Jet Substructure Measurements in Multijet Production with the ATLAS Experiment

BOOST 2024

Jingjing Pan (Jing), on behalf of the ATLAS collaboration

July 30th, 2024

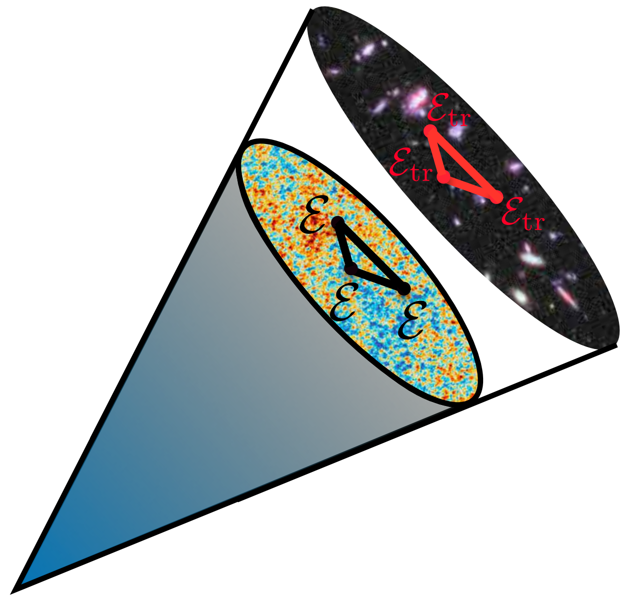
Measurement of Jet Track Functions

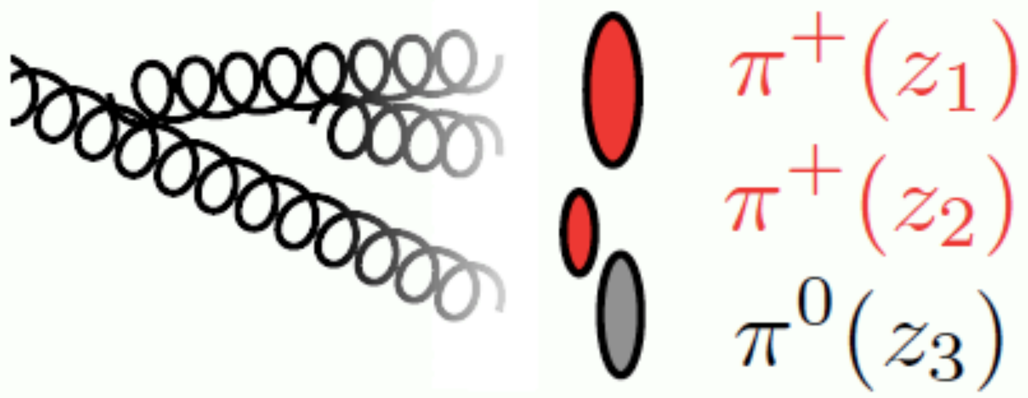
Jet Track Functions

ATLAS-CONF-2024-012

p_T fraction of a hard parton converted into all **charged hadrons**

- **Universal & non-perturbative**: foundational, not calculable, must be **measured**
- **Non-linear** Renormalization Group (RG) evolution **(Beyond DGLAP)**





$$z_1 + z_2 + z_3 = 1$$

Measure the $r_q \equiv p_T^{\text{tracks}}/p_T^{\text{all}}$ distribution & their moments & RG evolution

$$T_g(z) \propto \delta(\underline{z_1 + z_2 - z})$$

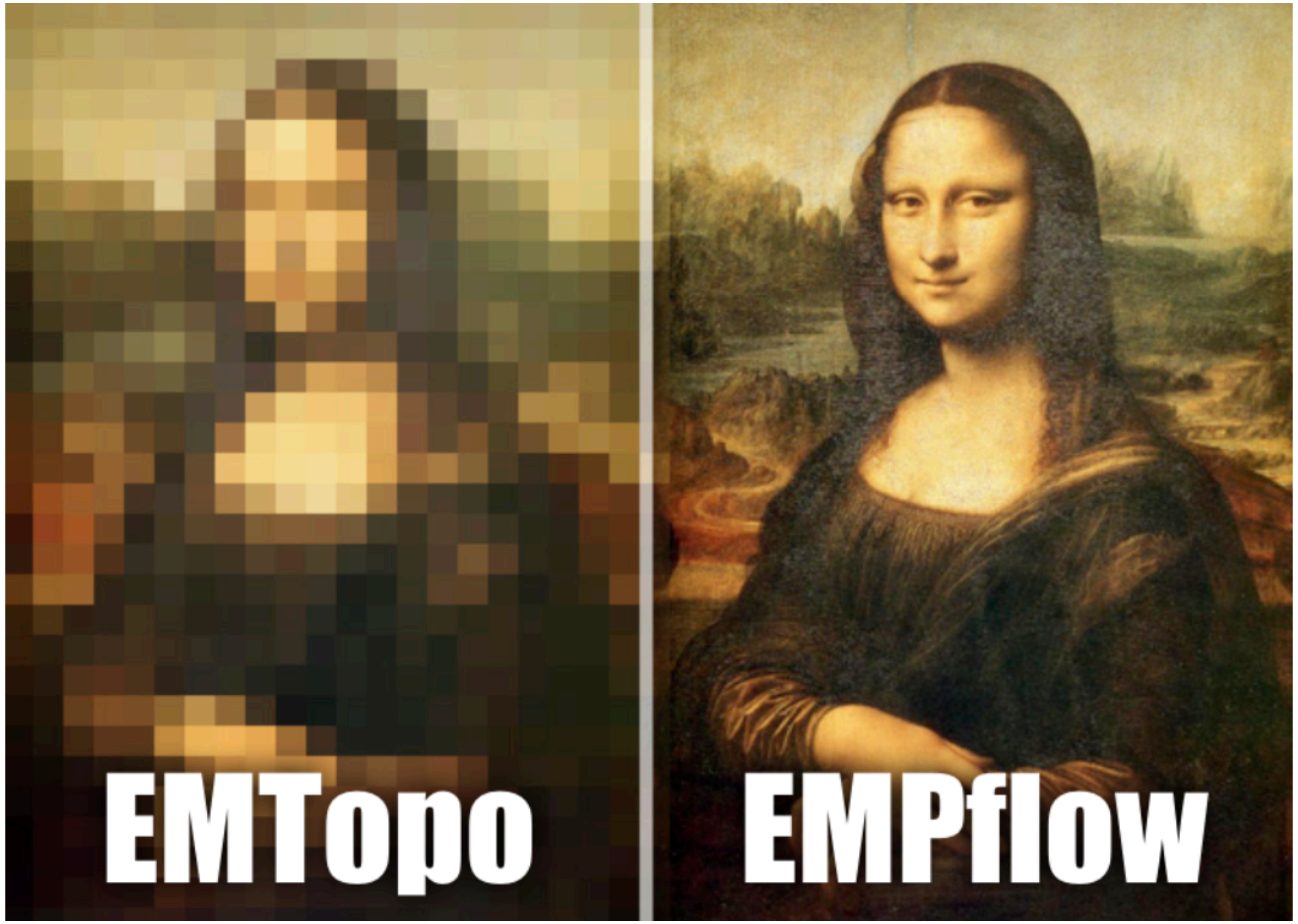
$$D_{g\to\pi^+}(z) \propto \underline{\delta(z_1 - z) + \delta(z_2 - z)}$$

Analogous to Fragmentation Functions (into a single hadron, linear RG)

1303.6637, 2108.01674, 2201.05166, 2210.10061, 2308.00028

On “Track(s)” to Higher Precision

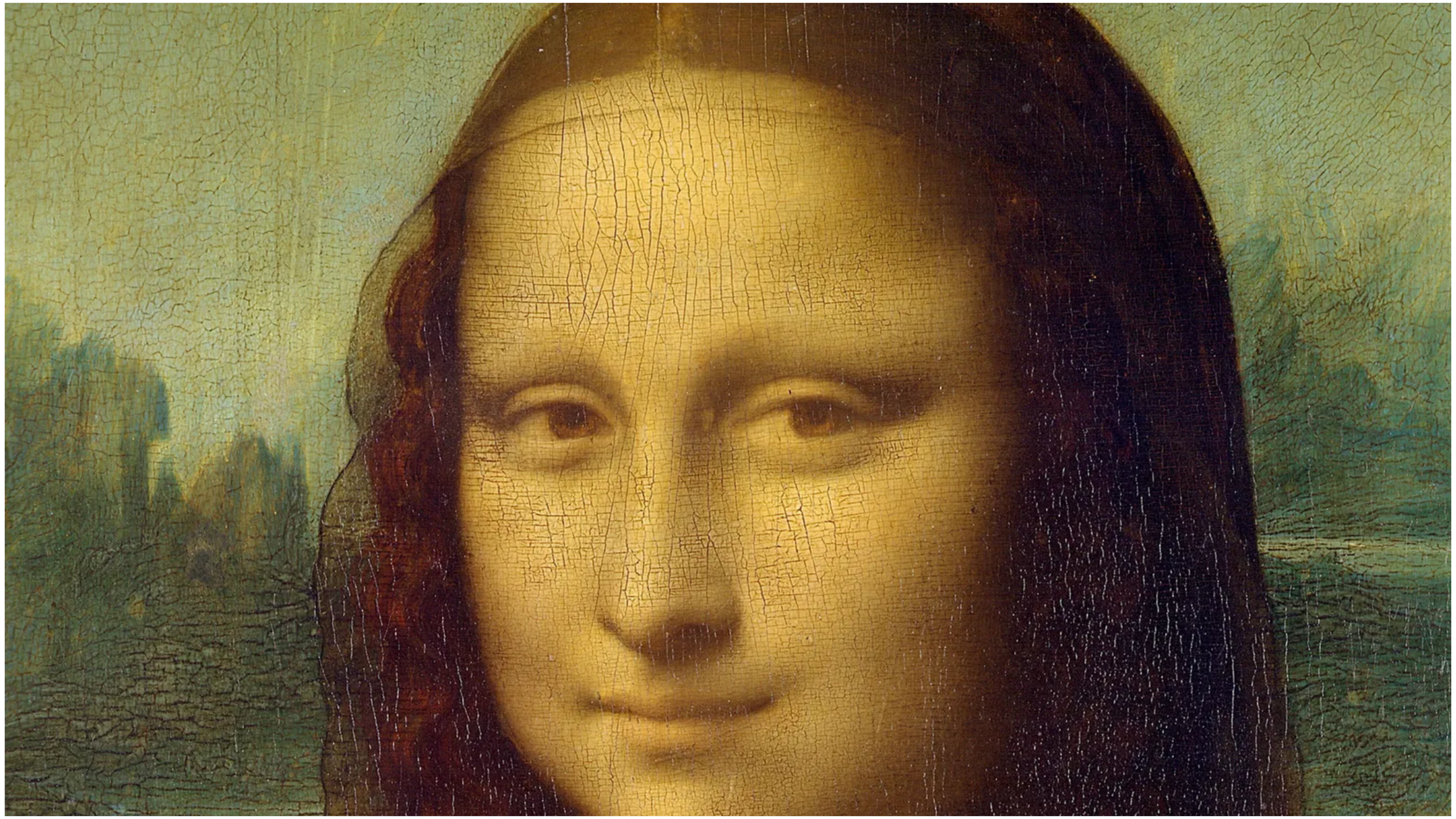
ATLAS-CONF-2024-012



1303.6637, 2108.01674, 15 track-based JSS measurements to date

On “Track(s)” to Higher Precision

ATLAS-CONF-2024-012

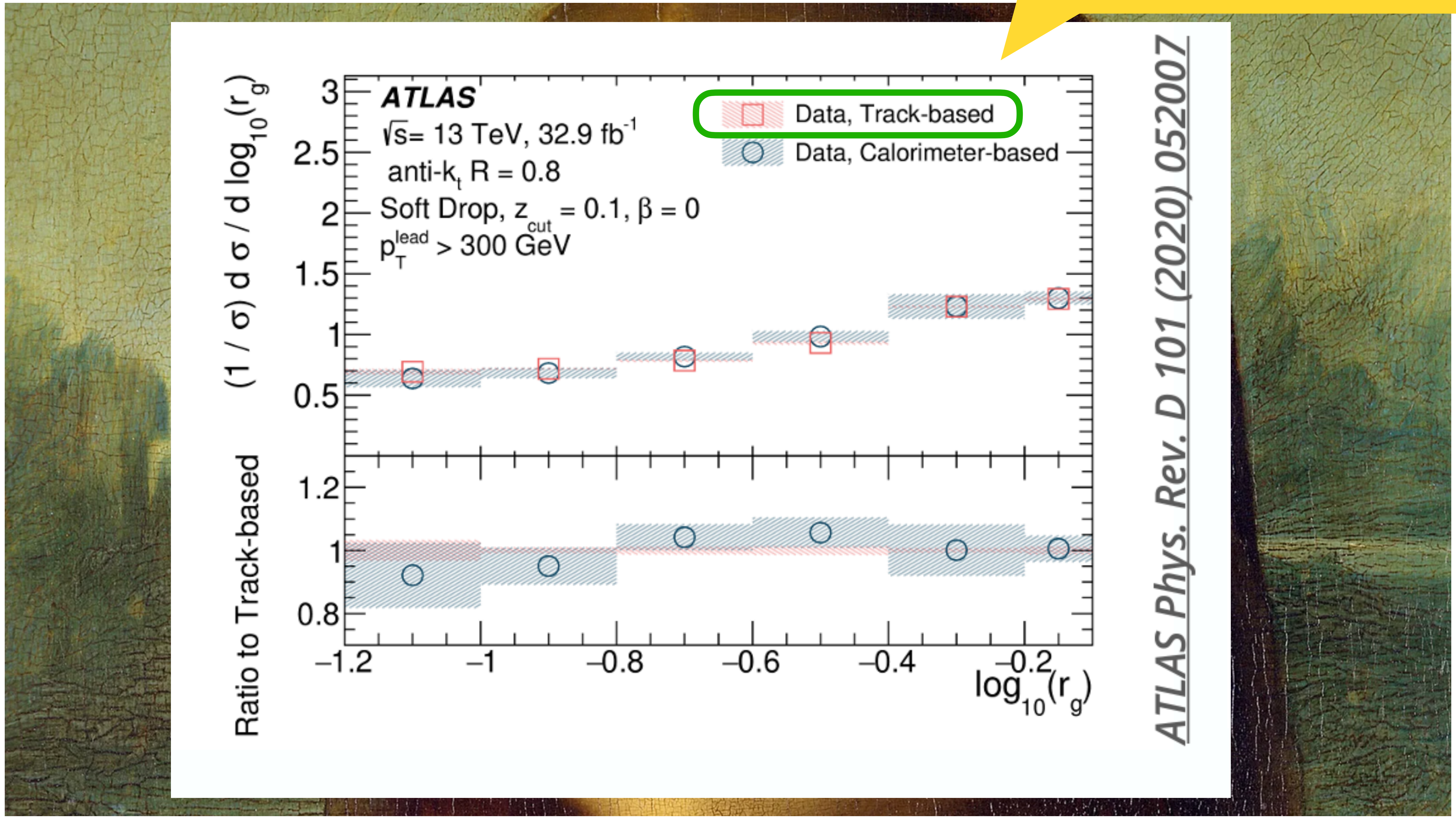


1303.6637, 2108.01674, 15 track-based JSS measurements to date

On “Track(s)” to Higher Precision

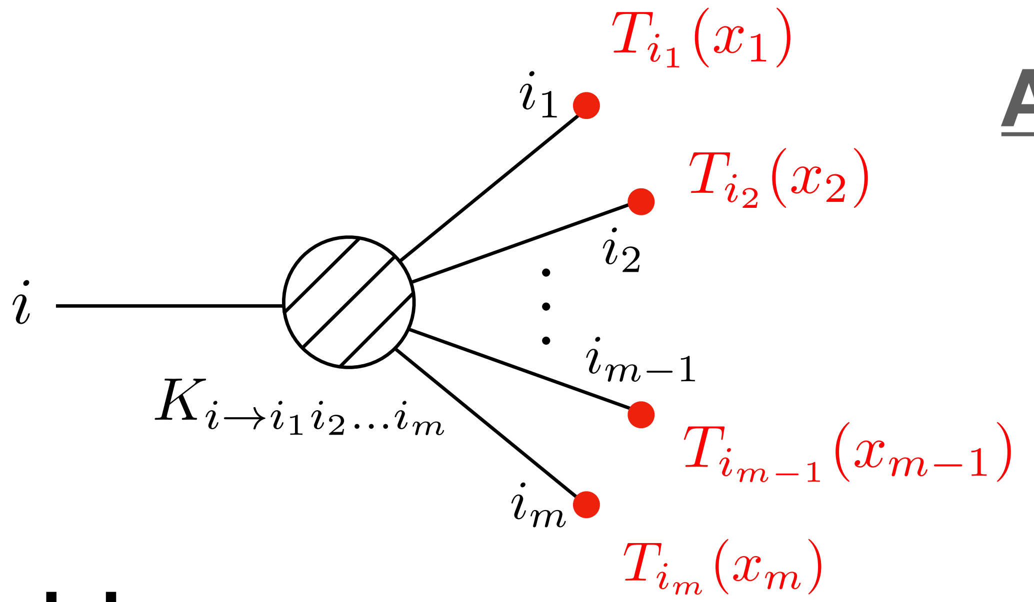
ATLAS-CONF-2024-012

Will enable jet substructure calculations on charged hadrons



ATLAS Phys. Rev. D 101 (2020) 052007

1303.6637, 2108.01674, 15 track-based JSS measurements to date



ATLAS-CONF-2024-012

Extracting the Moments

- The moments of a binned histogram are evaluated by:

Binned n^{th} moment: $m_n = \sum^{i} x_i^n h_i$

Here x_i^n is the center of the i^{th} bin → discretization effects

- e.g. 1^{st} moment the average, 2^{nd} related to the variance
- A new data-driven method with MultiFold corrected for the discretization effects:

Iterative Bayesian Unfolding (RooUnfold)

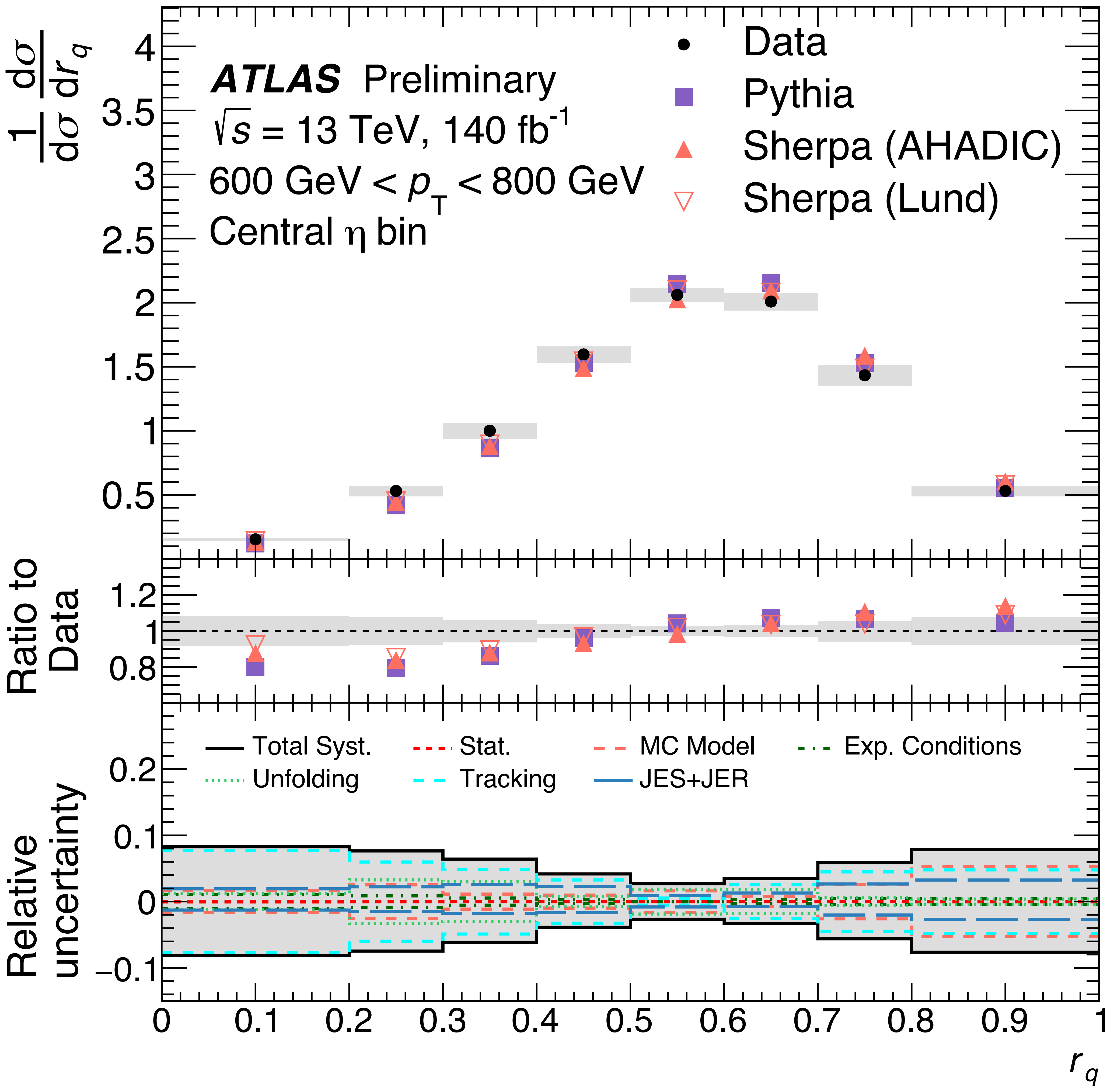
$$\frac{m^{\text{IBU}}}{m^{\text{corrected}}} = \frac{m^{\text{MultiFold}}_{\text{IBU binning}}}{m^{\text{MultiFold}}_{\text{unbinned}}} \Longrightarrow m^{\text{corrected}} = m^{\text{IBU}} \times \frac{m^{\text{MultiFold}}_{\text{unbinned}}}{m^{\text{MultiFold}}_{\text{IBU binning}}}$$

Get this ratio as the binning correction factor

1303.6637, 2108.01674, 2004.11381, 2201.05166

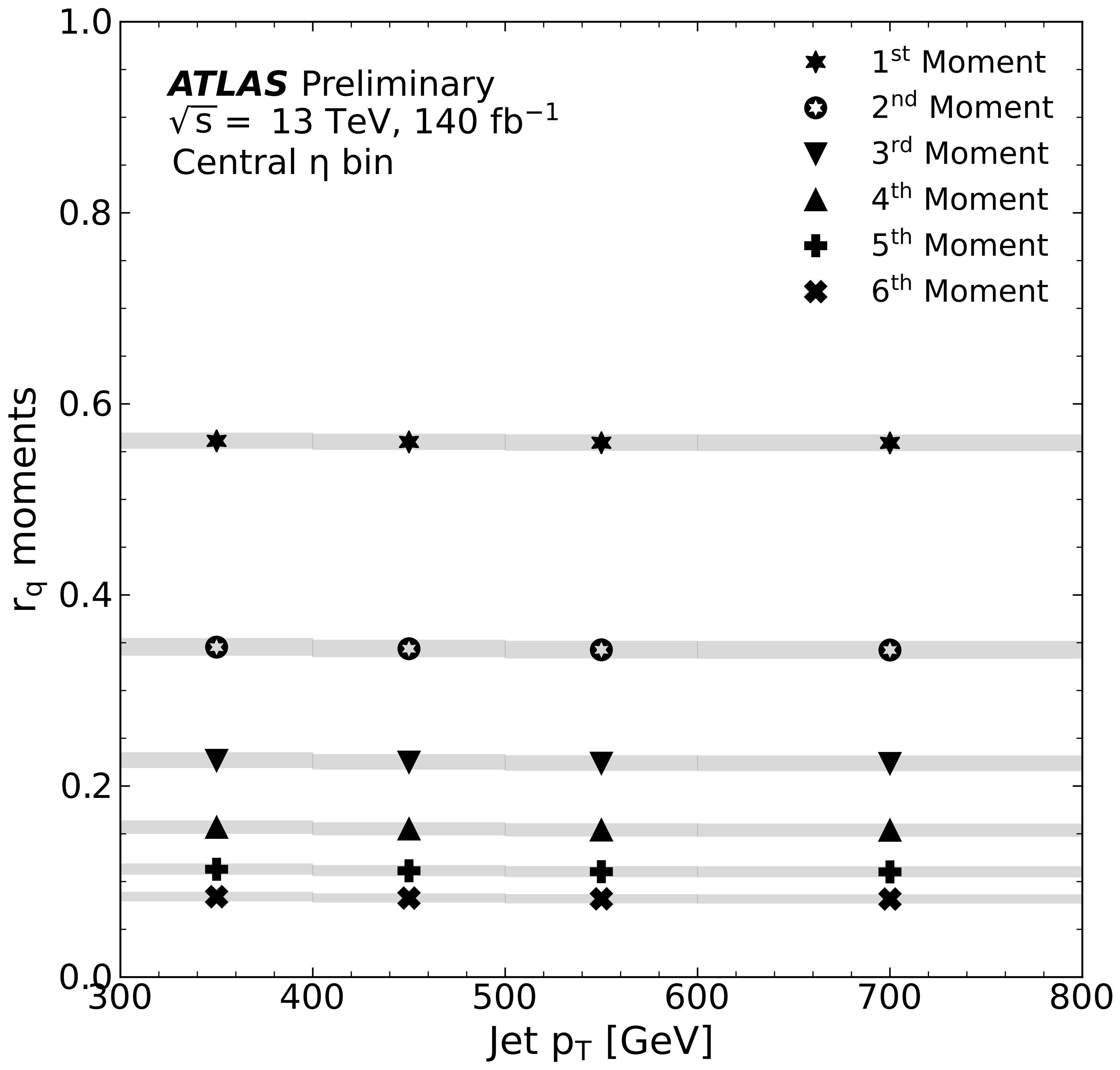
ATLAS-CONF-2024-012

Extracting the Moments



Corrected Binned n^{th} moment:

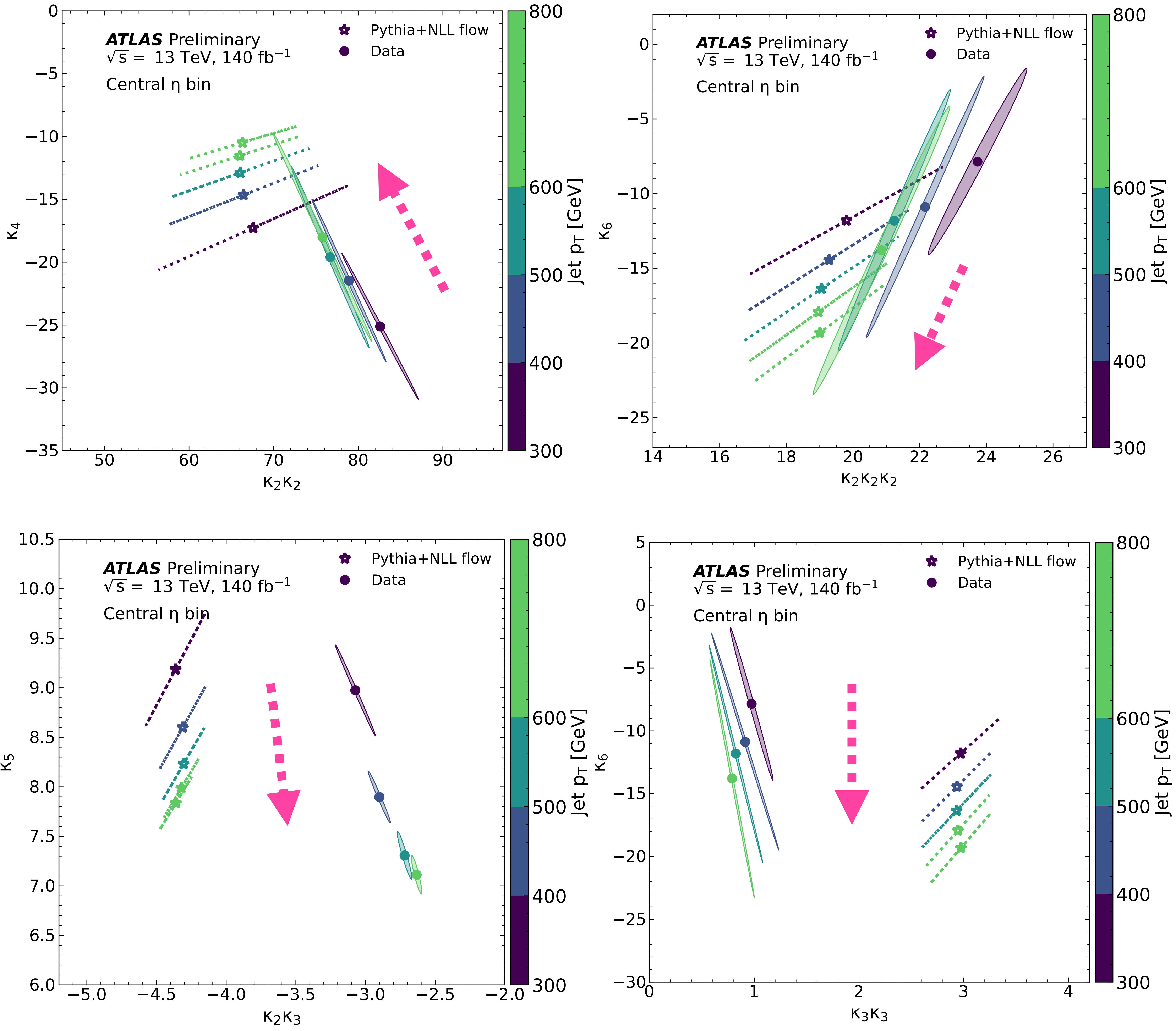
$$m_n = \sum^{i} x_i^n h_i \times \left(\frac{m^{\text{MultiFold}}_{\text{unbinned}}}{m^{\text{MultiFold}}_{\text{IBU binning}}} \right)^{\text{nom.}}$$



ATLAS-CONF-2024-012

Non-linear RG Evolution

- Non-linearity
  - From multi-hadron correlations
  - Qualitative first step beyond DGLAP
- The shape are computed by pQCD@NLL*
  - The exact starting point (i.e. the lowest p_T bin) of the predictions are currently from Pythia
  - → qualitative agreement



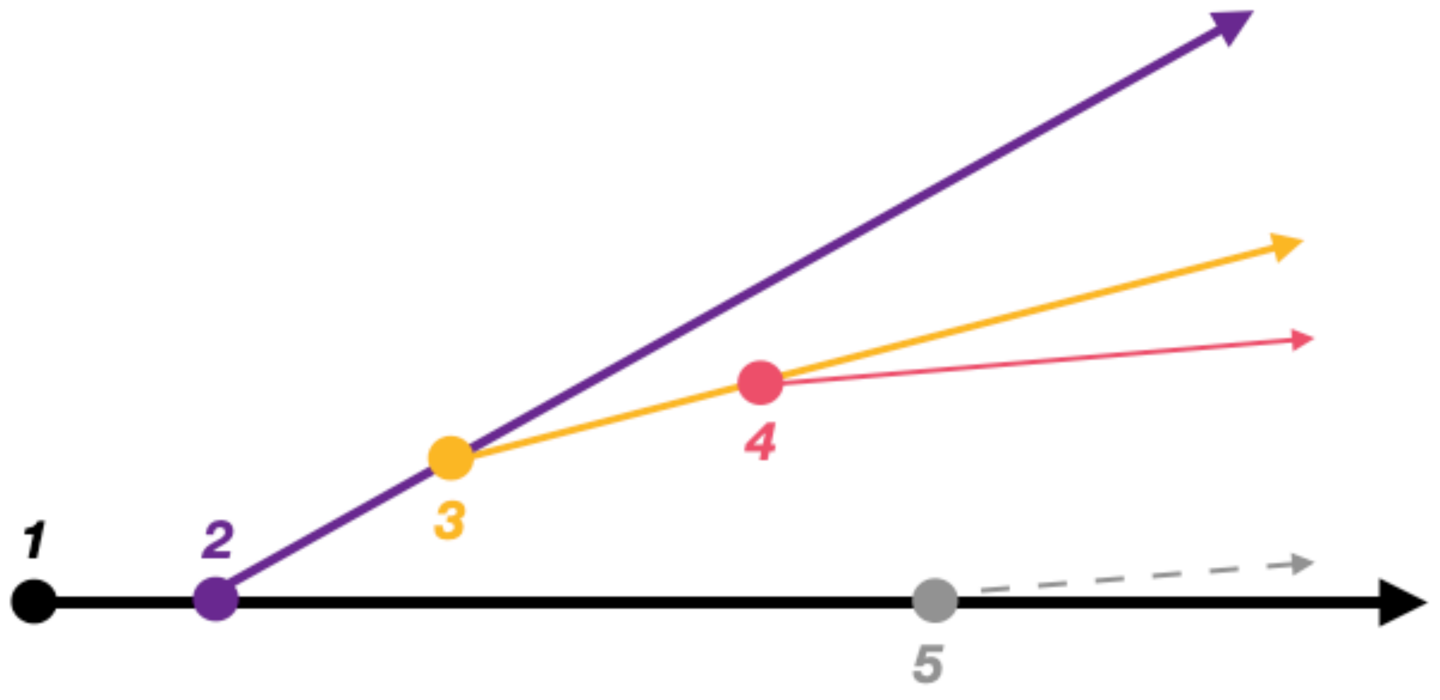
* Provided by the authors of 2201.05166

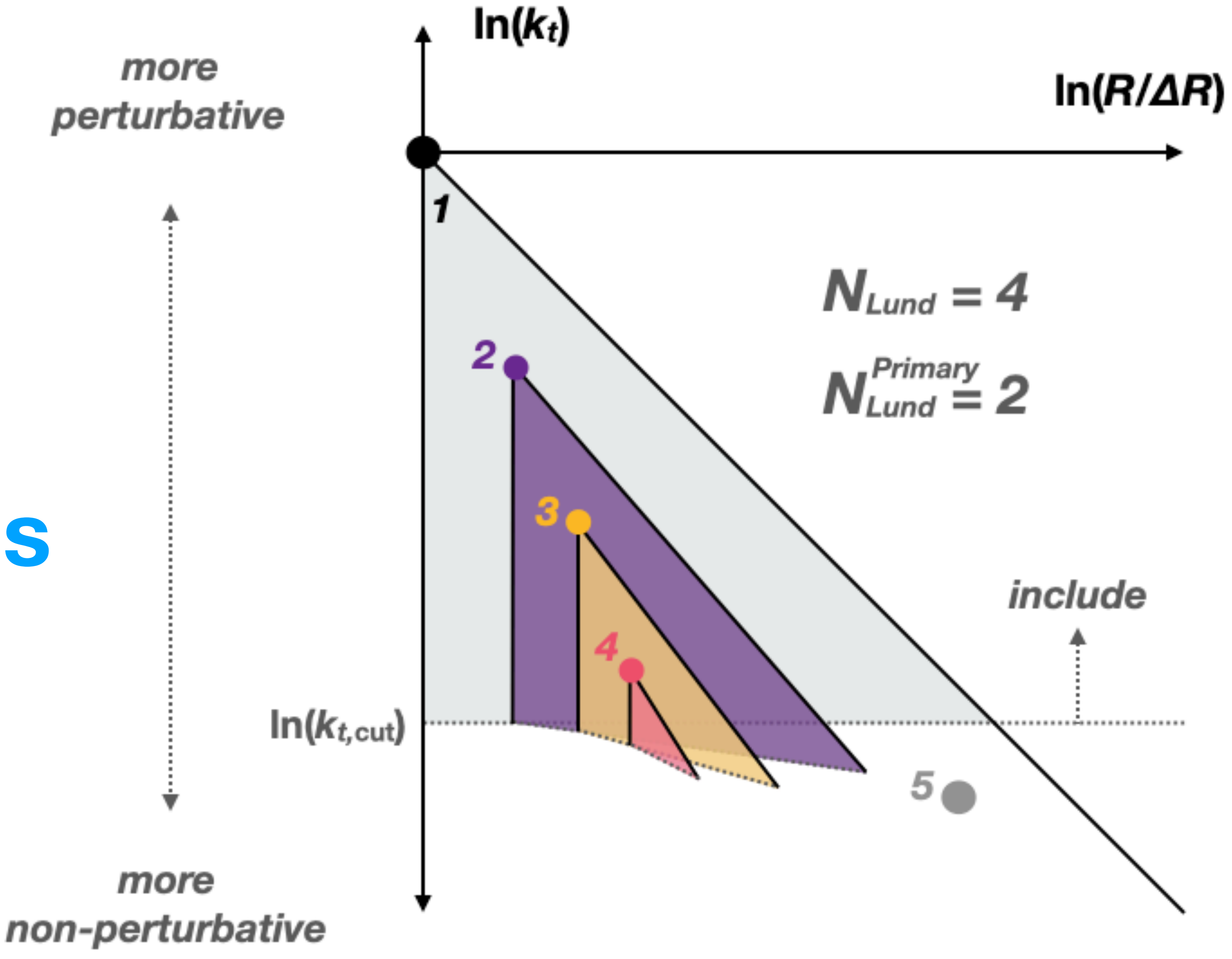
Measurement of Lund Subjet Multiplicities

Lund Subjet Multiplicities

CERN-EP-2024-029

- Lund & primary subjet multiplicities with $k_{t,\mathrm{cut}}$ = [0.5, 1, 2, 5, 10, 20, 50, 100] GeV
- where $k_t = p_T^{\mathrm{emission}} \Delta R$ (emission, core)
- core = harder branch & emission = softer branch
- Lund multiplicity utilizes **full structure** of the jet
- Primary Lund multiplicity looks **only at primary nodes**
- Average multiplicities extracted from **full distribution**





Full definitions at: 2205.02861

Motivation

CERN-EP-2024-029

- Parton Shower Monte Carlos (PSMC) are used extensively in LHC physics
  - Particularly affects the **precision** of results
  - Important to have PSMC with higher order effects & higher logarithmic accuracy
- Multiplicities are **sensitive to higher order effects in QCD**
  - double-soft & triple-collinear splittings are important
  - New analytical predictions used to benchmark PS models with higher-logarithmic accuracy [PanScales, Salam et al. 2002.11114]
- Important utility as input to higher accuracy PS development
  - Additional new shower development with the ALARIC algorithm [Höche *et al.*]

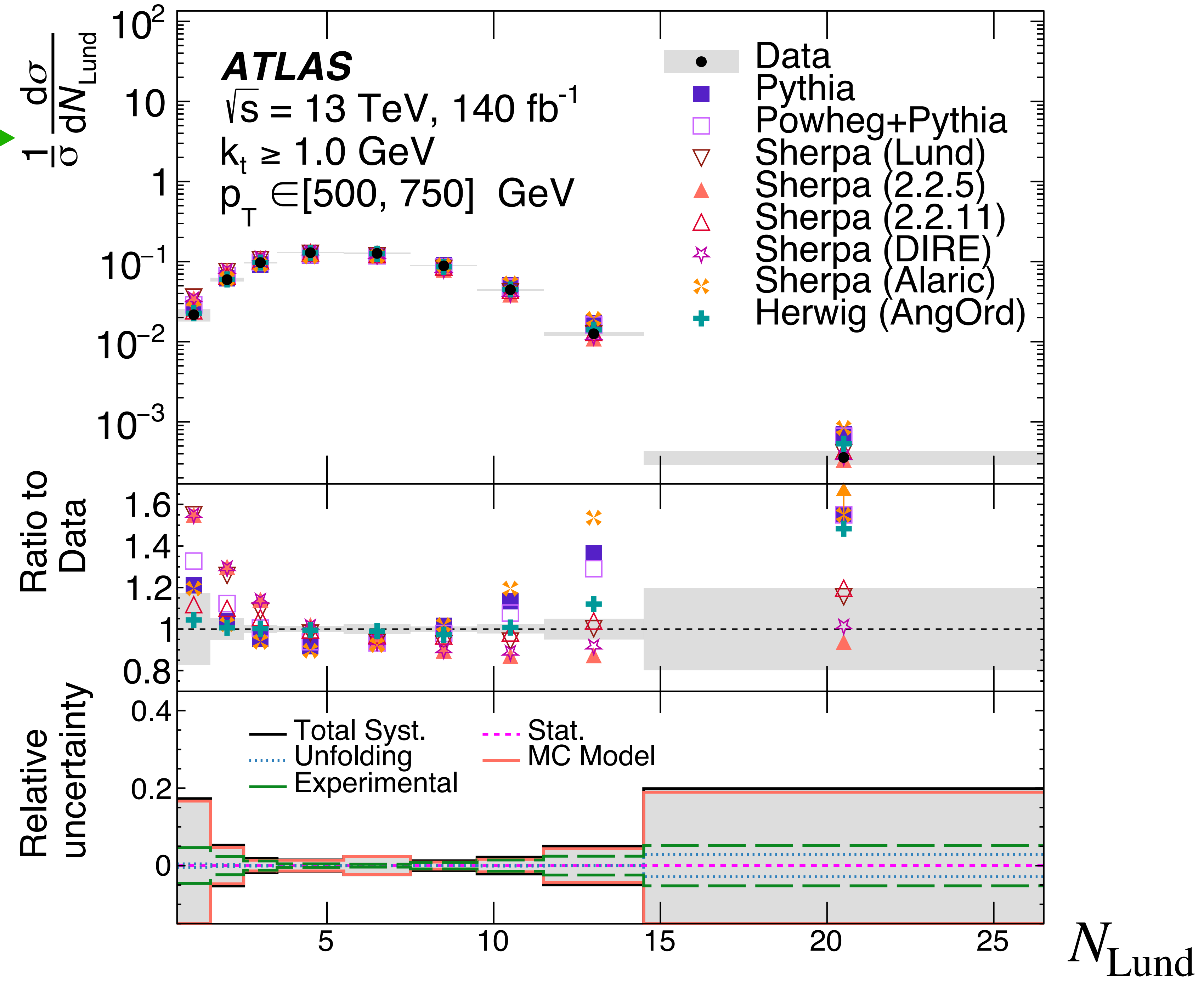
Observable Details

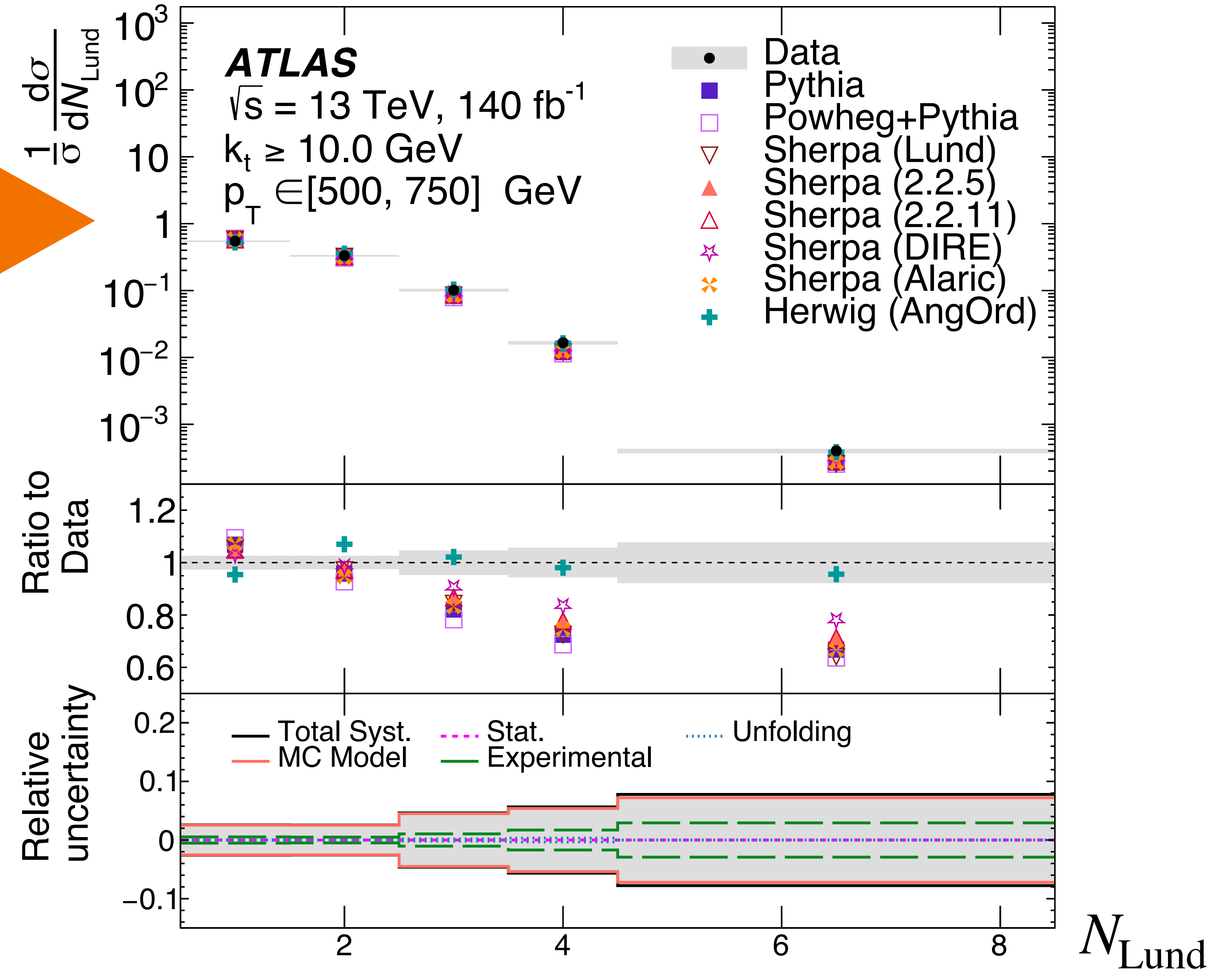
CERN-EP-2024-029

- Multiplicity is calculated by counting **the number of subjets above a specified relative transverse momentum k_t** in a jet's angular-ordered clustering history
  - **Only used charged tracks** associated with a jet for better resolution
  - Tracks are reclustered using **angular-ordered** Cambridge/Aachen algorithm
  - Primary multiplicity counts **only primary emissions** (directly from the core of the jet)
- Definition of k_t
  - The scale of each reconstructed emission **is restored from charged-particle to all-particle scale** to facilitate comparisons to theory calculations
  - i.e. $k_t = (p_T^{\text{jet}}/p_T^{\text{charged}})\, p_T^{\text{emission}} \Delta R$(emission, core)

Results - Lund Multiplicities

- Compared with **new ALARIC parton shower***
  - Herwig generally models the data best, especially in perturbative regions
  - Sherpa does very well in non-perturbative regions
- At **small** k_t, models **disagree** with the data at both high and low multiplicity
- At **higher** k_t, **a relative slope** between most models and the data is observed

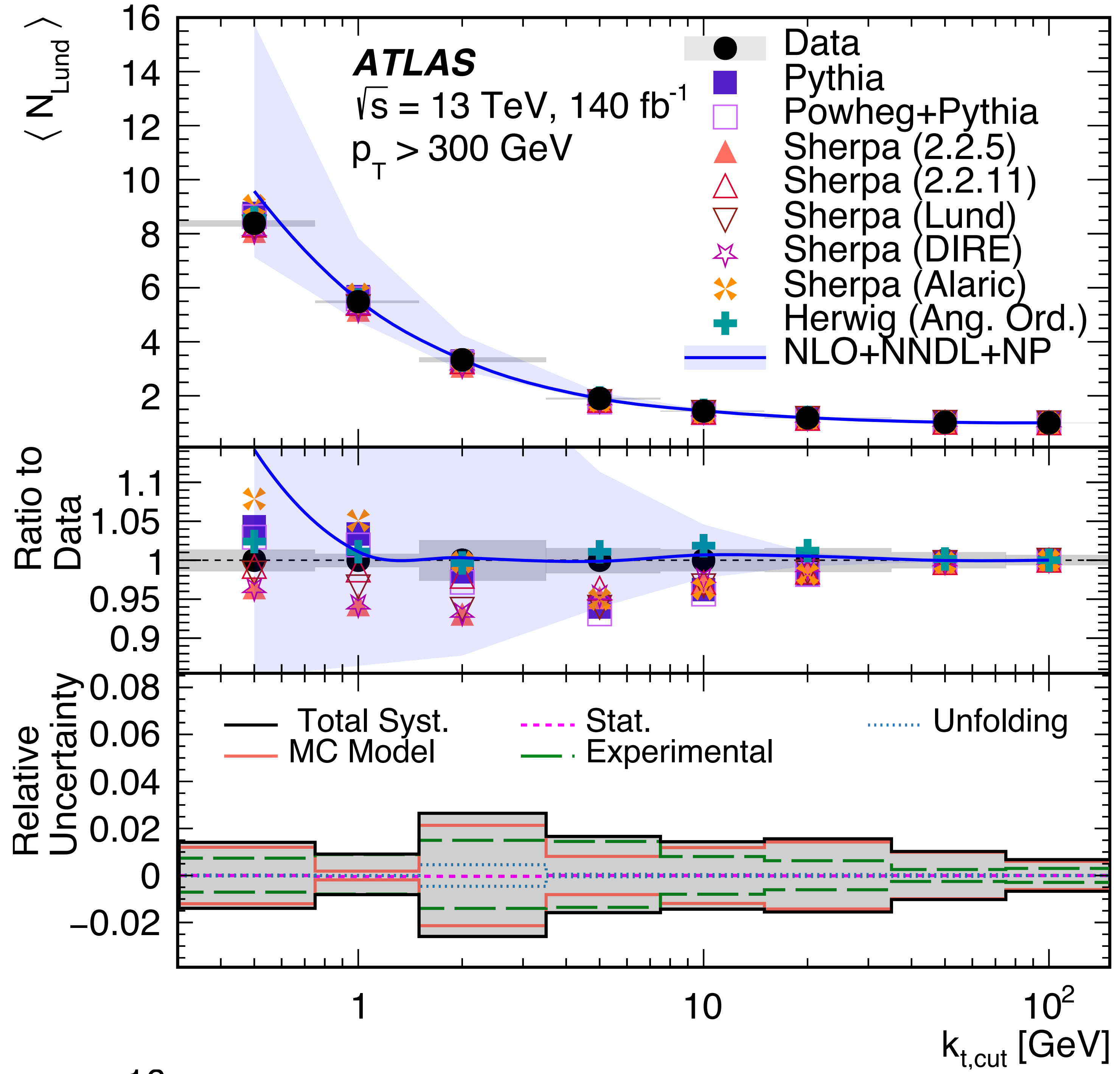


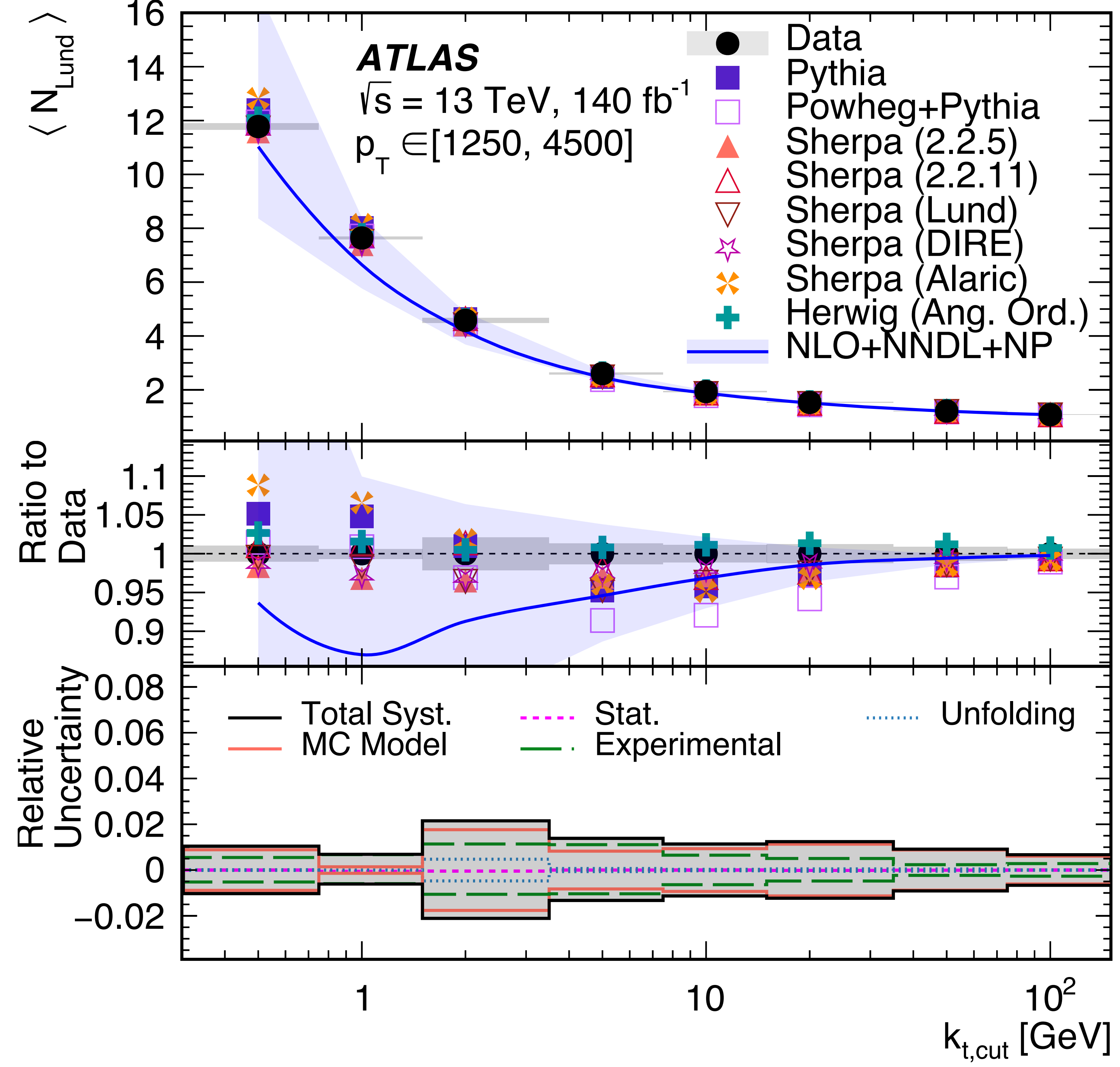


* Provided by authors
Höche *et al.* arXiv:2208.06057

Results - Average Multiplicities

- Average value as a function of k_t can be predicted analytically [NLO + NNDL]
  - This calculation is used to **benchmark new parton showers***
- Calculation **agrees** with the data within theory uncertainty
  - Non-perturbative effects large for $k_t < 5$ GeV
- Sherpa models using cluster vs. string based hadronization **differ** in the non-perturbative region





* PanScales, Salam et al, arXiv:2307.11142

A Simultaneous Unbinned Differential Cross Section Measurement of 24 Z+jets Kinematic Observables

Motivation

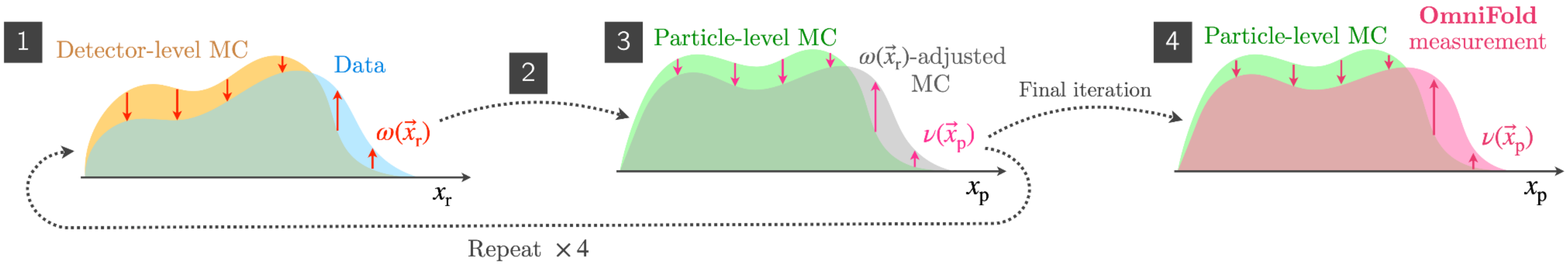
CERN-EP-2024-132

- Measurements of Z+jets - **clean** signal, minimum bias (from leptonic Z decays)
- To date only published as binned differential cross-sections
- Unlike previous results, **this analysis** is:
  - **High-dimensional**, a measurement of **24 observables unfolded simultaneously**
  - **Unbinned**, published as **a dataset of particle-level MC events with data-derived weights**
    - → significant flexibility & utility, both now & in the future

Observables & Unfolding Methodology

CERN-EP-2024-132

- Standard Z+jets selection* in **boost regime**: $p_T^{\mu\mu} > 200$ GeV
  - Leading & sub-leading jet: $p_T, y, \phi, \tau_1, \tau_2, \tau_3, m, n_{\text{tracks}}$
  - Leading & sub-leading muon: p_T, y, ϕ; Di-muon system: p_T, y
- The OmniFold (MultiFold) algorithm takes **2 sets** of inputs: MC samples at **both truth & reco levels**, and **the data**
- **Iteratively** proceed in **2 steps of reweighting** & **2 steps of truth-reco mapping**

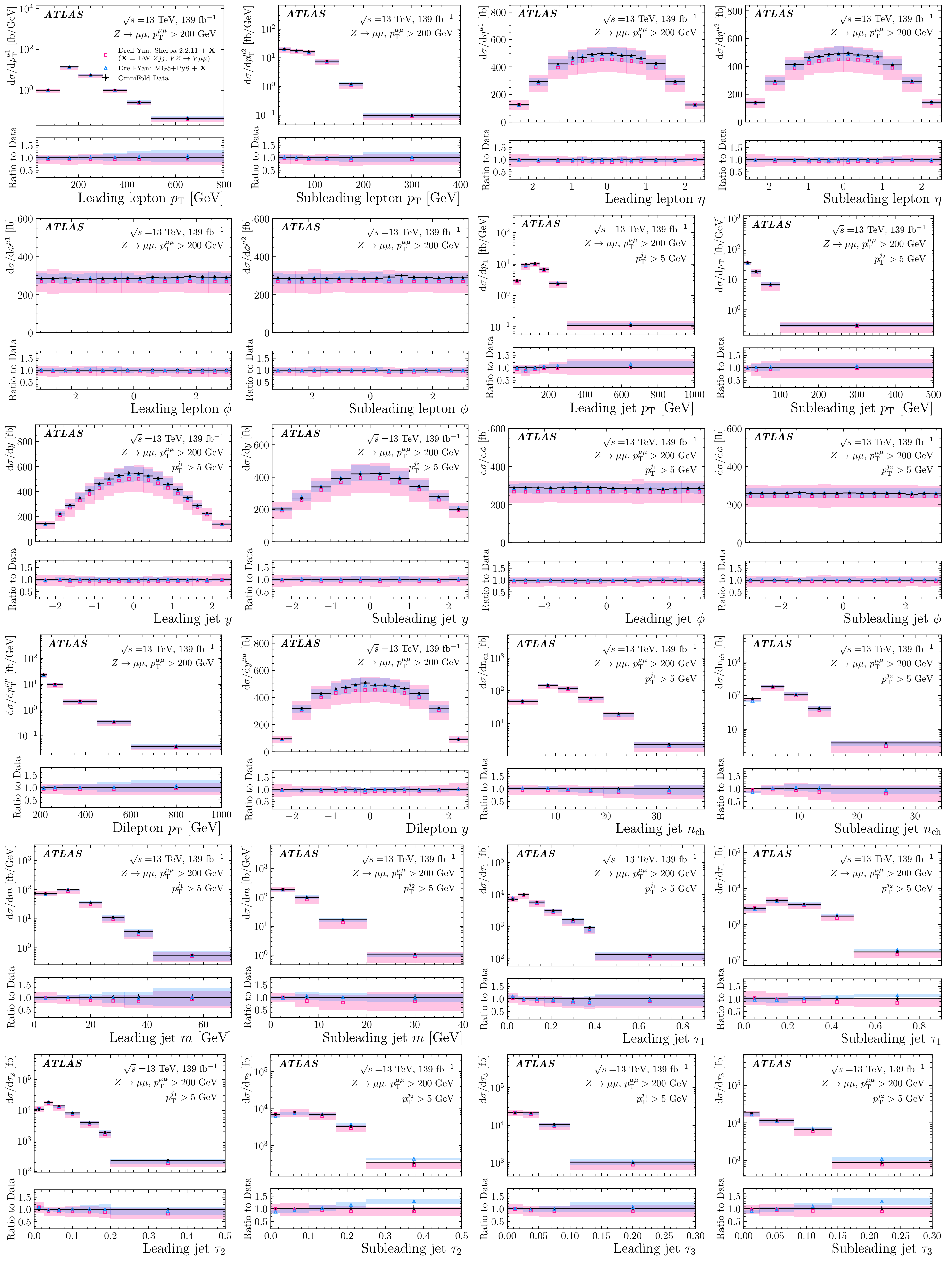


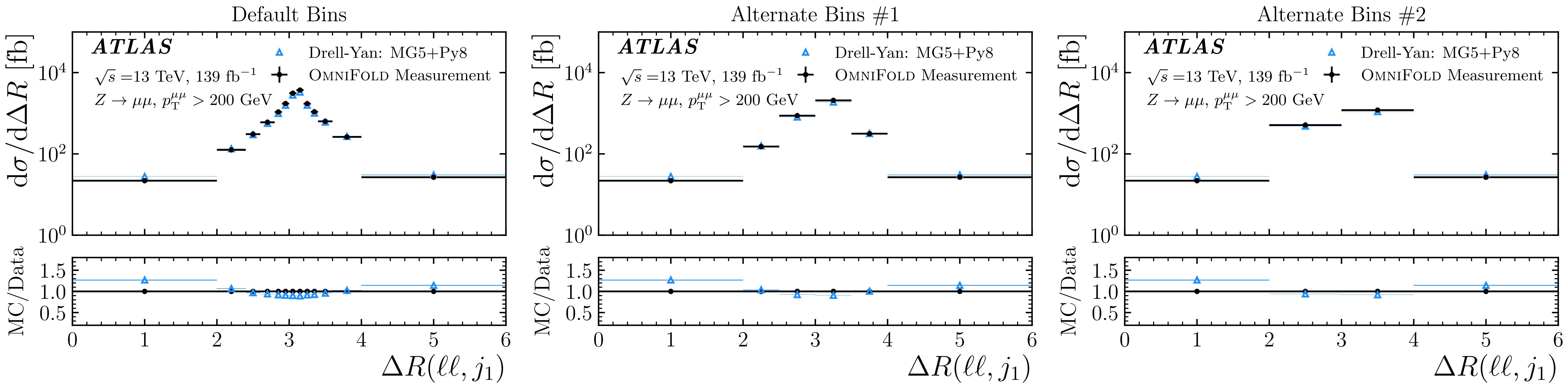
CERN-EP-2024-132

Results - 24 Unbinned Observables

- Right: measured differential cross sections of **the full set of 24 Z+jet observables** (with binning only for display purposes)
- Bottom: measured differential cross sections of $\Delta R(\text{ll}, j_1)$ using **3 different bin choices**.

(MC predictions shown with only statistical uncertainties)



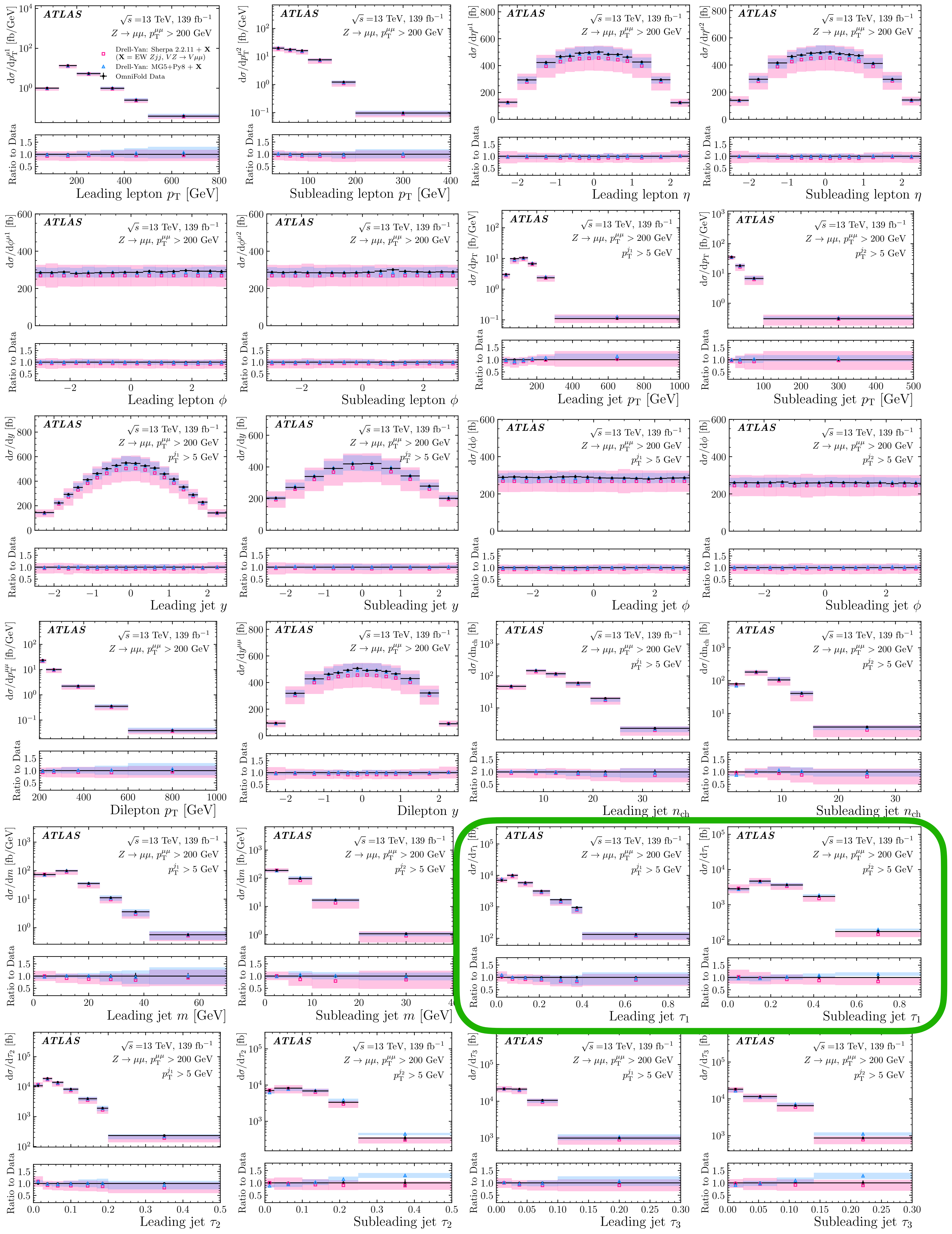


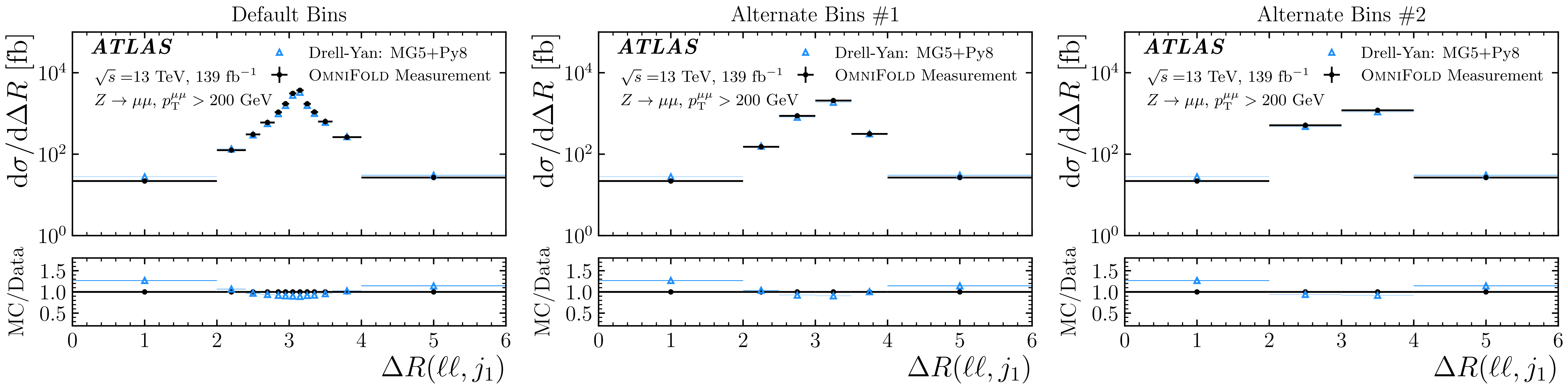
CERN-EP-2024-132

Results - 24 Unbinned Observables

- Right: measured differential cross sections of **the full set of 24 Z+jet observables** (with binning only for display purposes)
- Bottom: measured differential cross sections of ΔR(ll, j_1) using **3 different bin choices**.

(MC predictions shown with only statistical uncertainties)





Results - New Observables

This ratio of two jet substructure can be useful for e.g. classifying W vs. QCD jets

$$\tau_N = \frac{1}{d_0}\sum_k p_{T,k}\min_A\{\Delta R_{A,k}\}$$

↑ Sum over constituents

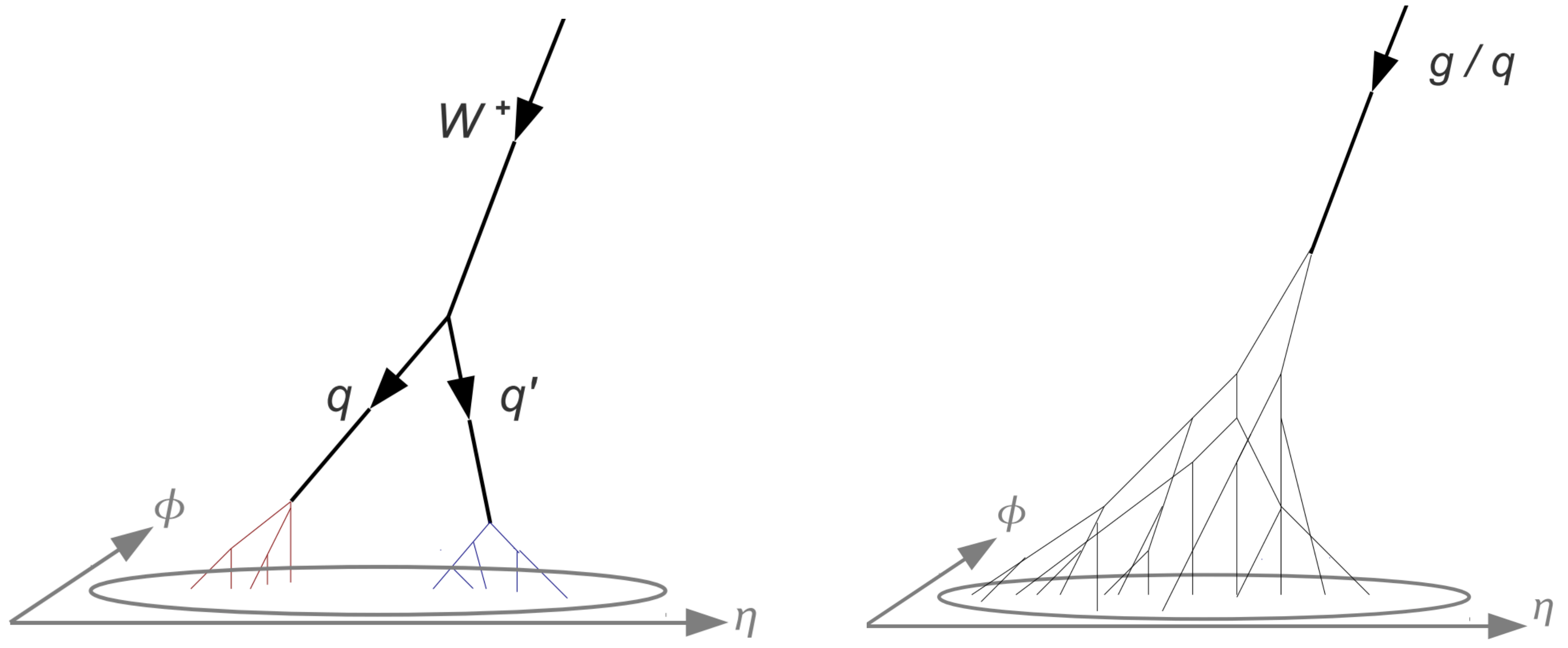
↑ Minimize distance to candidate subjet axes

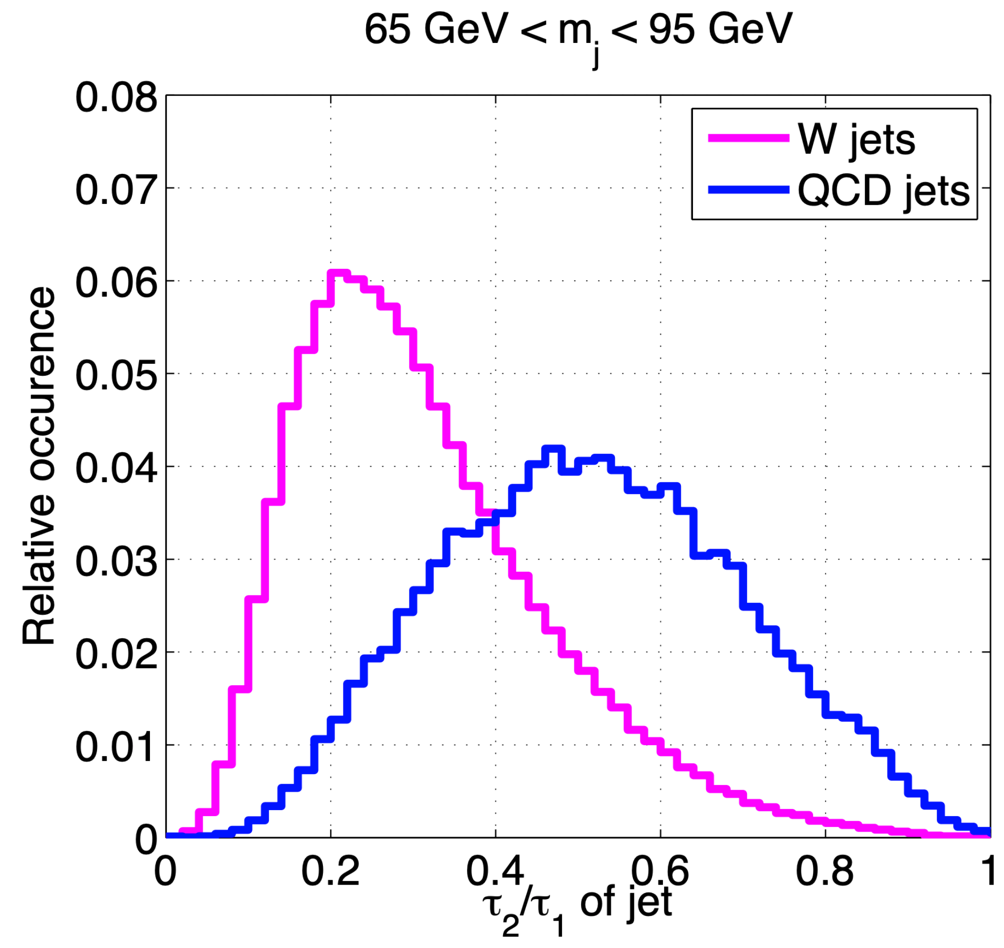
Jet shape that "counts" number of subjets!

subjets: ≤ N > N

τN: 0 1

Adapted from "N-jettiness" [Stewart, Tackmann, Waalewijn: 1004.2489]





Images from J. Thaler's talk and paper from 2011

CERN-EP-2024-132

Results - New Observables

This ratio of two jet substructure can be useful for e.g. classifying W vs. QCD jets

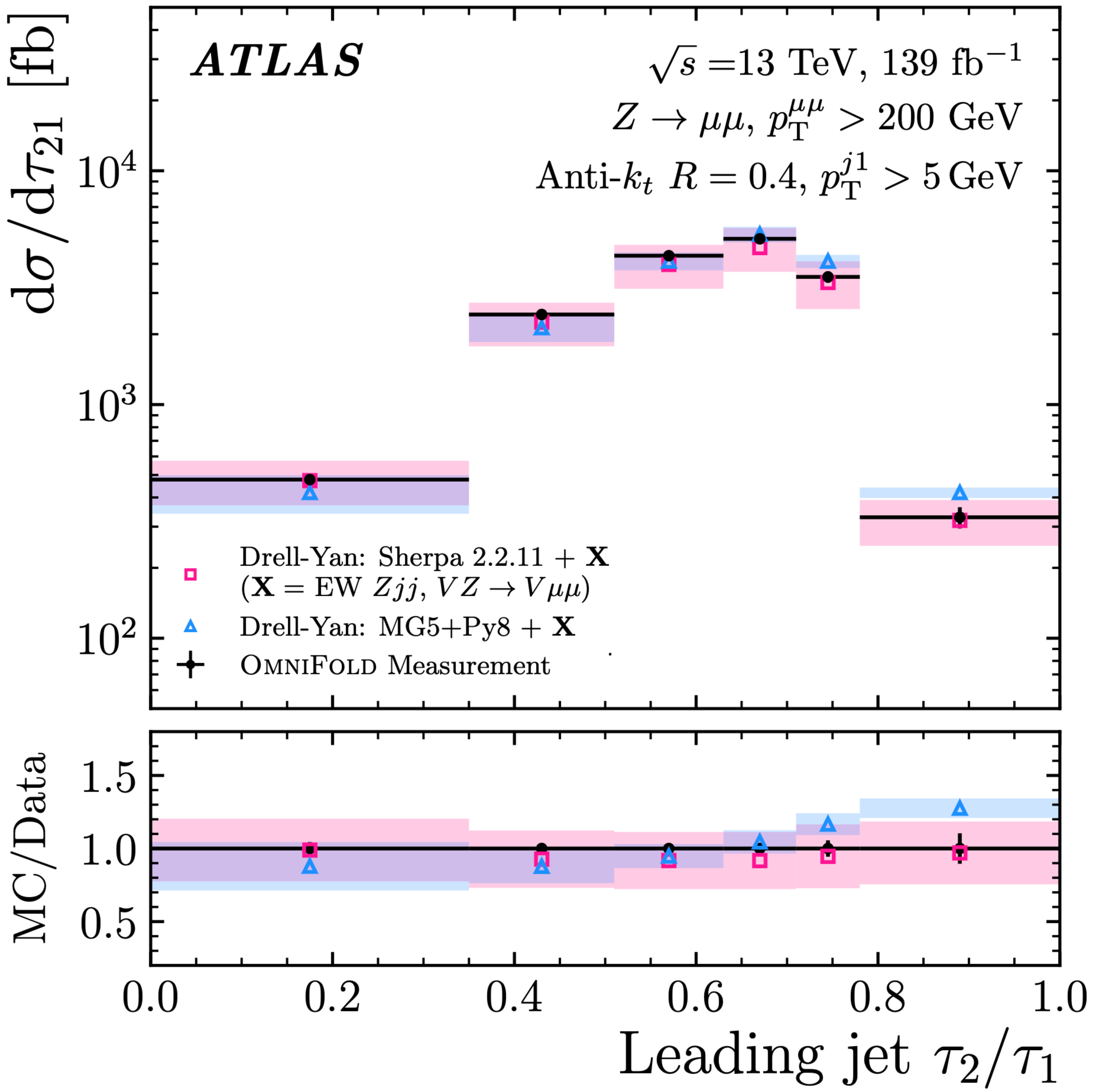
$$\tau_N = \frac{1}{d_0}\sum_k p_{T,k}\min_A\{\Delta R_{A,k}\}$$

Sum over constituents

Minimize distance to candidate subjet axes

Jet shape that "counts" number of subjets!

Adapted from "N-jettiness" [Stewart, Tackmann, Waalewijn: 1004.2489]



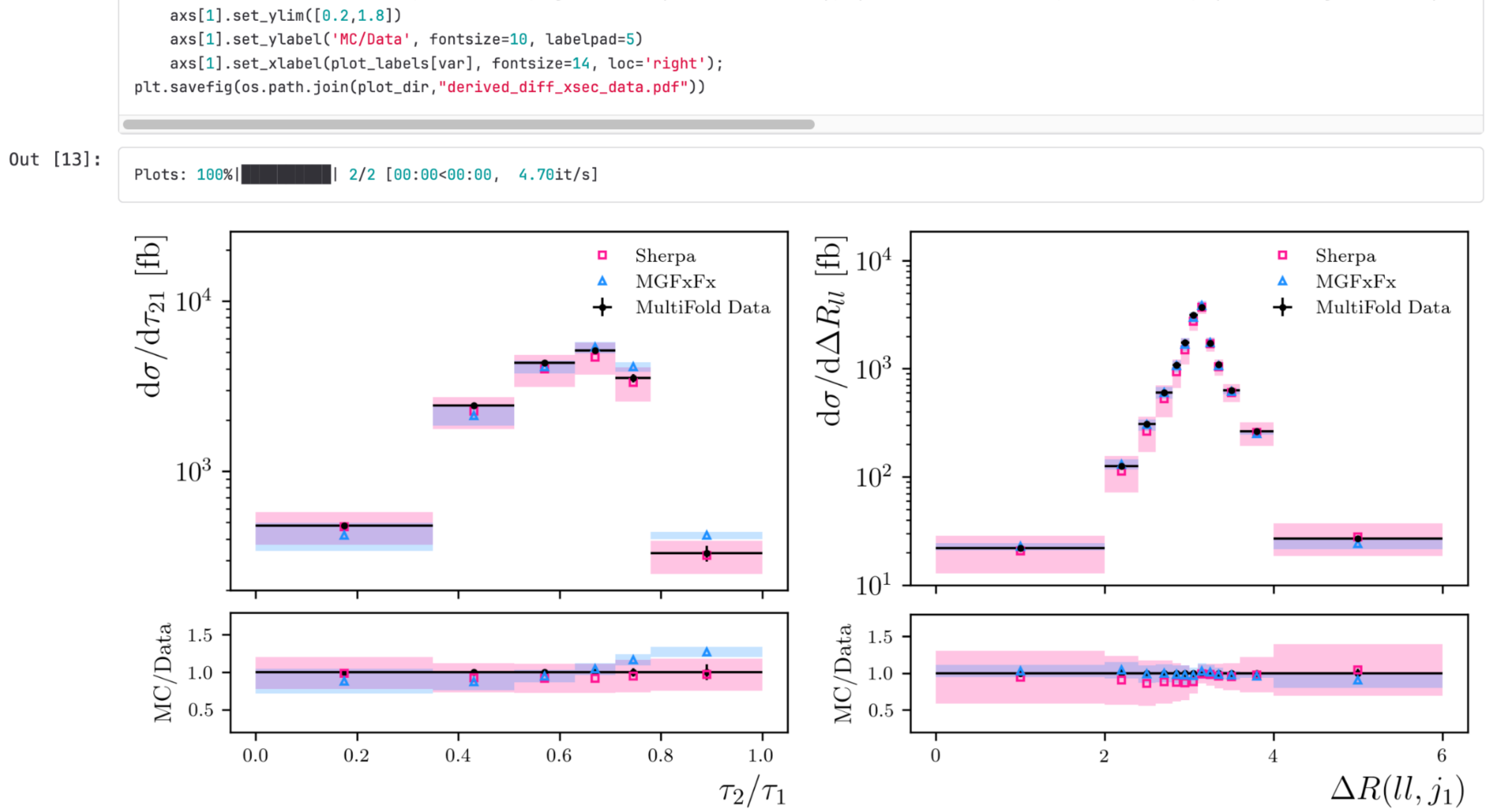
Images from J. Thaler's talk and paper from 2011

Dataset & Jupyter Notebooks

- Datasets:
  - https://zenodo.org/records/11507450
- Codebase:
  - https://gitlab.cern.ch/atlas-physics/public/sm-z-jets-omnifold-2024
- Notebooks:

1_basics.ipynb	Open in Colab
2_pseudo_results.ipynb	Open in Colab
3_results.ipynb	Open in Colab

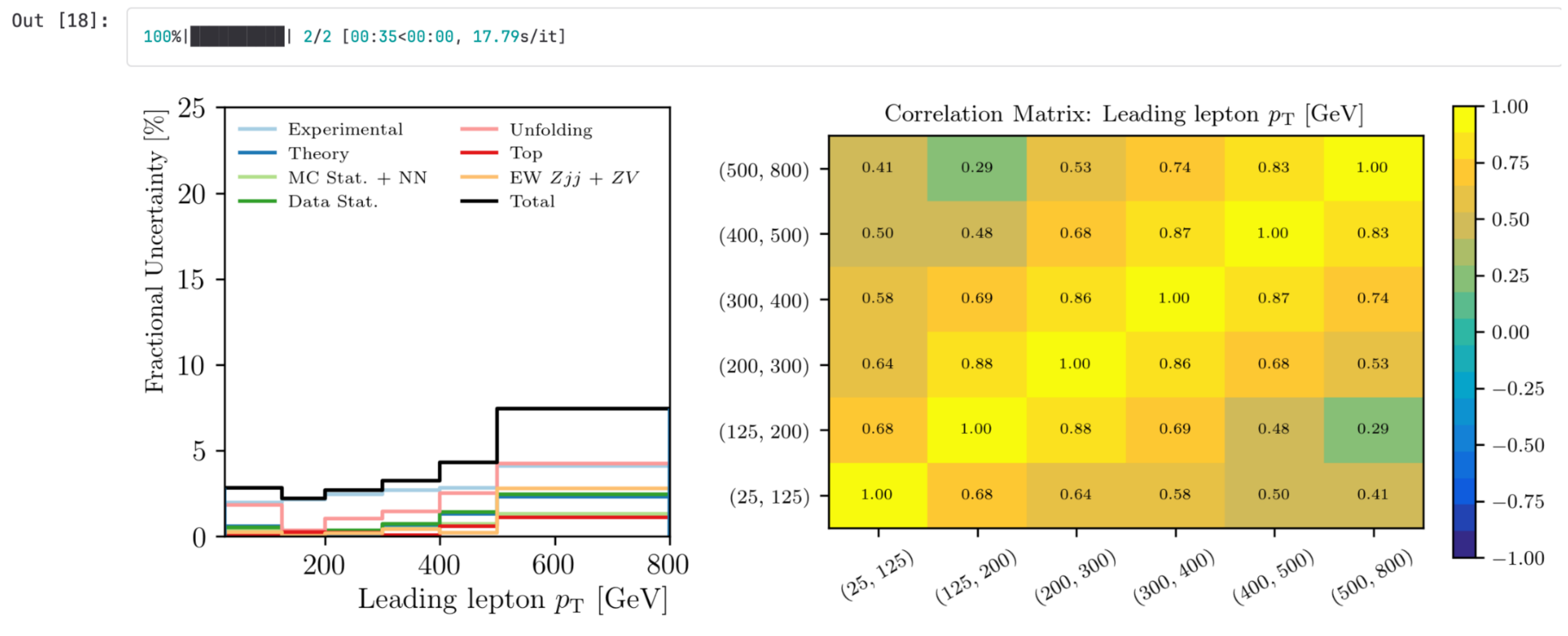
Construct **derived** observables



```
    axs[1].errorbar(bin_centers, np.ones(len(bin_centers)), xerr=bin_widths/2, yerr=uncertainties[var+"_total"]/100, marker=".", color="k", alpha=1, ec
    _ = make_error_boxes(axs[1], bin_centers, sherpa_density/multifold_density, np.vstack([bin_widths/2,bin_widths/2]), np.vstack([(sherpa_density/mult
    _ = make_error_boxes(axs[1], bin_centers, mgfxfx_density/multifold_density, np.vstack([bin_widths/2,bin_widths/2]), np.vstack([(mgfxfx_density/mult
    axs[1].set_ylim([0.2,1.8])
    axs[1].set_ylabel('MC/Data', fontsize=10, labelpad=5)
    axs[1].set_xlabel(plot_labels[var], fontsize=14, loc='right');
plt.savefig(os.path.join(plot_dir,"derived_diff_xsec_data.pdf"))
```

```
Out [13]:  Plots: 100%|████████| 2/2 [00:00<00:00,  4.70it/s]
```

Plot correlation matrices



```
In [18]:  duo_plots([0, 1])
```

```
Out [18]:  100%|████████| 2/2 [00:35<00:00, 17.79s/it]
```

Correlation Matrix: Leading lepton p_T [GeV]

	(25, 125)	(125, 200)	(200, 300)	(300, 400)	(400, 500)	(500, 800)
(500, 800)	0.41	0.29	0.53	0.74	0.83	1.00
(400, 500)	0.50	0.48	0.68	0.87	1.00	0.83
(300, 400)	0.58	0.69	0.86	1.00	0.87	0.74
(200, 300)	0.64	0.88	1.00	0.86	0.68	0.53
(125, 200)	0.68	1.00	0.88	0.69	0.48	0.29
(25, 125)	1.00	0.68	0.64	0.58	0.50	0.41

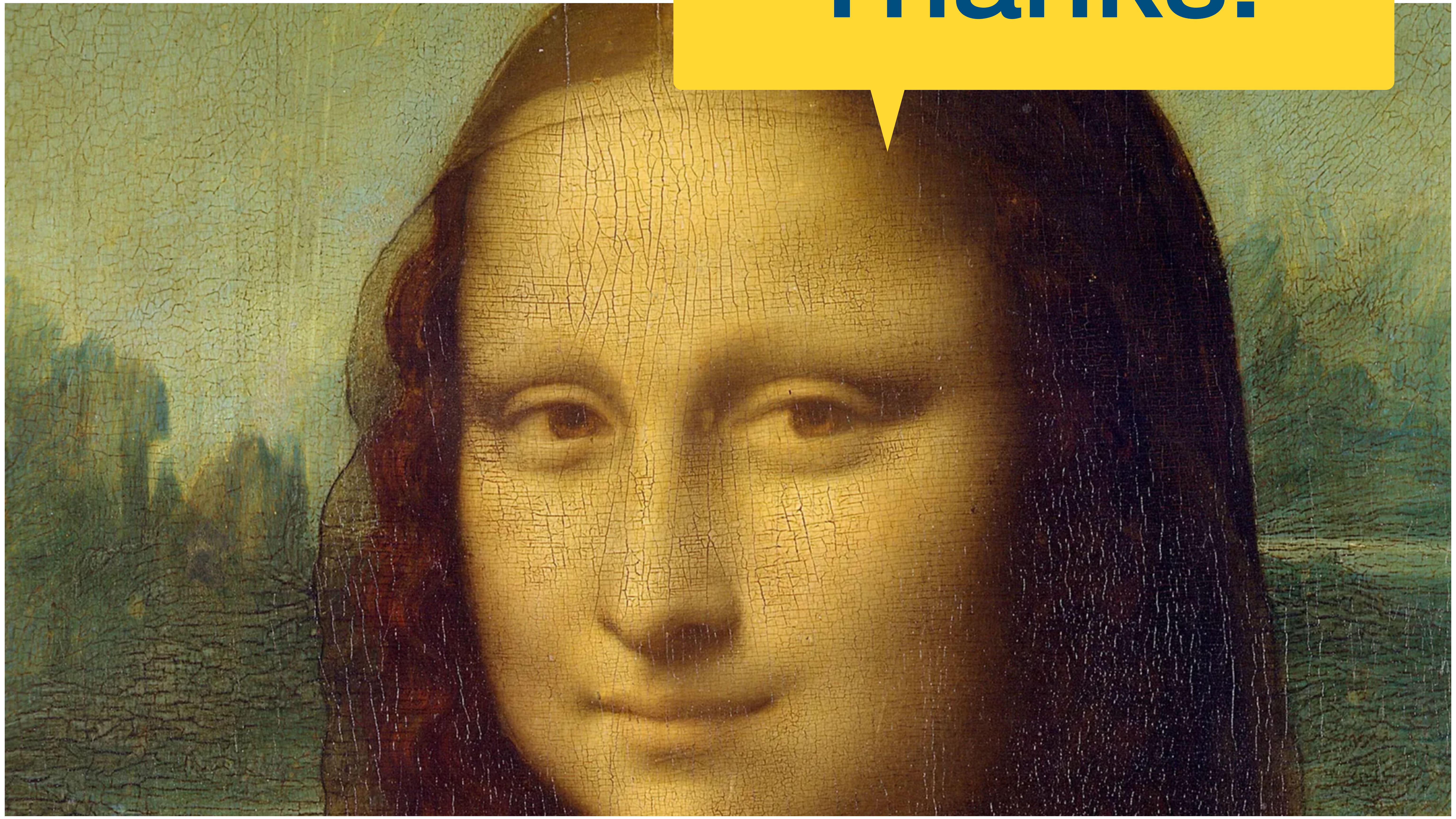
Summary

- The Jet Track Function analysis first extracts a **foundational quantity in QCD**
  - Also **the key missing piece** required to connect **track-based jet substructure measurements to theory calculations**, opening up a new chapter of precision jet substructure program at colliders
  - Provides a new test of pQCD - a qualitative first step **beyond the DGLAP paradigm**
- The Lund subjet multiplicity measurement provides key inputs for **validating the higher accuracy parton shower development**
  - **Broad impact** on improving the **precision** & enhancing the sensitivity of LHC results
- The OmniFold (MultiFold) analysis of Z+jets kinematic observables demonstrates a first **simultaneous unbinned 24-dimensional measurement**
  - Allows to **optimize binning post-unfolding** & **construct new observables** based on the 24 unfolded observables

References

- The Jet Track Function measurement: CONF-STDM-2024-14
  - 1303.6637, 2108.01674, 2201.05166, 2210.10061, 2308.00028
- The Lund subjet multiplicity measurement: CERN-EP-2024-029
  - 2002.11114, 2205.02861, 2208.06057, 2307.11142
- The OmniFold (MultiFold) Z+jets measurement: CERN-EP-2024-132
  - 1911.09107
  - https://gitlab.cern.ch/atlas-physics/public/sm-z-jets-omnifold-2024

Thanks!



Backup

Measurement of Jet Track Functions

Event Selection and Unfolding

CONF-STDM-2024-14

- Prescaled triggering on p_T^{leading}
- $R = 0.4$ anti-kt **PFlow**
- Select the two leading jets
- Fiducial selection: $p_T^{\text{leading}} > 240$ GeV, $|\eta| < 2.1, \ p_T^{\text{leading}}/p_T^{\text{subleading}} < 1.5$
- Unfolding in 3 dimensions: jet p_T, $r_q$relatively $|\eta|$

Binning Correction Uncertainty

- The uncertainty associated with the binning correction is then taken as:

$$\sigma = \left(\frac{m^{\text{MultiFold}}_{\text{unbinned}}}{m^{\text{MultiFold}}_{\text{IBU binning}}}\right)^{\text{nominal}} - \left(\frac{m^{\text{MultiFold}}_{\text{unbinned}}}{m^{\text{MultiFold}}_{\text{IBU binning}}}\right)^{\text{leadingUnc.}}$$

Inclusive tracking efficiency unc.

Table 3: Summary of the binning correction and associated uncertainty for the different moments. Here showing only for the central results and $300 < p_{\mathrm{T}} < 400$ GeV for brevity.

Moment order	Corr. from Nominal	Corr. from Incl. Eff.	Absolute Corr. Unc.	% Corr. Unc.
1	0.9969	0.9932	0.0037	0.34%
2	0.9873	0.9824	0.0049	0.43%
3	0.9746	0.9688	0.0058	0.50%
4	0.9597	0.9526	0.0071	0.57%
5	0.9437	0.9365	0.0072	0.58%
6	0.9277	0.9202	0.0075	0.58%

Very small as expected

Measurement of Lund Subjet Multiplicities

Event Selection and Unfolding

CERN-EP-2024-029

- Triggering on p_T^{leading}
- $R = 0.4$ AntiKt**PFlow**
- Select the two leading jets
- Fiducial selection: $p_T^{\text{leading}} > 240$ GeV $|\eta| < 2.1, \;\; p_T^{\text{leading}}/p_T^{\text{subleading}} < 1.5$
- Unfolding in 3 dimensions: jet p_T, multiplicity, relatively $|\eta|$

- **2 regularization methods** on the fine-binned results
- Rebinning
  - Ensures distributions **not sensitive to sub-detector resolution effects**
  - Avoids potential bias inherent in wide bins present in tails
- Averaging
  - Similar effect of regularization as rebinning
  - **Removes need for binning correction** on average

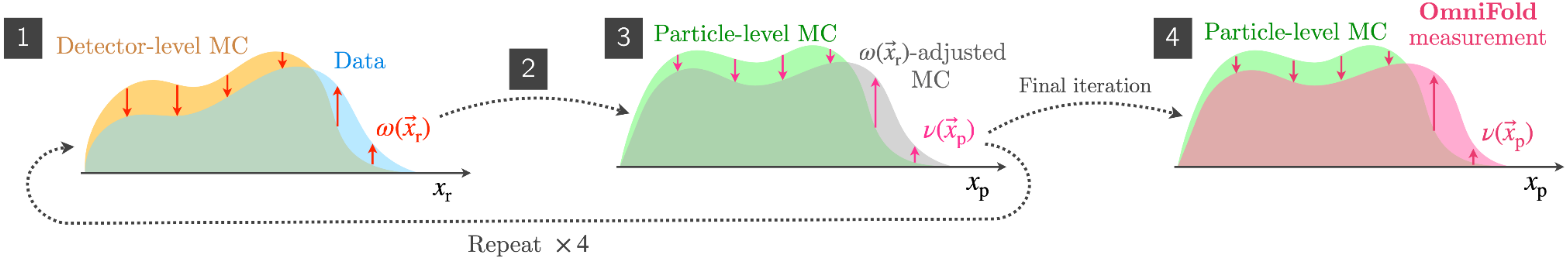
Regularization methods: M. Stanley, P. Patil & M. Kuusela 2111.01091

A Simultaneous Unbinned Differential Cross Section Measurement of 24 Z+jets Kinematic Observables

Unfolding Methodology

CERN-EP-2024-132

- The OmniFold (MultiFold) algorithm takes **2 sets** of inputs: MC samples at **both truth & reco levels**, and **the data**
- **Iteratively** proceed in **2 steps of reweighting** & **2 steps of truth-reco mapping**



- Extensive **validation** studied:
  - Closure test (make sure to target "pseudo-data"); χ^2 compatibility test (p-values>0.05); stress test (dramatic detector distortions) → usage recommendations*
  - A **blinded** measurement