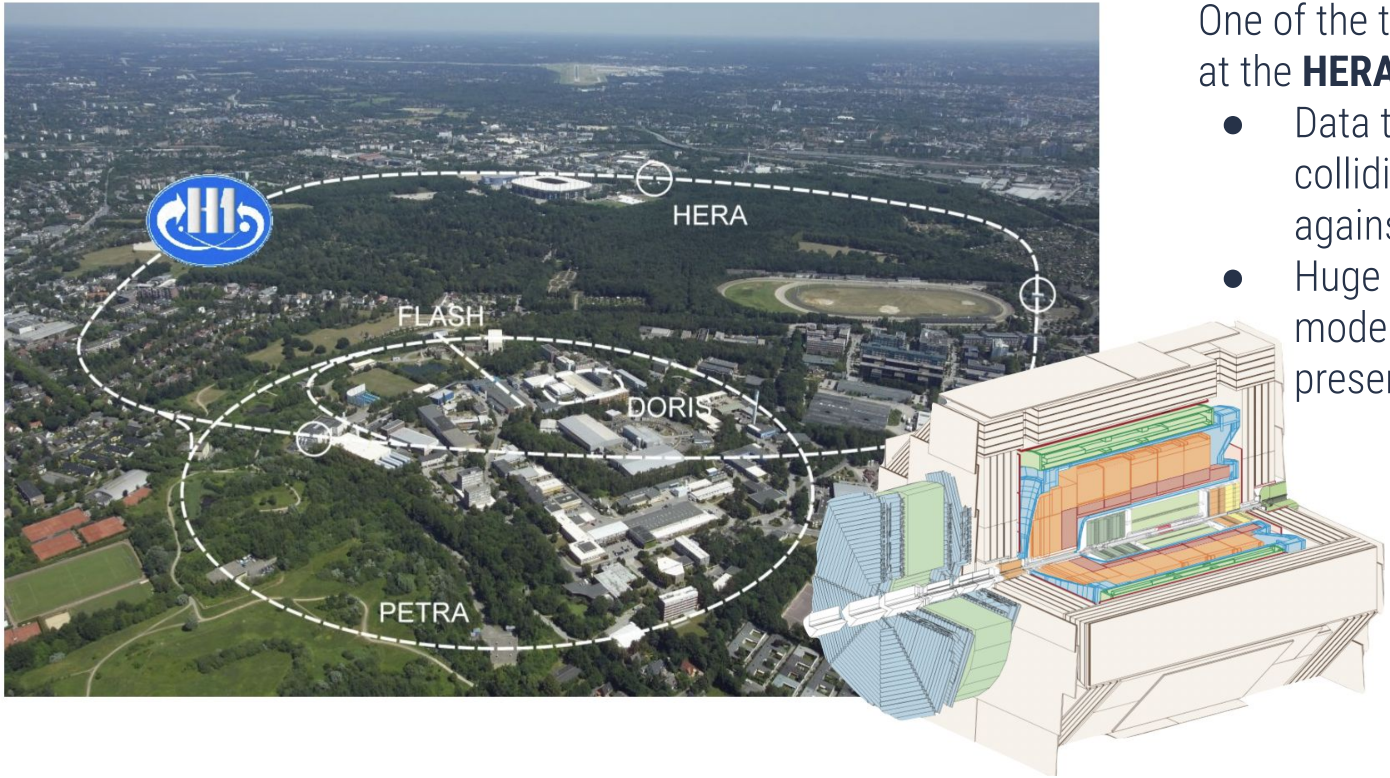
Machine Learning-Assisted Measurement of Lepton-Jet Azimuthal Angular Asymmetries and of the complete final state in Deep-Inelastic Scattering at HERA

Vinicius Mikuni on behalf of the H1 Collaboration

The H1 Detector

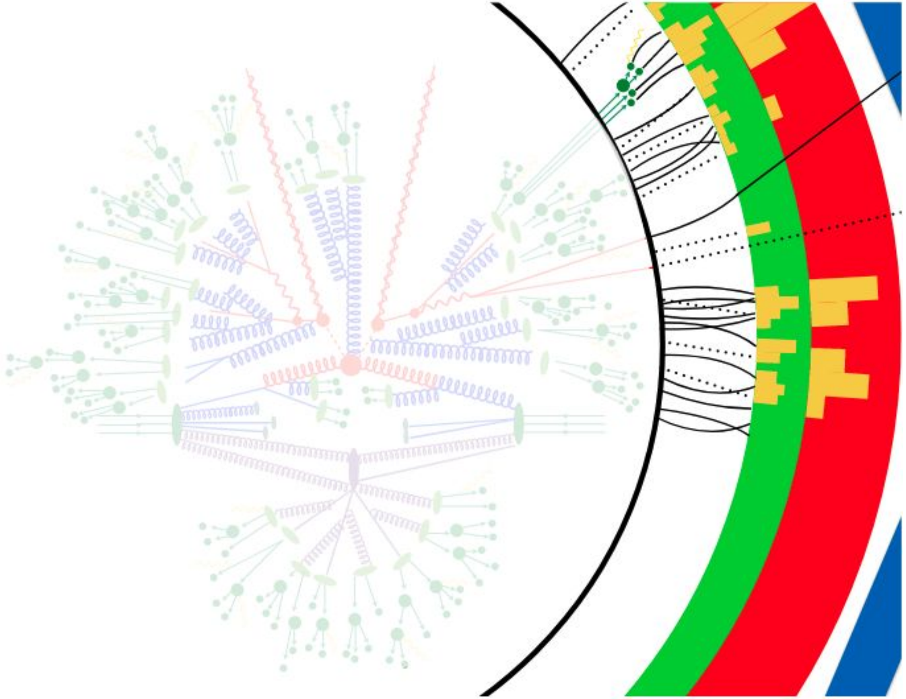


One of the two multipurpose detectors at the **HERA** accelerator facility

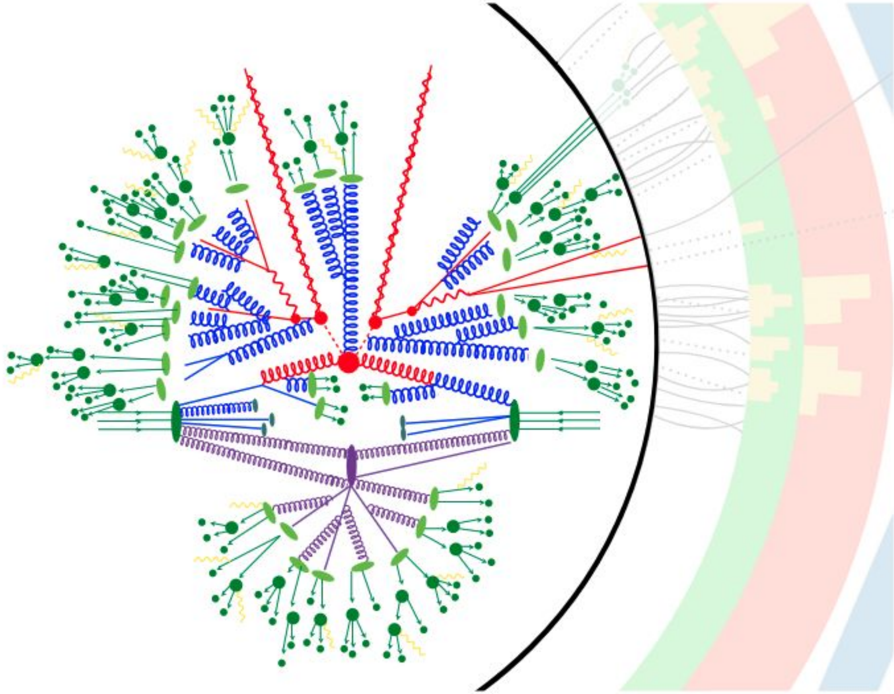
- Data taking from 1992 to 2007 colliding **electrons/positrons** against **protons**
- Huge **data preservation effort** to modernize the software and preserve the data

Unfolding

What we measure

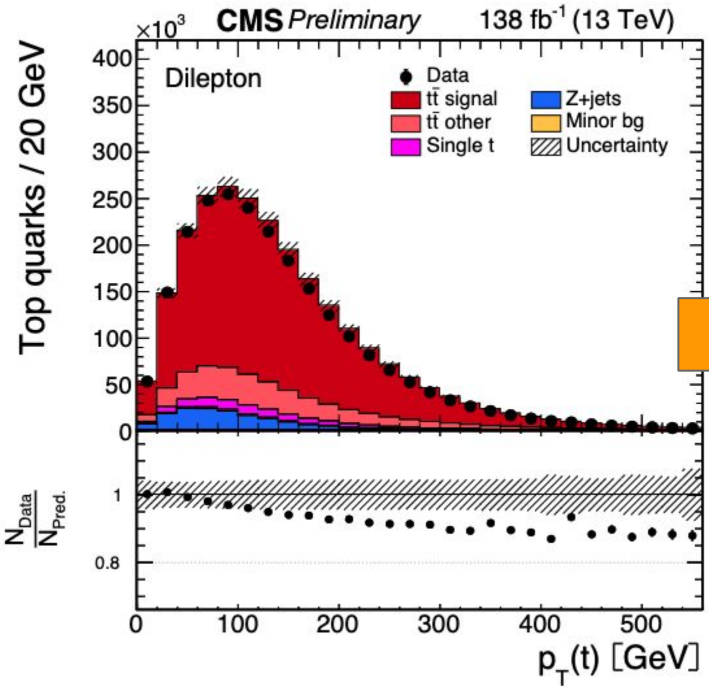


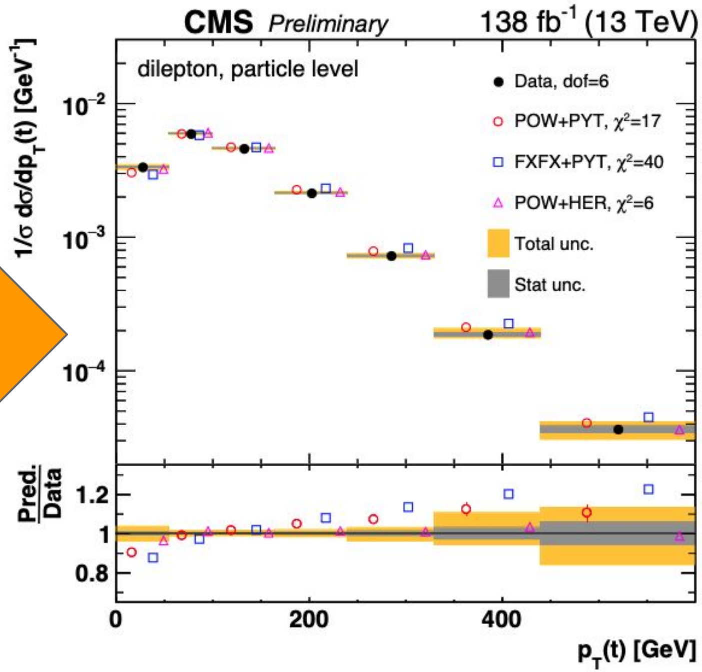
What we want



Unfolding

Source: CMS-PAS-TOP-20-006





Unfolding

How to define the **optimal binning**?

- Choice depends on the **distribution** and **phase space**
- Need to compromise when **combining** results from **different experiments**

Unfolding

How to define the **optimal binning**?

- Choice depends on the **distribution** and **phase space**
- Need to compromise when **combining** results from **different experiments**

How to include **multiple distributions**?

- Histograms are hard to scale: **curse of dimensionality**
- Unfolding uncertainties can be reduced using **additional observables**

Unfolding

How to define the **optimal binning**?

- Choice depends on the **distribution** and **phase space**
- Need to compromise when **combining** results from **different experiments**

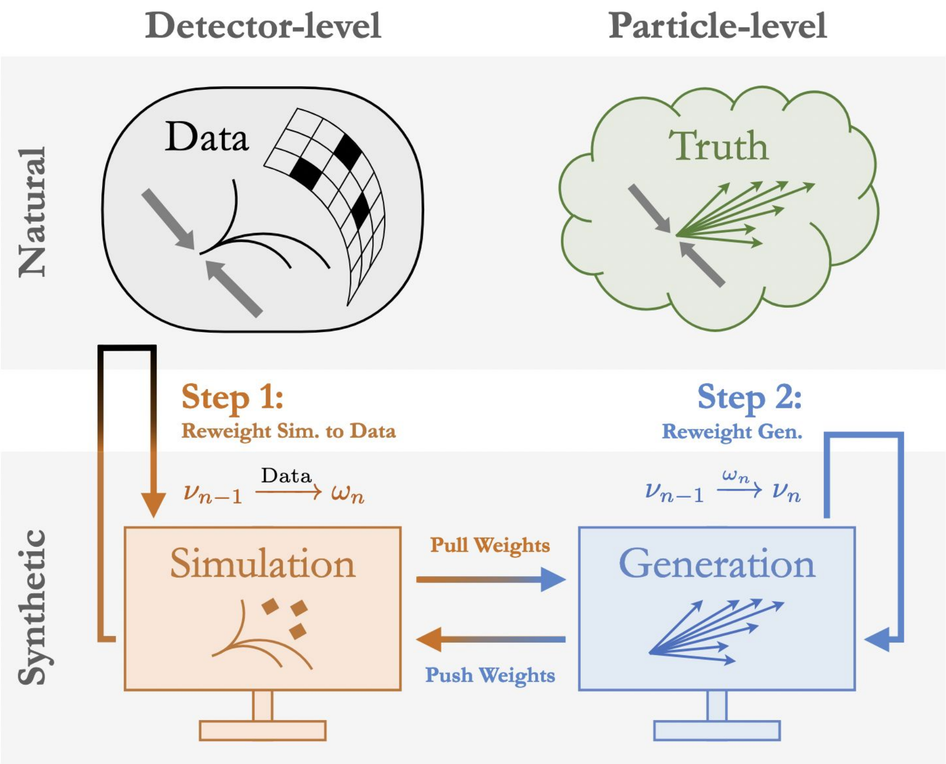
How to include **multiple distributions**?

- Histograms are hard to scale: **curse of dimensionality**
- Unfolding uncertainties can be reduced using **additional observables**

How to unfold distributions that are **not** defined for each event?

- Moments of distributions
- Energy Correlators

ML Based 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!

Source: Andreassen et al. PRL 124, 182001 (2020)

ML Based Unfolding

PHYSICAL REVIEW LETTERS **128,** 132002 (2022)

Measurement of Lepton-Jet Correlation in Deep-Inelastic Scattering with the H1 Detector Using Machine Learning for Unfolding

V. Andreev,[23] M. Arratia,[35] A. Baghdasaryan,[46] A. Baty,[16] K. Begzsuren,[39] A. Belousov,[23,*] A. Bolz,[14] V. Boudry,[31] G. Brandt,[13] D. Britzger,[26] A. Buniatyan,[6] L. Bystritskaya,[22] A. J. Campbell,[14] K. B. Cantun Avila,[47] K. Cerny,[28] V. Chekelian,[26] Z. Chen,[37] J. G. Contreras,[47] L. Cunqueiro Mendez,[27] J. Cvach,[33] J. B. Dainton,[19] K. Daum,[45] A. Deshpande,[38] C. Diaconu,[21] G. Eckerlin,[14] S. Egli,[43] E. Elsen,[14] L. Favart,[4] A. Fedotov,[22] J. Feltesse,[12] M. Fleischer,[14] A. Fomenko,[23] C. Gal,[38] J. Gayler,[14] L. Goerlich,[17] N. Gogitidze,[23] M. Gouzevitch,[42] C. Grab,[49] T. Greenshaw,[19] G. Grindhammer,[26] D. Haidt,[14] R. C. W. Henderson,[18] J. Hessler,[26] J. Hladký,[33] D. Hoffmann,[21] R. Horisberger,[43] T. Hreus,[50] F. Huber,[15] P. M. Jacobs,[5] M. Jacquet,[29] T. Janssen,[4] A. W. Jung,[44] H. Jung,[14] M. Kapichine,[10] J. Katzy,[14] C. Kiesling,[26] M. Klein,[19] C. Kleinwort,[14] H. T. Klest,[38] R. Kogler,[14] P. Kostka,[19] J. Kretzschmar,[19] D. Krücker,[14] K. Krüger,[14] M. P. J. Landon,[20] W. Lange,[48] P. Laycock,[41] S. H. Lee,[3] S. Levonian,[14] W. Li,[16] J. Lin,[16] K. Lipka,[14] B. List,[14] J. List,[14] B. Lobodzinski,[26] E. Malinovski,[23] H.-U. Martyn,[1] S. J. Maxfield,[19] A. Mehta,[19] A. B. Meyer,[14] J. Meyer,[14] S. Mikocki,[17] M. M. Mondal,[38] A. Morozov,[10] K. Müller,[50] B. Nachman,[5] Th. Naumann,[48] P. R. Newman,[6] C. Niebuhr,[14] G. Nowak,[17] J. E. Olsson,[14] D. Ozerov,[43] S. Park,[38] C. Pascaud,[29] G. D. Patel,[19] E. Perez,[11] A. Petrukhin,[42] I. Picuric,[32] D. Pitzl,[14] R. Polifka,[34] S. Preins,[35] V. Radescu,[30] N. Raicevic,[32] T. Ravdandorj,[39] P. Reimer,[33] E. Rizvi,[20] P. Robmann,[50] R. Roosen,[4] A. Rostovtsev,[25] M. Rotaru,[7] D. P. C. Sankey,[8] M. Sauter,[15] E. Sauvan,[21,2] S. Schmitt,[14] B. A. Schmookler,[38] L. Schoeffel,[12] A. Schöning,[15] F. Sefkow,[14] S. Shushkevich,[24] Y. Soloviev,[23] P. Sopicki,[17] D. South,[14] V. Spaskov,[10] A. Specka,[31] M. Steder,[14] B. Stella,[36] U. Straumann,[50] C. Sun,[37] T. Sykora,[34] P. D. Thompson,[6] D. Traynor,[20] B. Tseepeldorj,[39,40] Z. Tu,[41] A. Valkárová,[34] C. Vallée,[21] P. Van Mechelen,[4] D. Wegener,[9] E. Wünsch,[14] J. Žáček,[34] J. Zhang,[37] Z. Zhang,[29] R. Žlebčík,[34] H. Zohrabyan,[46] and F. Zomer[29]

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

V. Andreev[47], M. Arratia[31], A. Baghdasaryan[43], A. Baty[8], K. Begzsuren[37], A. Bolz[15], V. Boudry[27], G. Brandt[14], D. Britzger[11], A. Buniatyan[6], L. Bystritskaya[47], A.J. Campbell[15], K.B. Cantun Avila[44], K. Cerny[25], V. Chekelian[11], Z. Chen[33], J.G. Contreras[44], J. Cvach[29], J.B. Dainton[21], K. Daum[42], A. Deshpande[35,39], C. Diaconu[23], A. Drees[35], G. Eckerlin[15], S. Egli[40], E. Elsen[15], L. Favart[3], A. Fedotov[47], J. Feltesse[13], M. Fleischer[15], A. Fomenko[47], C. Gal[35], J. Gayler[15], L. Goerlich[19], N. Gogitidze[15], M. Gouzevitch[47], C. Grab[45], T. Greenshaw[21], G. Grindhammer[11], D. Haidt[15], R.C.W. Henderson[20], J. Hessler[11], J. Hladký[29], D. Hoffmann[23], R. Horisberger[40], T. Hreus[46], F. Huber[16] P.M. Jacobs[4], M. Jacquet[26], T. Janssen[3], A.W. Jung[41], J. Katzy[15], C. Kiesling[11], M. Klein[*,21], C. Kleinwort[15], H.T. Klest[18], R. Kogler[15], P. Kostka[21], J. Kretzschmar[21], D. Krücker[15], K. Krüger[15], M.P.J. Landon[22], W. Lange[15], P. Laycock[39], S.H. Lee[36], S. Levonian[15], W. Li[17], J. Lin[17], K. Lipka[15], B. List[15], J. List[15], B. Lobodzinski[11], O.R. Long[31], E. Malinovski[47], H.-U. Martyn[1], S.J. Maxfield[21], A. Mehta[21], A.B. Meyer[15], J. Meyer[15], S. Mikocki[19], V.M. Mikuni[4], M.M. Mondal[24], K. Müller[46], B. Nachman[4], Th. Naumann[15], P.R. Newman[6], C. Niebuhr[15], G. Nowak[19], J.E. Olsson[15], D. Ozerov[47], S. Park[35], C. Pascaud[26], G.D. Patel[21], E. Perez[12], A. Petrukhin[34], I. Picuric[28], D. Pitzl[15], V. Radescu[16], N. Raicevic[28], T. Ravdandorj[37], D. Reichelt[12], P. Reimer[29], E. Rizvi[22], P. Robmann[46], R. Roosen[3], A. Rostovtsev[47], M. Rotaru[7], D.P.C. Sankey[9], M. Sauter[16], E. Sauvan[23,2], S. Schmitt[**,15], B.A. Schmookler[35], G. Schnell[5], L. Schoeffel[13], A. Schöning[16], F. Sefkow[15], S. Shushkevich[11], Y. Soloviev[15], P. Sopicki[19], D. South[15], A. Specka[27], M. Steder[15], B. Stella[32], U. Straumann[46], C. Sun[35], T. Sykora[30], P.D. Thompson[6], F. Torales Acosta[4], D. Traynor[22], B. Tseepeldorj[37,38], Z. Tu[39], G. Tustin[35], A. Valkárová[30], C. Vallée[23], P. Van Mechelen[3], D. Wegener[10], E. Wünsch[15], J. Žáček[30], J. Zhang[33], Z. Zhang[26], R. Žlebčík[30], H. Zohrabyan[43], F. Zomer[26]

Physics Letters B 844 (2023) 138101

ELSEVIER

Contents lists available at ScienceDirect

Physics Letters B

journal homepage: www.elsevier.com/locate/physletb

PHYSICS LETTERS B

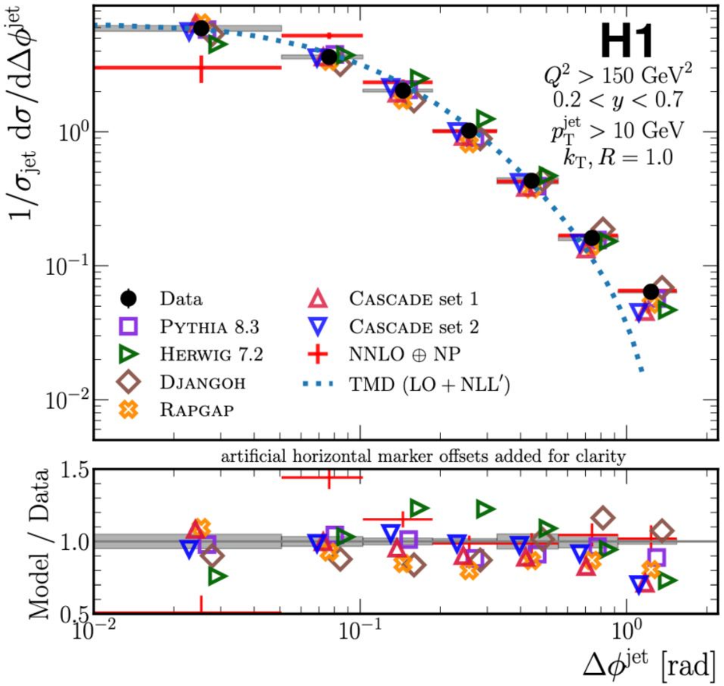
Unbinned deep learning jet substructure measurement in high Q^2 ep collisions at HERA

Check for updates

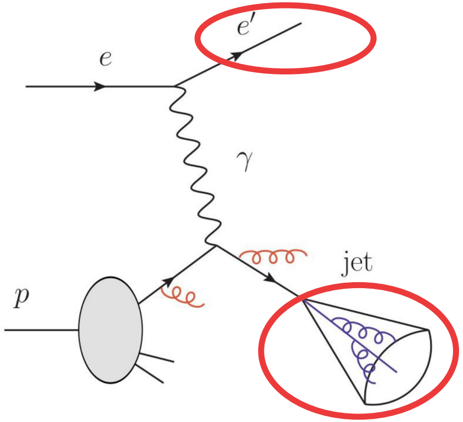
V. Andreev[ar], M. Arratia[ac], A. Baghdasaryan[an], A. Baty[p], K. Begzsuren[ah], A. Bolz[n], V. Boudry[y], G. Brandt[m], D. Britzger[v], A. Buniatyan[g], L. Bystritskaya[ar], A.J. Campbell[n], K.B. Cantun Avila[ao], K. Cerny[w], V. Chekelian[v], Z. Chen[ae], J.G. Contreras[ao], J. Cvach[aa], J.B. Dainton[s], K. Daum[am], A. Deshpande[ag,aj], C. Diaconu[u], A. Drees[ag], G. Eckerlin[n], S. Egli[ak], E. Elsen[n], L. Favart[d], A. Fedotov[ar], J. Feltesse[l], M. Fleischer[n], A. Fomenko[ar], C. Gal[ag], J. Gayler[n], L. Goerlich[q], N. Gogitidze[n], M. Gouzevitch[ar], C. Grab[ap], T. Greenshaw[s], G. Grindhammer[v], D. Haidt[n], R.C.W. Henderson[r], J. Hessler[v], J. Hladký[aa], D. Hoffmann[u], R. Horisberger[ak], T. Hreus[aq], F. Huber[o], P.M. Jacobs[e], M. Jacquet[x], T. Janssen[d], A.W. Jung[al], J. Katzy[n], C. Kiesling[v], M. Klein[s], C. Kleinwort[n], H.T. Klest[ag], R. Kogler[n], P. Kostka[s], J. Kretzschmar[s], D. Krücker[n], K. Krüger[n], M.P.J. Landon[t], W. Lange[n], P. Laycock[aj], S.H. Lee[b], S. Levonian[n], W. Li[p], J. Lin[p], K. Lipka[n], B. List[n], J. List[n], B. Lobodzinski[v], O.R. Long[ac], E. Malinovski[ar], H.-U. Martyn[a], S.J. Maxfield[s], A. Mehta[s], A.B. Meyer[n], J. Meyer[n], S. Mikocki[q], V.M. Mikuni[e], M.M. Mondal[ag], K. Müller[aq], B. Nachman[e], Th. Naumann[n], P.R. Newman[g], C. Niebuhr[n], G. Nowak[q], J.E. Olsson[n], D. Ozerov[ar], S. Park[ag], C. Pascaud[x], G.D. Patel[s], E. Perez[k], A. Petrukhin[af], I. Picuric[z], D. Pitzl[n], R. Polifka[ab], S. Preins[ac], V. Radescu[o], N. Raicevic[z], T. Ravdandorj[ah], P. Reimer[aa], E. Rizvi[t], P. Robmann[aq], R. Roosen[d], A. Rostovtsev[ar], M. Rotaru[h], D.P.C. Sankey[i], M. Sauter[o], E. Sauvan[u,c], S. Schmitt[n,*], B.A. Schmookler[ag], G. Schnell[f], L. Schoeffel[l], A. Schöning[o], F. Sefkow[n], S. Shushkevich[v], Y. Soloviev[n], P. Sopicki[q], D. South[n], A. Specka[y], M. Steder[n], B. Stella[ad], U. Straumann[aq], C. Sun[ag], T. Sykora[ab], P.D. Thompson[g], F. Torales Acosta[e], D. Traynor[t], B. Tseepeldorj[ah,ai], Z. Tu[aj], G. Tustin[ag], A. Valkárová[ab], C. Vallée[u], P. Van Mechelen[d], D. Wegener[j], E. Wünsch[n], J. Žáček[ab], J. Zhang[ae], Z. Zhang[x], R. Žlebčík[ab], H. Zohrabyan[an], F. Zomer[x]

3 papers on ML-based unfolding using H1 data

Azimuthal Asymmetries



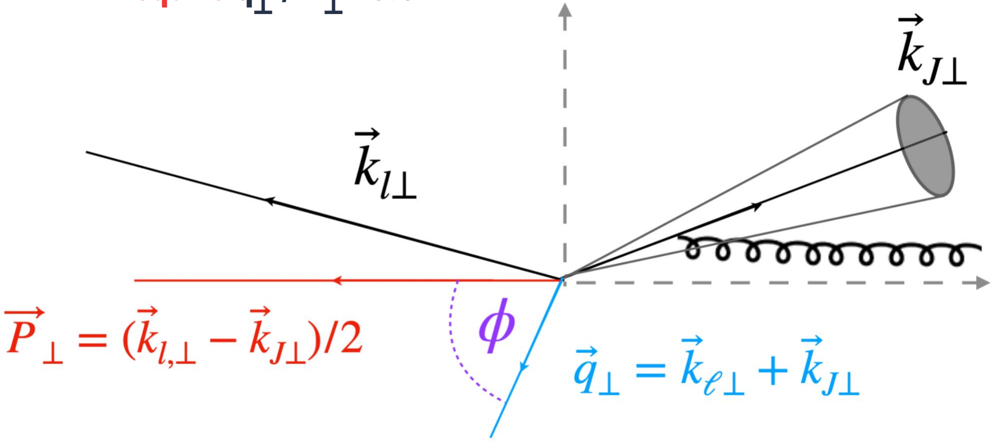
Study of correlations between the scattered lepton and jet



Phys.Rev.Lett. 128 (2022) 13, 132002

Azimuthal Asymmetries

Require $q_\perp / P_\perp > 0.3$



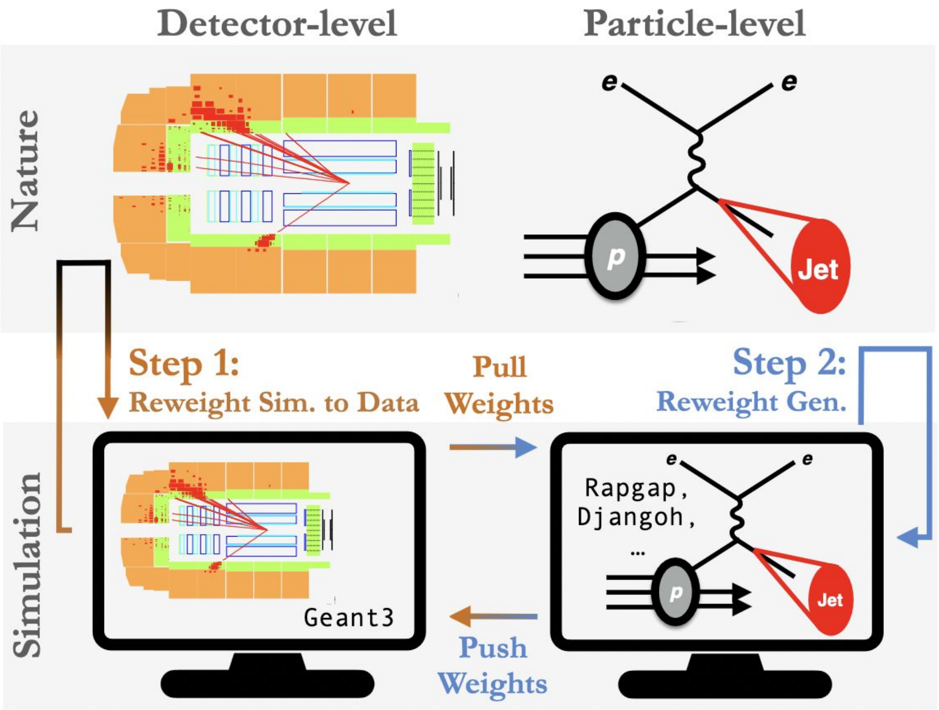
- $k_{l\perp}$: transverse momentum of the scattered lepton
- $k_{J\perp}$: transverse momentum of the jet

Final state lepton and jet are **mostly** back-to-back

- **Imbalance** can arise from perturbative initial/final state radiation
- Target the region where the asymmetry is dominated by **soft gluon emissions:** $P_\perp \gg q_\perp$
- Provide information for **TMD PDF measurements** where the soft gluon contribution can be factorized

Measure: **cos(ϕ), cos(2ϕ), cos(3ϕ)**

Azimuthal Asymmetries



Reuse the results presented at **PRL. 128, 132002** on the measurement of lepton-jet correlations

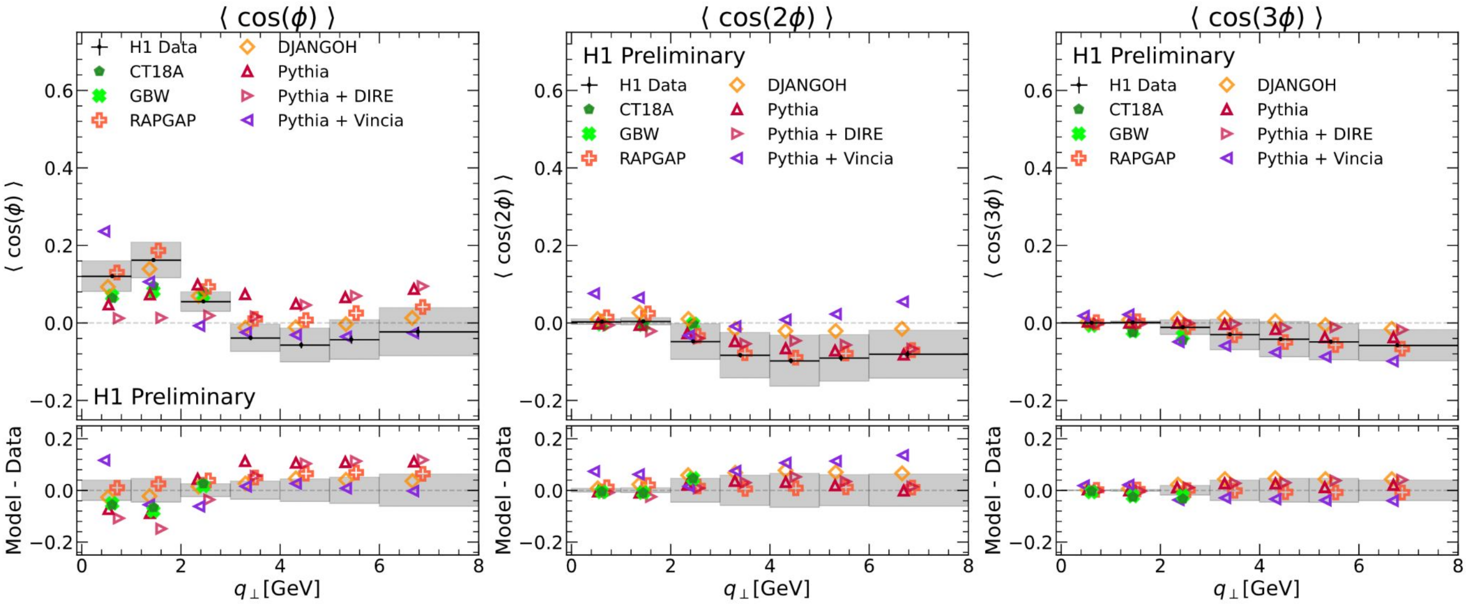
- Quantities previously unfolded

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

$$\cos(\phi) = \frac{\vec{q_\perp}.\vec{P_\perp}}{|\vec{q_\perp}|.|P_\perp|}$$

Results

arXiv:2412.14092, submitted to PLB



Dedicated DIS generators do a good job **everywhere**, especially **Rapgap**

Pythia predictions not tuned to this data

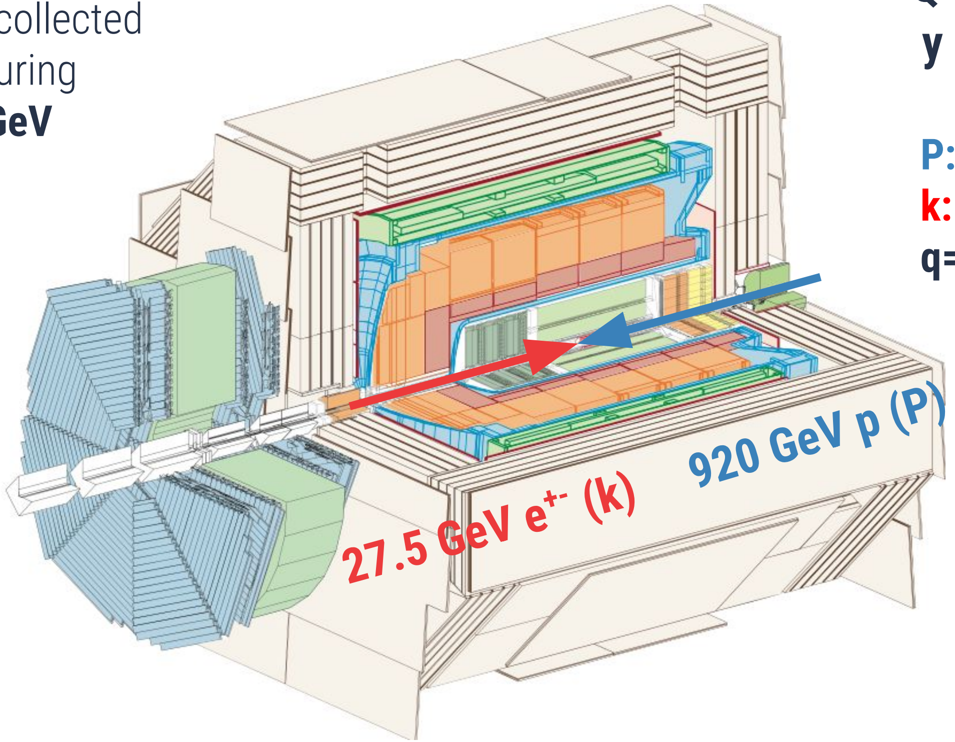
GBW Includes gluon saturation effects while **CT18A** uses NLO TMD calculations with collinear PDFs, both currently available only for low $q_\perp$

What if we unfolded everything?

Experimental setup

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

Goal: Include the information of **all reconstructed particles + scattered lepton** in the collision



$Q^2 = -q^2$

$y = Pq / pk$

P: incoming proton 4-vector

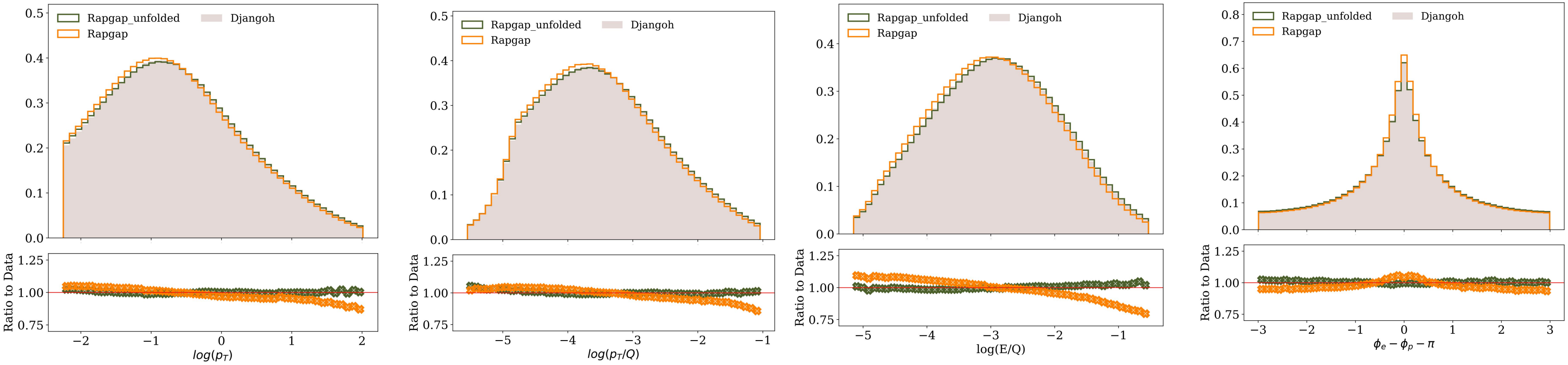
k: incoming electron 4-vector

q=k-k' : 4-momentum transfer

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

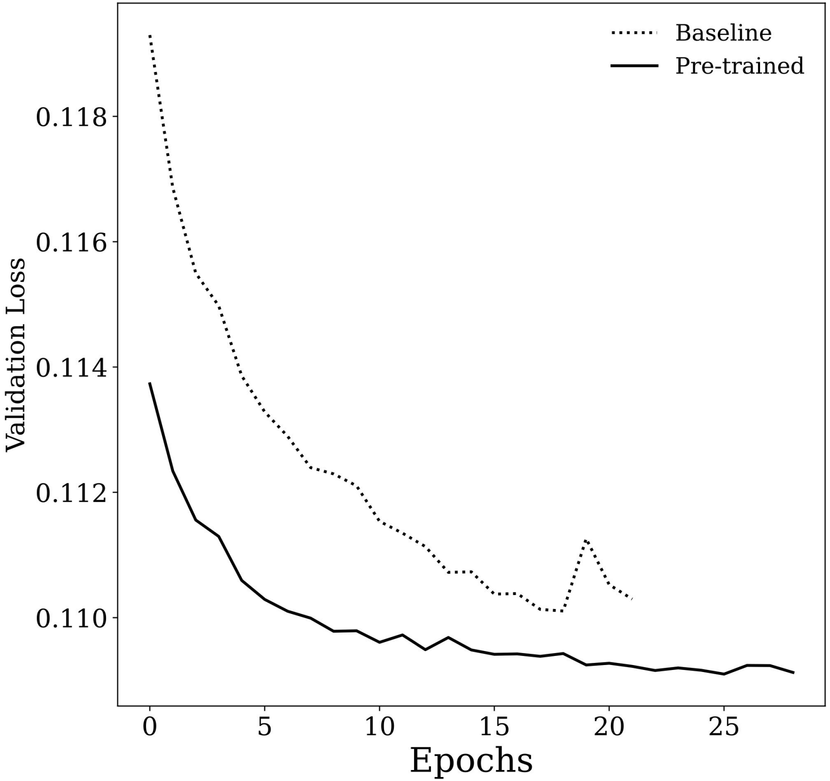
Closure test

- Use **Djangoh** as the **pseudo-data** and unfold **Rapgap**



- Features used during the unfolding:
  - Kinematic information of all hadrons and scattered lepton

Pretraining



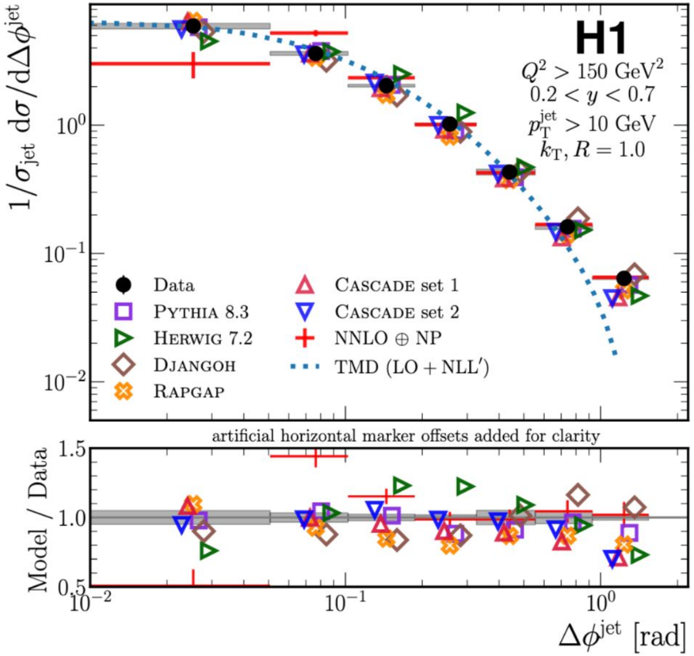
We would like to unfold up to 130*3 = 390 features simultaneously: **requires lots of data**

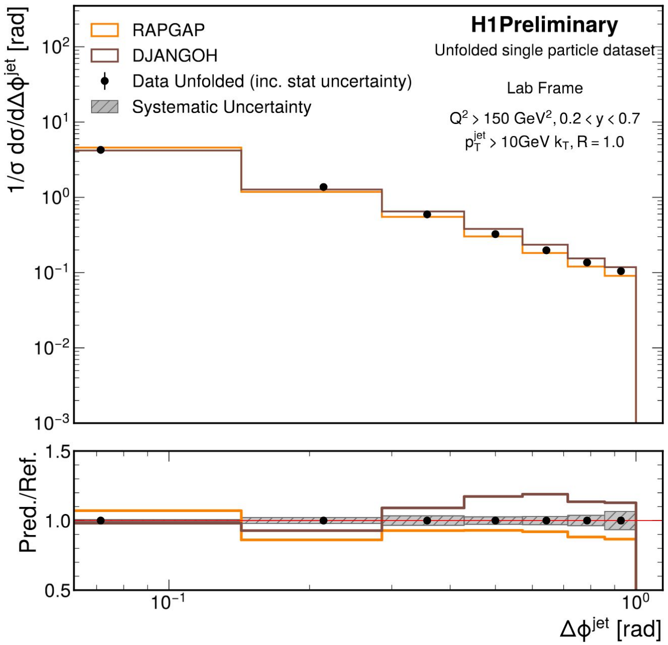
- Our data size is around 500k events, but we have 20M simulations for 2 different simulators
- **Idea**: Pretrain a model using only simulations and then **fine-tune** this model with data
- Use this model as the starting point for the rest of all trainings needed for the unfolding

Results

Cluster unfolded jets using kT algorithm with radius of 1.0

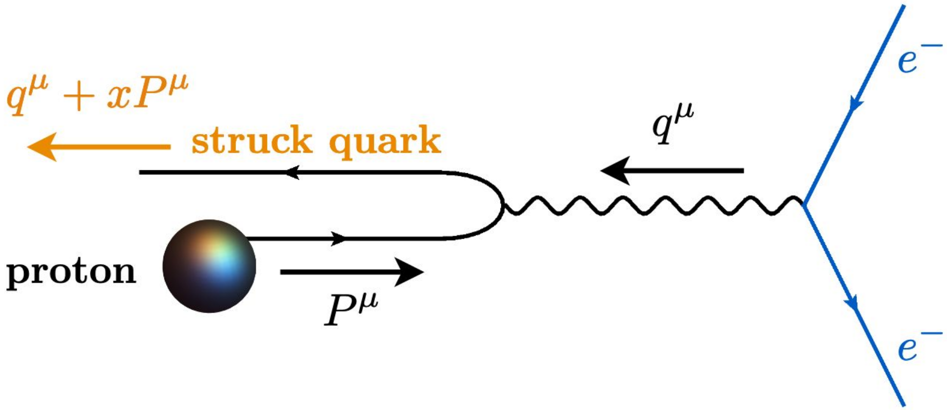
We are able to re-derive **past results**





Phys. Rev. Lett.(128) 132002

Results

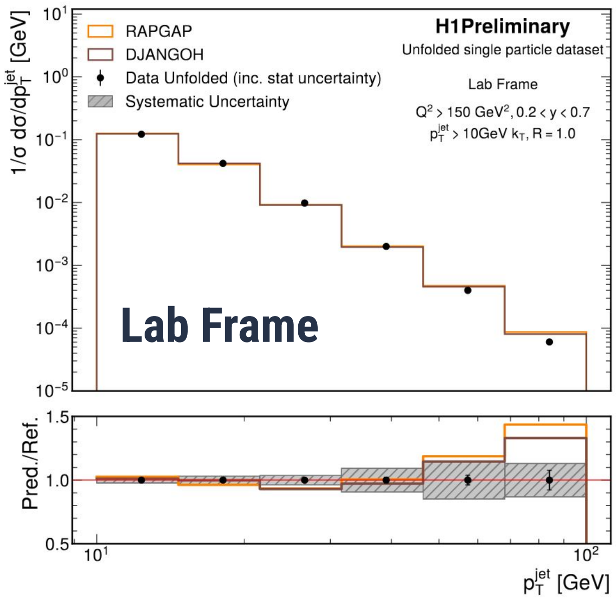


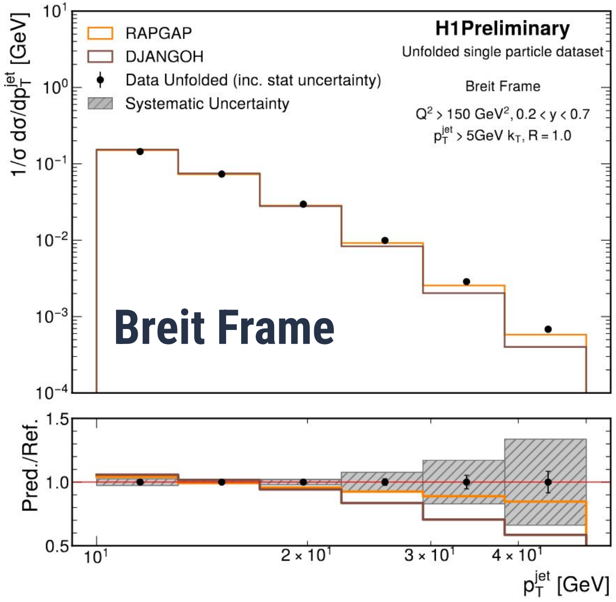
Breit Frame provides a natural frame to study ep collisions, where the struck quark forms a jet opposite from the proton beam: useful for jet and TMD studies

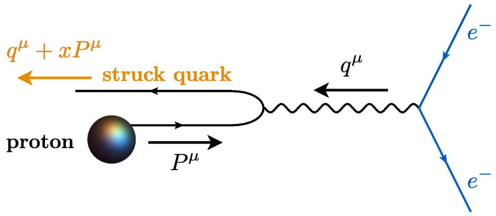
- Starting from the Lab frame, we need to boost the system: not trivial in terms of unfolding

Results

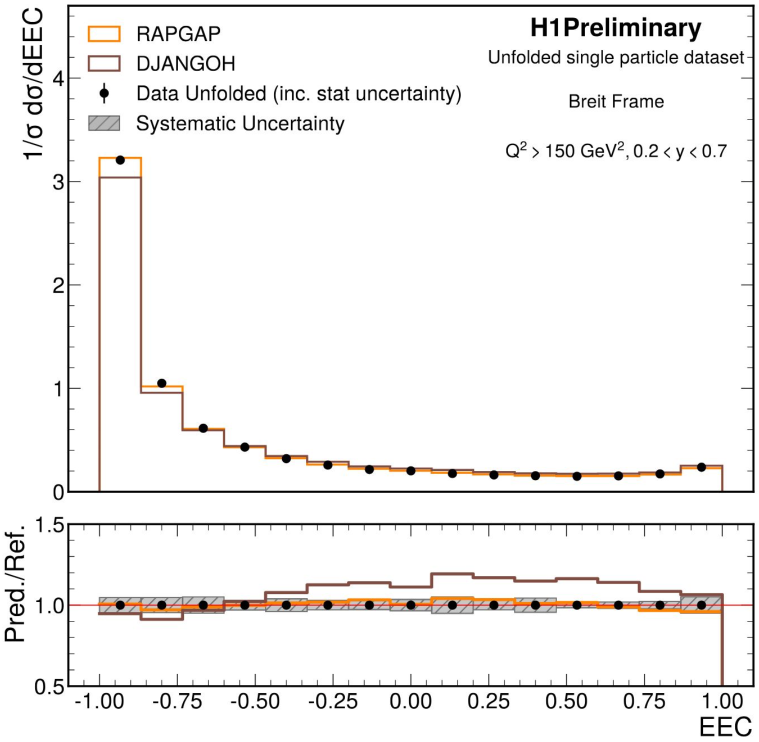
Cluster jets using kT algorithm with radius of 1.0
We can study observables in **different frames**!







Unfold observables that are hard to unfold without machine learning: **Energy Correlators**



Sensitive to transverse momentum dependent parton distribution functions and fragmentation functions

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

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

Eq. from Phys.Rev.D 103 (2021) 9, 094005

Conclusions and Next steps

- Unfolding is the task of removing detector effects from physics observables
- Building on previous results, we **unfold the information of all reconstructed hadrons and scattered lepton in high Q^2 DIS events** at the H1 experiment
- Preliminary results with more details available at: https://www-h1.desy.de/h1/www/publications/htmlsplit/H1prelim-25-031.long.html
- Working on the publication draft with more details, observables, and comparisons with predictions!

Backup

Systematic uncertainties

Systematic uncertainties

- **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

Unfolding uncertainties

Phi dependence

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

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

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)

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

See more in this talk given by Fernando Torales

Experimental setup

Fiducial Phase space definition:

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

Particle selection:

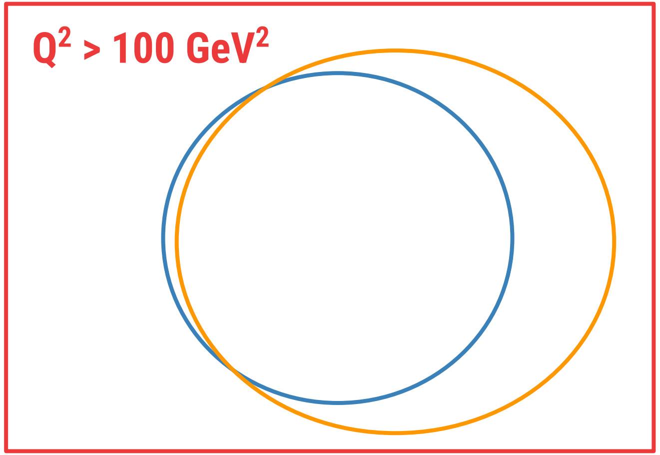
- $p_T > 0.1$ GeV
- $-1 < \eta_{lab} < 2.75$
- Charge information used if $\eta_{lab} < 2$

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
- $-1 < \eta_{lab} < 2.75$

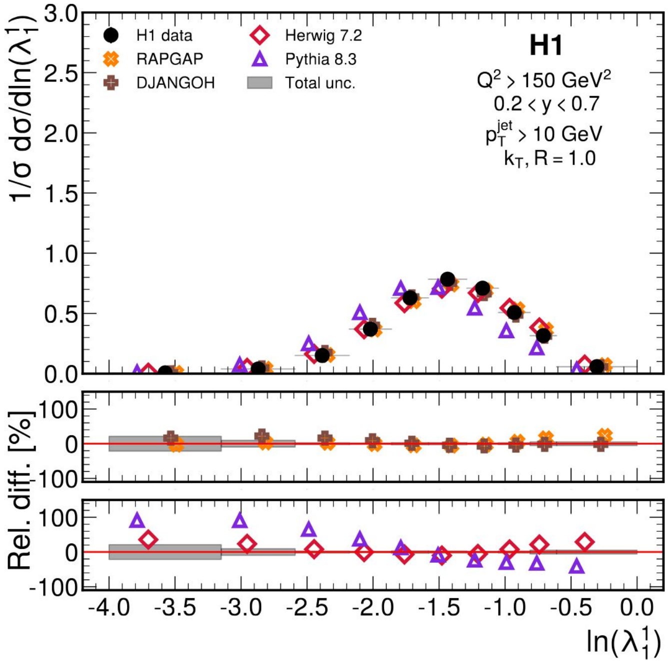


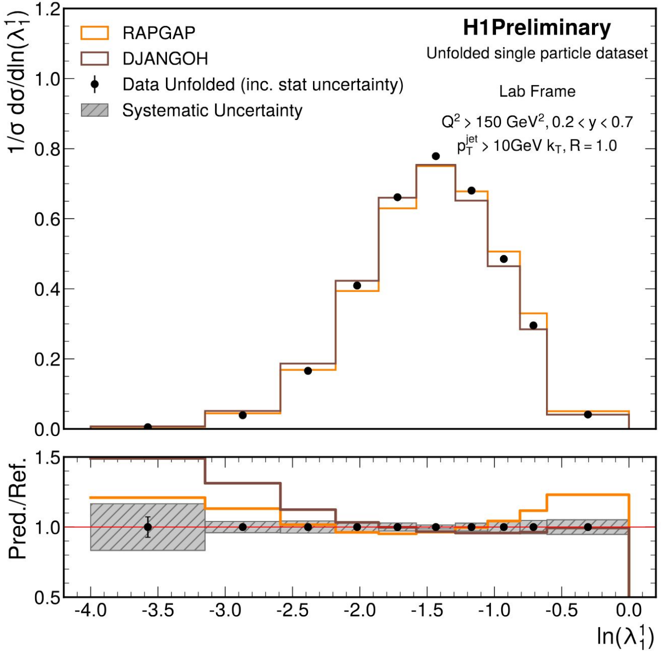
- Pass reco selection: **Red -> Orange**: **77%**
- Pass fiducial selection: **Red -> Blue**: **58%**
- Pass fiducial and reco selection: **Blue -> Orange**: **96%**
- Don't pass fiducial but pass reco: **Red -> Orange** (without **blue**): **50%**

Results

Cluster unfolded jets using kT algorithm with radius of 1.0

We are able to re-derive **past results**





Phys.Lett.B 844 (2023) 138101

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