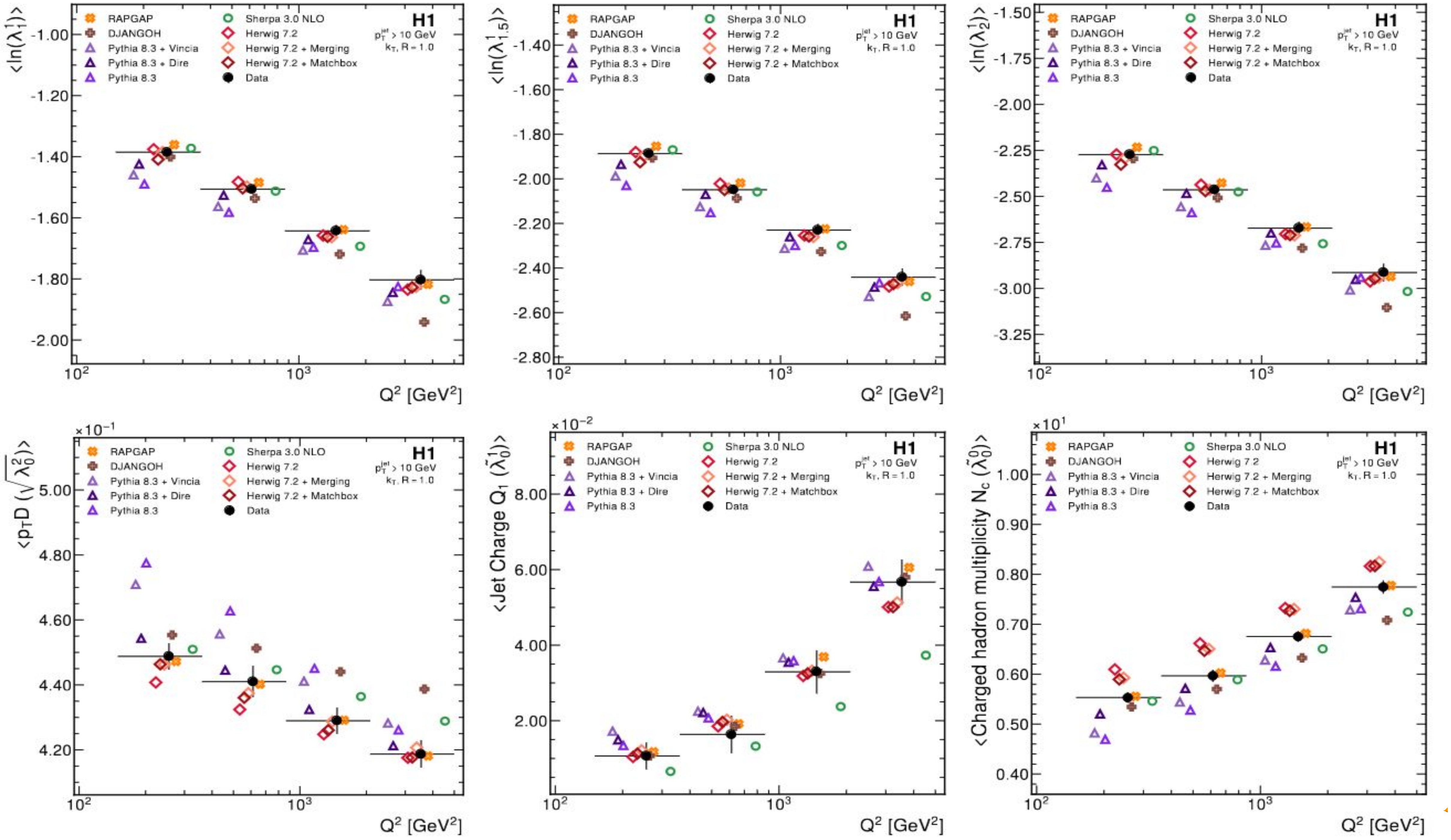
$$\tilde{\lambda}^\kappa_0 = Q_\kappa = \sum_{i \in \text{jet}} q_i \times z_i^\kappa .$$

Multi-Differential H1 Jet Substructure



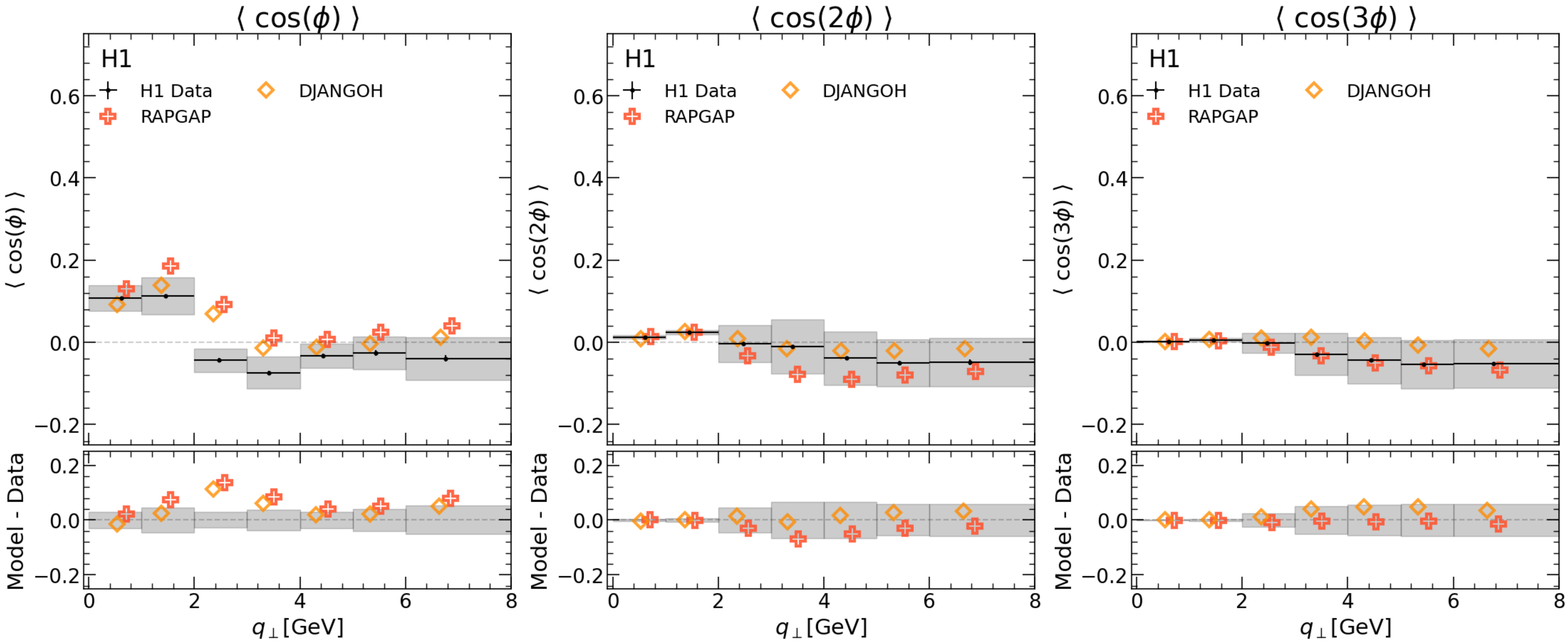
- Q^2distribution simultaneously unfolds, displaying the energy scale dependance of observables, and yielding 30 unfolded distributions!

Moments of jet substructure



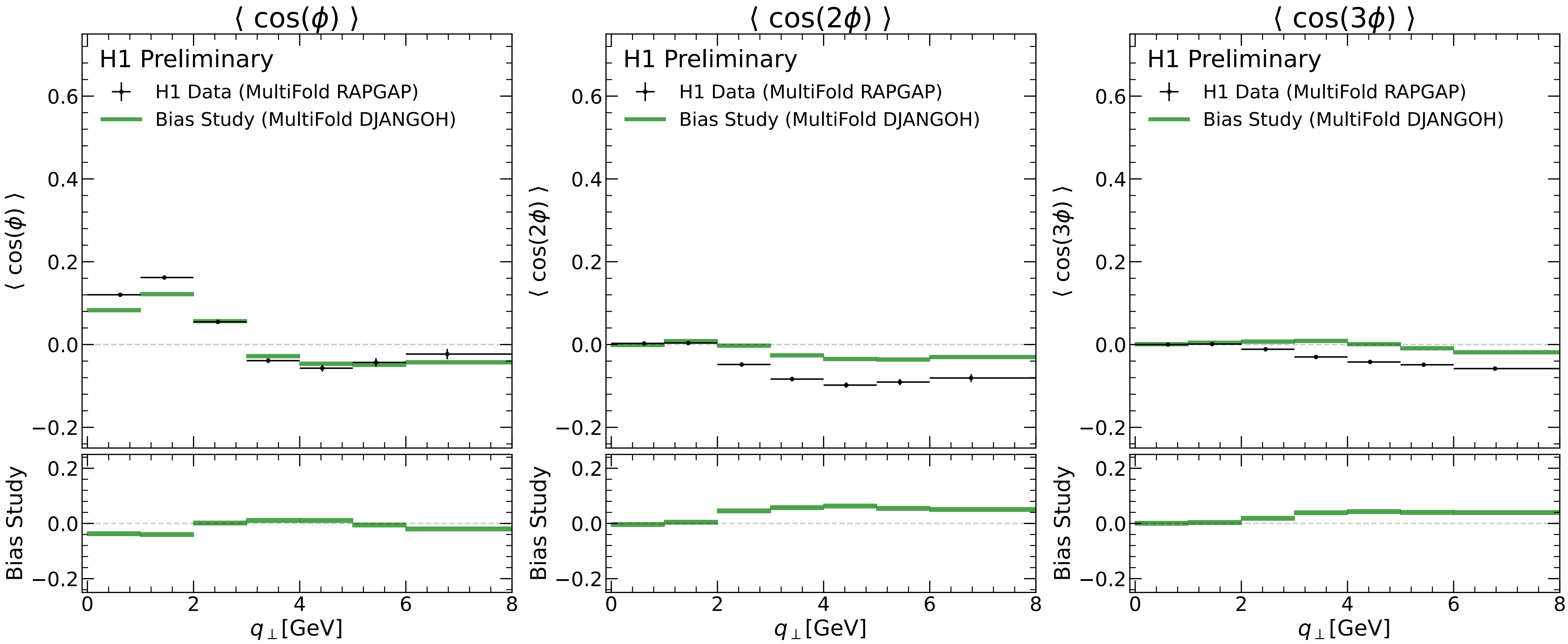
- Mean value of distributions also unfolded for free!
- Better agreement w/ generators at higher Q^2

H1 Unfolded Data + MC



- Leading moment is $\langle \cos(\phi) \rangle$, expected in lepton-jet events
- All harmonics approach 0.0 at higher $q_\perp$, *may* compromise $P_\perp \gg q_\perp$
- Rapgap and Django, tuned to HERA II, good agreement
- Breakdown of systemics next slide

Investigation of Model Bias vs. $q_\perp$ [GeV]



- Leading uncertainty is model bias in the unfolding for $\cos(2\phi)$ and $\cos(3\phi)$
- Difference in the result when unfolding using RAPGAP and DJANGO
- Reporting Abs. Errors; <u>central values are very close to 0.0</u>
- The Total Uncertainty is quite stable between harmonics

Systematic Uncertainties

- Model Dependance:
  - The bias of the unfolding procedure is determined by taking the difference in the result when unfolding using RAPGAP and DJANGO
  - The two generators have different underlying physics, thus providing a realistic evaluation of the procedure bias

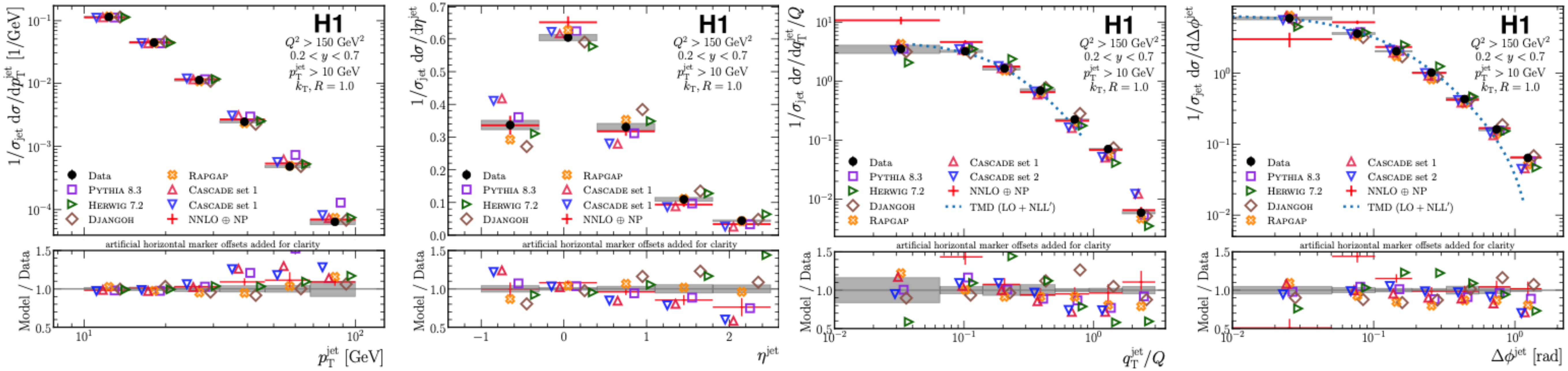
- QED Radiation Corrections
  - Difference of correction between RAPGAP and DJANGO
  - Take RAPGAP with and without QED corrections
  - Take DJANGO with and without QED corrections

- Systematic uncertainties are determined by varying an aspect of the simulation and repeating the unfolding
  - These values detail the magnitude of variation:
  - HFS-object energy scale: $\pm 1\,\%$
  - HFS-object azimuthal angle: ± 20 mrad
  - Scattered lepton azimuthal: ± 1 mrad
  - Scattered lepton energy: $\pm 0.5 - 1.0\,\%$

Further Background

- Machine learning (OmniFold) is used to perform an 8-dimensional, unbinned unfolding. Present four, binned results:

- Use the 8-dimensional result to explore the Q^2 dependence and any other observables that can be computed from the electron-jet kinematics



Extracted from the same phase-space as Yao's analysis, but reporting a different observable

OmniFold

1. $\omega_n(m) = \nu_{n-1}^{\text{push}}(m) L[(1,\text{Data}), (\nu_{n-1}^{\text{push}}, \text{Sim.})](m)$

$$\omega_n^{\text{pull}}(t) = \omega_n(m)$$

   - Detector level simulation is weighted to match the data
   - $L[(1,\text{Data}), (\nu_{n-1}^{\text{push}}, \text{Sim.})](m)$ approximated by classifier trained to distinguish the *Data* and *Sim*.

2. $\nu_n(t) = \nu_0(t) L[(\omega_n^{\text{pull}}, \text{Gen.}), (\nu_0, \text{Gen.})](t)$

   - Transform weights to a proper function of the generated events to create a new simulation
   - $L[(\omega_n^{\text{pull}}, \text{Gen.}), (\nu_{n-1}, \text{Gen.})](t)$ approximated by classifier trained to distinguish Gen. with *pulled* weights from Gen. using $\text{weights}_{\text{old}}$ / $\text{weights}_{\text{new}}$

Each iteration of step 2 learns the correction from the original ν_0 weights

Advantage: Easier implementation, no need to store previous ν_n model

Disadvantage: Learning correction from ν_0 is more computationally expensive

IBU Generalization

$$t_j^{(n)} = \sum_i \Pr_{n-1}(\text{truth is } j | \text{measure } i) \Pr(\text{measure } i)$$

$$= \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i,$$

$$L[(w, X), (w', X')](x) = \frac{p_{(w,X)}(x)}{p_{(w',X')}(x)},$$

Differential Cross Section

- Back-to-back electron-jet production from *ep* collision,

$$e(l) + p(P) \to e(l') + J_q(p_J) + X$$

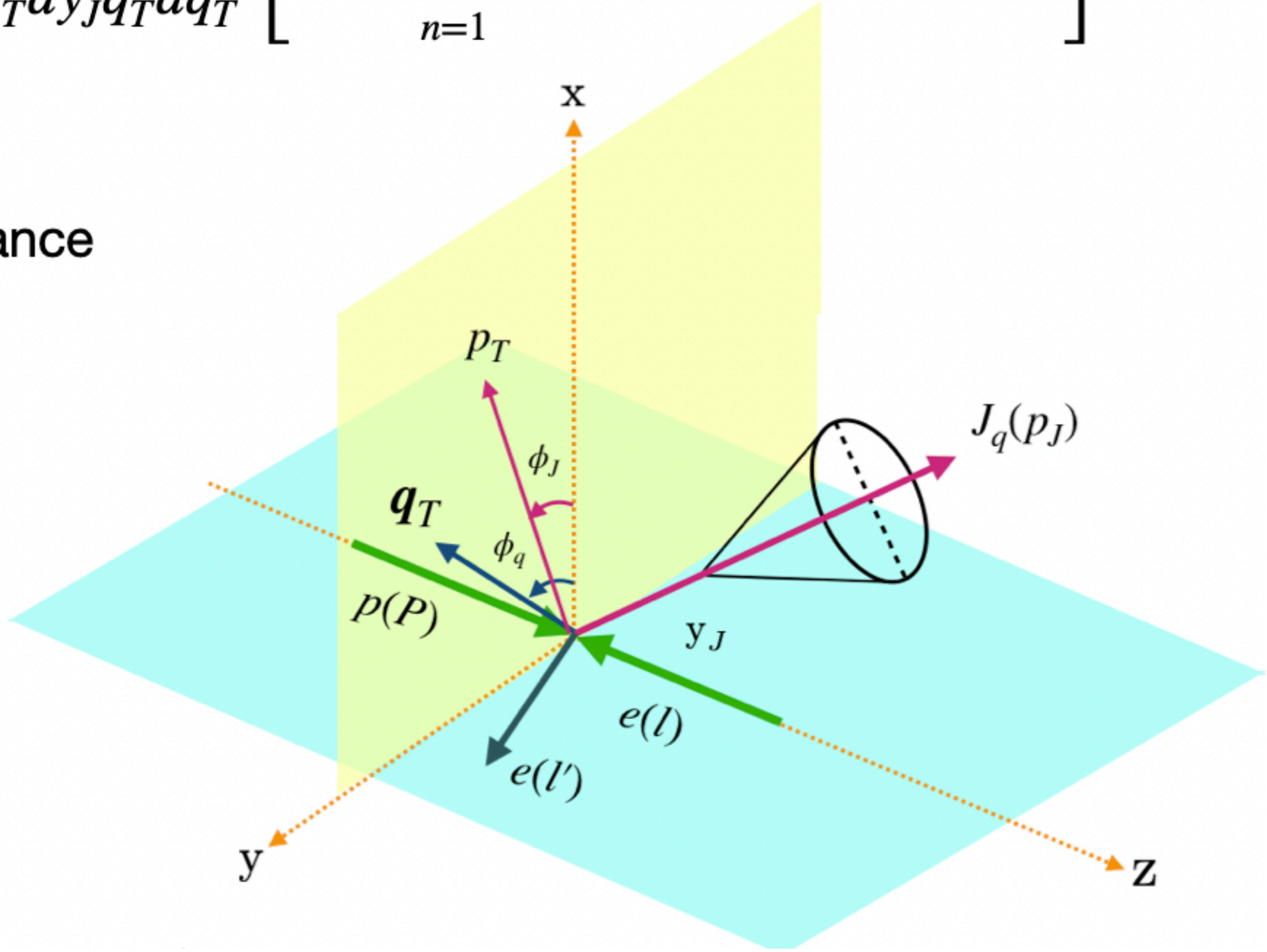
$$\frac{d\sigma}{d^2\boldsymbol{p}_T dy_J d\phi_J d^2\boldsymbol{q}_T} = \frac{d\sigma}{2\pi d^2\boldsymbol{p}_T dy_J q_T dq_T}\left[1 + 2\sum_{n=1}^{\infty} v_n(p_T, y_T)\cos(n(\phi_q - \phi_J))\right]$$

q_T : transverse momentum imbalance

$$\boldsymbol{q}_T = \boldsymbol{l}'_T + \boldsymbol{p}_{JT}$$

p_T : jet transverse momentum

y_J : jet rapidity



Note: slightly different angle definition, but background still applies]

Credit: Fanyi Zhao