Analysis progress:

Measurement of Q/G Fractions and Topics Extractions in Dijet Events

ATLAS-HD Meeting Anweiler am Trifels

Monday, May 30, 2022

Matt Leblanc[1], Christof Sauer[2], André Schöning[3] (supervisor)

[1] CERN, Geneva, Switzerland; [2,3] Ruprecht-Karls-Universität Heidelberg

[1] mleblanc@cern.ch, [2] christof.sauer@cern.ch, [3] schoning@mail.desy.de

Analysis – What to measure?

– Incl. triple-differential dijet cross section

- We'd like to measure the inclusive triple-differential dijet cross section

$$\frac{1}{\sigma^{\text{fid}}} \frac{\mathrm{d}^3\sigma}{\mathrm{d}p_{\text{T}}\mathrm{d}y\mathrm{d}z_{\text{NN}}} = \frac{1}{N^{\text{tot}}} \frac{N}{\Delta p_{\text{T}}\Delta y\Delta z_{\text{NN}}}$$

- The number of dijet events is measured in bins of ...
  - ... **transverse momentum** p_{T}
  - ... **rapidity** y (more central and more forward jets)
  - ... a **NeuralNetwork score** z_{NN}[a]
- **Only the leading two jets** in each event **are considered**, separated in being "more central" and "more forward", i.e., $|y_{\text{central}}| < |y_{\text{forward}}|$
- We use **tracks** matched to fully calibrated ($R = 0.4$) `EMTopo` jets for the measurement
- For the event selection, the two leading jets satisfy $p_{\text{T}}^{\text{lead}}/p_{\text{T}}^{\text{sublead}} < 1.5$, $|\eta^{(\text{sub})\text{lead}}| < 2.1$ as well as $p_{\text{T}} > 60\,\text{GeV}$

[a] The same measurement was performed wrt the Soft Drop multiplicity n_{SD}

Motivation

At colliders like the LHC, it is nearly impossible to kinematically isolate pure samples of different jets (e.g., quark jets, gluon jets etc)

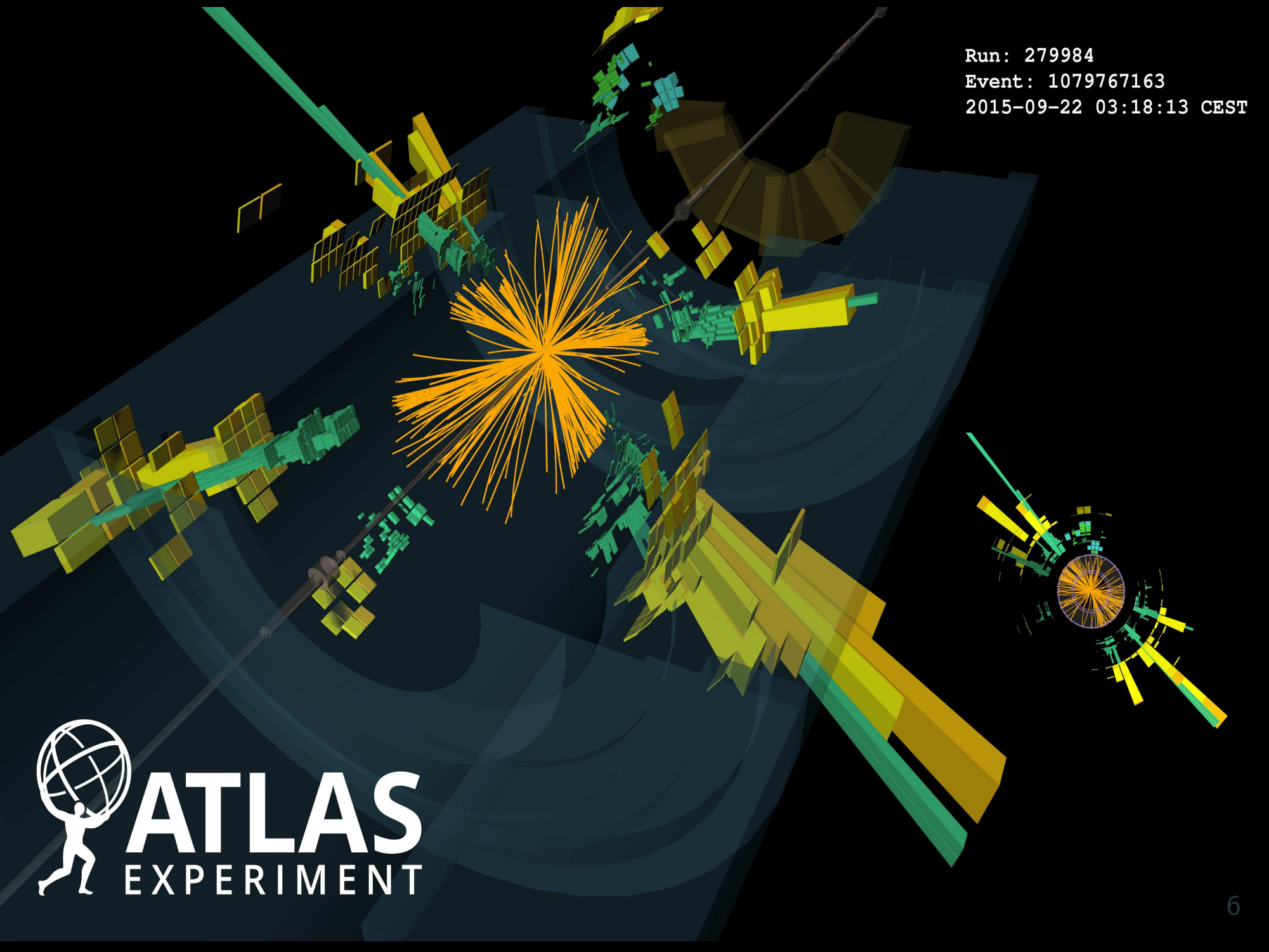
Multijet samples

- This analysis makes use of several MC samples
- They are used to derive systematic uncertainties (especially model-related ones)

MC Generator	Matrix Elements	Parton Shower	Hadronisation	PDFs
Pythia 8.230 (Nominal)	**2→2 LO**	Dipole-style p_T-ordered	Lund string	NNPDF 2.3
Powheg V2 +Pythia 8.230	**2→2 NLO**	Dipole-style p_T-ordered	Lund string	NNPDF30NLO
Herwig 7.1.3	2→2 NLO 2→3 LO	**Angle-ordered**	Cluster	MMHT2014NLO
Herwig 7.1.3	2→2 NLO 2→3 LO	**Dipole**	Cluster	MMHT2014NLO
Sherpa 2.2.5	2→2 LO	CSS	**Cluster (AHADIC)**	CT14NNLO
Sherpa 2.2.5	2→2 LO	CSS	**Lund string (via Pythia 6.4)**	CT14NNLO
MG5_aMC@NLO +Pythia 8.212	**2→2,3,4 LO**	Dipole-style p_T-ordered	Lund string	NNPDF30NLO
Powheg+Herwig 7	2→2 NLO	**Angle-ordered**	**Cluster**	NNPDF30NLO

- Following results are for campaign MC16 a+d+e and Data 1516+17+18 (Run2)

What's a jet?


Run: 279984
Event: 1079767163
2015-09-22 03:18:13 CEST
ATLAS
EXPERIMENT

Jets in a probabilistic language

- A jet can be described as an unordered set of particles with variable length

$$J(\{p_1^\mu, ..., p_M^\mu\}) = J(\{p_{\pi(1)}^\mu, ..., p_{\pi(M)}^\mu\}), \quad M \geq 1, \quad \forall \pi \in S_M$$

- p_i^μ denotes the four-momentum of the ith constituents of the jet
- The jet is unordered due to quantum-mechanical indistinguishability
- It is of variable length due to the probabilistic nature of its formation

- Why is this important?

→ **Because we should respect those symmetries when dealing with jets**

→ **Deep Sets** [arXiv:1703.06114] (EFN/PFN [arXiv:1810.05165])

Architectures

– EFNs and PFNs

$$z_{\text{EFN}}(\{p_1, ..., p_M\}) = F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right)$$

$$z_{\text{PFN}}(\{p_1, ..., p_M\}) = F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$

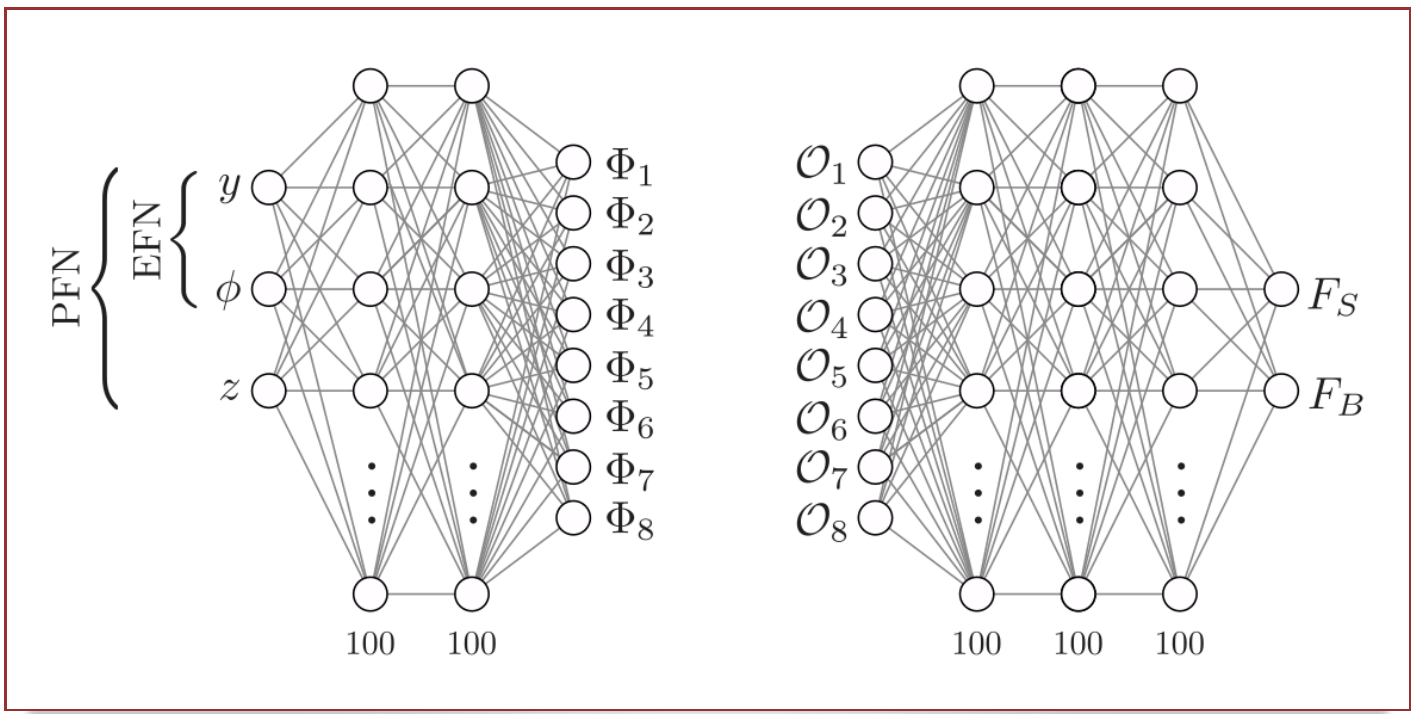
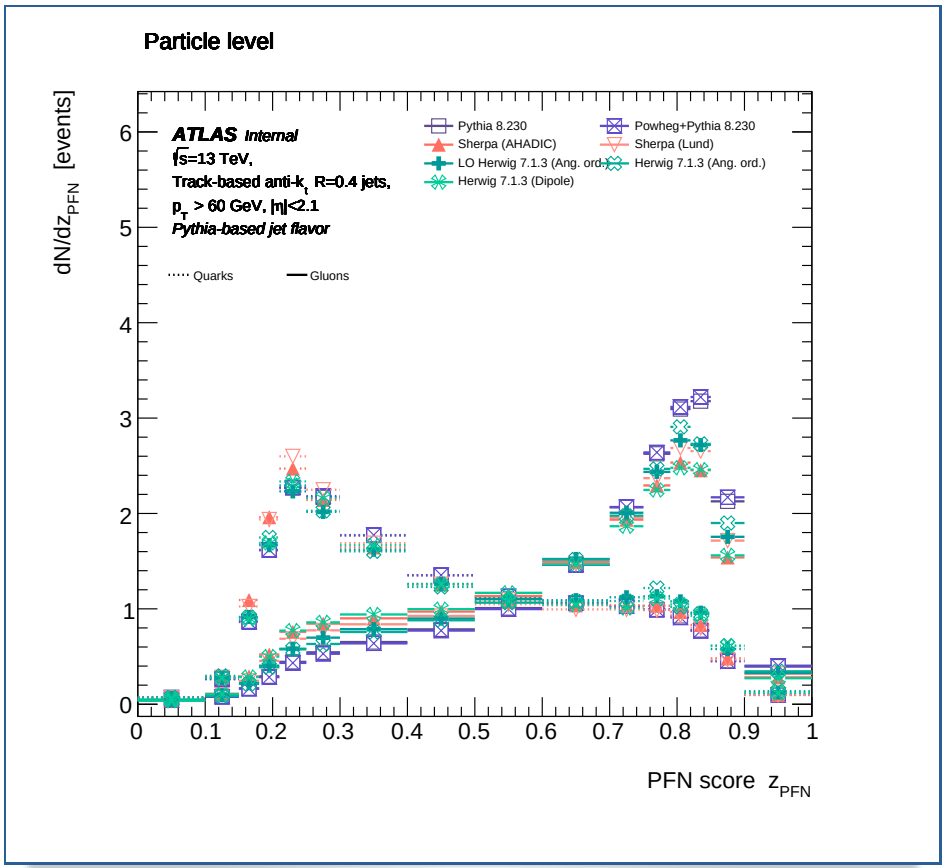


Image adapted and modified from *Energy Flow Networks: Deep Sets for Particle Jets*, P. T. Komiske *et al.*

- Both architectures implement functions that are invariant wrt permutations of its inputs
- **Input features:**

  momentum fraction $z_j = p_{\text{T},j}/p_{\text{T},J}$, centered $\delta\eta_j = \eta_j - \eta_J$ and $\delta\phi_j = \phi_j - \phi_J$

Training
– Quark/Gluon NN score z_{NN} distributions



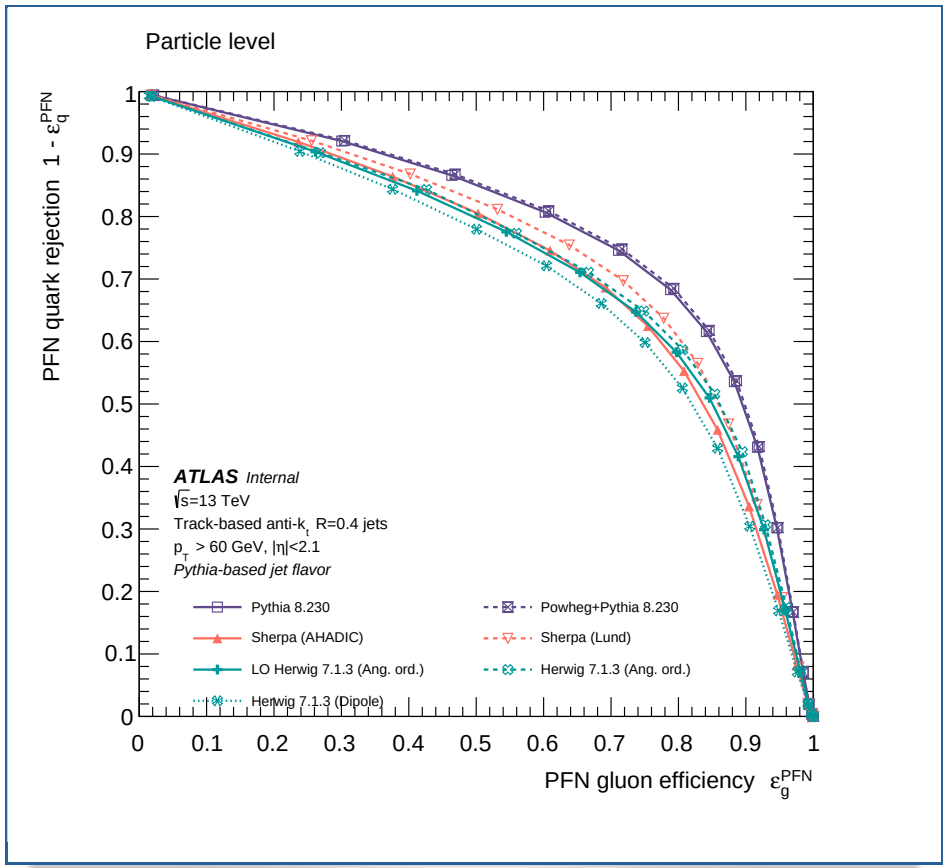


Fig. Score for Q and G jets for different generators (left) and ROC curve (right)

- Plot the predicted score of the NN for quark and gluon jets respectively
- Distributions are inclusive in transverse momentum $[60, 4500]$ (GeV)
- Good discrimination power between quarks and gluons!

Distributions
– p_{T} Spectra for MC and Data

$$\frac{1}{N_{\mathrm{tot}}} \frac{N}{\Delta p_{\mathrm{T}} \Delta y \Delta z_{NN}}$$

- Figure shows the transverse momentum for the binning used in the measurement
- The lower (sub-)plot shows the ratio of all MCs wrt data
- All distributions have been normalized to data (see resp. scale factor for each MC in the legend)
- Spectra look nice and smooth → re-weighting of JZ slices and application of trigger scale factors works

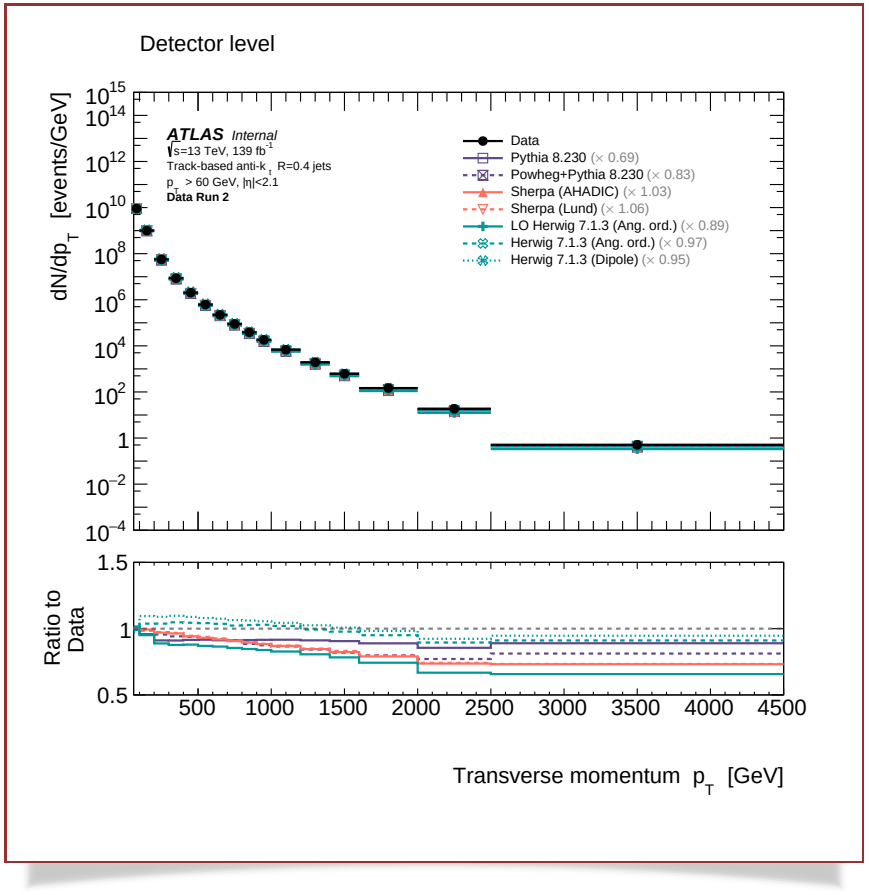


Fig. Transverse momentum p_{T}.

Distributions
– NN score (PFN in this case)

$$\frac{1}{N^{tot}} \frac{N}{\Delta p_T \Delta y \Delta z_{NN}}$$

- Figure shows the NN(=PFN) score for different generators and data for the binning used in the measurement (binning not optimal yet)
- Gluon-initialized jets tend towards larger values of z_{NN}, while (light) quarks are more likely to get a smaller score
- The lower (sub-)plot shows the ratio of all MCs wrt data
- All distributions have been normalized to data (see resp. scale factor for each MC in the legend)

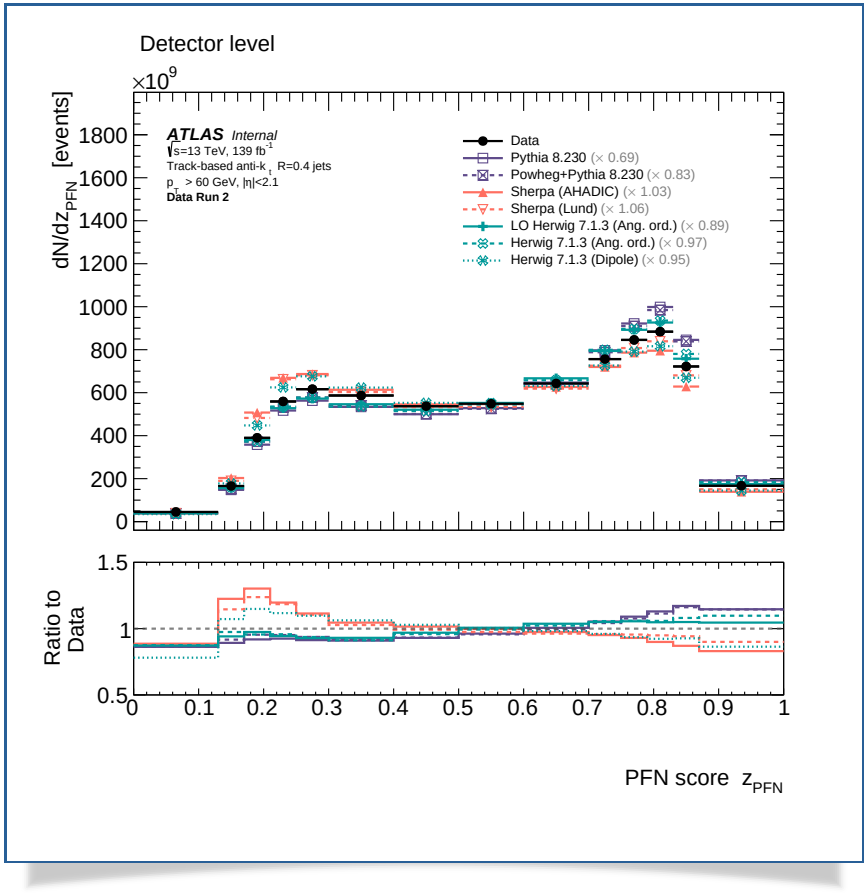
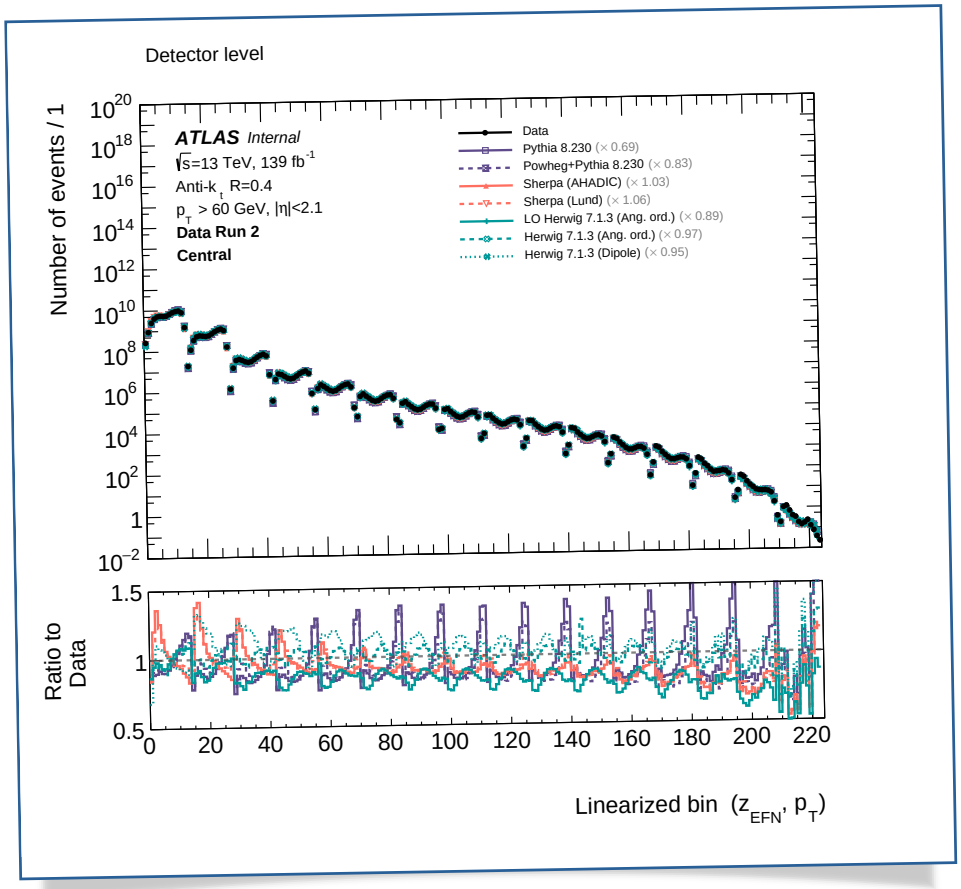
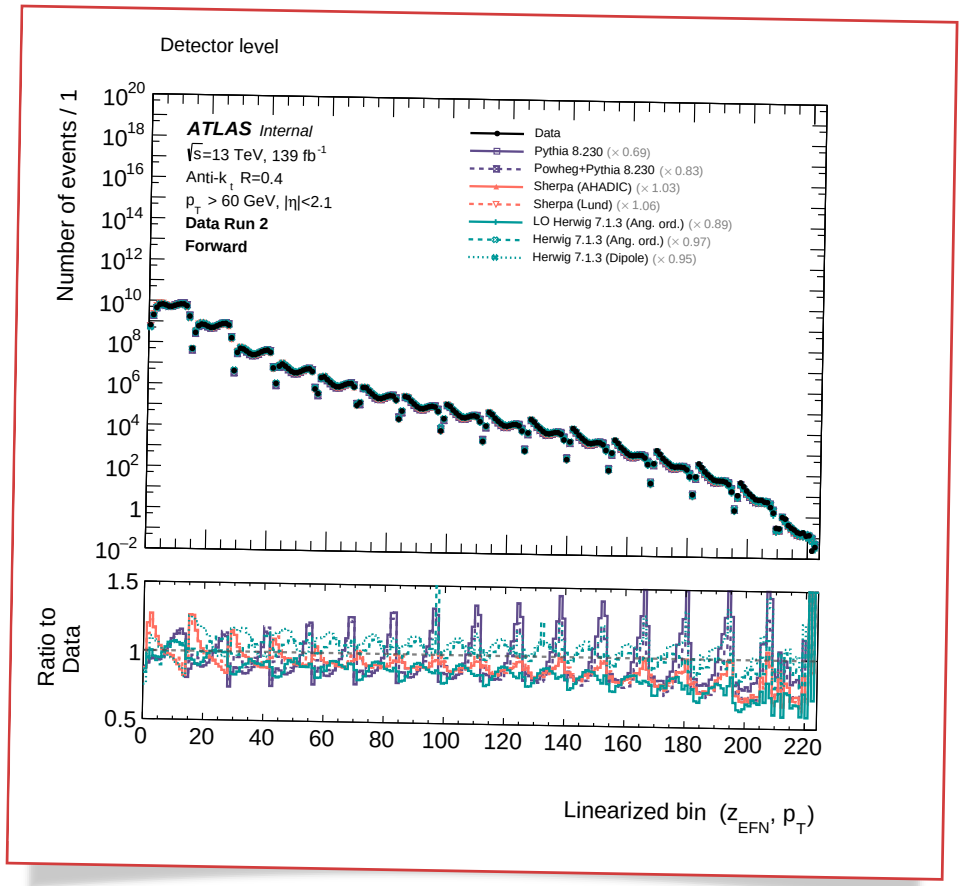


Fig. Incl. NN score z_{PFN}

Full measurement at Detector Level

- Plot the NN score in each p_{T} bin for central (left) and forward (right) jets





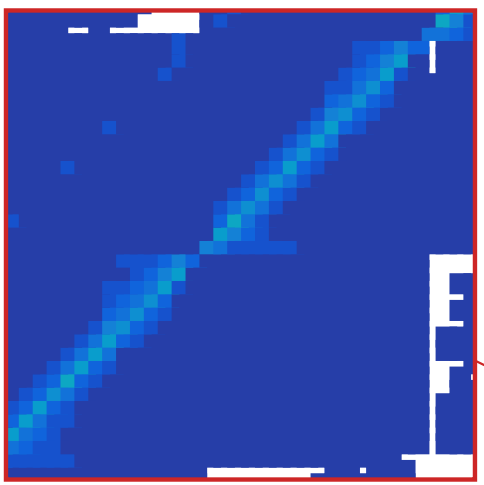
- Nmber of quark initialized jets increases towards higher p_{T}
- Furthermore, there even is a clear difference between forward and central jets (latter one is gluon enriched)

Unfolding

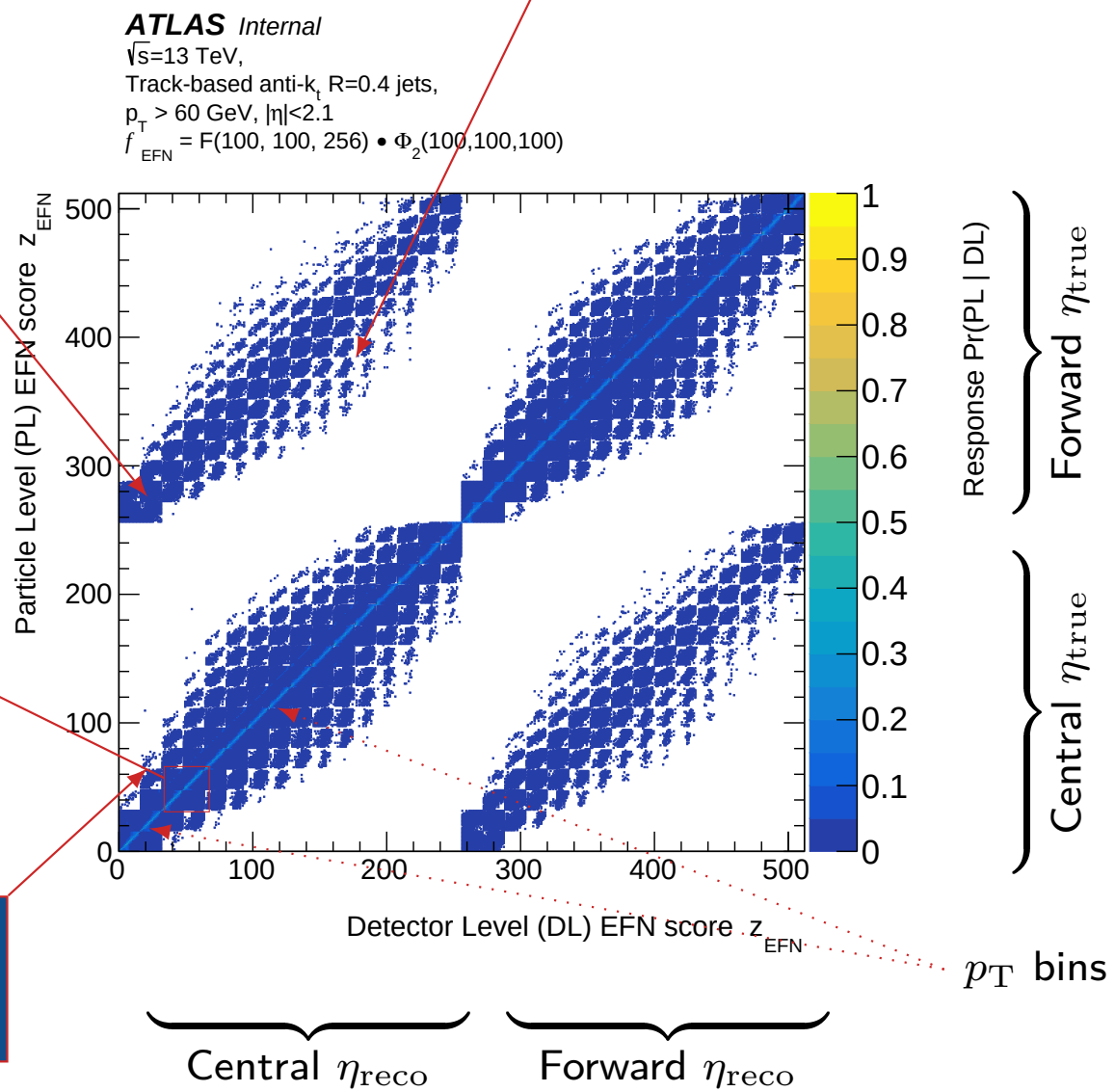
– Features of the resp. matrix

Forward truth jet with $p_{\mathrm{T}}^{\mathrm{true}} \in [400, 760]$ being reconstructed as central jet with $p_{\mathrm{T}}^{\mathrm{true}} \in [760, 1000]$

Forward truth jet being reconstructed as central jet

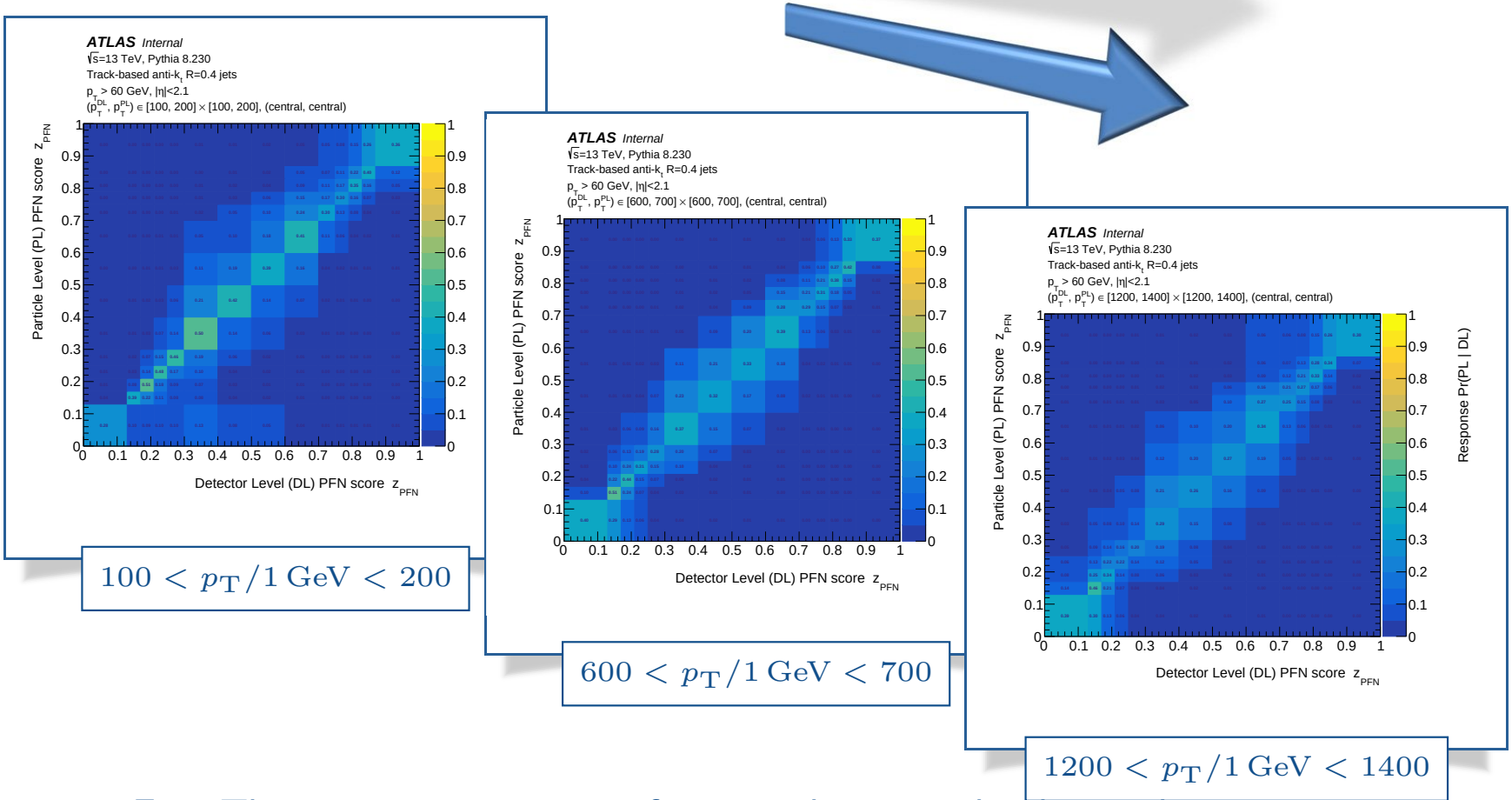


Truth jets with $p_{\mathrm{T}}^{\mathrm{true}} \in [175, 270]$ but reconstructed with $p_{\mathrm{T}}^{\mathrm{reco}} \in [120, 175]$



Response matrices – three examples

- To make the interpretation a bit easier: look at some examples
- The three figures show the resp. matrices for the nominal sample (Pythia8.230)
- Not perfect, but keep in mind that the model has been trained on truth level information solely!

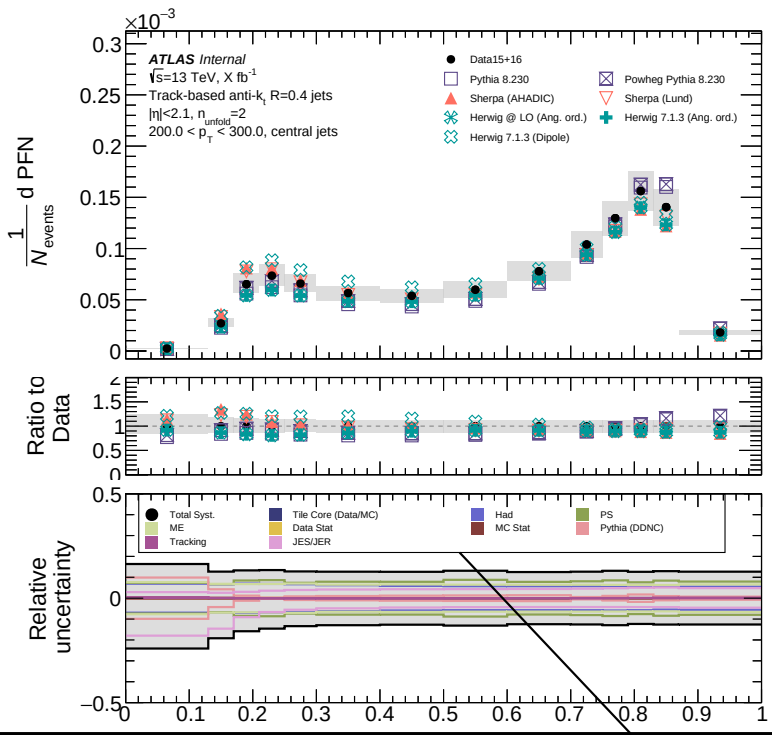


Figs. Three response matrices for central jets on the diagonal.

Uncertainties

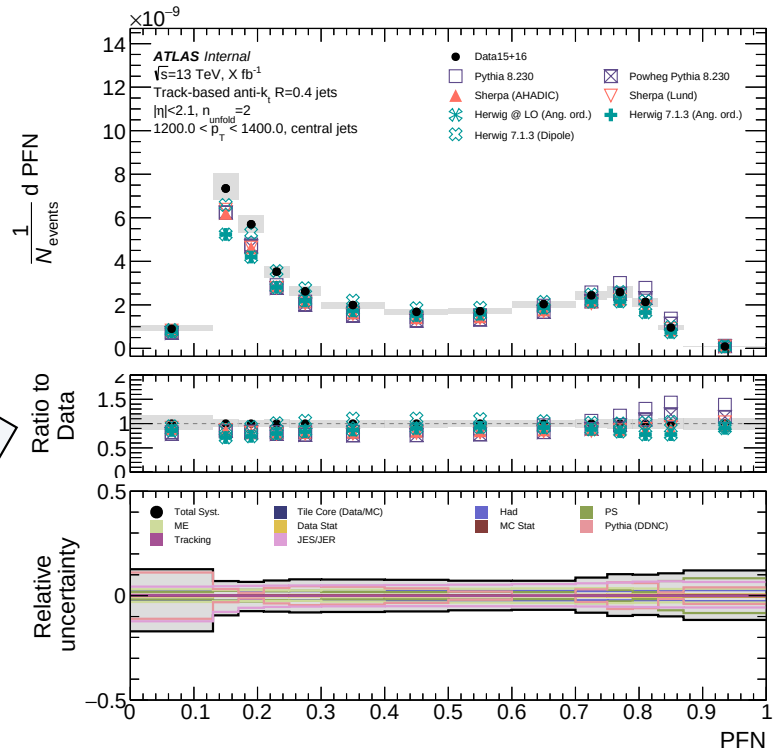
– Different p_{T} bins, same region (central jets)

$200 < p_{\mathrm{T}}/1\,\mathrm{GeV} < 300$



More quarks!!!

$1200 < p_{\mathrm{T}}/1\,\mathrm{GeV} < 1400$

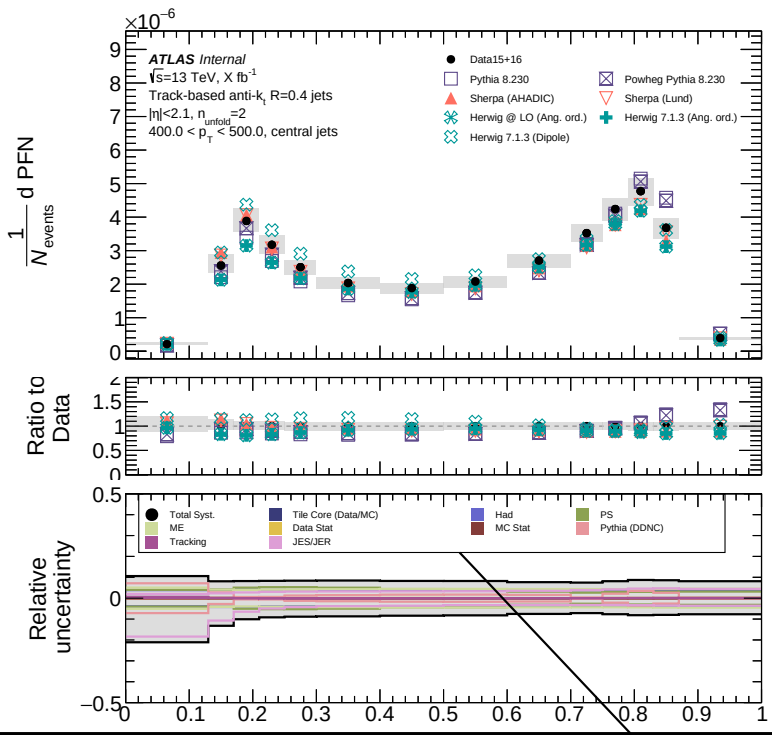


- Strong p_{T} dependence observed
- As expected from the multijet XS, fraction of quark-initialized jets increases

Uncertainties

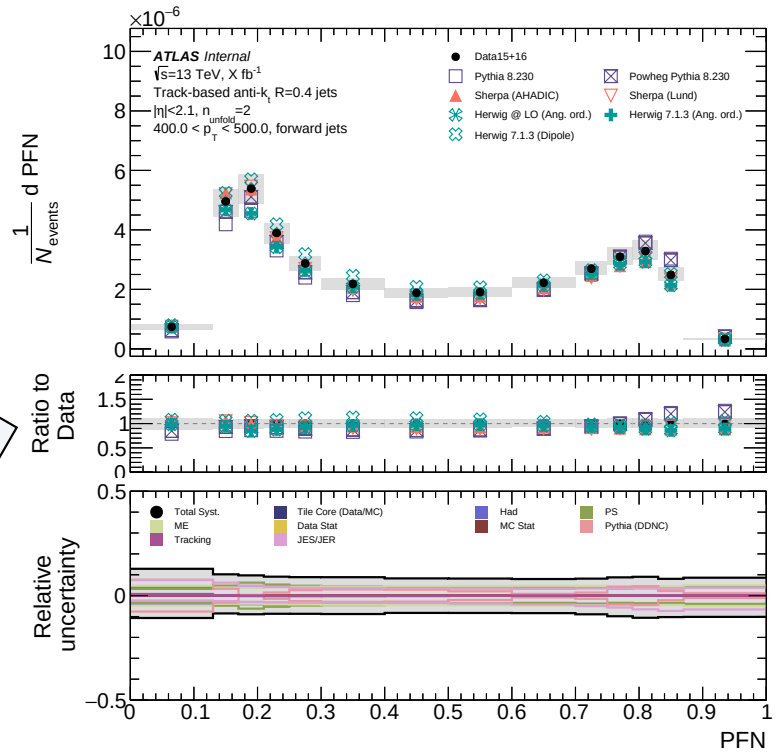
– Same p_{T} bin, different region

Central



More quarks!!!

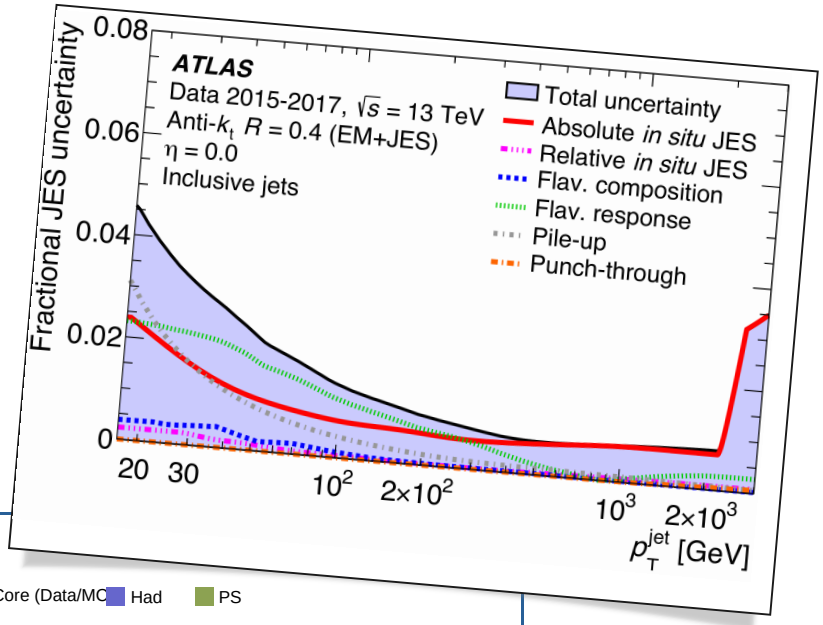
Forward

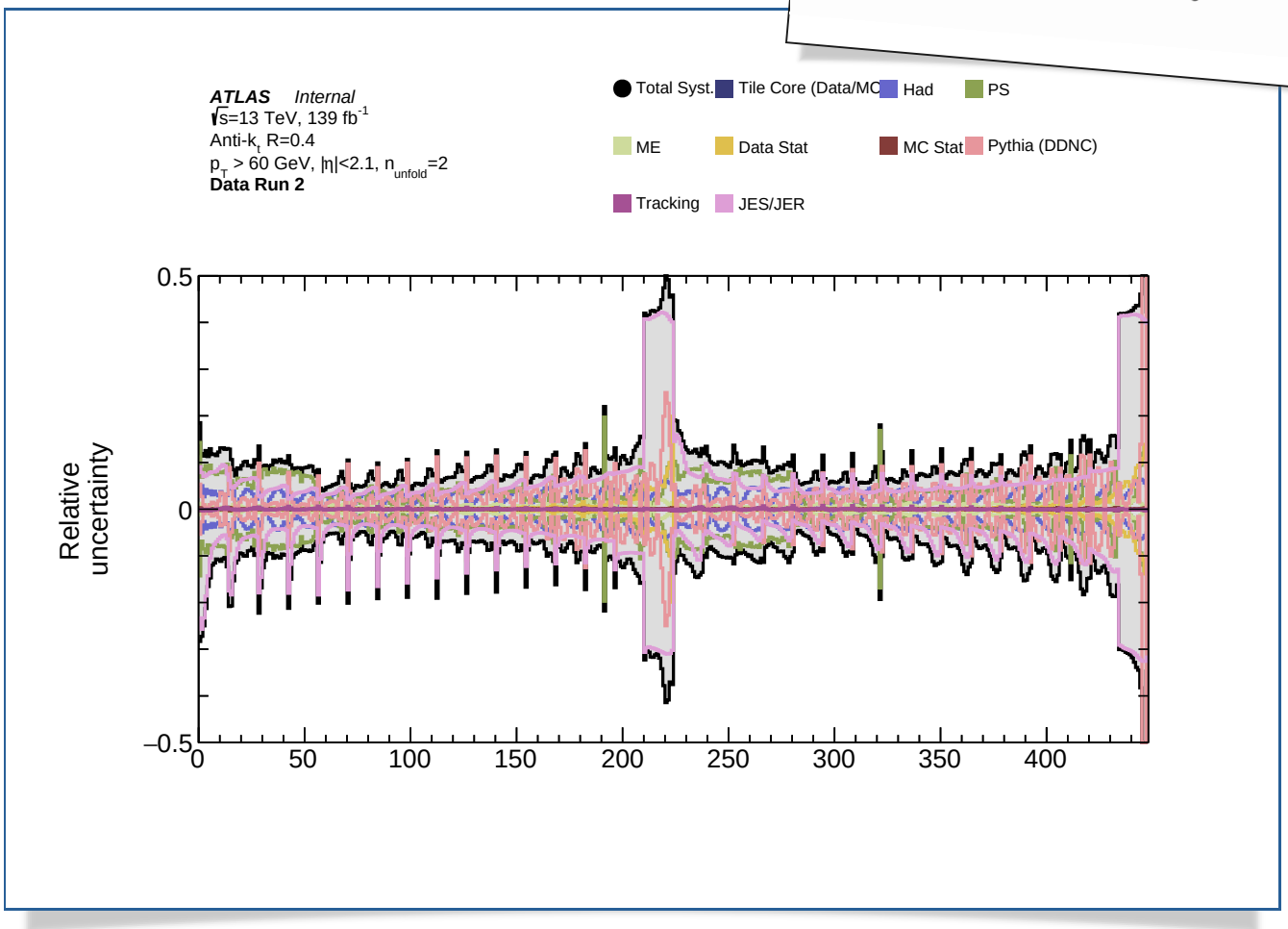


- The NN is capable of distinguish between quarks and gluons in the different regions

Full Measurement
– Uncertainties

- Figure shows the uncertainties for each bin
- Modeling uncertainties mostly dominant
- Highest p_{T} dominated by single particle NP





"Jet Topics"
– Disentangeling Quark and Gluon Distributions

On the Topic of Jets: Disentangling Quarks and Gluons at Colliders, Eric M. Metodiev, Jesse Thaler

- The basic idea is to measure quark and gluon distributions in a label-agnostic way (**no** need for MC templates!)
- The assumtion is that the two samples $p_{M_{1/2}}(x)$ are a mixture of the underlying Q/G distributions

$$p_{M_1}(x) = f_1^Q p_Q(x) + (1 - f_1^Q) p_G(x)$$
$$p_{M_2}(x) = f_2^Q p_Q(x) + (1 - f_2^Q) p_G(x)$$

- The objective is to compute the Q/G distributions $p_{Q/G}$ as well as the fractions $f_{Q/G}$

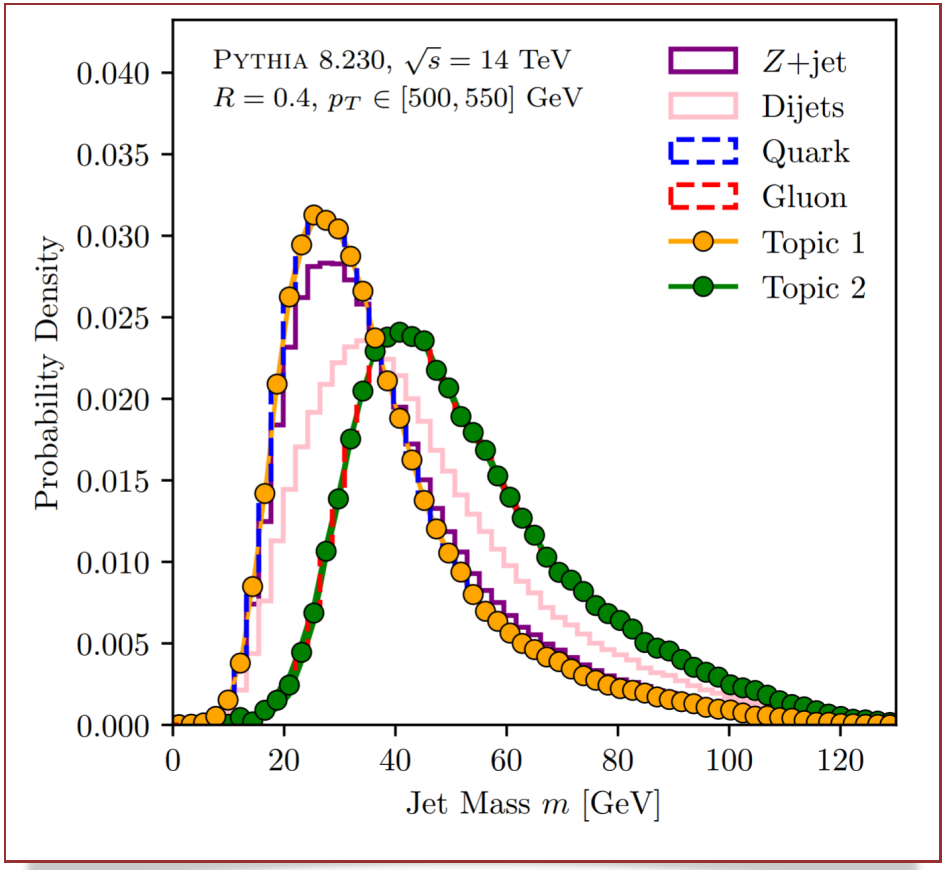
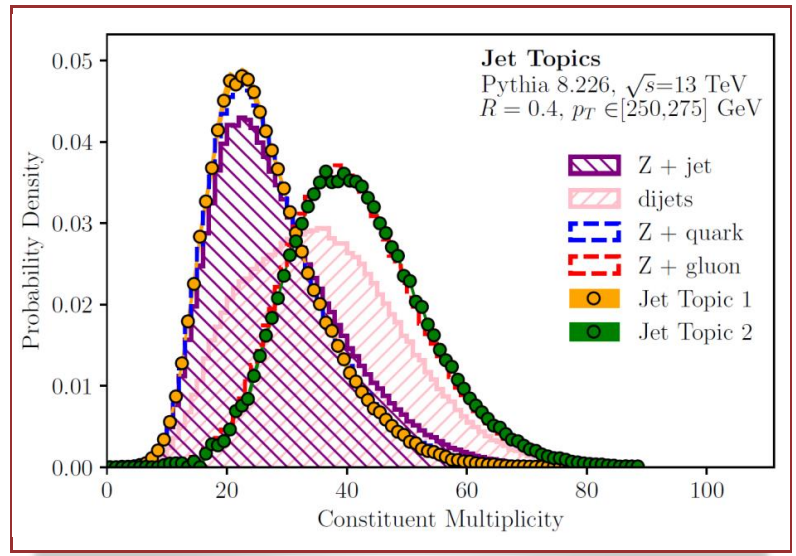


Fig. "Jet Topics"; figure taken from [arXiv:1802.00008].

"Jet Topics"
– Disentangeling Quark and Gluon Distributions

- The method works provided that the observable is *mutually-irreducible*
- To get the fractions and the topic distributions, the so-called *reducibility factors* κ must be determined
- With reducibility factors
  $\kappa_{1/2} = \min_x \frac{p_{M_{1/2}}(x)}{p_{M_{2/1}}(x)}$
- Once the reducibility factors are known, the Q/G fraction as well as the underlying Q/G distributions can be computed

$\rightarrow \; f_1^Q = \frac{1-\kappa_1}{1-\kappa_1\kappa_2},\; f_2^Q = \kappa_2 \frac{1-\kappa_1}{1-\kappa_1\kappa_2}$

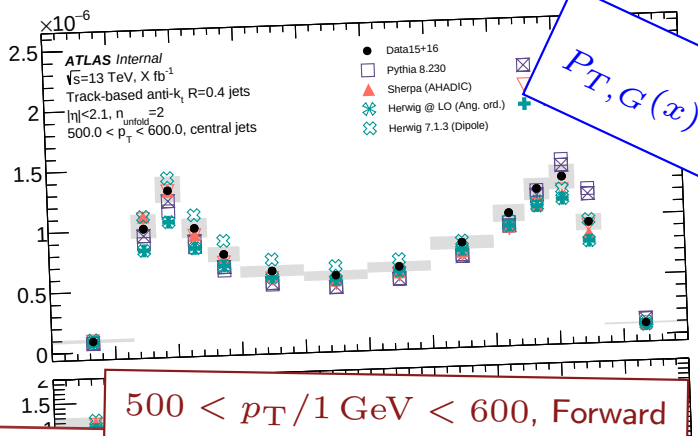


$$p_Q = \frac{p_{M_1} - \kappa_1 p_{M_2}}{1-\kappa_1} \qquad p_G = \frac{p_{M_2} - \kappa_2 p_{M_2}}{1-\kappa_2}$$

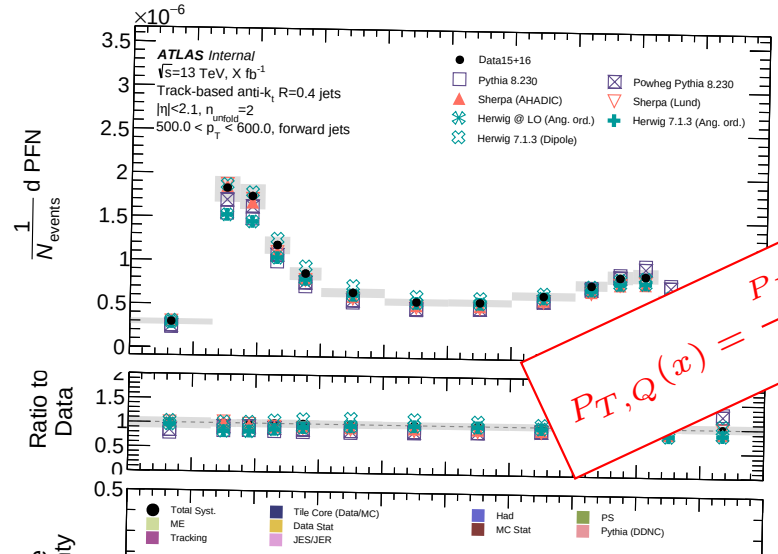
“Jet Topics“ – One example

- Determine separate quark and gluon jet distributions using “jet topics”
- The topics are extracted directly from the hadron-level multi-differential XS with **no** input from simulation

$500 < p_{\mathrm{T}}/1\,\mathrm{GeV} < 600$, Central

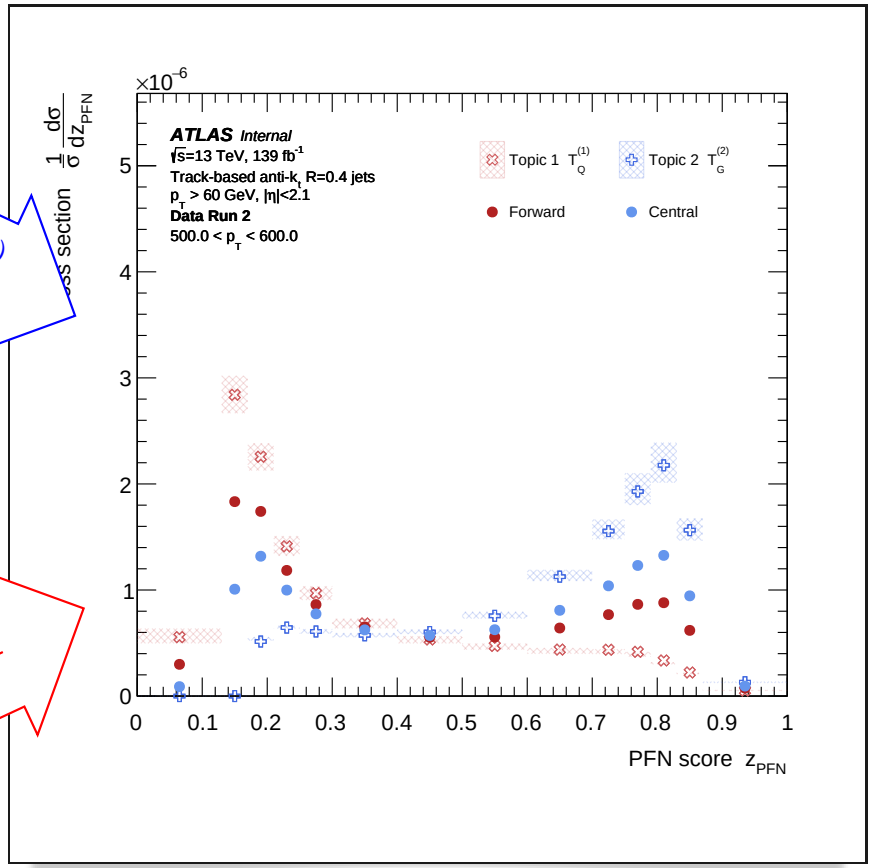


$500 < p_{\mathrm{T}}/1\,\mathrm{GeV} < 600$, Forward



$$P_{T,G}(x) = \frac{PM_c(x) - \kappa_c PM_f(x)}{1-\kappa_c}$$

$$P_{T,Q}(x) = \frac{PM_f(x) - \kappa_f PM_c(x)}{1-\kappa_f}$$

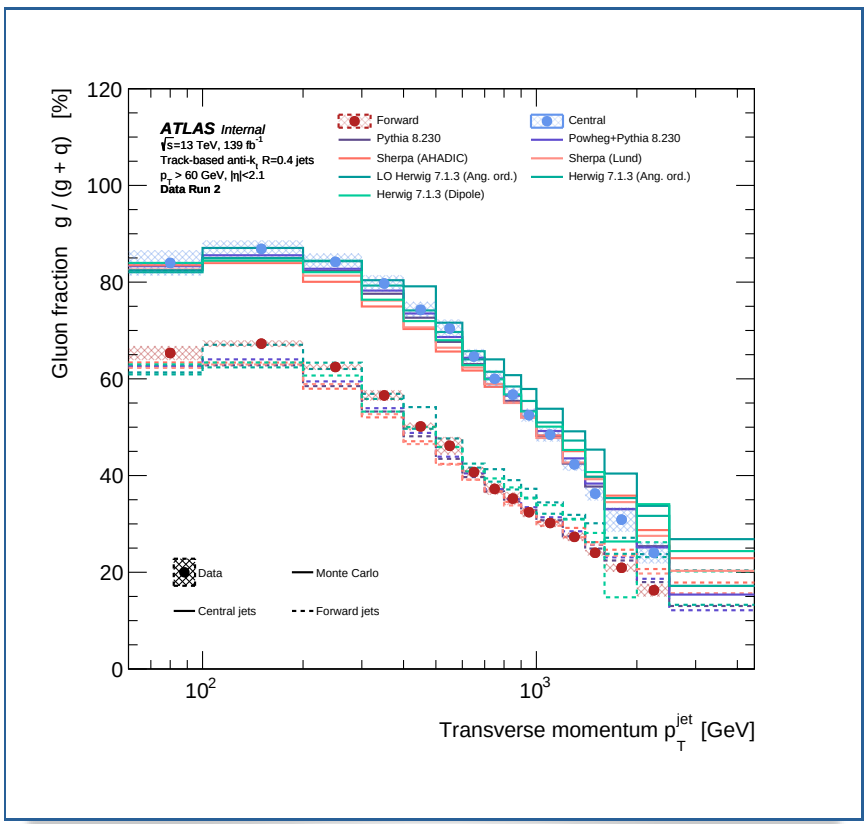


On the Topic of Jets: Disentangling Quarks and Gluons at Colliders, Eric M. Metodiev, Jesse Thaler

“Jet Topics“

– Gluon fractions from data (Run 2)

- Given the jet topics (Q/G distributions), the Q/G fraction can be computed
- Figure shows the gluon fraction in different p_{T} bins
- Clear differences wrt to the gluon fraction given by arbitrary jet label (right)
- Difference mostly pronounced for low-p_{T} jets



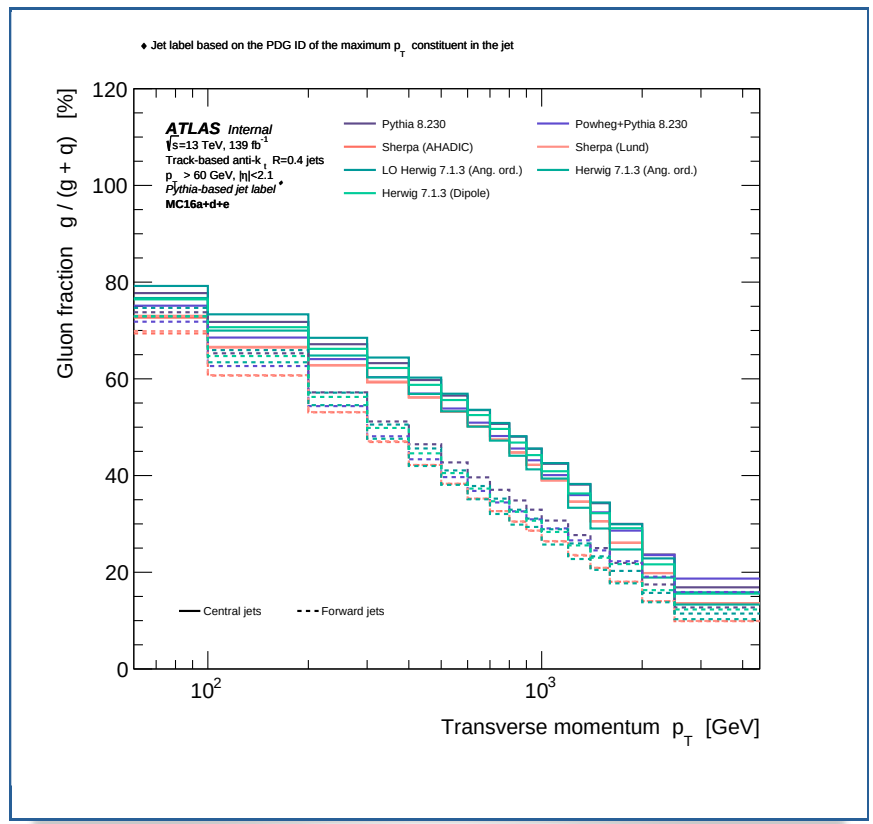


Fig. **Gluon fraction in different p_{T} bins from topics (left) and MC (right).**

p_{T} topics

– Jet topics for non-mutual irreducible observables

- Topic distributions for non-mutual irreducible observables y can be extracted based on the simultaneous measurement $P(x, y)$ with a mutual irreducible one x
- The jet topics of y are then simply given by re-weighting $P_{Q/G}(y) = \int \mathrm{d}x\, P(y|x) P_{Q/G}(x)$
- Figure shows the extracted Q/G distributions for the transverse momentum
- Since the reconstruction of the p_{T} is based on calo-clusters, the obtained topics are also IRC-safe

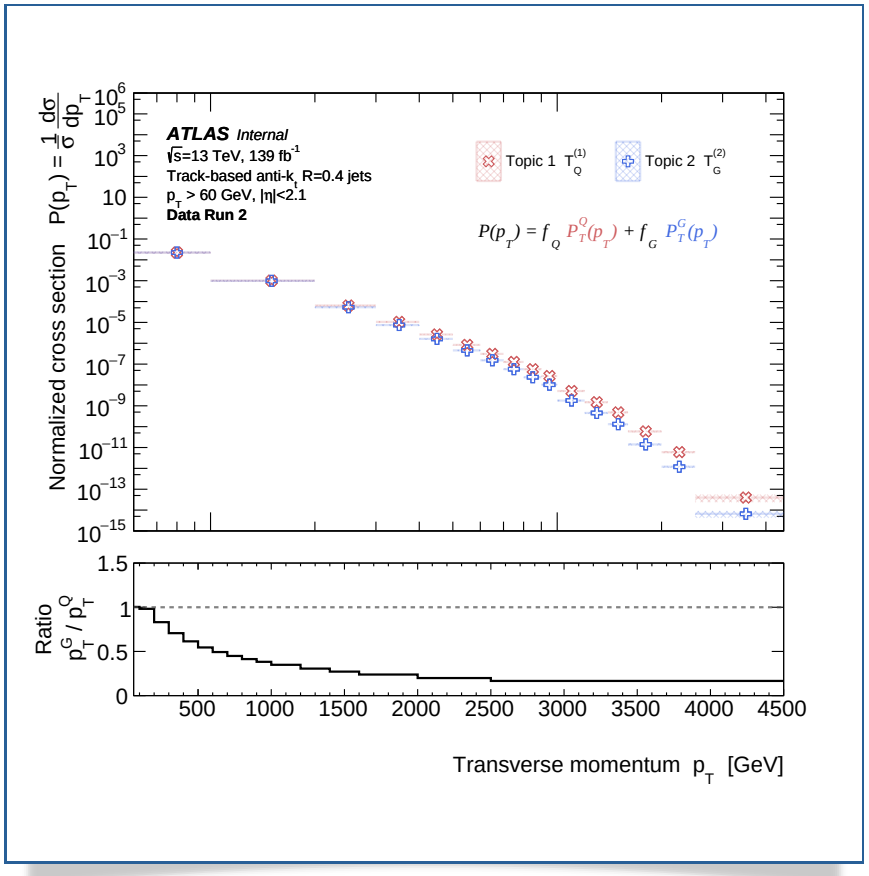


Fig. Q/G p_{T} distributions

Summary

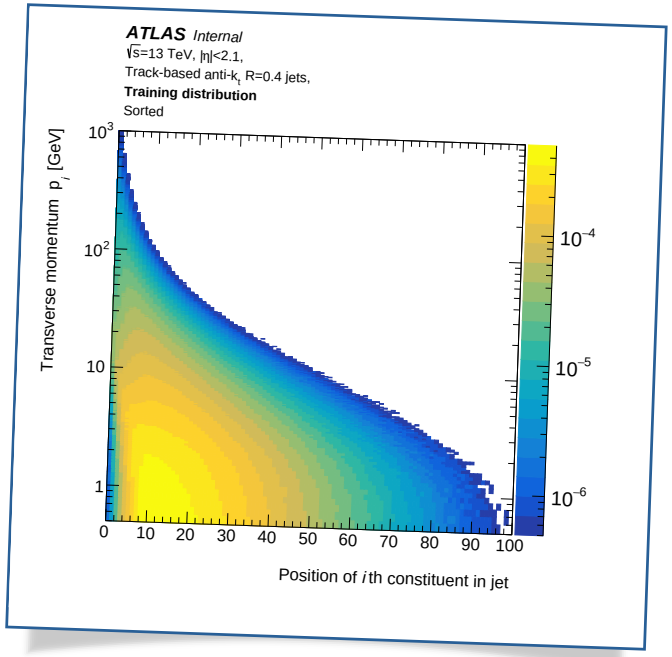
- First results for the measurement of the inclusive triple-differential dijet cross section $\frac{1}{\sigma}\frac{\mathrm{d}^3\sigma}{\mathrm{d}p_\mathrm{T}\,\mathrm{d}\eta\,\mathrm{d}z_\mathrm{NN}}$ was presented
- z_NN is given by to the EFN/PFN score and corresponds to the probability of the current jet being initialized by a gluon
- Binning still isn't optimal and needs further optimization
- Many uncertainties have been studied; as expected, modeling (PS in particular) uncertainties are dominant
- Jet topics are used to extract the corresponding quark- and gluon related distributions
- Based on the Q/G distributions, the gluon fraction in each p_T bin is computed
- The measurement has been done for all MC campaigns and the full Run 2 data

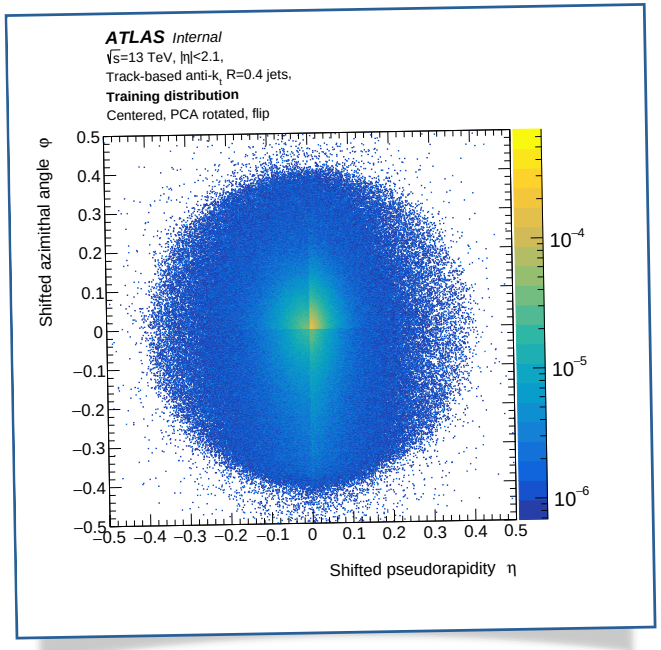
Outlook

- Further optimize the binning of the NN score
- Finalize uncertainties
- For jet topics, take correlations between uncertainties into account
- Get the jet topics and gluon fractions for the SD multiplicity measurement

Closer look at data
– Preprocessing

- Model is trained on particle-level info. only
- Preprocessing consists of several steps
- Process each jet on an event-on-event base
- Exploit as much symmetries as possible to ...
  - ... improve classification performance
  - ... speed up convergence
  - ... reduce size of training set



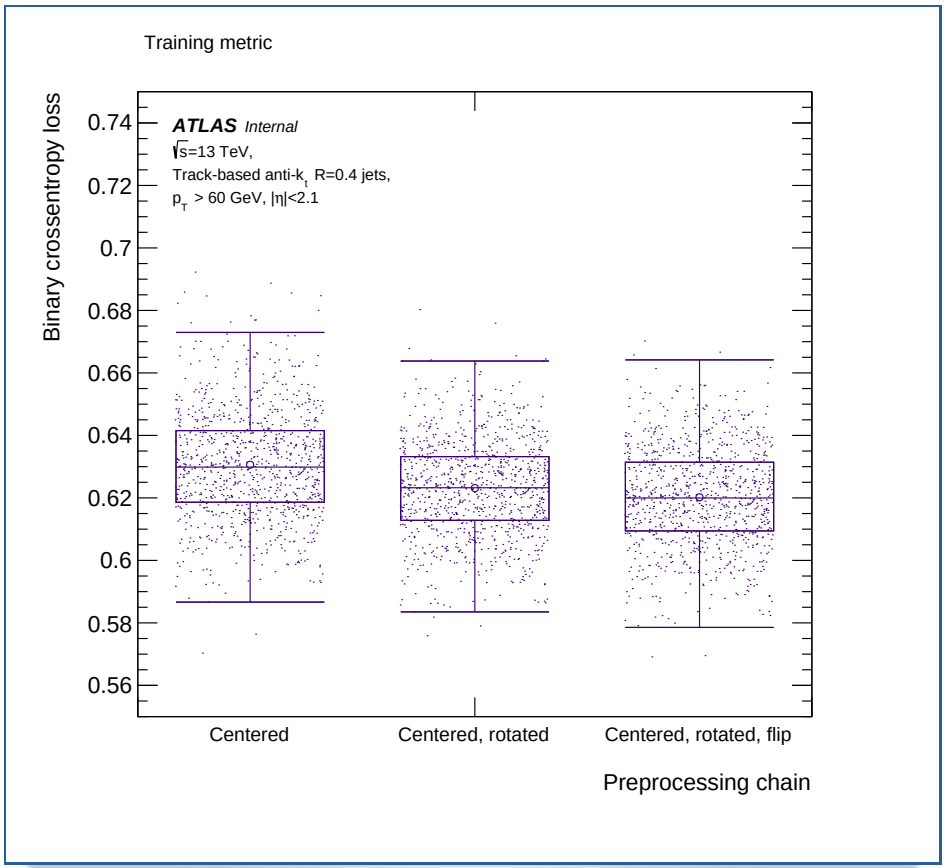


Preprocessing of jets

1. Sort constituents such that $p_{\mathrm{T},i} > p_{\mathrm{T},i+1}$
2. Center jet and its constituents
3. The PrincipalComponent axis of the jet is rotated such that it points at 12 o'clock
4. Flip the jet such that $E_{\mathrm{left}} < E_{\mathrm{right}}$

Training
– Optimization based on training metrics

- Check how some metrics depend on the choice of the preprocessing chain



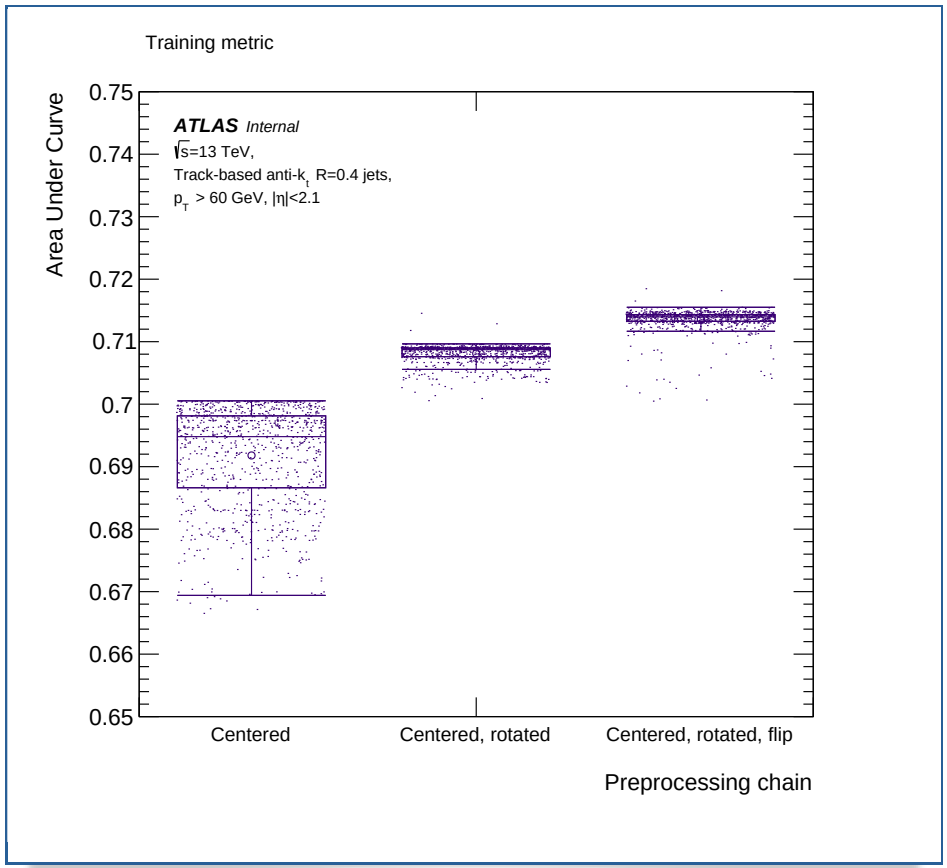


Fig. **Loss (left) and AUC (right) for different preprocessings.**

- Using as much symmetries as possible (center, rotation, flip) improves AUC
- Hence, I use the full preprocessing chain