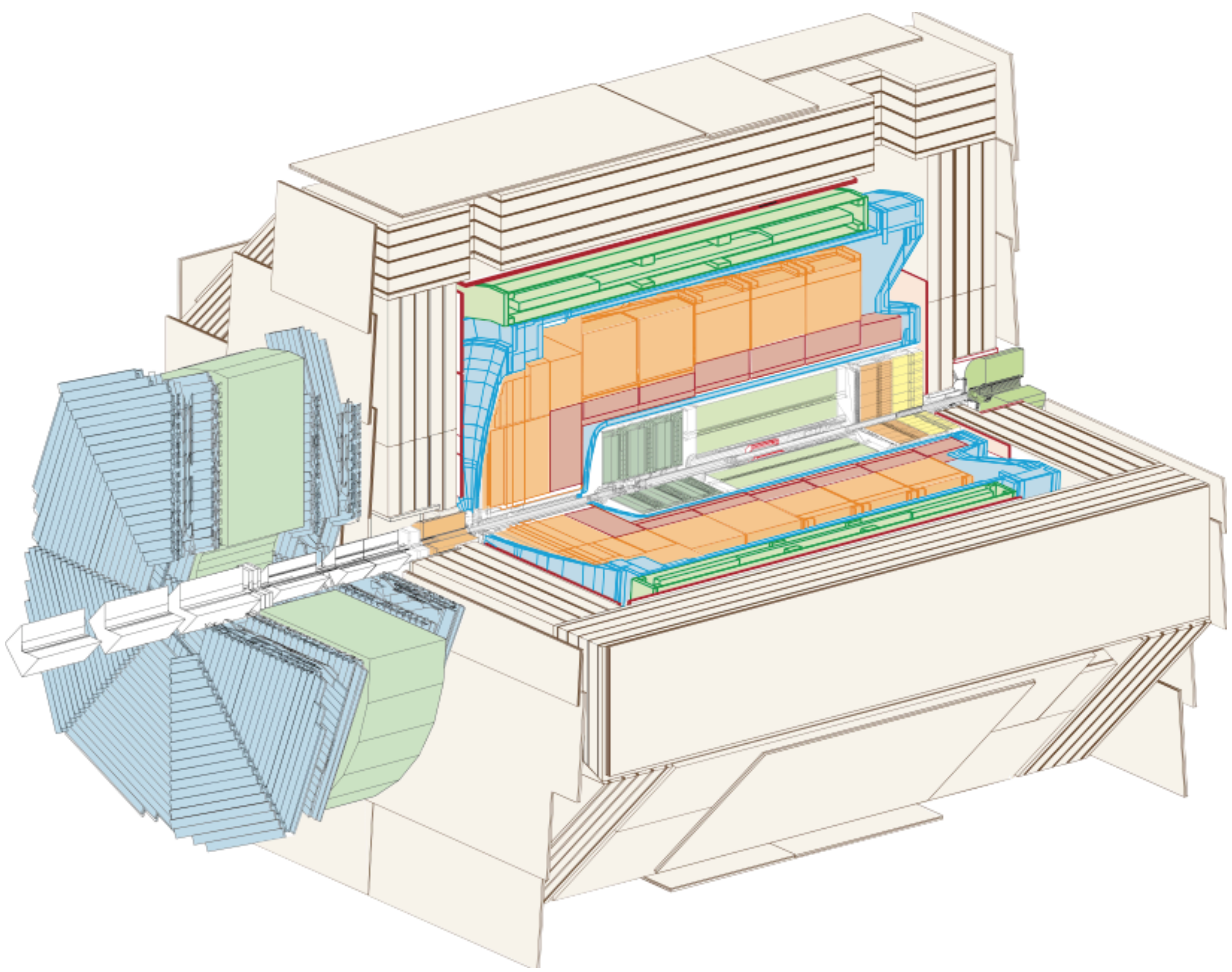
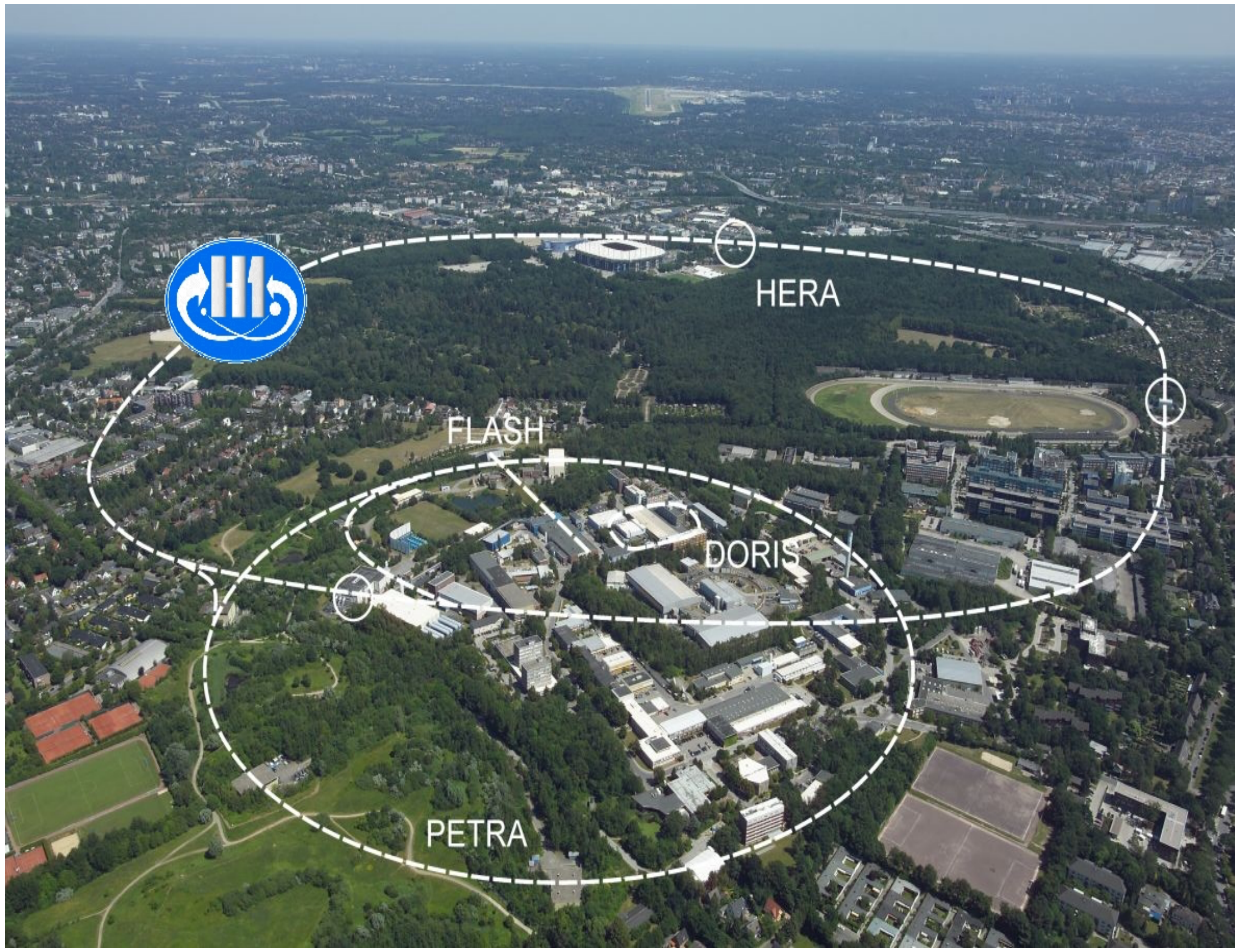
Lepton-Jet Azimuthal Asymmetry in H1 using MultiFold

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**

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

Lepton Jet Asymmetry

Key Ingredients:

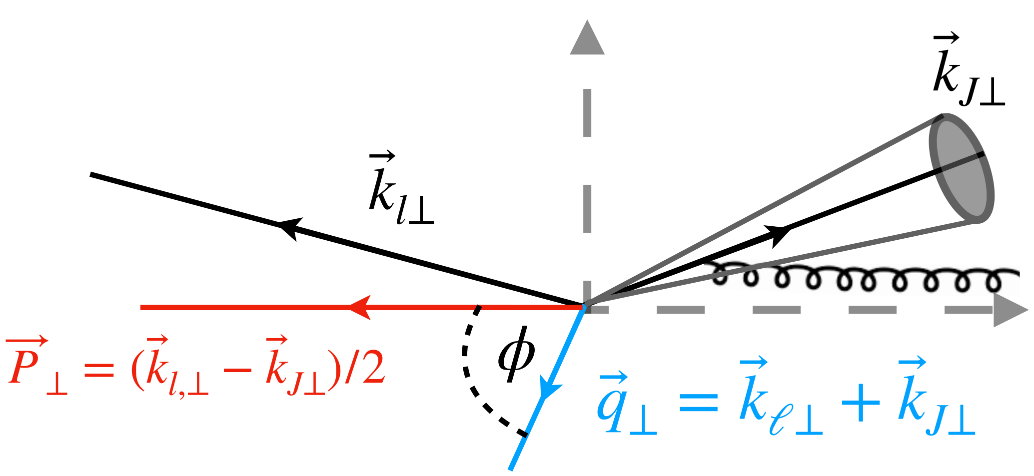
- $q_\perp$ = ***Total*** transverse momentum
- $P_\perp$ = Transverse momentum ***difference***
- ϕ = Angle between $q_\perp$ and $P_\perp$

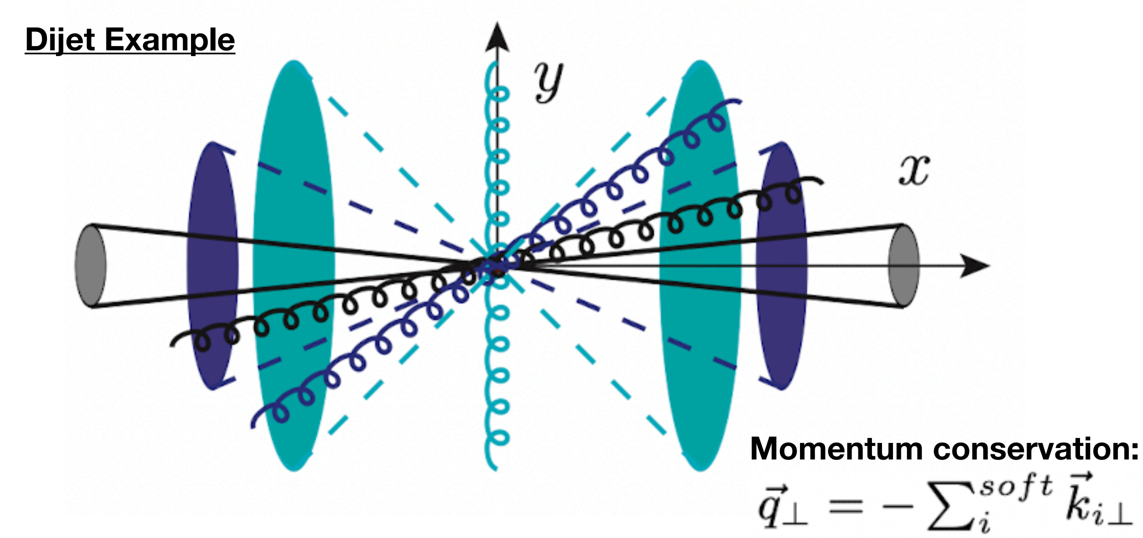
$$\vec{q}_\perp = \vec{k}_{\ell\perp} + \vec{k}_{J\perp}$$

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

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

$$\cos(\phi) = (\vec{q}_\perp \cdot \overrightarrow{P_\perp}) \ / \ \ \vec{q}_\perp \ \ \overrightarrow{P_\perp}$$



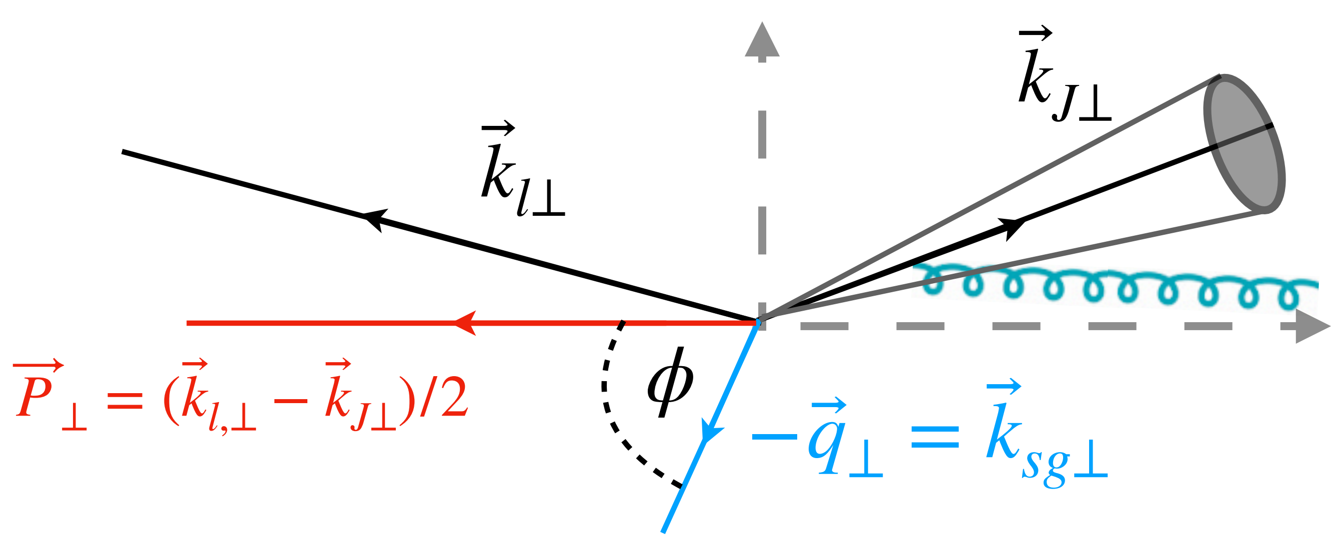


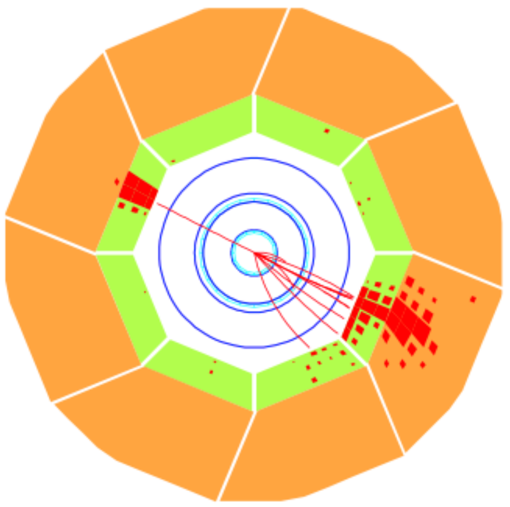
k_i, and therefore $q_\perp$ will tend to point in the direction of the jet
Darker colors indicate probability of gluon emission

Lepton Jet Measurement

- Total transverse momentum of the outgoing system $\vec{q}_\perp = \vec{k}_{\ell\perp} + \vec{k}_{J\perp}$, is typically *small but nonzero*
  - Significant interest in studying transverse momentum dependent (TMD) parton distributions
- Imbalance can come from soft gluon radiation
  - 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 $(\alpha_s \ln^2 P_\perp^2/q_\perp^2)^n$

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





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

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"

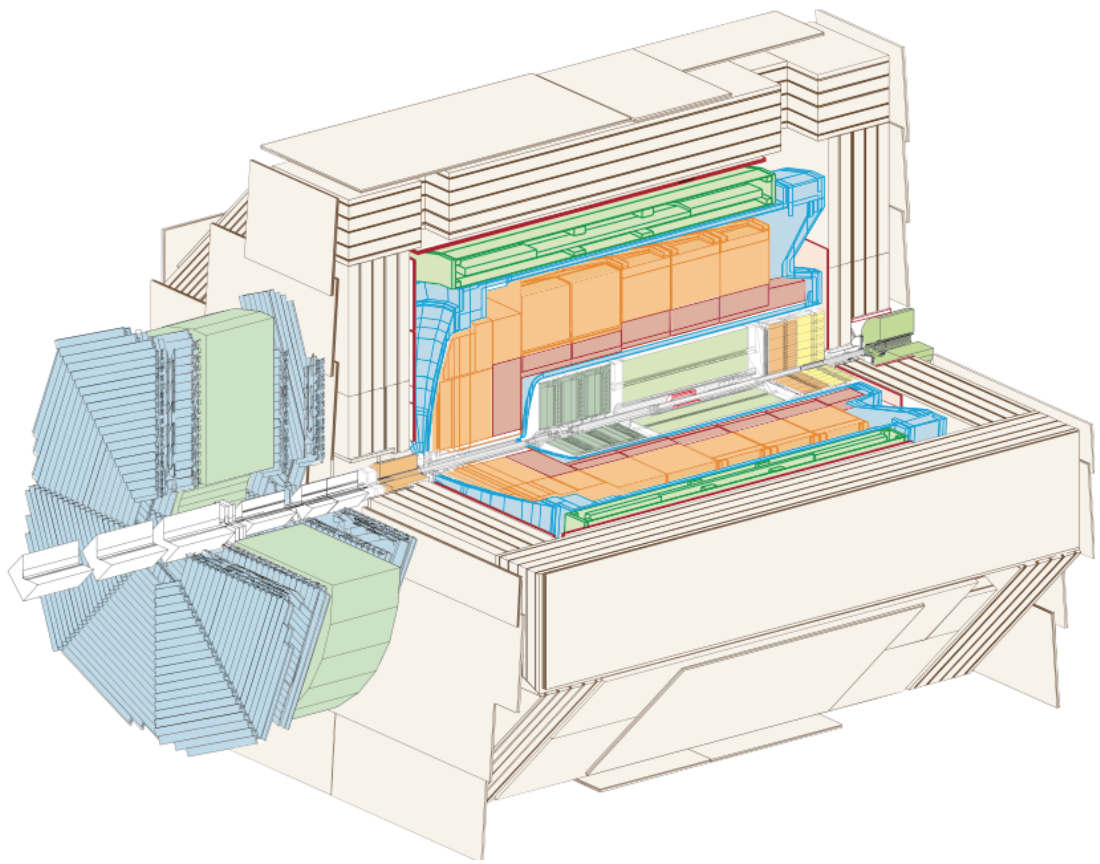
- H1 Data from 2006 and 2007 periods at 130 pb^{-1}
  - Positron-proton collisions

- Fiducial Cuts:
  - $0.2 < y < 0.7$
  - $Q^2 > 150\ \mathrm{GeV}^2$
  - $p_T^{\mathrm{jet}} > 10\ \mathrm{GeV}$
  - $-1 < \eta_{\mathrm{lab}} < 2.5$
  - $\mathrm{k_T}, R = 1.0$
  - $q_\perp/Q < 0.25$
  - $q_\perp/p_{\mathrm{T},jet} < 0.3$

Taking the *leading jet*

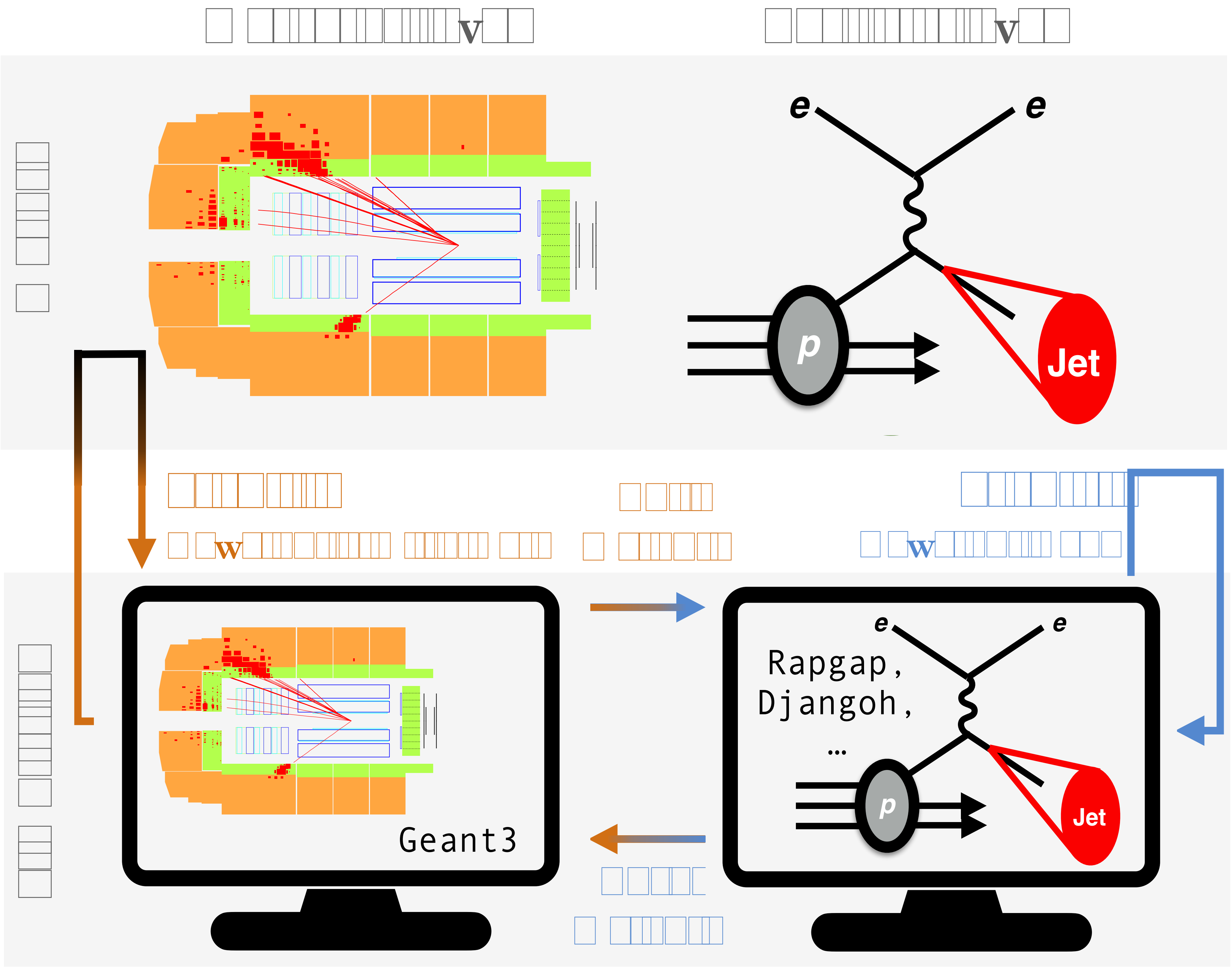
Cut on $q_\perp/p_{\mathrm{T},jet}$ to satisfy $P_\perp \gg q_\perp$:

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



$\sqrt{s} = 320\ \mathrm{GeV}$

MultiFold



2 step iterative approach

- Simulated events after detector interaction are re-weighted to match the data
- Create a “new simulation” by transforming weights to a proper function of the generated events

Machine learning is used to approximate 2 likelihood functions:

Reco MC to **Data** reweighting

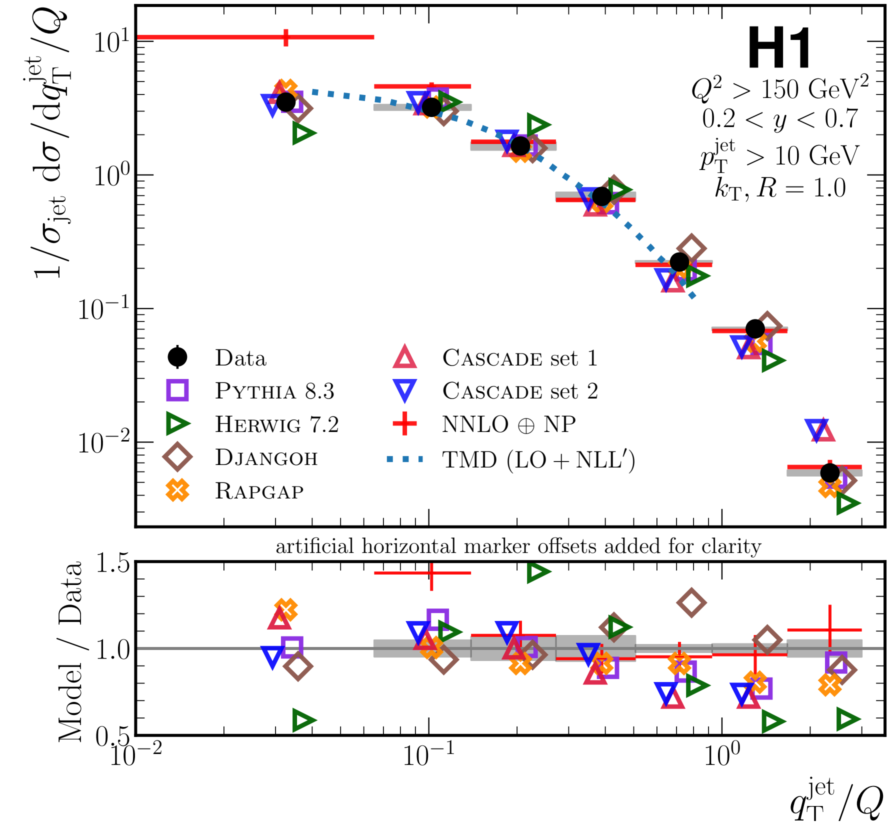
Previous and **new Gen** reweighting

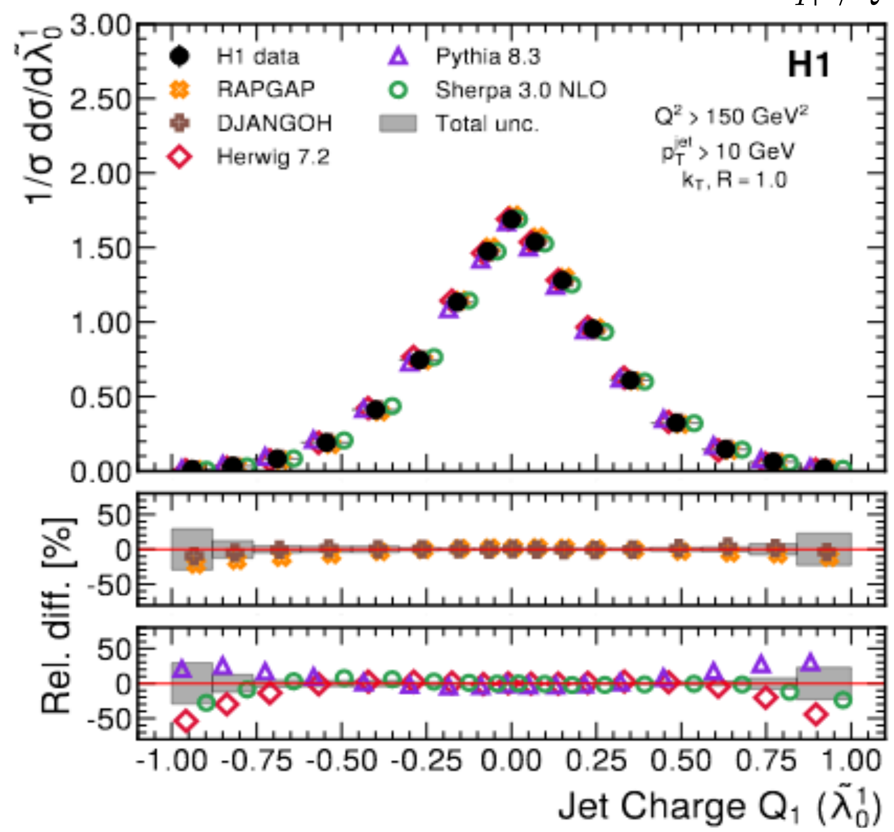
MultiFold Overview

- Multi-dimensional, un-binned unfolding result
  - Lepton-Proton momentum imbalance
  - PhysRevLett.128.132002
- Jet constituent-level unfolding
  - Un-binned Deep Learning unfolding of Jet Substructure
  - arXiv 2303.13620
- Recycling of unfolded event weights
  - And measure *moments*

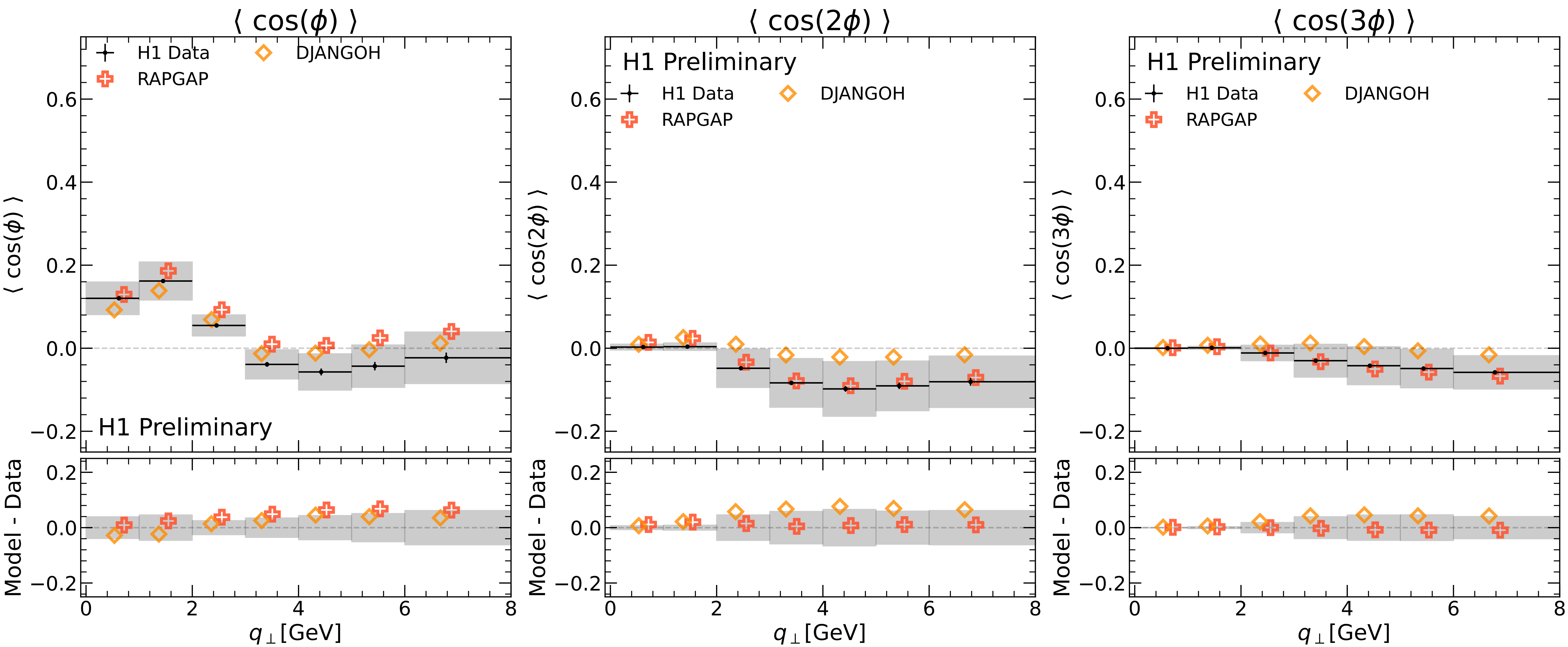
Multifold previously used to unfold:

$$p_x^e,\ p_y^e,\ p_z^e,\ p_T^{jet},\ \eta^{jet},\ \phi^{jet},\ \Delta\phi^{jet},\ q_T^{jet}/Q$$



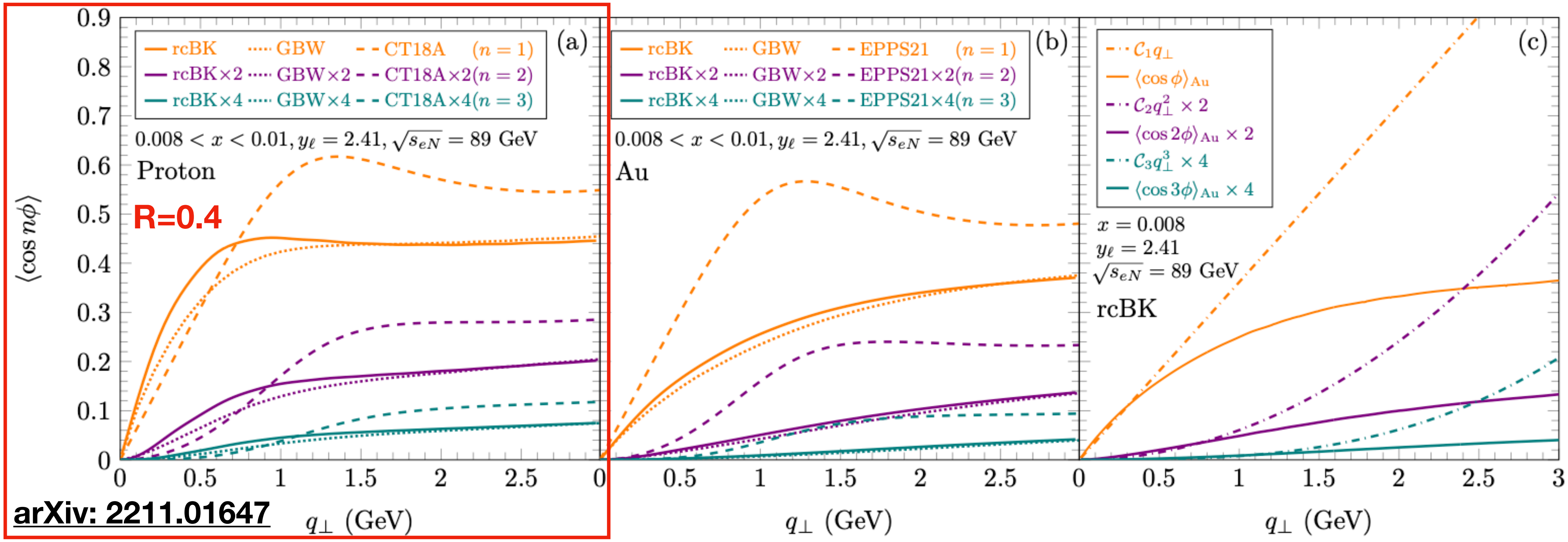


H1 Unfolded Data



- 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 data, exhibit good agreement
- Note *small absolute* value of central values

Even more interesting at EIC!

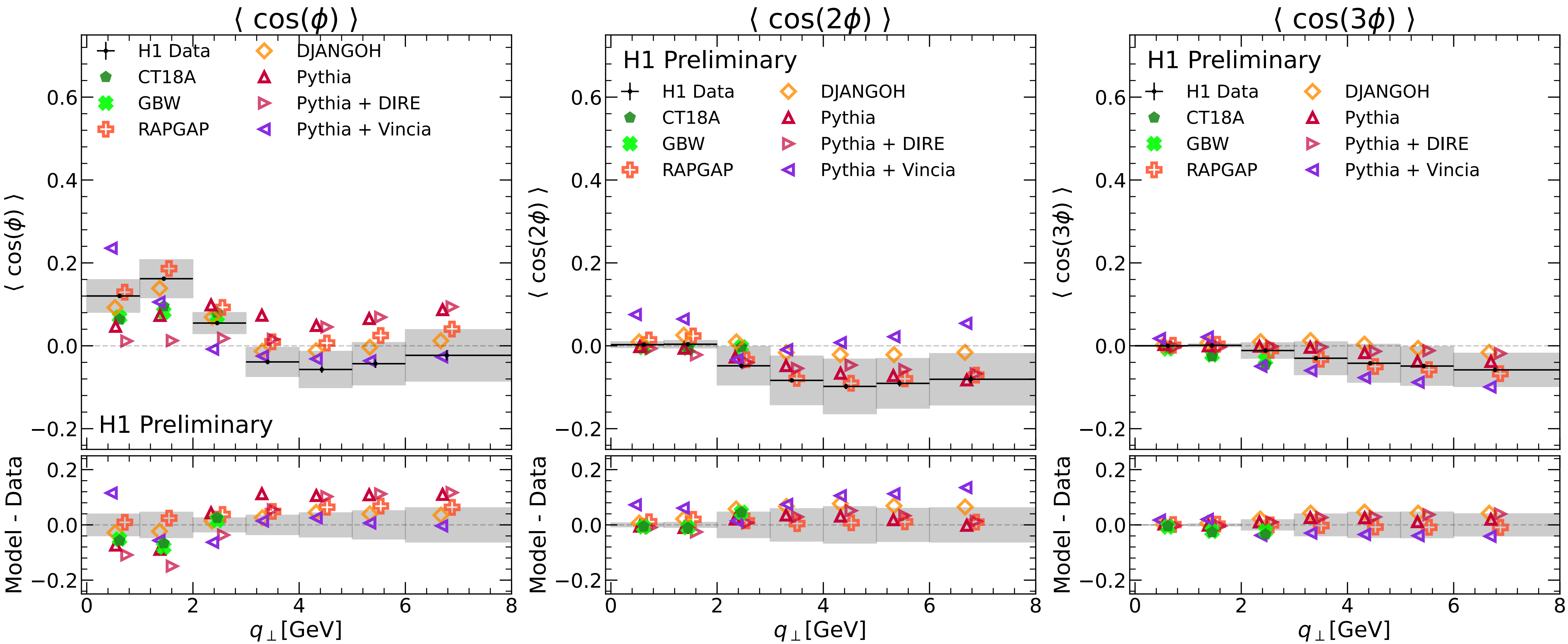


arXiv: 2211.01647

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

- **Asymmetry may be sensitive to Parton saturation effects (EIC)**
- GBW — Three parameter model fit to HERA data, input to $f(b, x)$
- Calculation in TMD framework with CT18A PDF
- **Recalculated to match HERA kinematics, with jet R=1.0**

H1 Unfolded Data



- **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.**
- **PYTHIA, not tuned to HERA II, performs inconsistently**

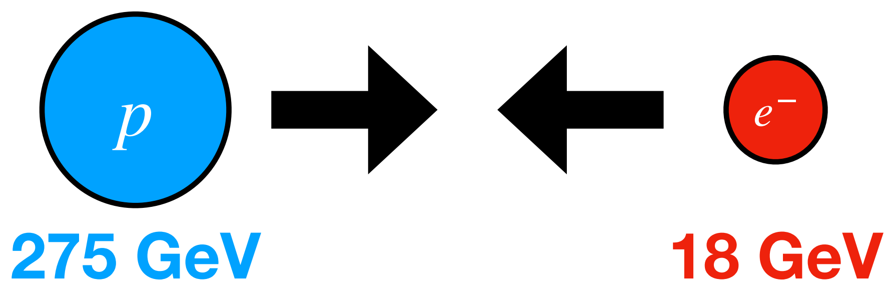
Conclusions

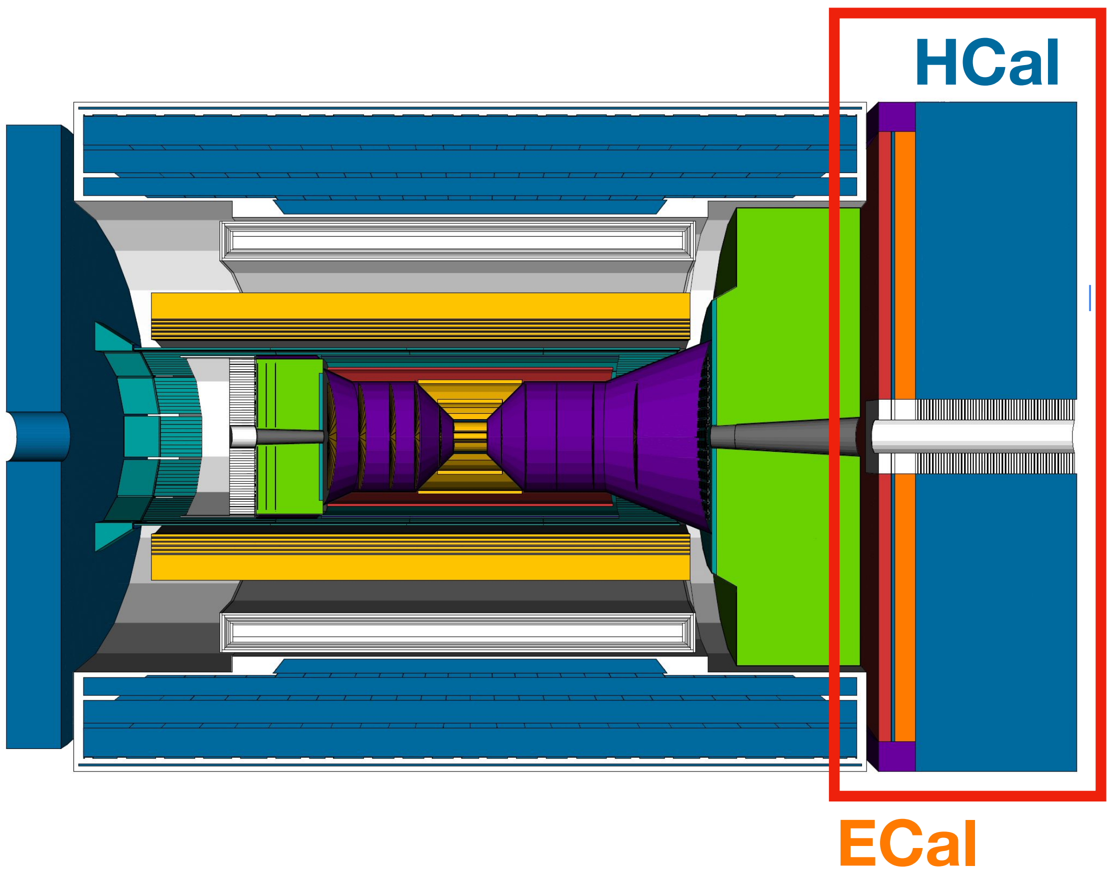
- Promising measurement probing soft gluon radiation
  - Test of pQCD calculations
  - Important reference for lepton-jet DIS measurements
  - Reasonable agreement with Rapgap + Djangoh
- MultiFold
  - First recycling of unfolded event weights! Reusability is key
  - measurement of *moments*, requiring the *unbinned unfolding!*
- Outlook:
  - Because of H1's data + simulation conservation, we can use recent insights and advances in methodology to analyze ~15 year old data
  - Important Implications for studies at EIC, both in observable and methods

ML-Assisted Detector Optimization

Goal:
best experimental design suited for the best detector reconstruction

Forward Hadronic Calorimeter

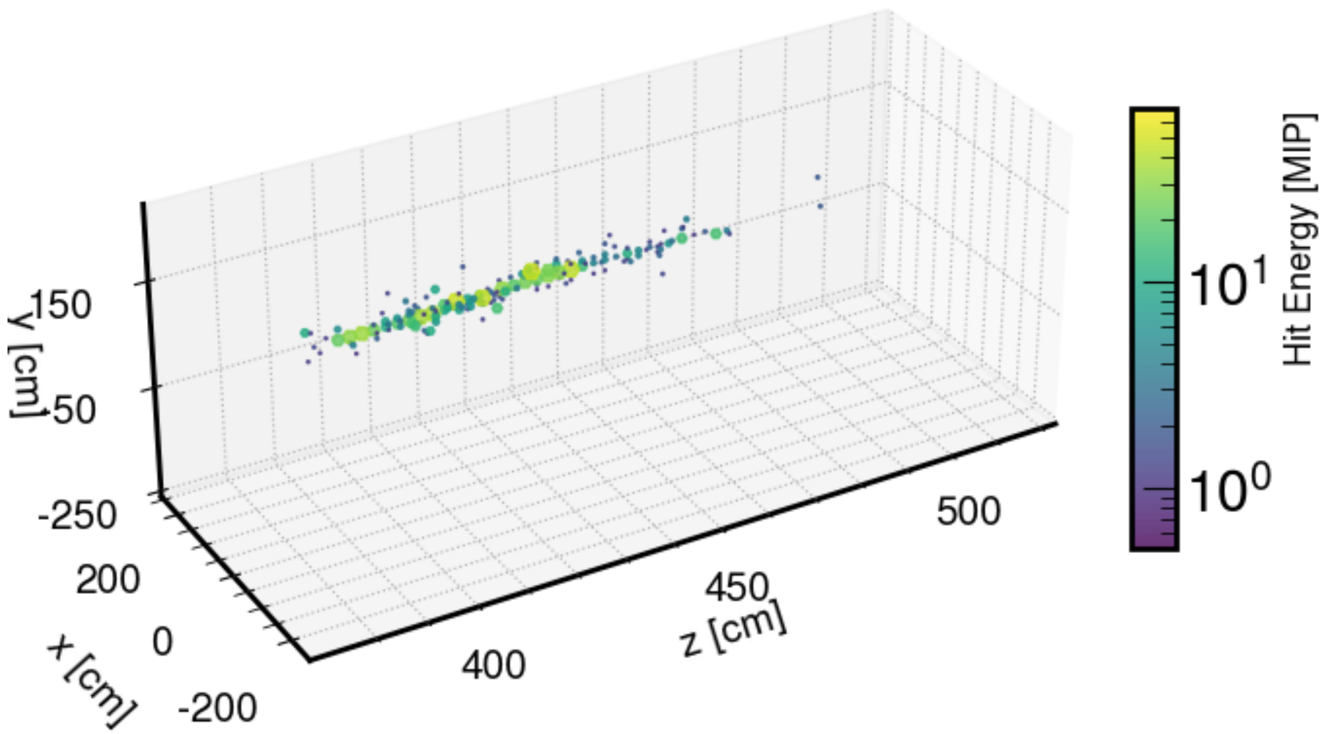


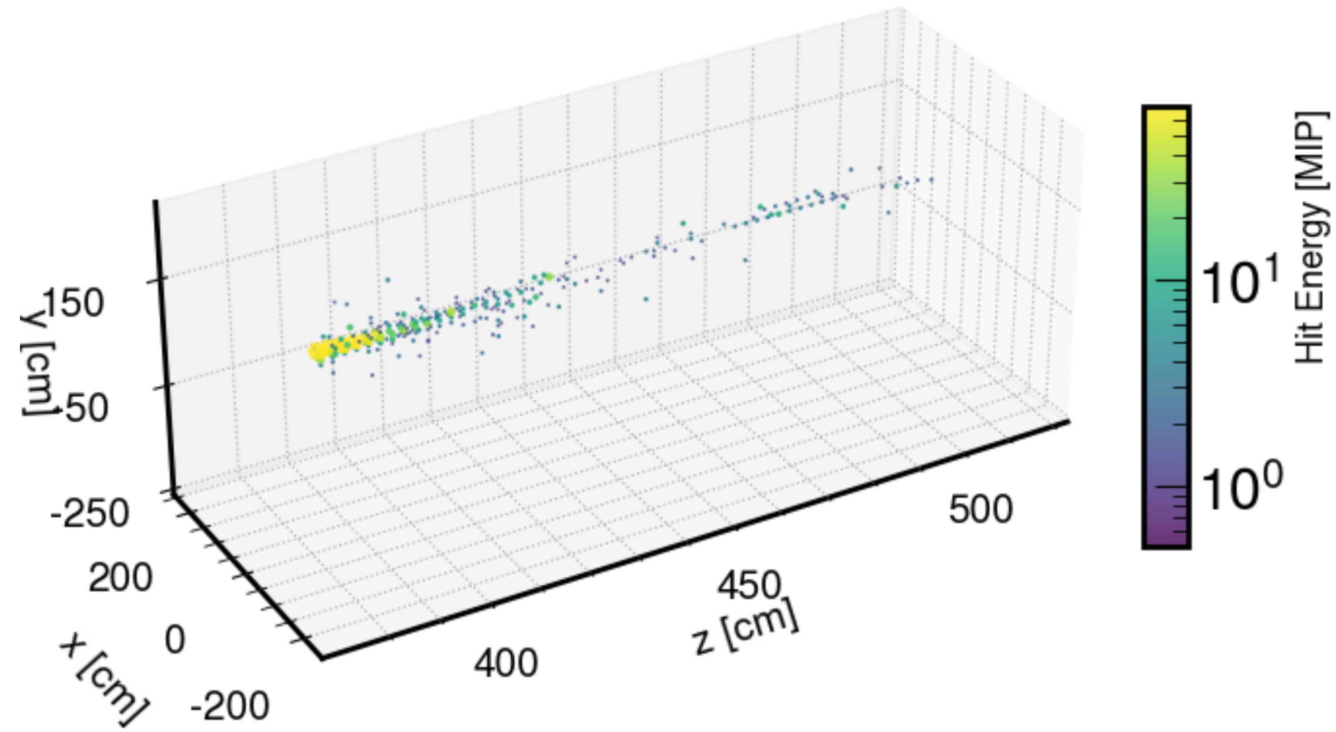


- The incoming proton/ion has a significantly larger kinetic energy than the incoming electron.
- If we want to measure *jets*, we need a granular, forward calorimeter
  - Forward region, $1.2 < \eta < 3.5$
- Deep Sets and GNNs for pion energy regression
- G4 approximation of *ePIC*

Figure Courtesy ePIC

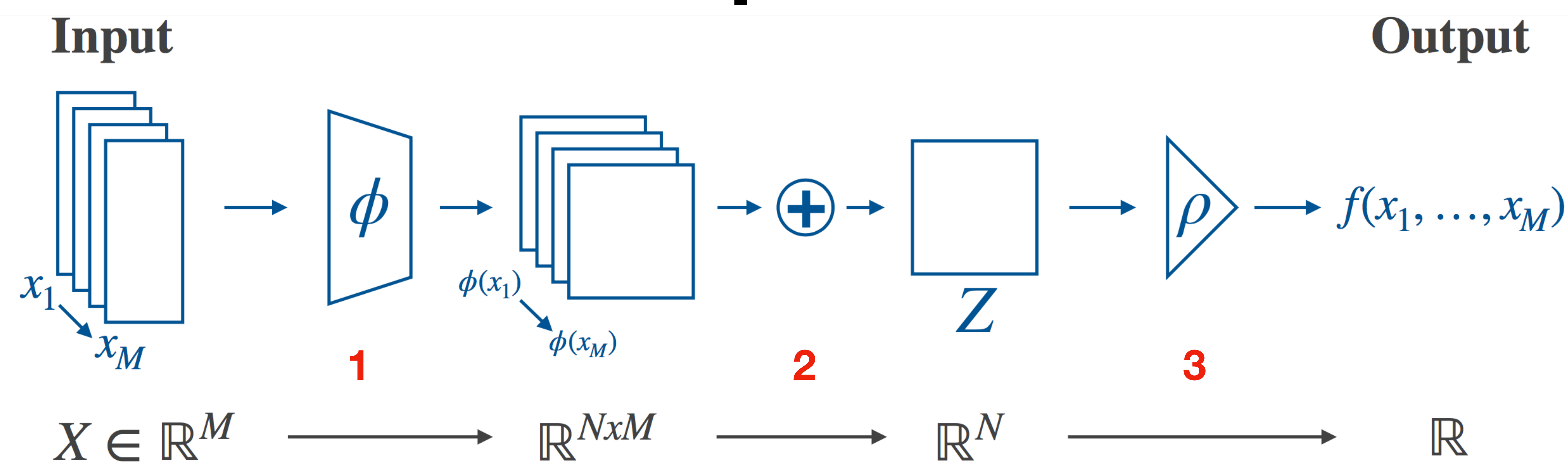
Detector Simulation and Reconstruction

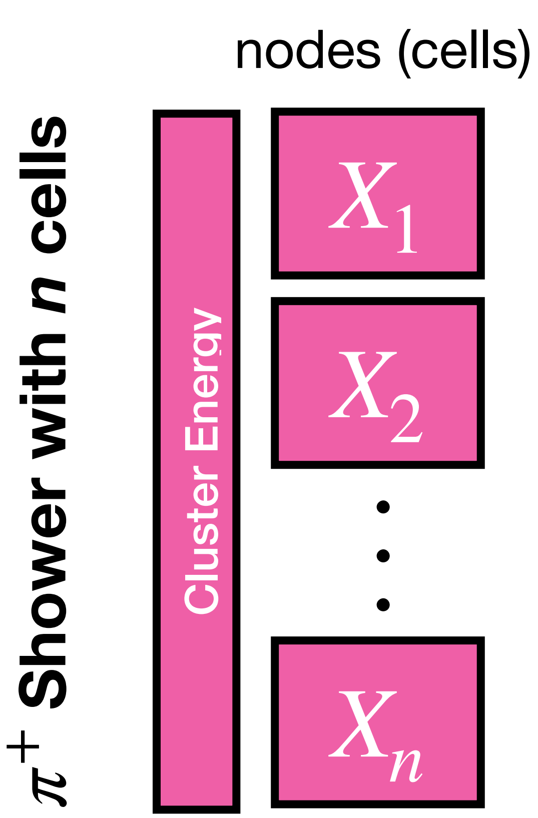




- Geant4 Simulation of single π^+ showers
  $1 < P_{\text{Gen.}} < 125\ \text{GeV}/c$
- $\mathcal{O}100 - 1000$ Cell Hits per shower, ***point clouds***
- Establish a model to predict $P_{\text{Gen.}}$ given cell information
- Condition model on position of longitudinal segmentation

Deep Sets





$$X_i = \left\{ \begin{matrix} E \\ X \\ Y \\ Z \end{matrix} \right\} \in \mathbb{R}^4$$

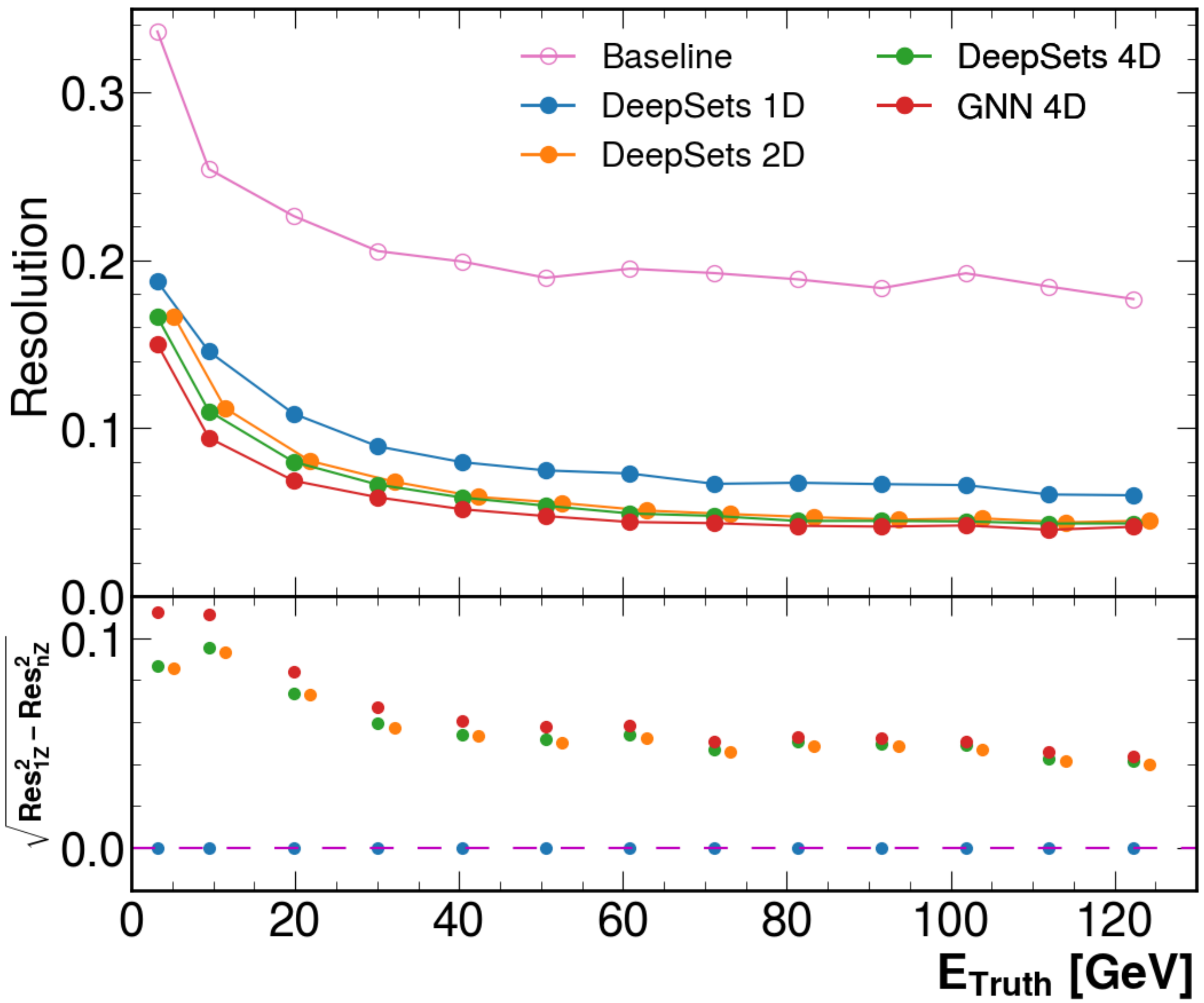
1. Latent
2. Aggregation
3. Final Output

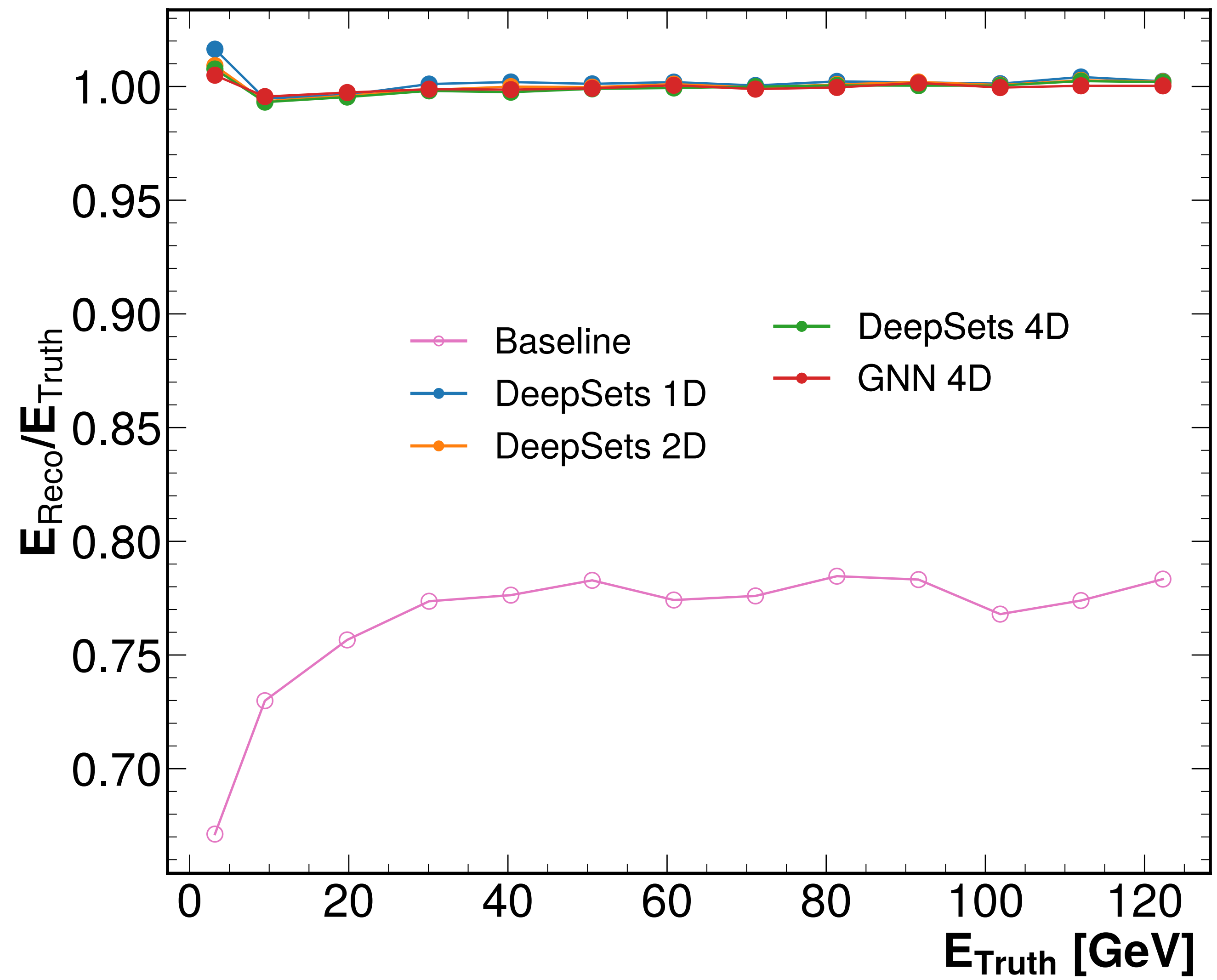
Goal:

π^+ Energy Regression

Model uses energy and position information for energy regression

Energy Regression: Feature Dimension

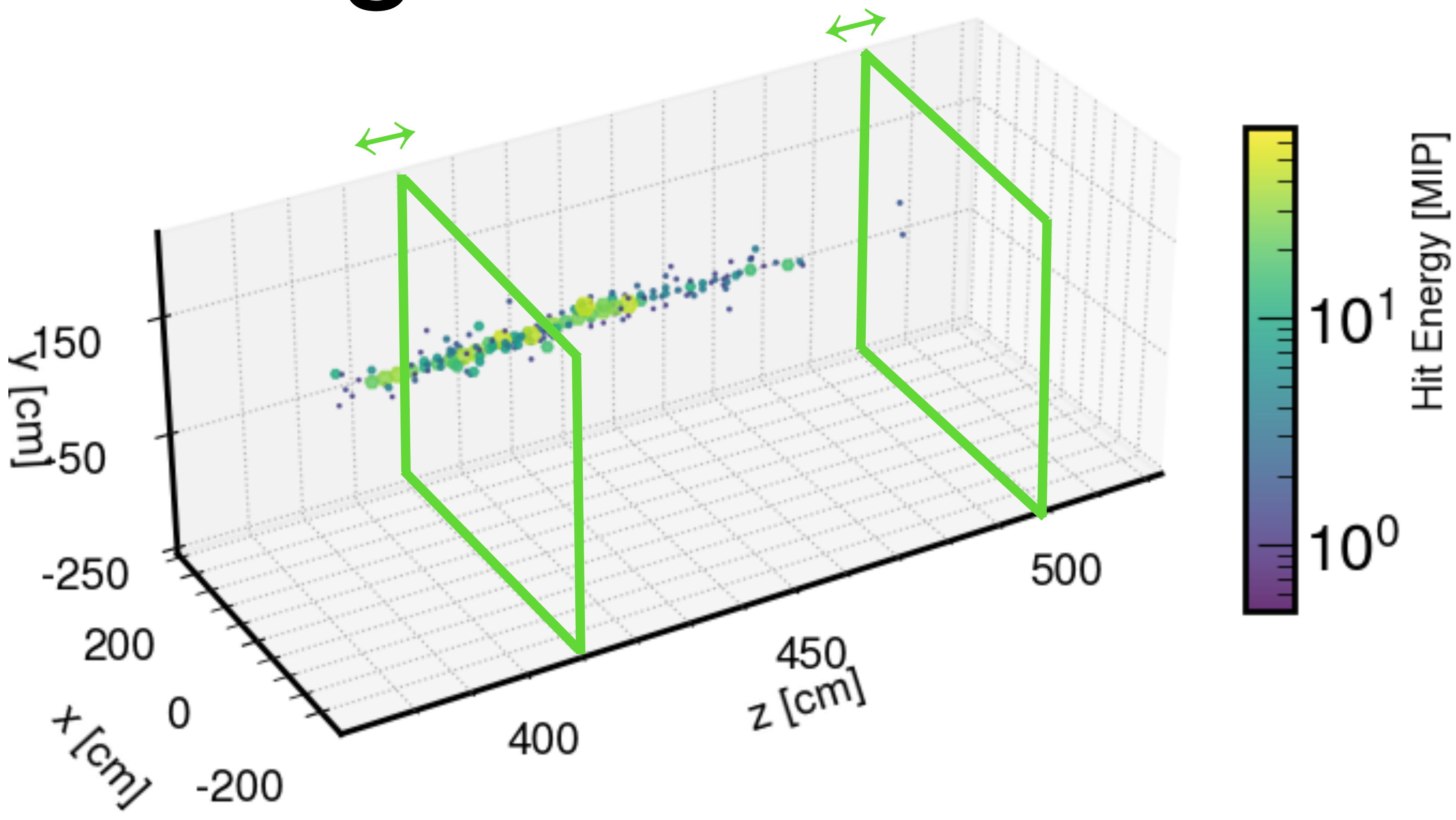




- Using DeepSets for regression, biggest improvement is the inclusion of cell-Z information
- Additional transverse information (2D →4D) less impactful
- Energy scale within %2 of truth

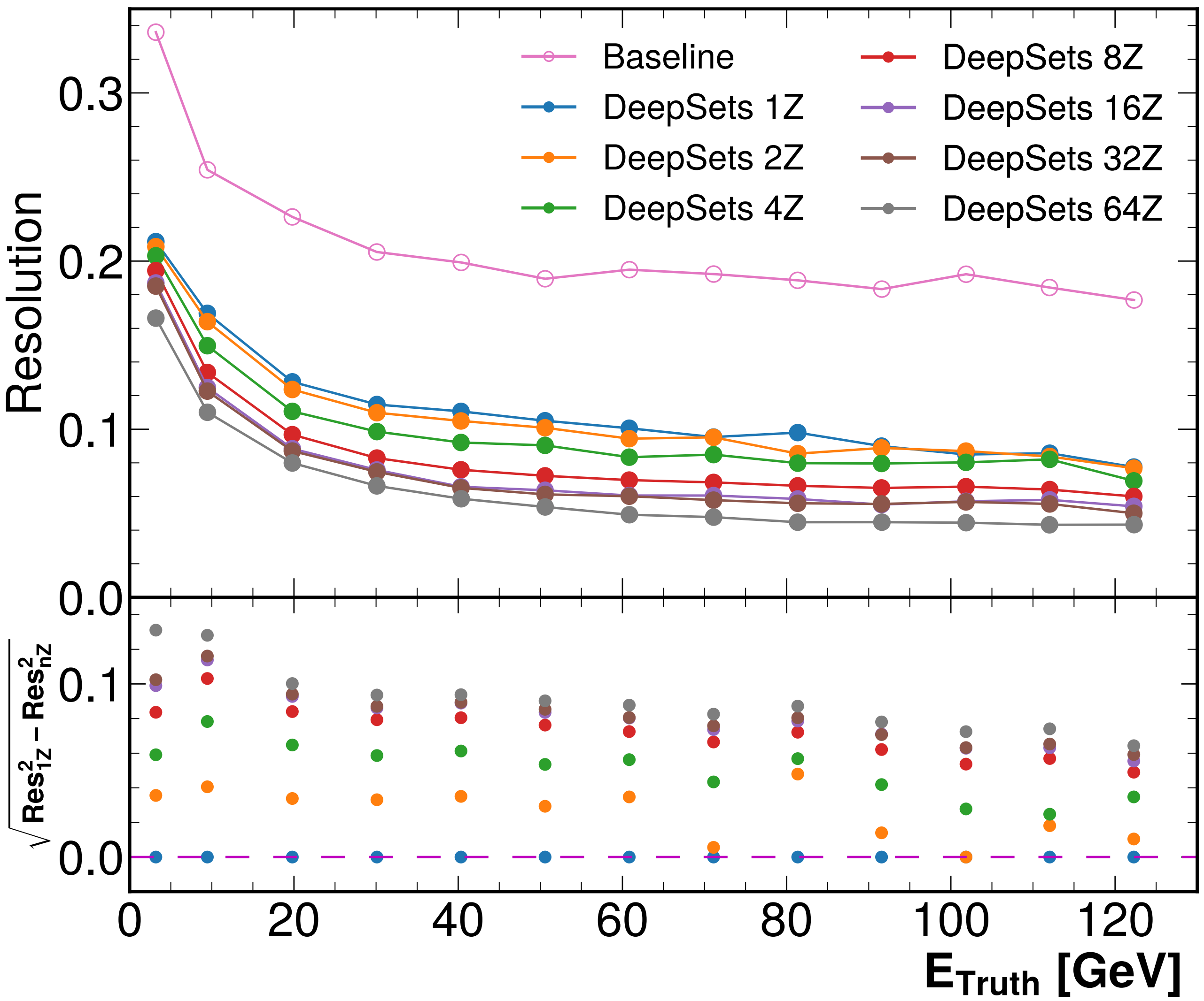
Data Processing for Models

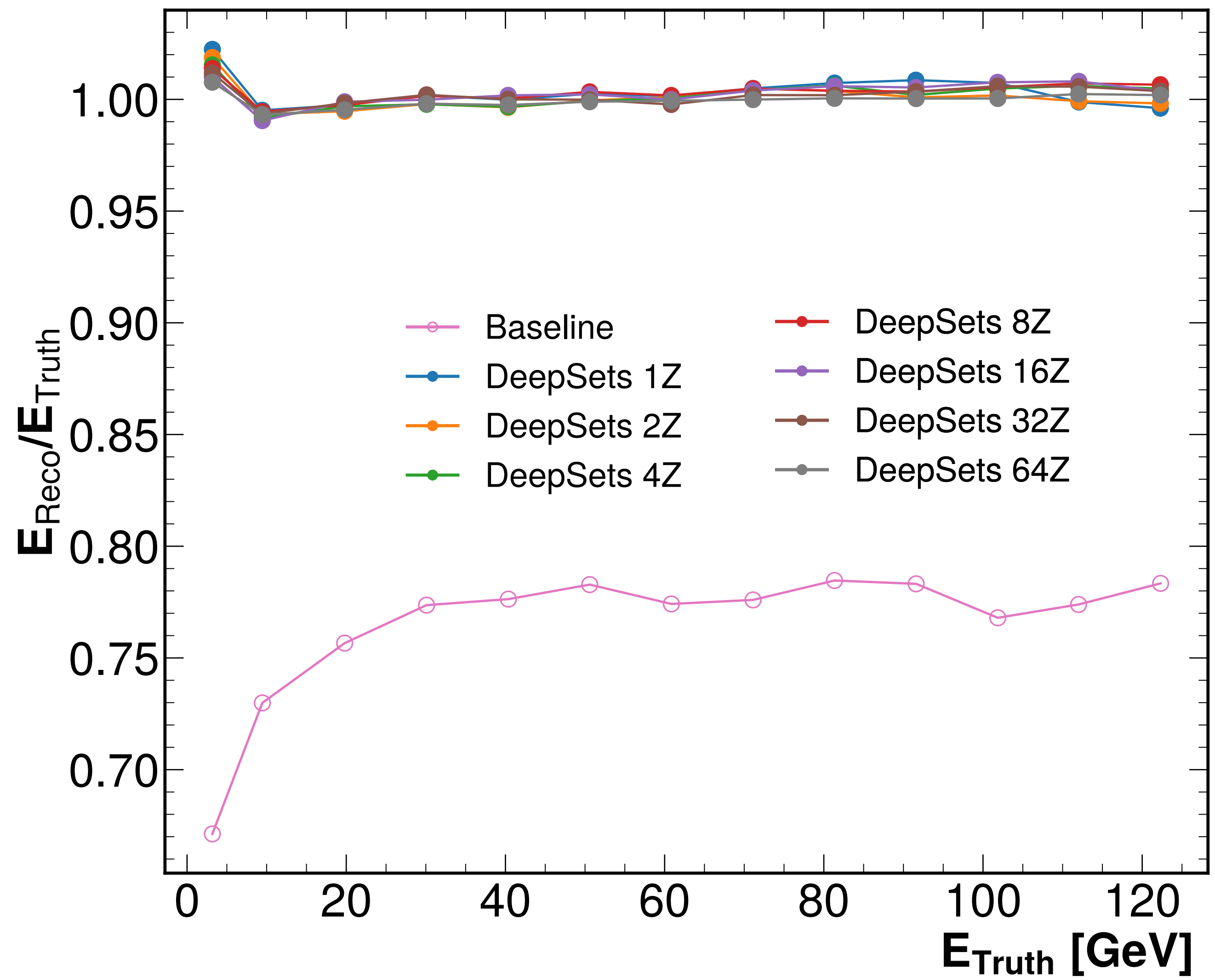
4-Layer Configuration



- Full point cloud readout is *unrealistic* for final detector
- Segment the calorimeter N=1-64 layers
- Run regression, identifying optimal longitudinal configuration

Energy Regression: Number of Layers





- 1-Layer configuration w/ Deepsets outperforms baseline
- Intuitive increase in performance as N_z increases
- Scale is insensitive N_z

Conclusions and Outlooks

- Longitudinal cell information yield greatest improvement. Resolution less sensitive to transverse segmentation

- Paper out this week!
  - *The Optimal use of Segmentation for Sampling Calorimeters*

- Next Step: Model conditioned directly on detector configuration

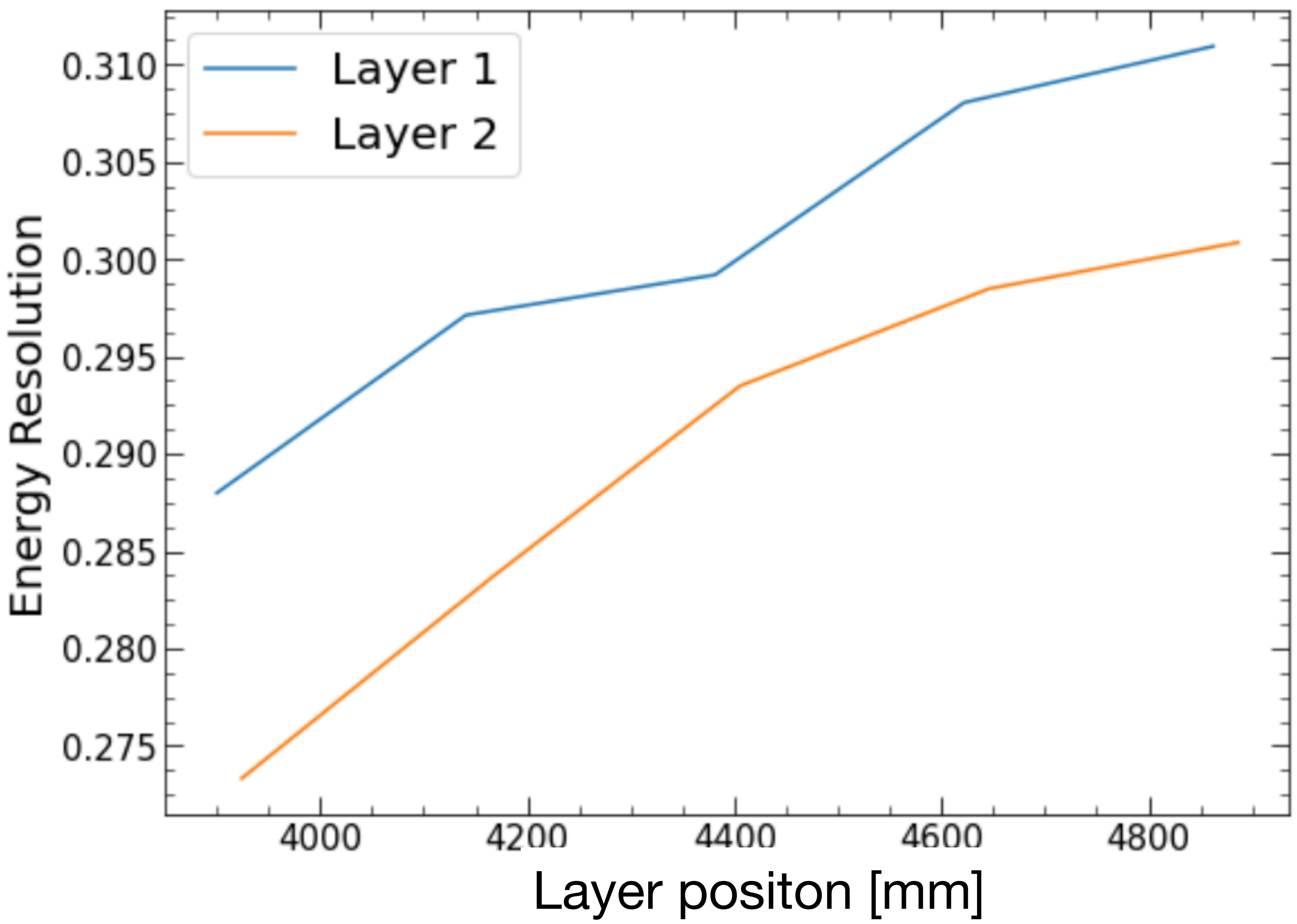
$$\sigma_E = f(z_1,\ z_2, \vec{x})$$

END… END

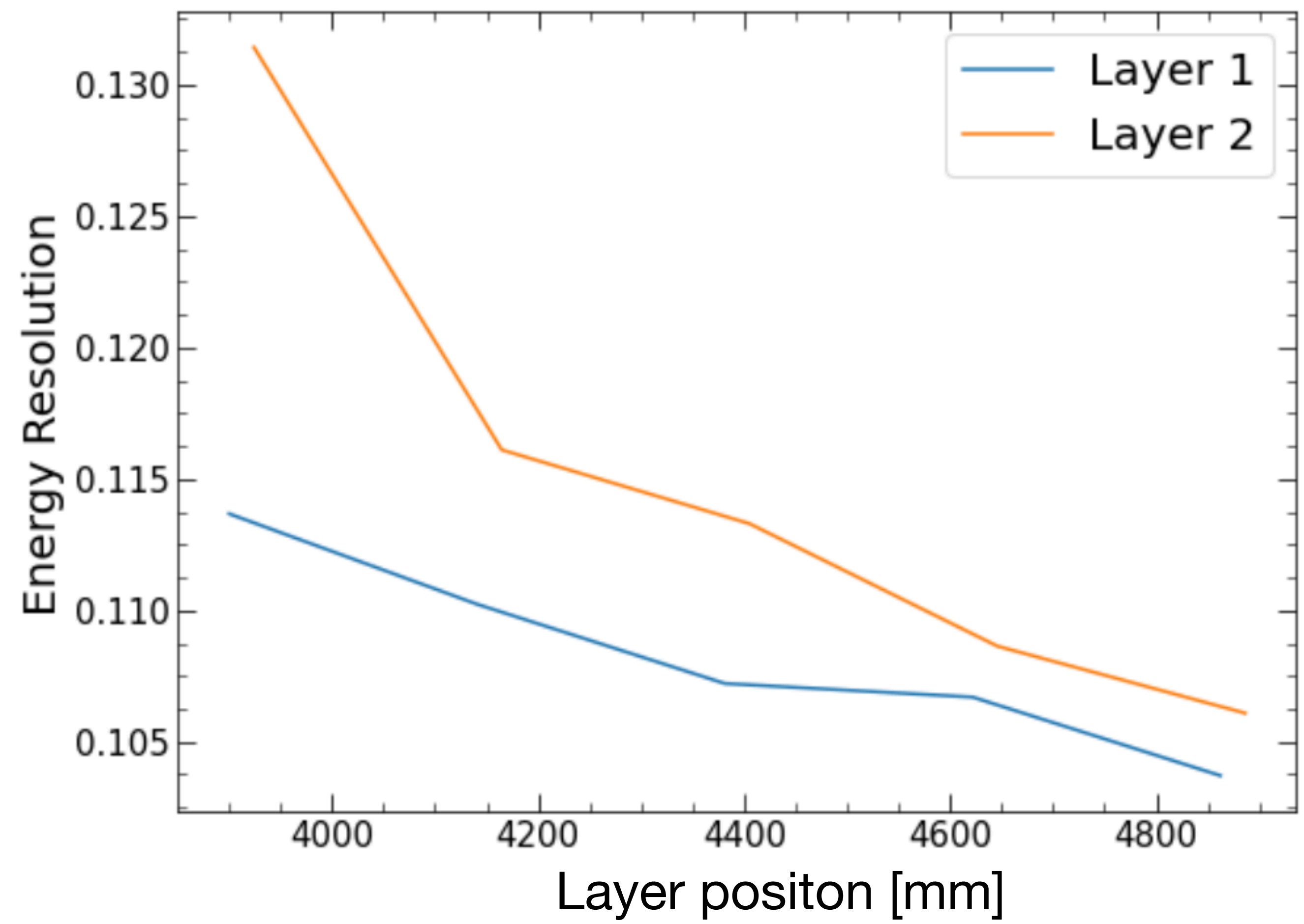
Backup

$$\sigma_E = f(z_1,\ z_2, \vec{x})$$

$P_{\text{Gen.}} < 10.0$ GeV/c

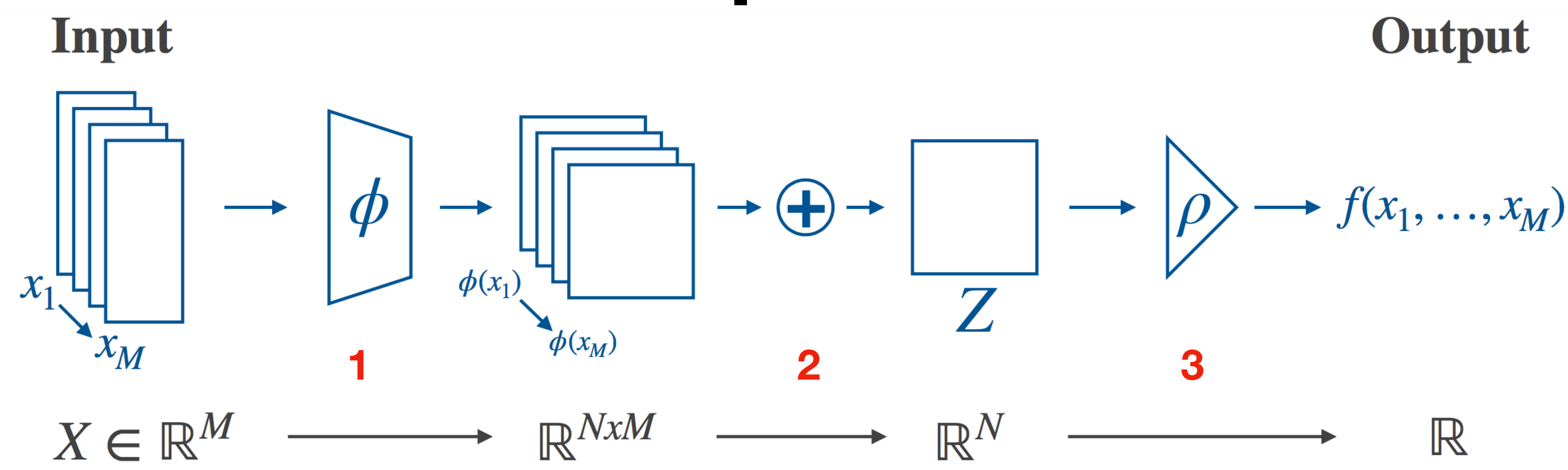


$P_{\text{Gen.}} > 50.0$ GeV/c



We have a differentiable function for energy resolution

Deep Sets



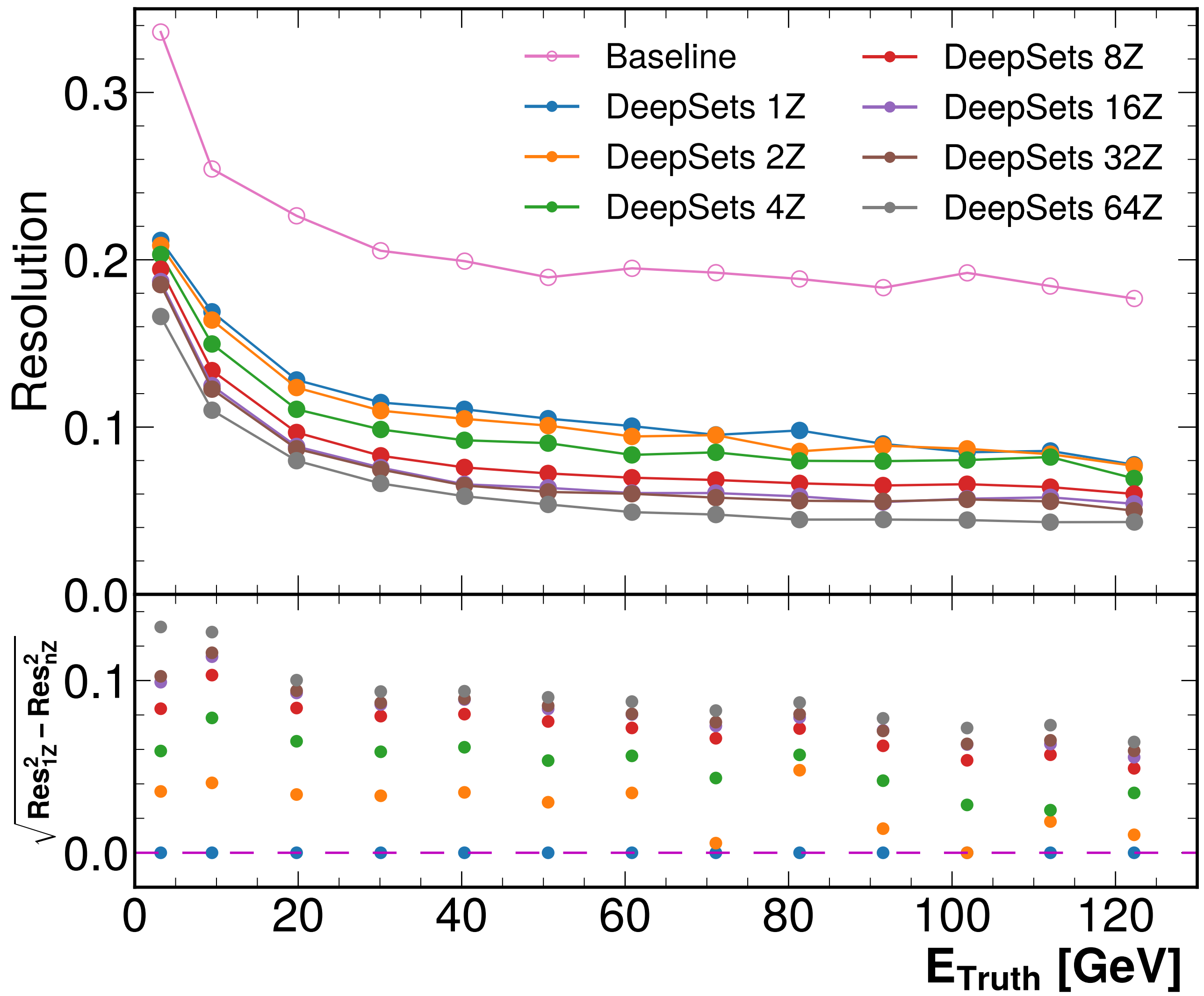
1. Transform inputs into some latent space
2. Destroy the ordering information in the latent space (+, μ)
3. Transform from the latent space to the final output

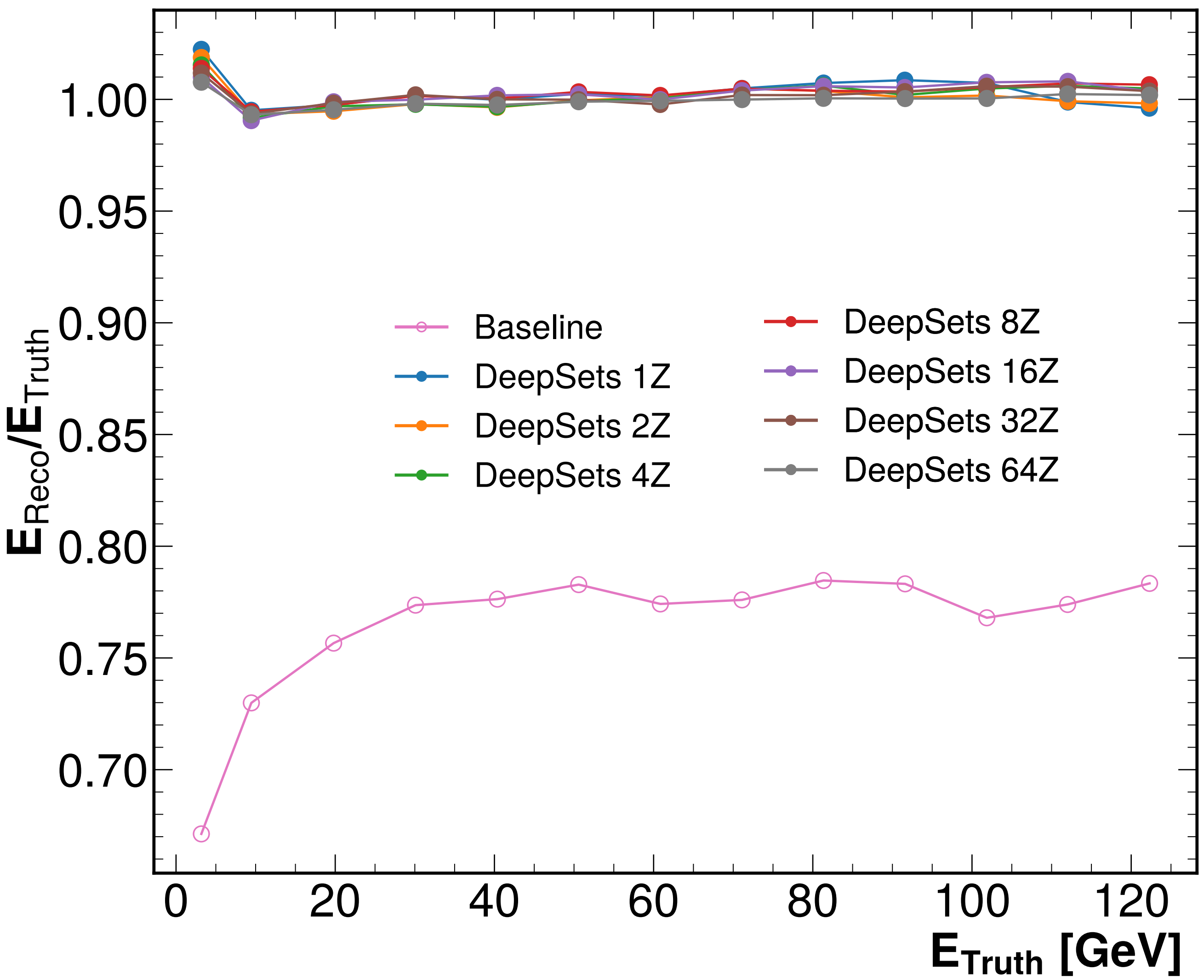
<u>Permutation Invariant</u>
Works well with point clouds
A GNN without edges

<u>arXiv: 1703.06114</u>

<u>arXiv:1810.05165</u>

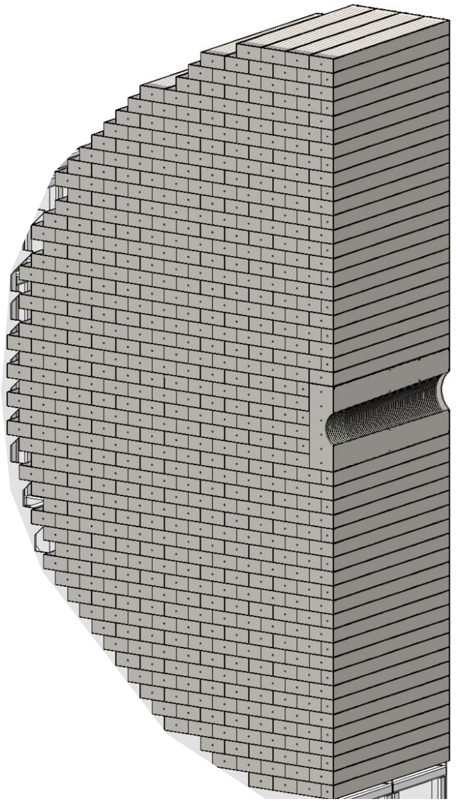
Energy Regression Results



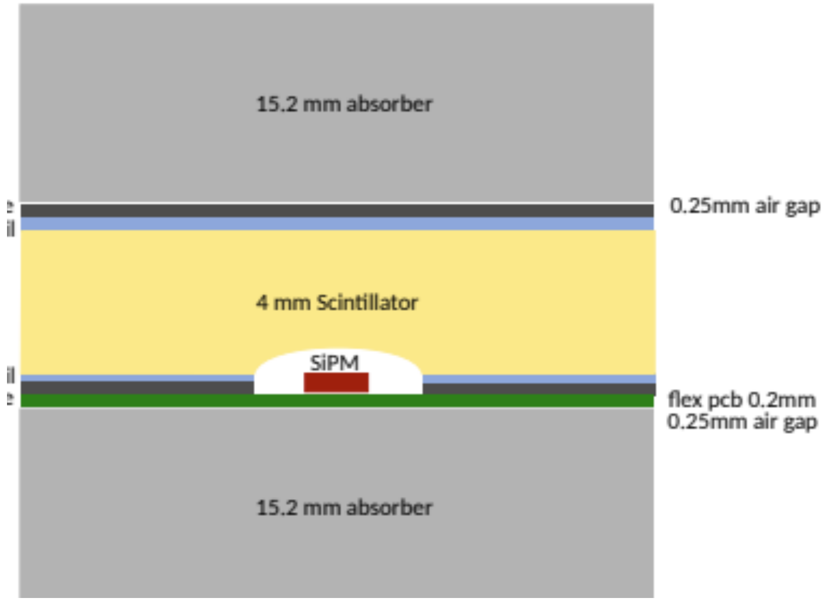


- Geant4 Simulation of single π^+ showers
- Condition model on position of longitudinal segmentation

Forward HCal



- High-granularity iron-scintillator calorimeter
- Forward region, 1.2 < η < 3.5
- Sampling calorimeter comprised of 0.3 cm scintillator tiles sandwiched between 2.0 cm steel plates



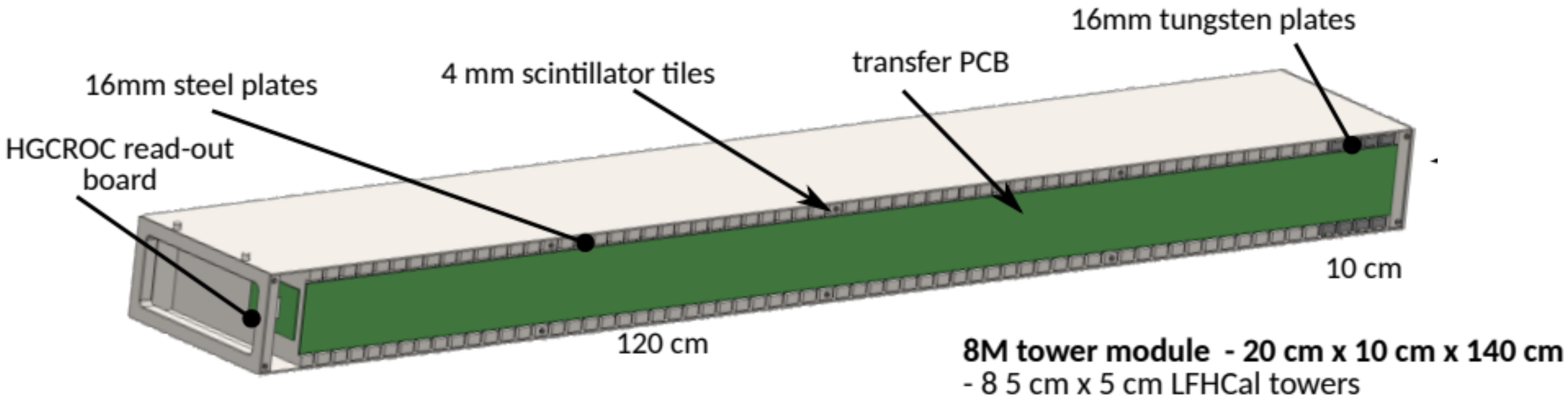


Figure Courtesy ePIC

Observable Motivation

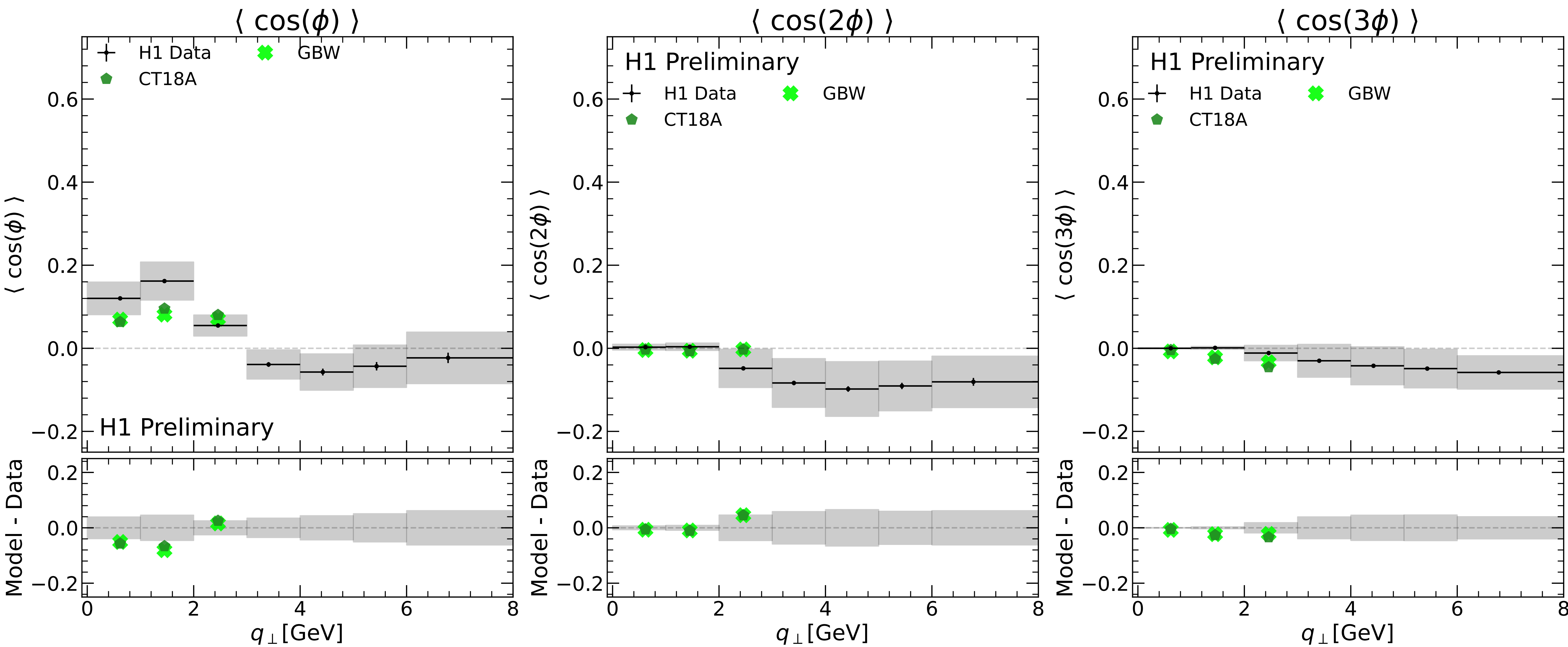
1. Probes soft gluon radiation $S(g)$
   - Soft gluon radiation can be the primary contribution to asymmetry for certain kinematics
   - Asymmetry is Perturbative, test pQCD calculations

2. May represent a vital reference for other signals, in particular TMD PDF measurements
   - Large interest in Lepton-Jet Correlations to probe TMDs
   - In TMD factorization framework, one can factorize contributions from transverse momentum dependent (TMD) PDFs and Soft gluon radiation

3. Observable is sensitive to gluon saturation phenomena, potentially measurable at the EIC

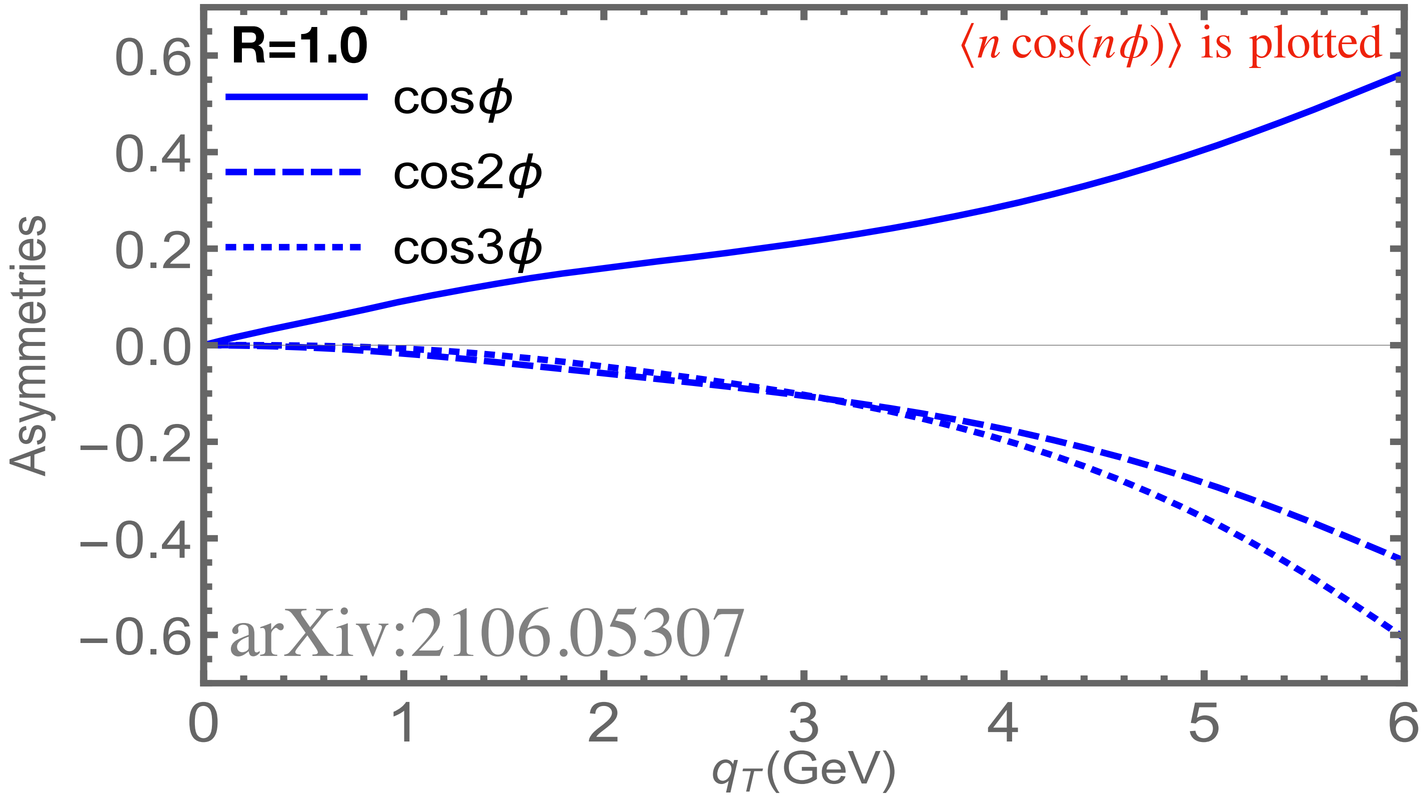
$\langle \cos(n\phi) \rangle$ for n = 1, 2, 3

H1 Unfolded Data



- **Note: Calculations done $q_{\perp} \leq 3.0$ GeV**
- **Differences could be due to sample bin average within the fiducial cuts**
- **CT18A is also a TMD calculation, disagreement could also be in kinematics constraints**

Two Sets of Calculations (Compare 2nd)



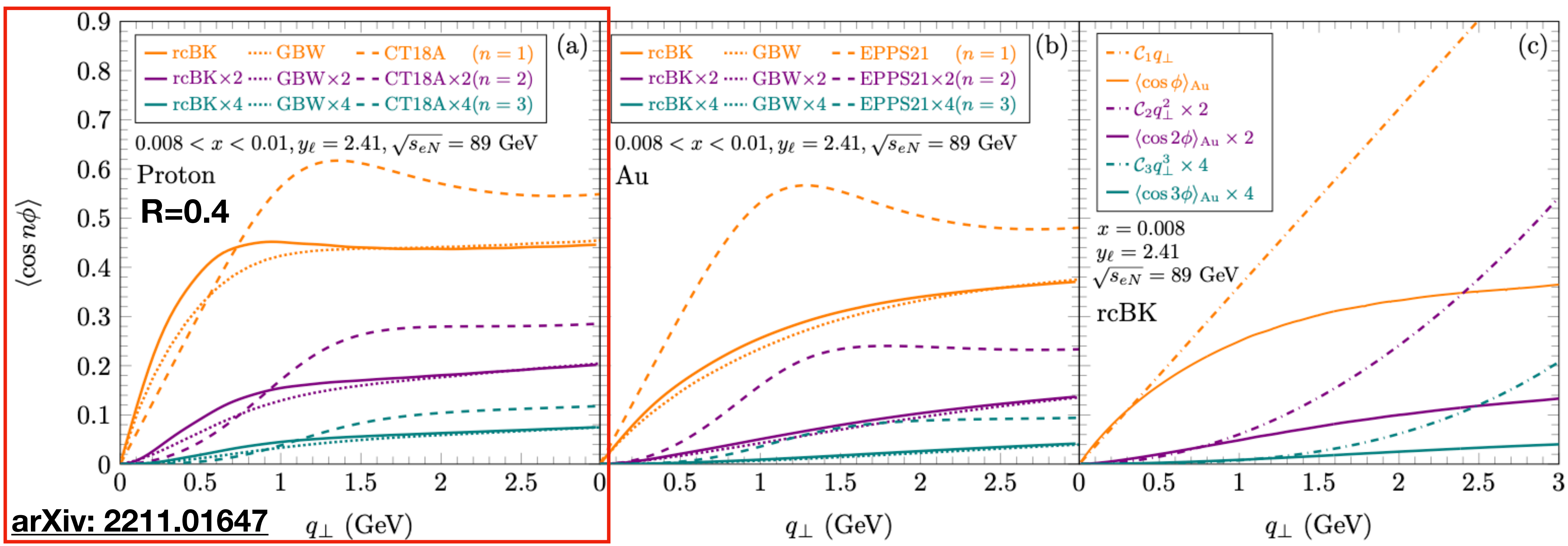
$$\vec{q}_\perp = \vec{k}_{\ell\perp} + \vec{k}_{J\perp}$$

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

$\sqrt{s}$ =140 GeV, $P_\perp$ = 20 GeV, y_l = 1.5, Q = 25 GeV

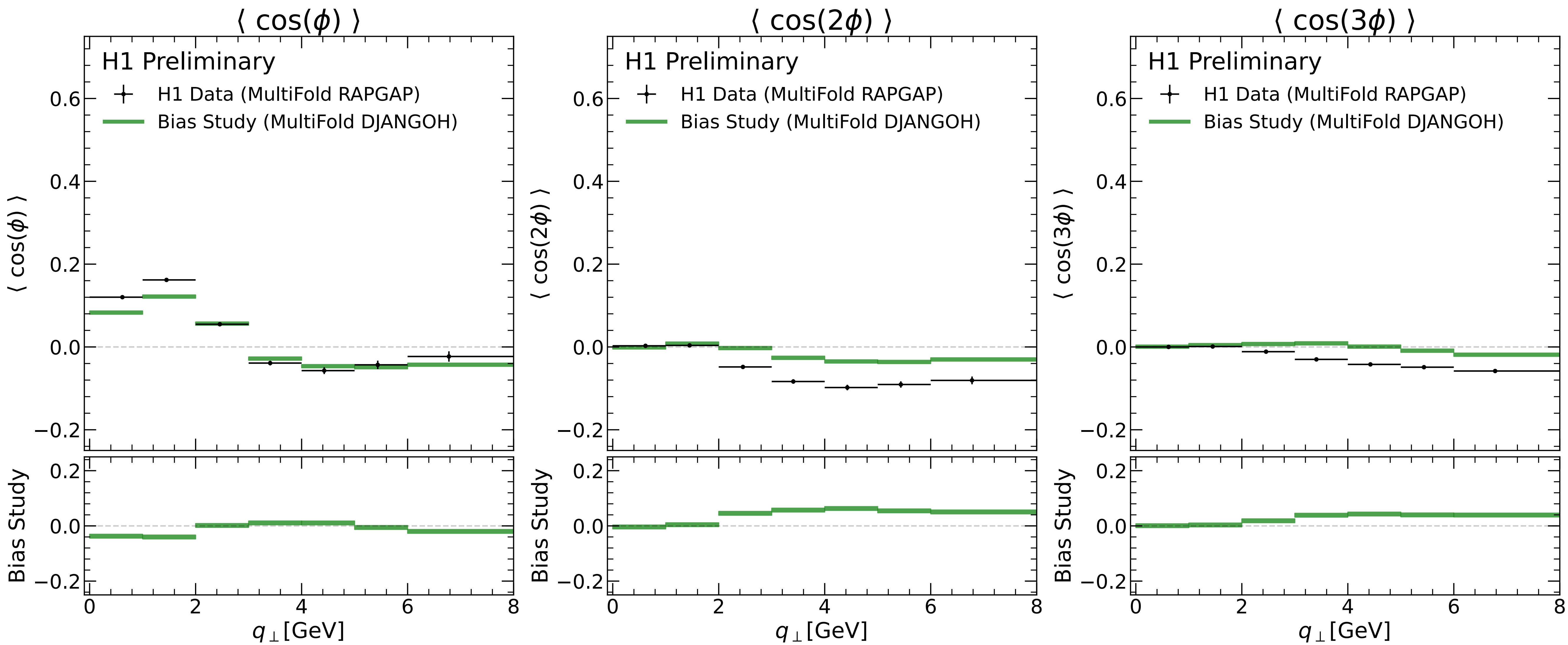
Radiative corrections enhanced $\propto (\alpha_s \ln^2 P_\perp^2/q_\perp^2)^n$

Soft Gluon Resummation



Harmonics of saturation with inputs from *GBW* model and CT18A PDF

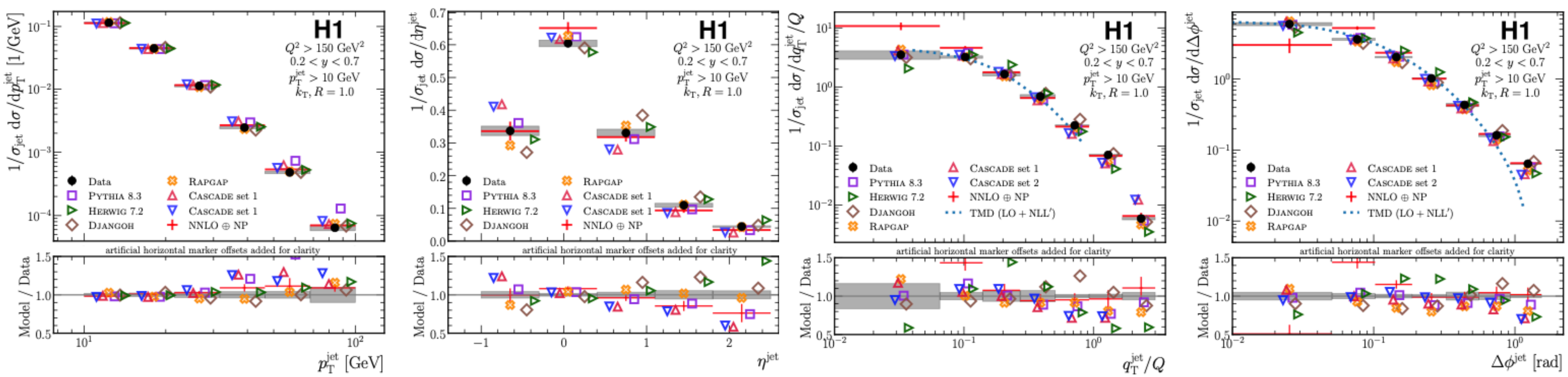
Investigation of Model Bias vs. $q_\perp$ [GeV]



- 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

Backup Further Background

- Machine learning (OmniFold) is used to perform an 8-dimensional, unbinned unfolding.
- 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. $$\omega_n(m) = \nu_{n-1}^{\text{push}}(m)\underline{L[(1,\text{Data}),(\nu_{n-1}^{\text{push}},\text{Sim.})](m)}$$

$$\nu_{n-1}(t) = \nu_n^{\text{push}}(m)$$

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

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

- 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

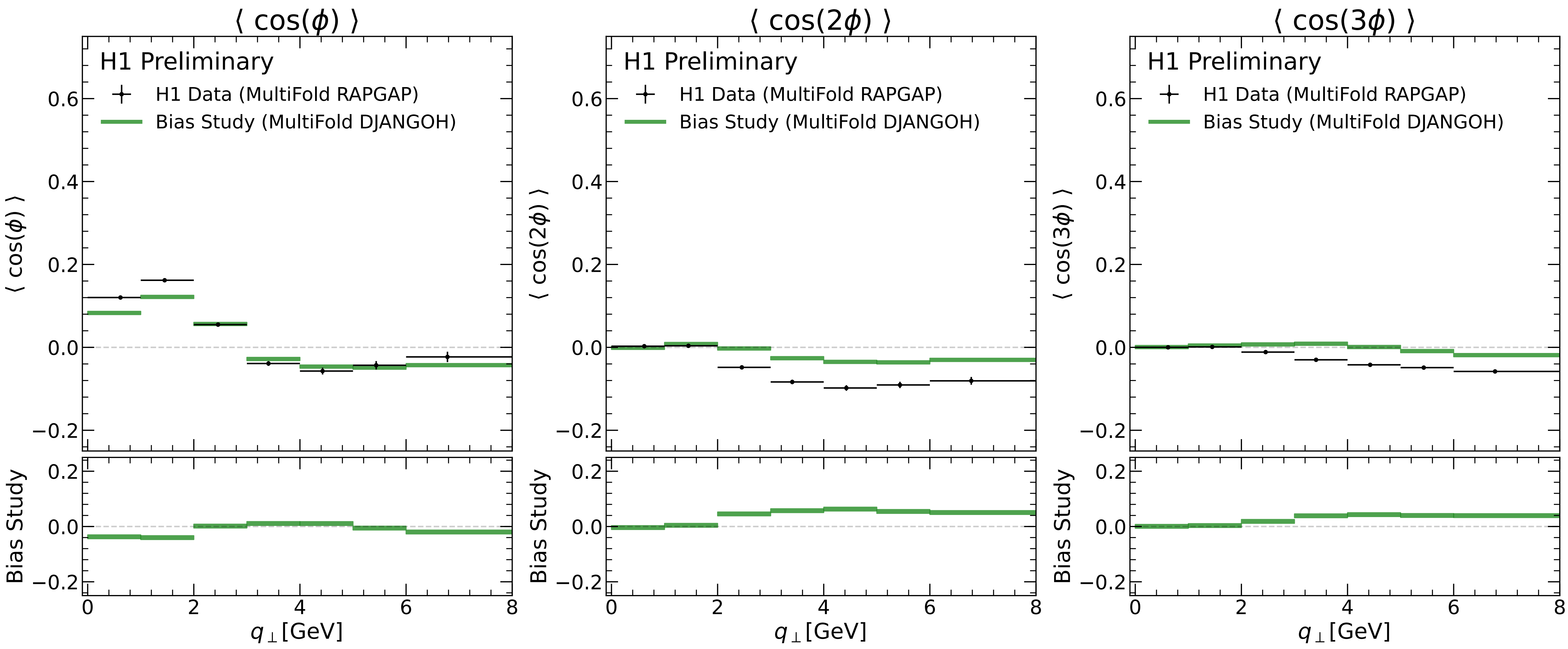
Systematic Uncertainties

- 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: $\pm 0.5 - 1.0\,\%$

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

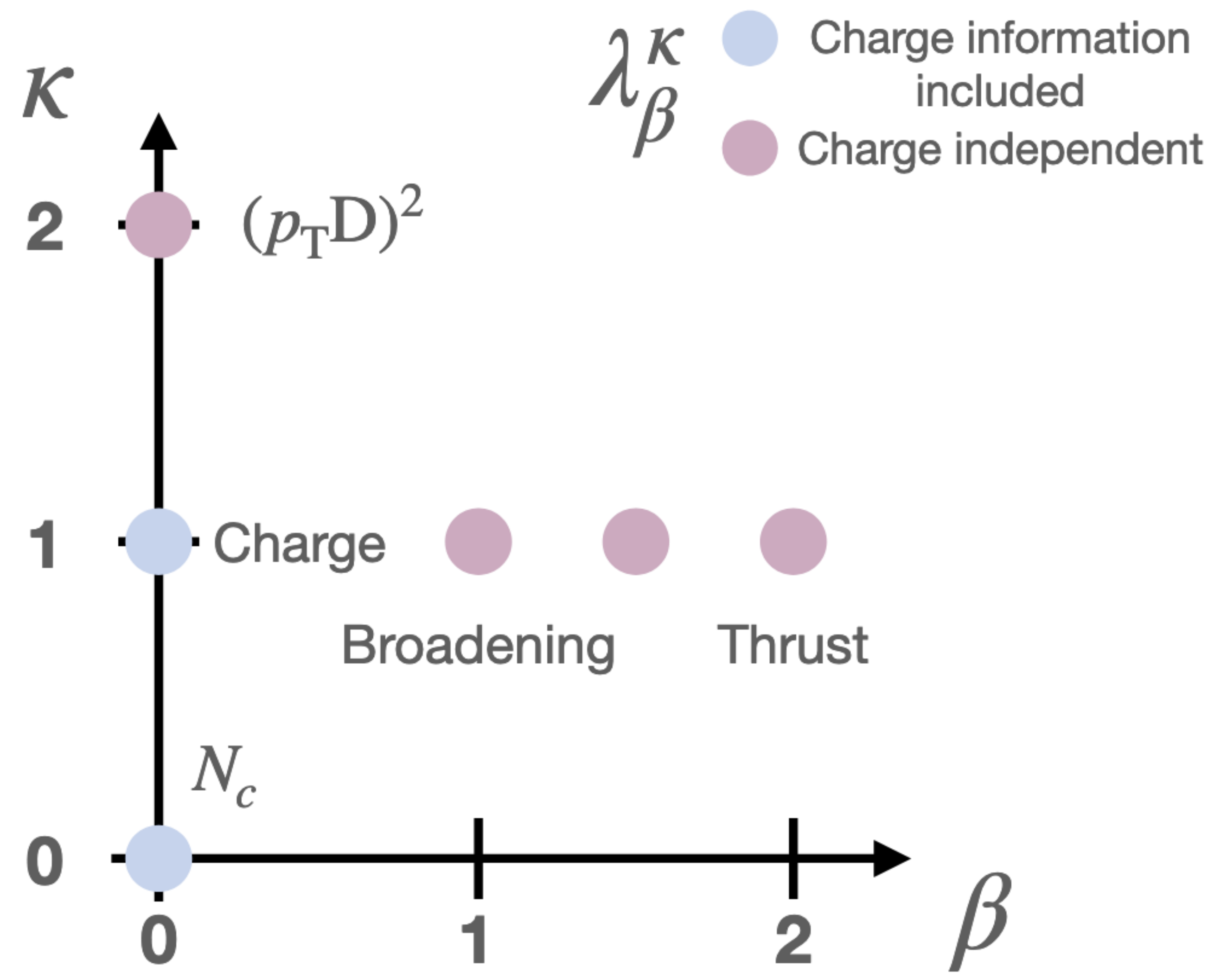


- Leading uncertainty is model bias in the unfolding for $\cos(2\phi)$ and $\cos(3\phi)$
- Difference in the result when unfolding using RAPGAP and DJANGO
- Reporting Abs. Errors; <u>central values are very close to 0.0</u>
- The Total Uncertainty is quite stable between harmonics

Jet Substructure Observables

Description of the jet substructure observables measured in this work.

Name/Symbol	Observable definition	Charge used
Logarithm of jet broadening	$\ln(\lambda_1^1)$	No
Intermediate observable	$\ln(\lambda_{1.5}^1)$	No
Logarithm of jet thrust	$\ln(\lambda_2^1)$	No
Momentum dispersion $p_T D$	$\sqrt{\lambda_0^2}$	No
Charged particle multiplicity N_c	$\tilde{\lambda}_0^0$	Yes
Jet charge Q_1	$\tilde{\lambda}_0^1$	Yes



IBU Generalization

IBU

$$t_j^{(n)} = \sum_i \Pr_{n-1}(\text{truth is } j | \text{measure } i) \Pr(\text{measure } i)$$
$$= \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i$$

Continuous Generalization

$$\nu_1(t) p_{\text{Gen}}(t) = \int dm' \, p_{\text{Gen}|\text{Sim}}(t|m') \, p_{\text{Data}}(m')$$

Using Classifiers that approximate the Likelihood ratio

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

Both converge to maximum likelihood estimate of particle-level distribution

Cross Section & ϕ

$$\frac{d^5\sigma^{ep\to e'qX}}{dy_\ell d^2P_\perp d^2q_\perp} = \sigma_0^{eq} x f_q(x)\delta^{(2)}(q_\perp)$$

Gluon Matrix Element

$$\mathcal{M}^{\mu\nu}(x,k_\perp) = \int \frac{d\xi^- d^2\xi_\perp}{P^+(2\pi)^3} e^{-ixP^+\xi^- + i\vec{k}_\perp\cdot\vec{\xi}_\perp} \qquad (
\times\langle P|F_a^{+\mu}(\xi^-,\xi_\perp)\mathcal{L}^\dagger_{vab}(\xi^-,\xi_\perp)\mathcal{L}_{vbc}(0,0_\perp)F_c^{\nu+}(0)|P\rangle$$

Integration over emitted gluon phase space

$$g^2\int\frac{d^3k_g}{(2\pi)^3 2E_{k_g}}\delta^{(2)}(q_\perp + k_{g\perp})C_F S_g(k_J,p_1)
= \frac{\alpha_s C_F}{2\pi^2 q_\perp^2}\left[\ln\frac{Q^2}{q_\perp^2} + \ln\frac{Q^2}{k_{\ell\perp}^2} + c_0 + 2c_1\cos(\phi) + 2c_2\cos(2\phi) + \cdots\right],$$

$$c_n = \ln\frac{1}{R^2} + f(n) + g(nR)\,,$$

Fourier Coefficient (Introduces ϕ dependance)

$$f(n) = \frac{2}{\pi}\int_0^\pi d\phi(\pi-\phi)\frac{\cos\phi}{\sin\phi}(\cos n\phi - 1)\,,$$

$$g(nR) = \frac{4}{\pi}\int_0^1\frac{d\phi}{\phi}\tan^{-1}\frac{\sqrt{1-\phi^2}}{\phi}\left[1-\cos(nR\phi)\right]
= \frac{n^2R^2}{4}\,{}_2F_3\left(1,1;2,2,2;-\frac{n^2R^2}{4}\right).$$

Differential Cross Section

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

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

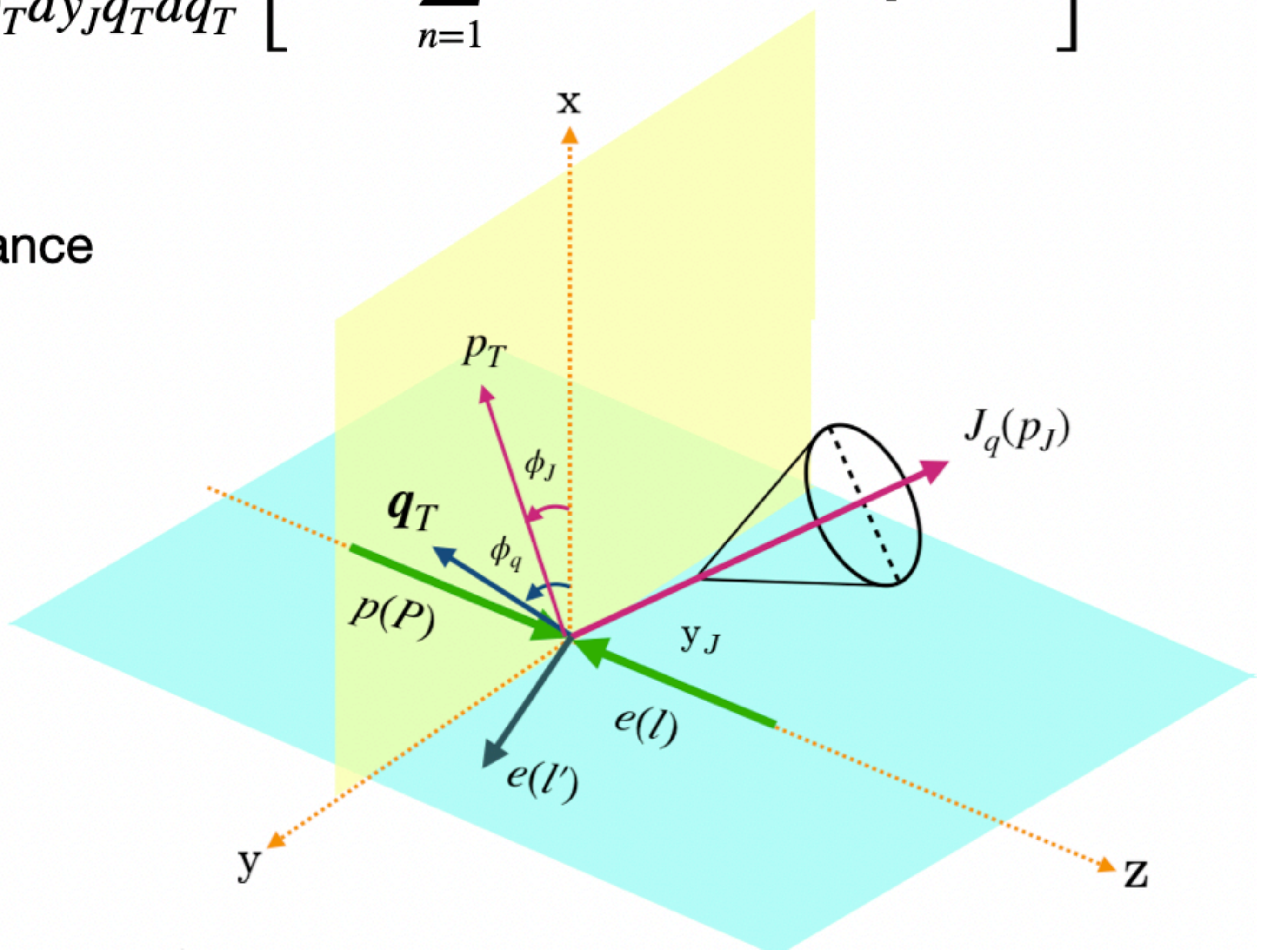
$$\frac{d\sigma}{d^2\boldsymbol{p}_T dy_J d\phi_J d^2\boldsymbol{q}_T} = \frac{d\sigma}{2\pi d^2\boldsymbol{p}_T dy_J q_T dq_T}\left[1 + 2\sum_{n=1}^{\infty} v_n(p_T, y_T)\cos(n(\phi_q - \phi_J))\right]$$

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