Classifying hadronic objects in ATLAS with ML/AI algorithms

Shudong Wang (IHEP, CAS)

on behalf of the ATLAS Collaboration

EPS-HEP 2025, Marseille, France

July 8, 2025

Introduction

- Hadronic objects are **ubiquitous** at LHC and are final state objects of various processes of interest.
- Identification of hadronic objects is essential for many physics analyses at the LHC.
- Some recent results on hadronic object identification in ATLAS are covered:
  - Quark/Gluon tagging.
  - Boosted W boson tagging.
  - Boosted top quark tagging.
- These advances enhance the potential for **precise measurements of the SM** (e.g. VBS, VBF) and **discovering new physics** (e.g. BSM heavy resonance search)

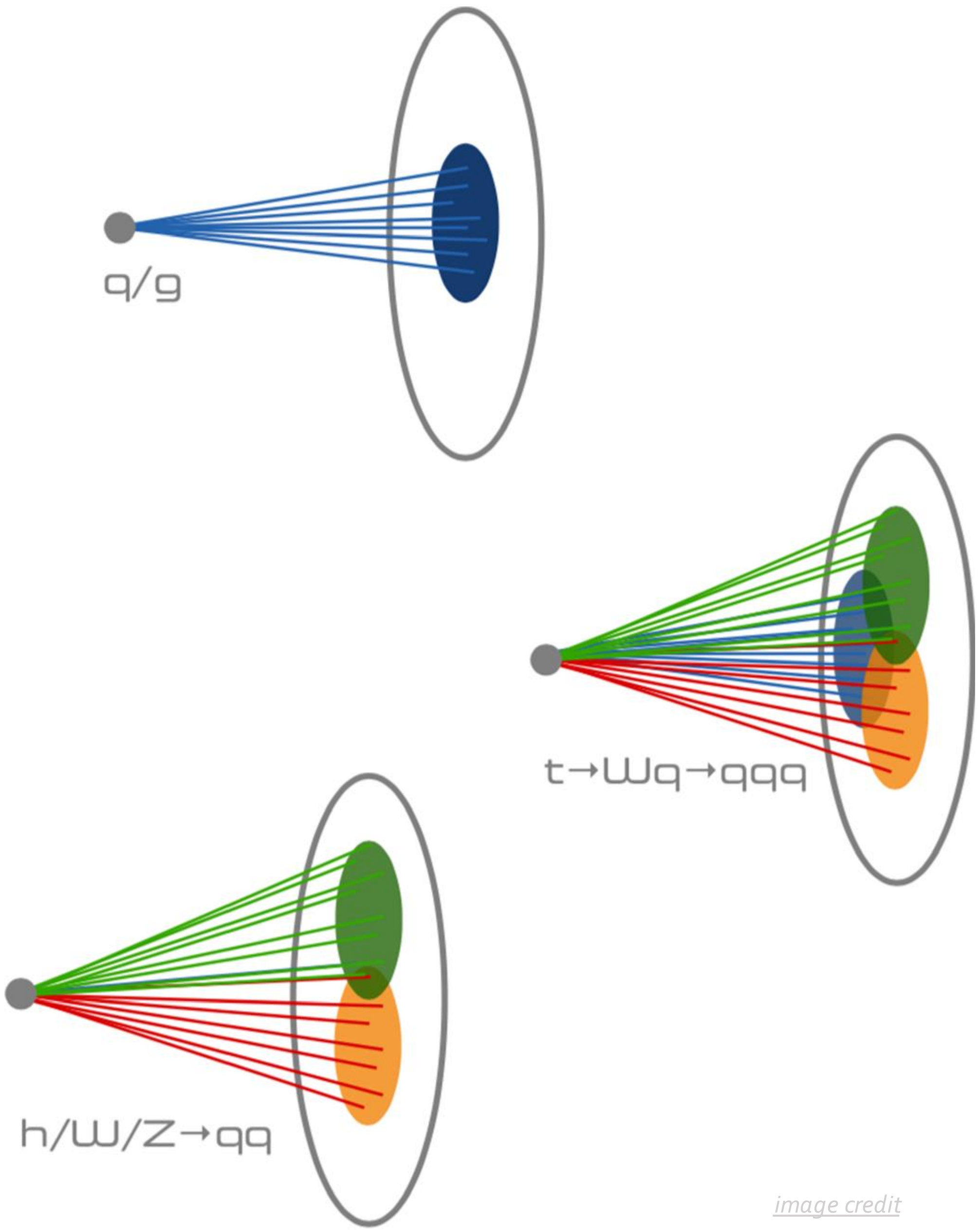


image credit

▪ W boson/top quark with large transverse momentum compared to its mass is referred to as *boosted* W boson/top quark.

Hadronic objects as jets

- **Quark/gluon, W->qq' and t->qq'b**, after showering and hadronization, will eventually become collimated sprays of particles called ***jets.***

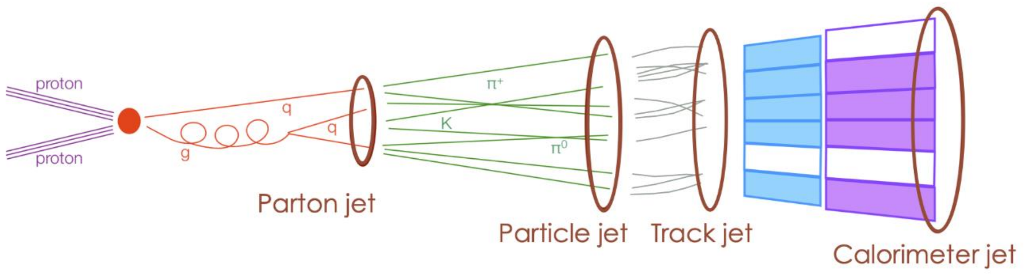
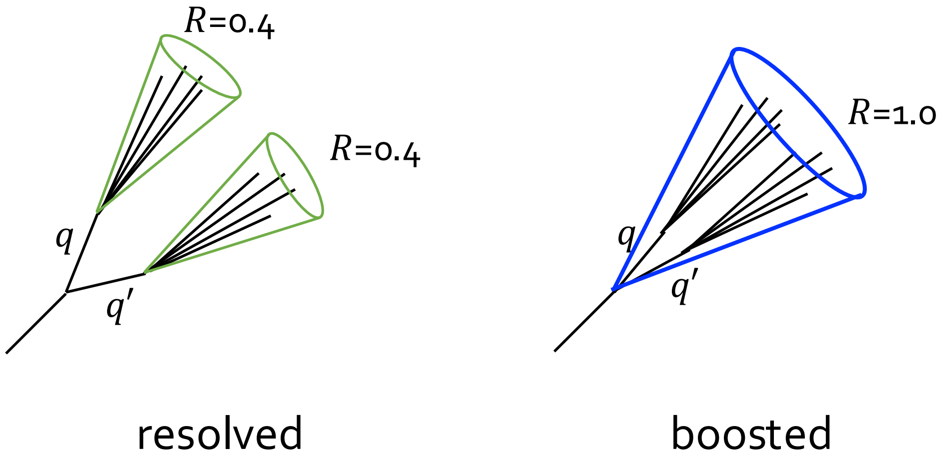


image credit

- The sprays of particles are clustered using **anti-k_t** algorithm with different jet radius parameter, R.

- Quark/gluon jets are **small-R (R=0.4)** jets.

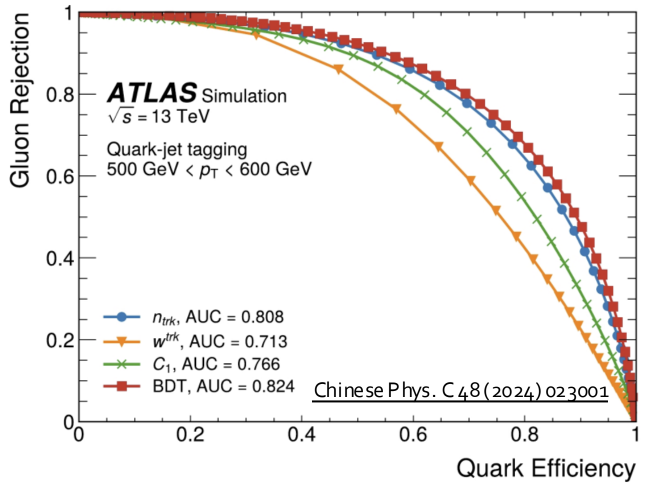
- Boosted W boson/top quark jets are **large-R (R=1.0)** jets.

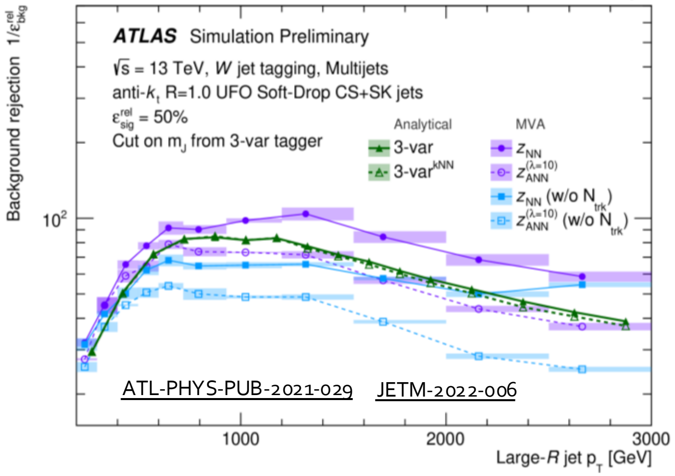


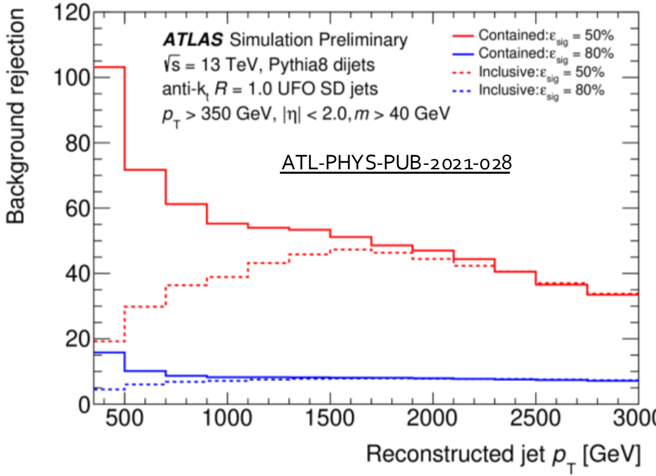
Jet tagging baselines

- Baseline taggers for quark/gluon, W and top jets are based on **high-level variables**.
  - Track multiplicity, jet track width, energy correlation function, N-subjettiness, k_t splitting scale, etc.

- **Baseline quark/gluon taggers:**
  - Single variable n_{trk} (number of tracks ghost-associated with the jet) q/g tagger.
  - BDT-based tagger: combines 3 jet substructure variables.
- **Baseline W-taggers:**
  - "3-var" cut-based tagger: cut on m_J (jet mass), D_2 (energy-correlation function ratio), n_{trk}.
  - DNN-based tagger (Z_{NN}): combines 10 jet substructure variables.
  - ANN tagger (Z_{ANN}): mass-decorrelated DNN tagger.
- **Baseline top-taggers:**
  - DNN-based tagger: combines 15 jet substructure variables.
    - "contained top tagger" and "inclusive top tagger"





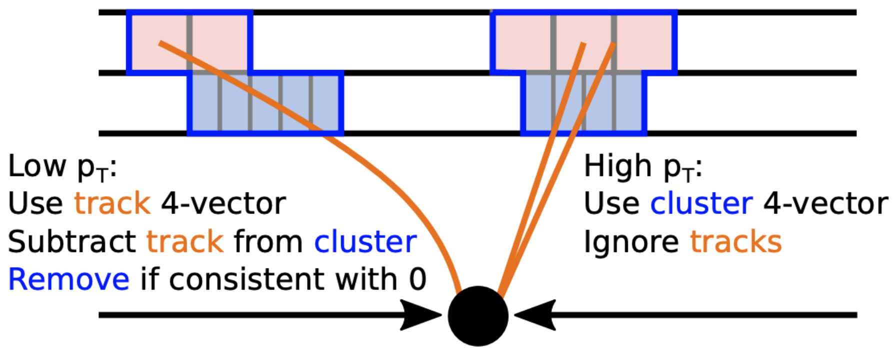


- The DNN-based W tagger is mass-correlated, it tends to tag jets with mass around W mass as W jet.
- ANN stands for adversarial neural network, a method to realize the mass-decorrelation of the DNN-based W tagger.

Jet tagging advances

- Recent jet tagging advances of ATLAS have focused on using **constituent-level variables** as inputs to taggers.
  - **Richer information** than manually defined high-level variables.
  - Allows taggers to **learn to their best capacity**.
- Jet constituents definitions in ATLAS:
  - Particle Flow Objects (PFOs)
    - Calorimeter + track 4-vector
    - Pileup mitigation using track matched to the primary vertex in jet reconstruction → good jet energy resolution at low p_{T}.
    - Used for **small-R** jet clustering.
  - Track-CaloClusters (TCCs)
    - Calorimeter energy, track angle.
    - Good performance at high p_{T}.
  - Unified Flow Objects (UFOs)
    - Combines desirable aspects of PFO and TCC reconstruction.
    - Optimal overall performance across the full kinematic range.
    - Used for **large-R** jet clustering.

Particle Flow Objects (PFOs)



Track-CaloClusters (TCCs)

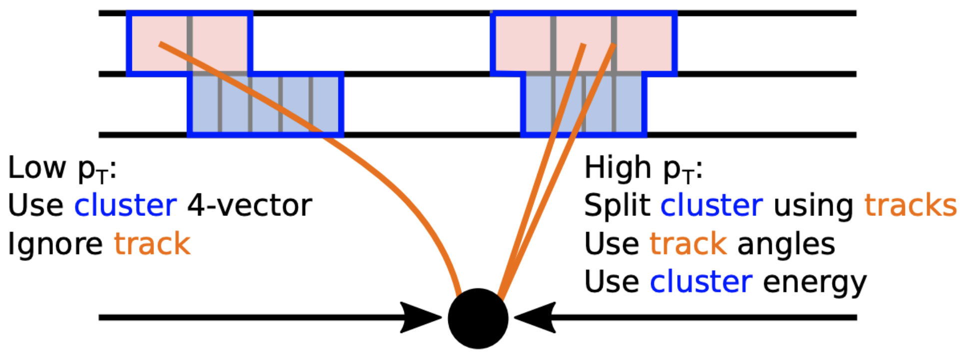


image credit

Machine learning algorithms for constituent-level jet taggers

Jets as vectors

- Jet constituents are sorted by decreasing p_{T}, then **flattened** into a vector.
- The vector is fed to a *simple DNN* with a few hidden layers.
- **No inductive bias.**
- Cannot handle the variable number of constituents naturally.
- Jets are **padded or truncated** to have a **fixed** number of constituents.

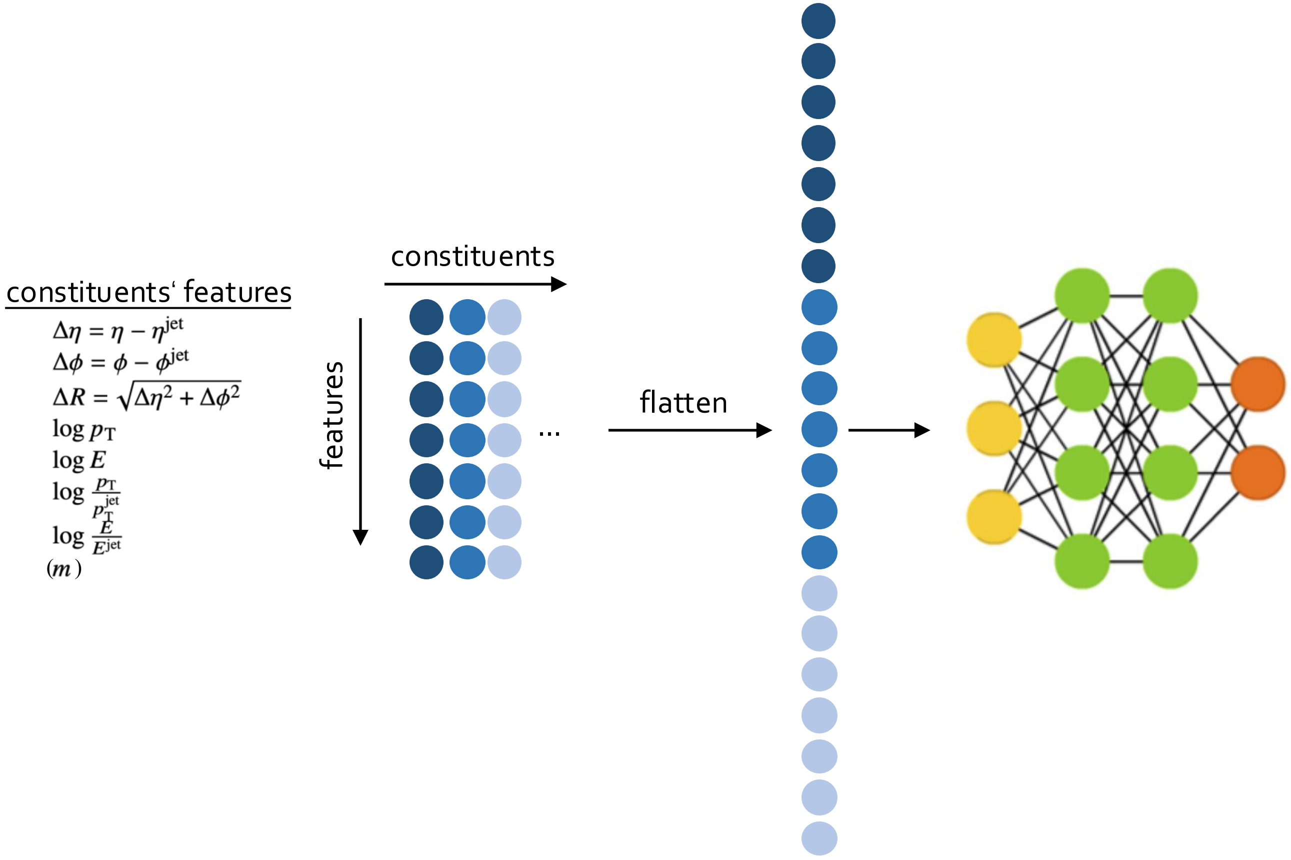
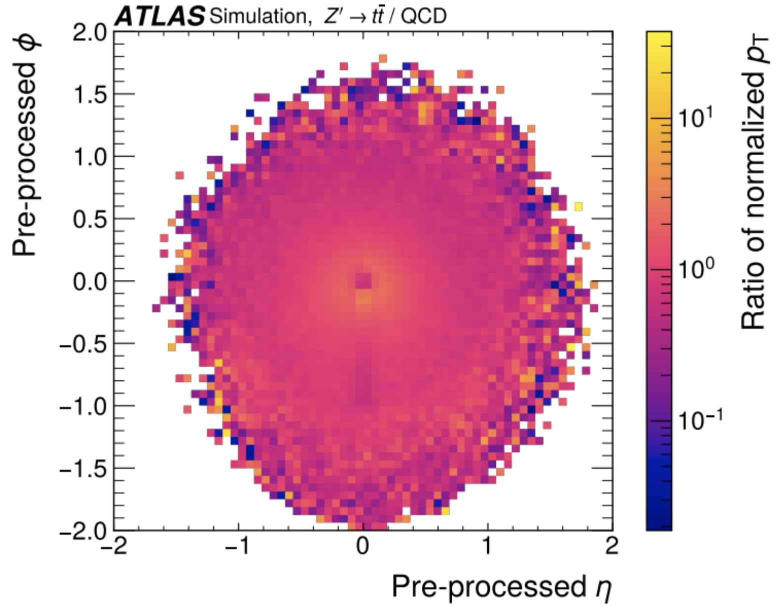


image credit

Jets as images

- Jets are converted into "**jet images**" by **binning** each constituent's η and ϕ coordinates.
- A large-scale CNN - ***ResNet 50*[1]** is used.
- Pixel values are the **sum of the raw p_T of the constituents within the pixel** (bin), **normalized** by the sum of the pixel values over the image, then **rescaled** by log(1 + 100x) to make the lower p_T patterns in the jet substructure visible.
- Handles **variable number of constituents**.
- Enforces **permutation invariance**.

JINST 19 (2024) P08018



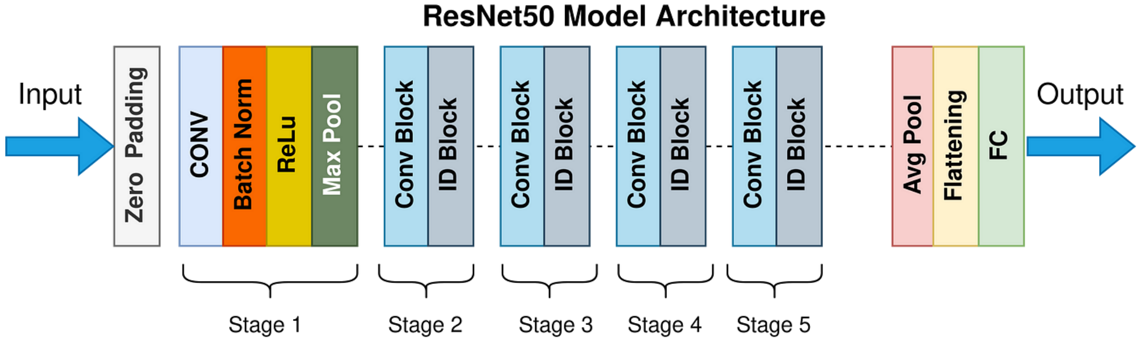


image credit

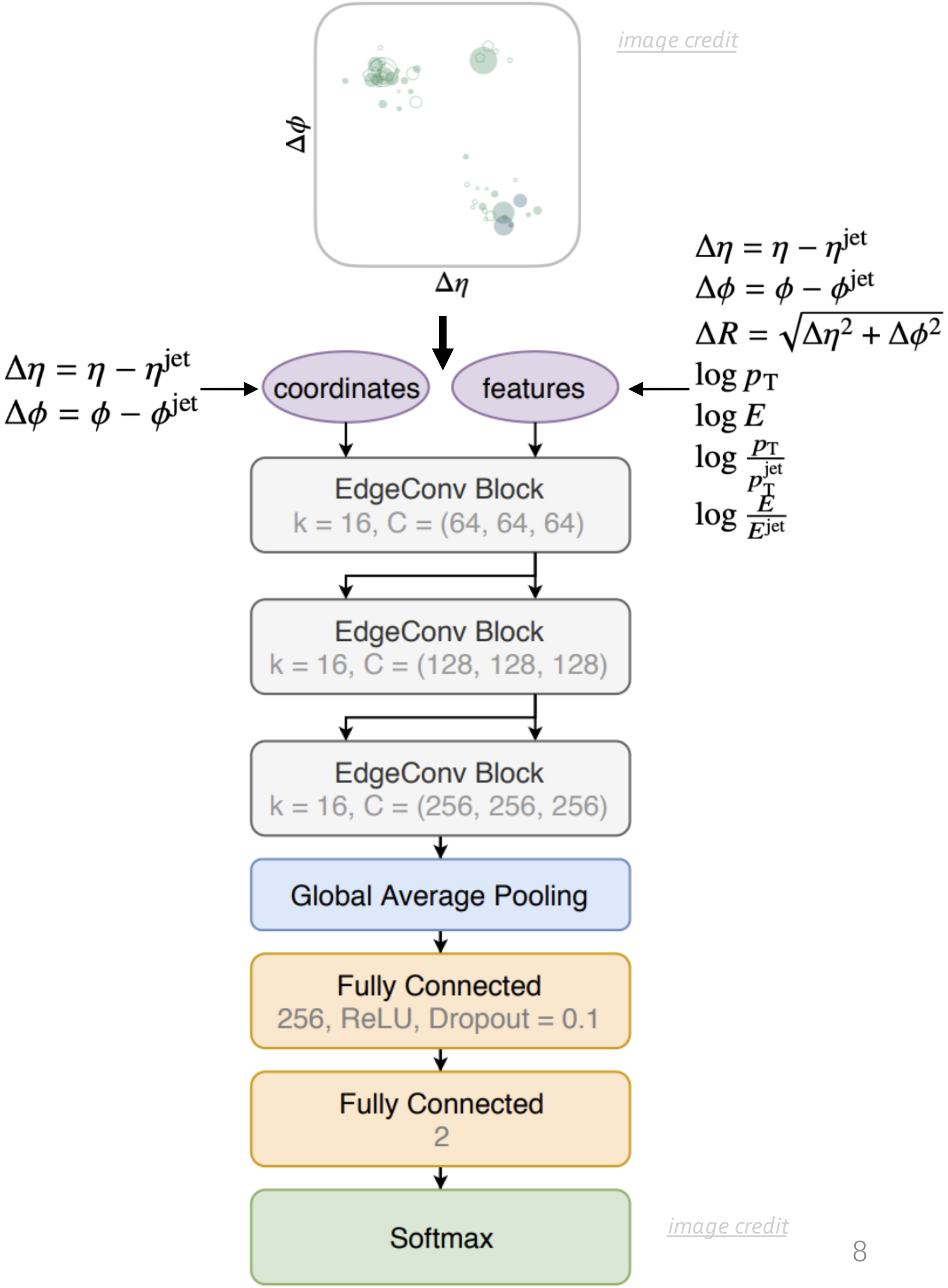
Jets as point clouds

- Jet represented as a **2D point cloud** on constituents' η-ϕ plane.

- ***ParticleNet***[2] is a GNN acting on the 2D point cloud, which dynamically construct and updates a graph composed of nodes (associated with constituents) and edges.

- Handles **variable number of constituents**.
- Enforces **permutation invariance**.

- EdgeConv operation acts on constituent pairs that are spatially close to each other, and can be stacked.
  - Learning both local and global structures, in a hierarchical way



Jets as Lund jet planes

- **Lund jet planes** are constructed by iteratively declustering the jet clustered using **Cambridge/Aachen (C/A)** algorithm.

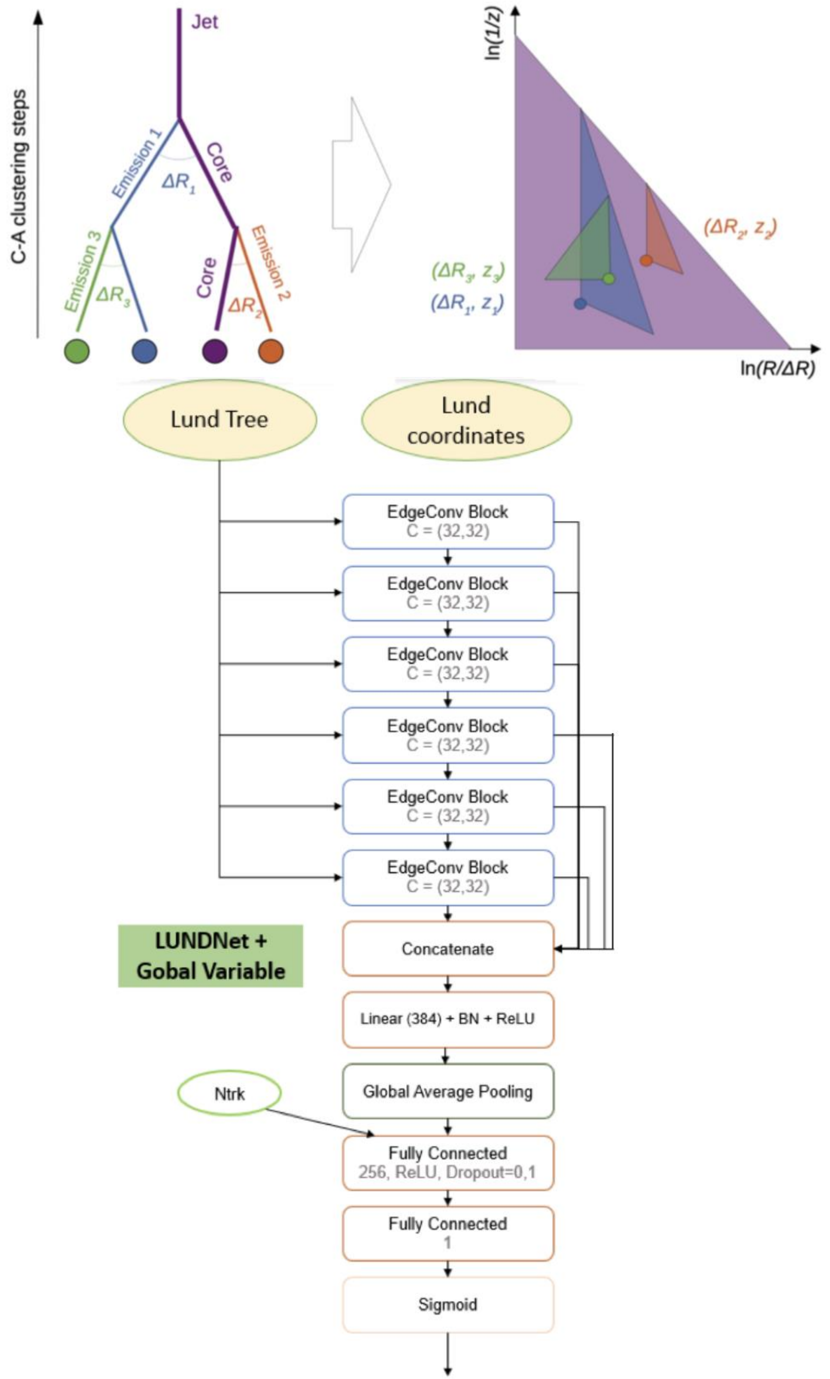
- ***LundNet[3]*** is a GNN acting on a graph representing the sequence of jet emissions, with structure similar to ParticleNet.

- Each node has 3 Lund kinematic variables: $\mathbf{ln(1/\Delta R)}$, $\boldsymbol{ln(k_t)}$, $\boldsymbol{ln(1/z)}$.

$$\Delta R_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta\phi_{ij}^2}, \qquad z = \frac{p_{\mathrm{T}}^j}{p_{\mathrm{T}}^i + p_{\mathrm{T}}^j}, \qquad k_t = p_{\mathrm{T}}^j \Delta R_{ij}$$

- $\boldsymbol{n_{\mathrm{trk}}}$ added as a global feature.

ATL-PHYS-PUB-2023-017



Jets as sets

- Jet represented as an **unordered set** of constituents.
- *Energy Flow Network (EFN) / Particle Flow Network (PFN)*[4] applies the DeepSet structure.
  - EFN uses **IRC-safe** constituent variables only, while PFN allows more input variables.
- *ParticleTransformer (ParT)*[5] is a Transformer designed with physics inspirations.
  - Takes advantage of **attention mechanism**.
- *Dynamically Enhanced Particle Transformer (DeParT)* is an expansion upon ParT.
- **Handles variable number of constituents and enforces permutation invariance.**

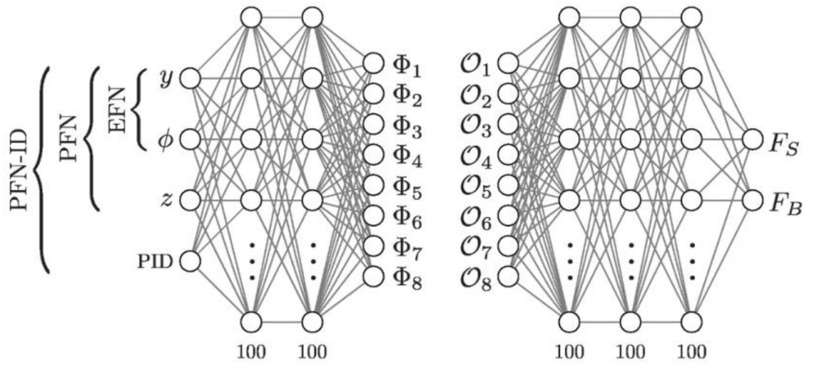
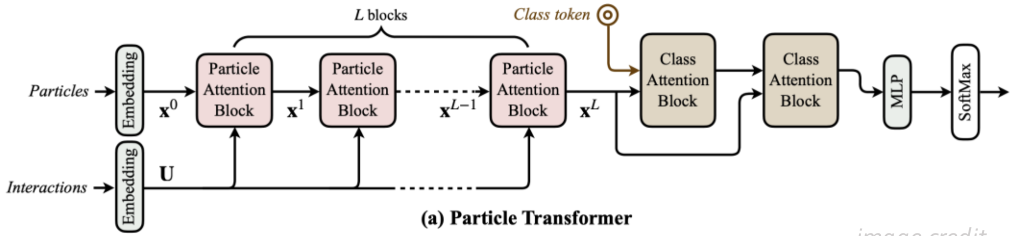


image credit

IRC-safe features

$$\Delta\eta = \eta - \eta^{\text{jet}}$$
$$\Delta\phi = \phi - \phi^{\text{jet}}$$
$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$
$$\frac{p_{\text{T}}}{p_{\text{T}}^{\text{jet}}}$$



(a) Particle Transformer

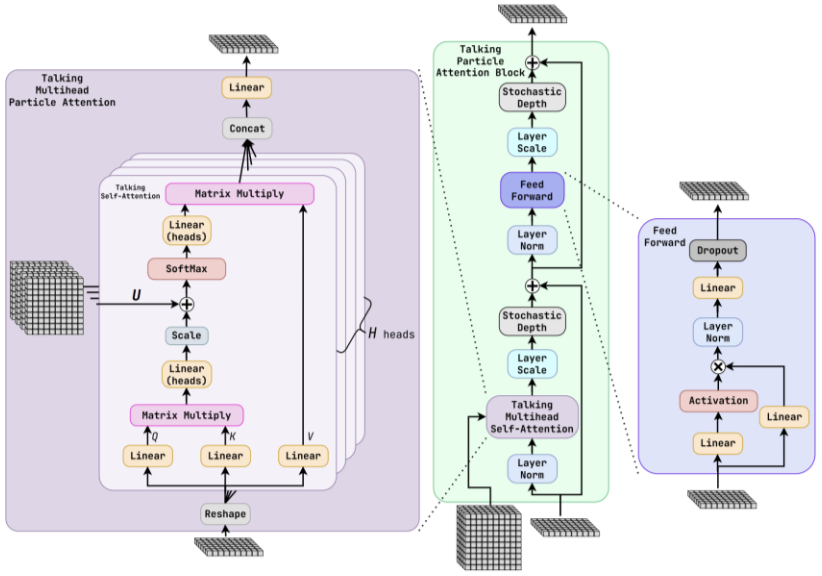
image credit

constituents' features

$$\Delta\eta = \eta - \eta^{\text{jet}}$$
$$\Delta\phi = \phi - \phi^{\text{jet}}$$
$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$
$$\log p_{\text{T}}$$
$$\log E$$
$$\log \frac{p_{\text{T}}}{p_{\text{T}}^{\text{jet}}}$$
$$\log \frac{E}{E^{\text{jet}}}$$
$$(m)$$

constituents' interaction features

$$\log \Delta^{ab} = \log \sqrt{(\eta^a - \eta^b)^2 + (\phi^a - \phi^b)^2}$$
$$\log k_{\text{T}}^{ab} = \log\left(\min\left(p_{\text{T}}^a, p_{\text{T}}^b\right)\Delta^{ab}\right)$$
$$z^{ab} = \min\left(p_{\text{T}}^a, p_{\text{T}}^b\right)/\left(p_{\text{T}}^a + p_{\text{T}}^b\right)$$
$$\log m^{2,ab} = \log\left(p^{\mu,a} + p^{\mu,b}\right)^2$$



ATL-PHYS-PUB-2023-032

Performance of constituent-based jet taggers

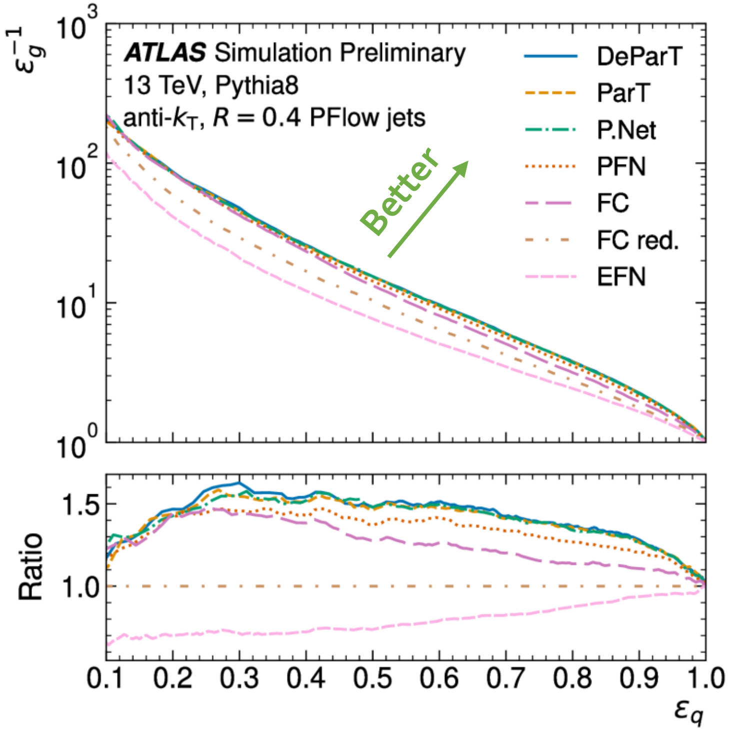
Quark/gluon tagging

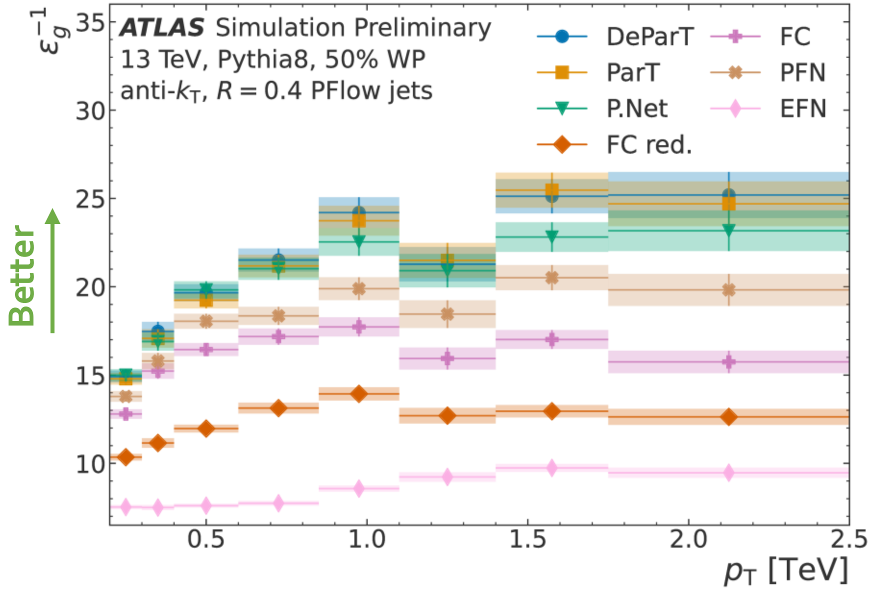
- Most constituent-level taggers outperforms baseline high-level tagger.
  - Baseline ***FC red.*** : a DNN with 5 high-level variables as input.

- Transformer-based taggers (***ParT, DeParT***) provides the best performance.

- ***EFN*** provides worse performance than baseline high-level tagger, as it's constrained to IRC-safe inputs only.

- Taggers perform better at higher jet p_{T}.
  - similar trends for all taggers.





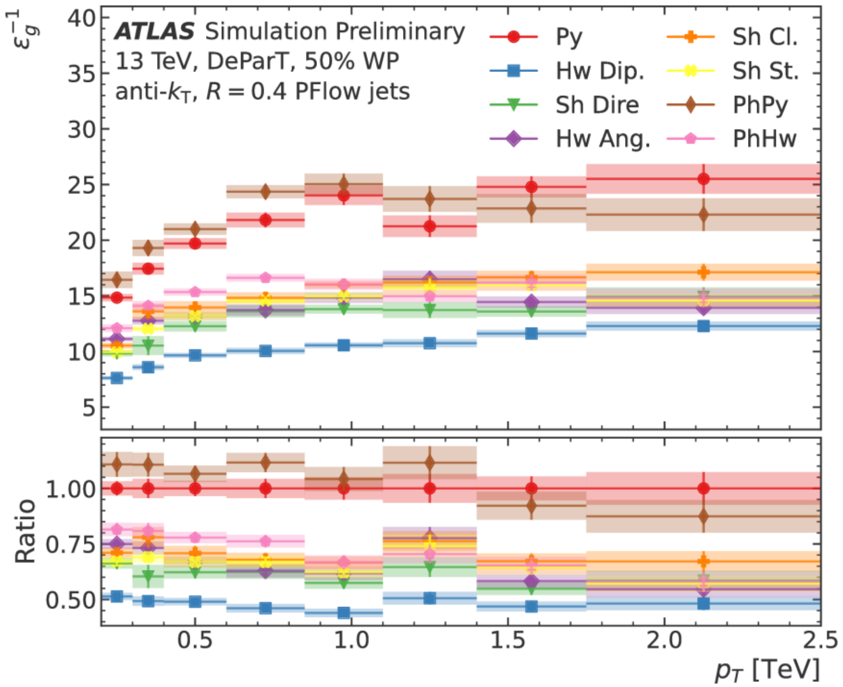
ATL-PHYS-PUB-2023-032

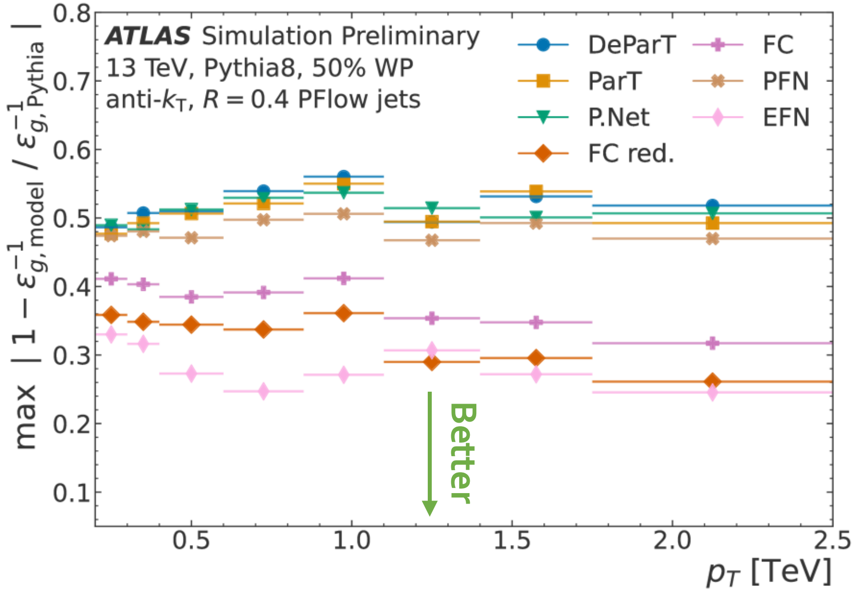
Quark/gluon tagging

- Tagger performance **varies significantly** as the MC modeling of Parton-Shower and hadronization changes.

- The differences in gluon rejection between the taggers, excluding ***EFN***, are **less pronounced** than the gluon rejection variations of a given tagger.

- Better performance, higher modeling dependence.
    - ***DeParT*** is the most sensitive one.
    - ***EFN*** is the least sensitive one, could be the result of its IRC-safe characteristic, which reduces sensitivity to non-perturbative effects

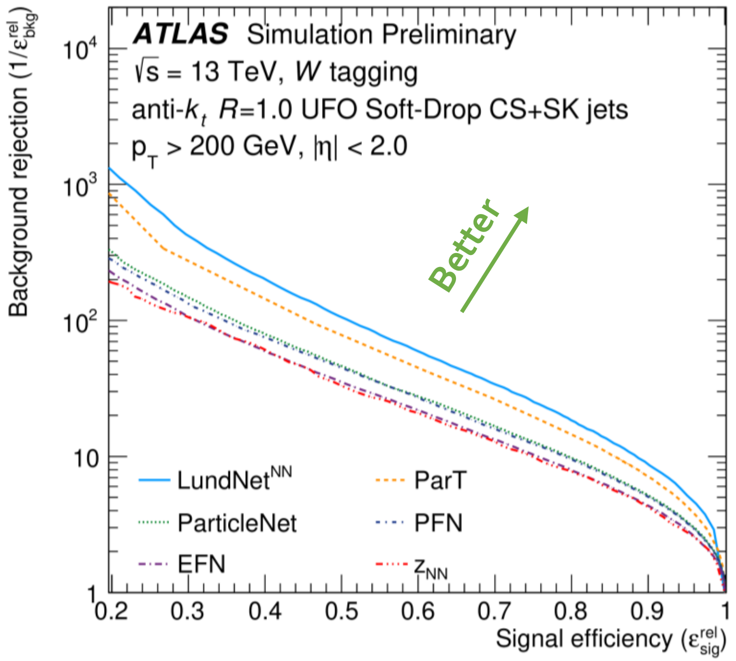


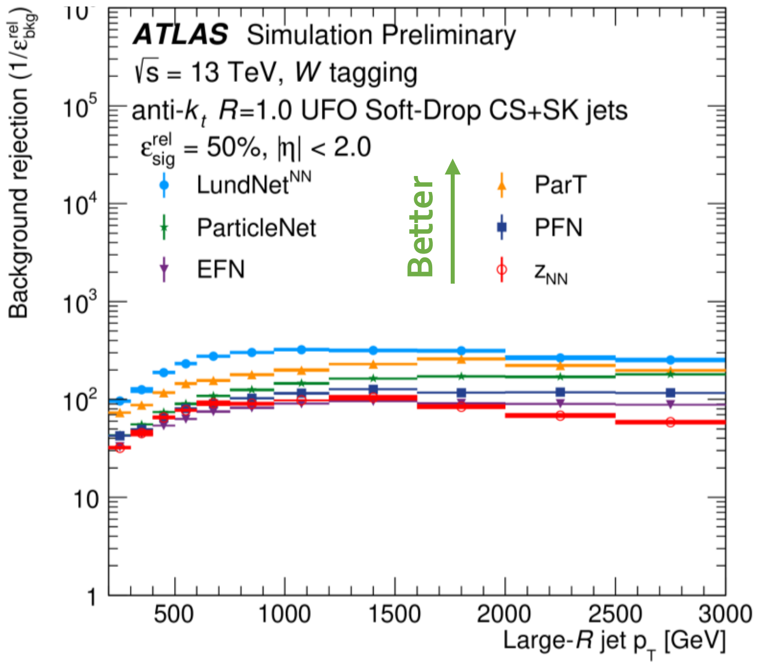


ATL-PHYS-PUB-2023-032

W tagging

- Constituent-level taggers outperforms baseline high-level tagger Z_{NN}.
- ***LundNet*** becomes the most performant one, even better than the sophisticated ***ParT.***
- Similar to quark/gluon tagging, ***EFN*** became the worst constituent-level tagger, but still slightly outperforms baseline high-level tagger.
- Taggers perform better at higher jet p_{T}.
    - similar trends for all taggers.



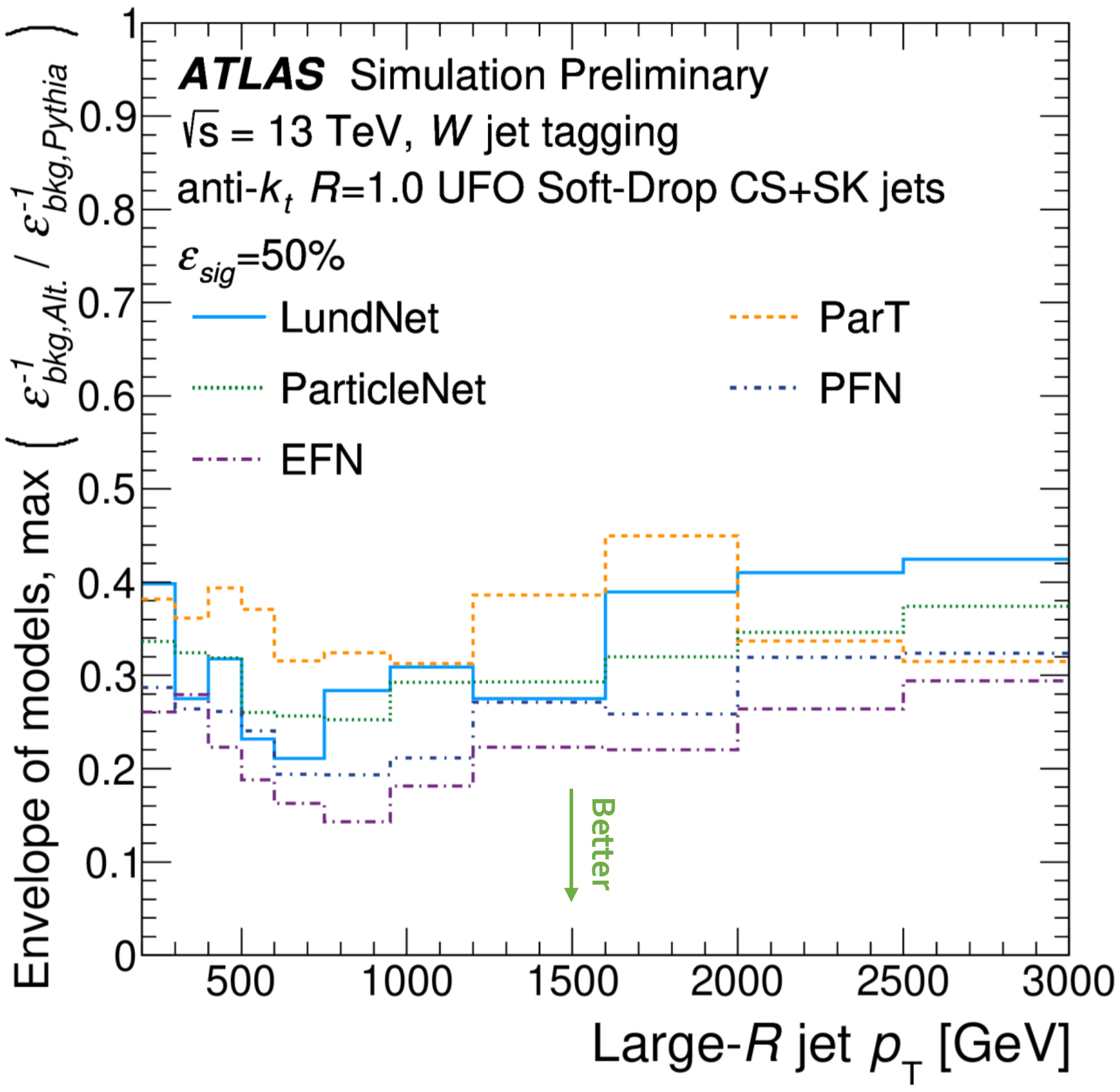


ATL-PHYS-PUB-2023-017 ATL-PHYS-PUB- 2023-020 JETM-2023-003

W tagging

- Tagger performance also tested on MC samples with different modeling setup.

- Sophisticated ML architectures are more likely to show higher model dependence.

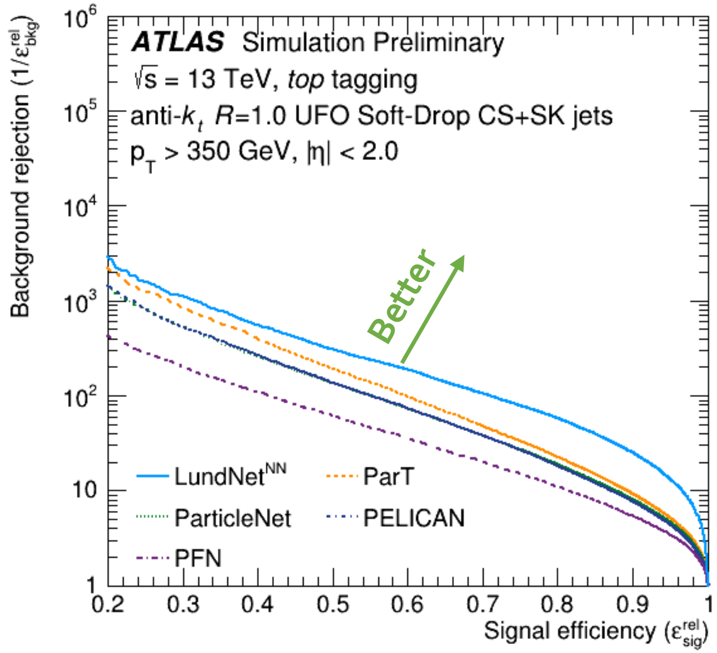


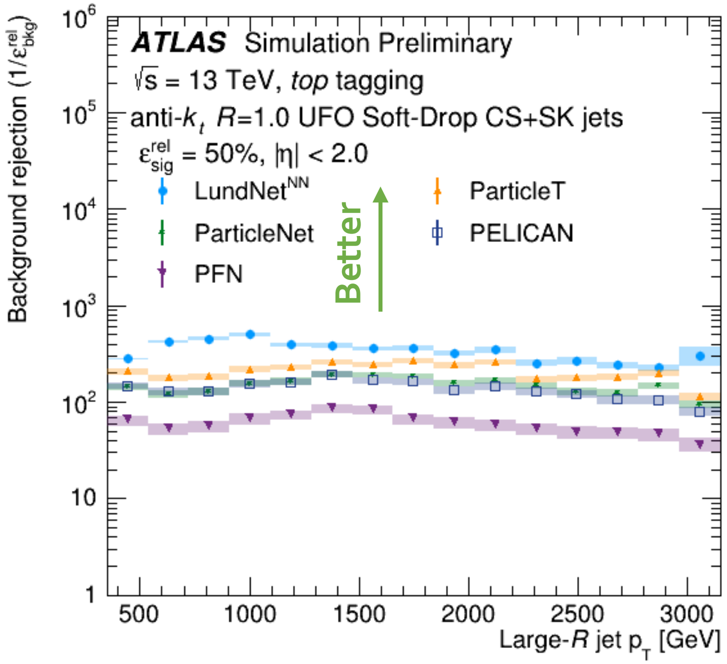
ATL-PHYS-PUB-2023-017 ATL-PHYS-PUB-2023-020 JETM-2023-003

Top tagging

- Similar performance ranking of constituent-level taggers as the ones seen in quark/gluon and W tagging cases.
  - ***LundNet*** becomes the top performer again.
  - **Physics-inspired** algorithm design has the potential to significantly enhance the performance.

- Relatively **flat tagger performance** across the whole supporting p_{T} range.
  - Different from what's seen in quark/gluon and W tagging.

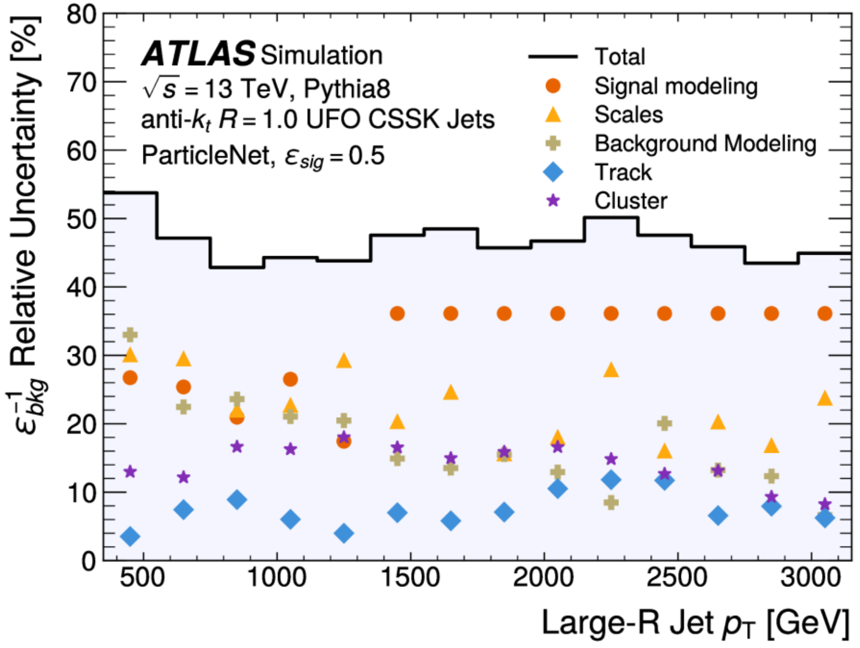


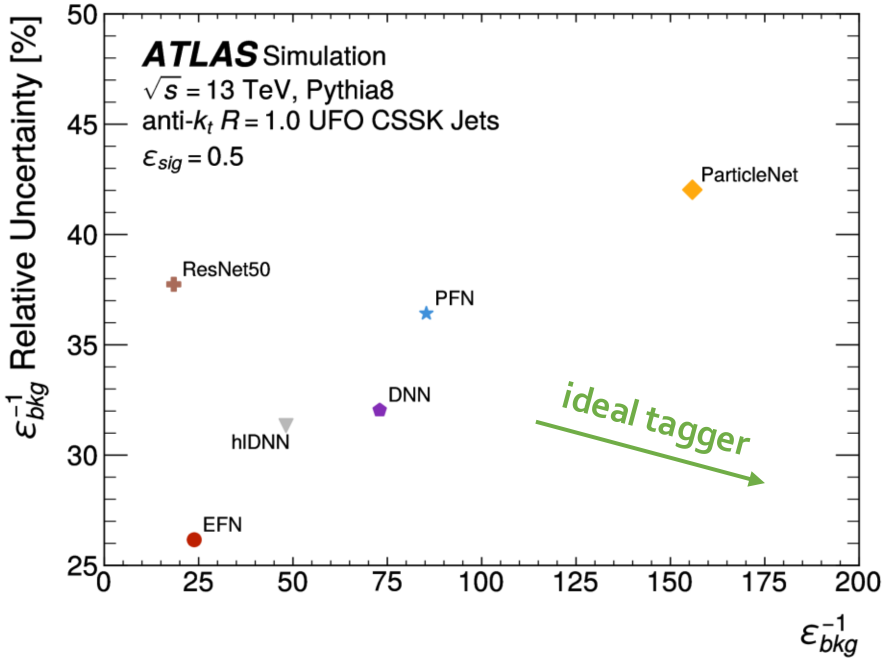


ATL-PHYS-PUB-2022-039 JETM-2024-02

Top tagging

- Comprehensive study evaluating **approximate** bottom-up experimental uncertainties and theory uncertainties on top-taggers.
- **Positive correlation** between performance and uncertainty.
  - Better taggers have higher uncertainties by almost factor 2.
  - ResNet50 is an outlier, with high relative uncertainty and poor performance.
- Further study needed to addressing the uncertainty issue, pushing ideal taggers from dream to life.





JINST 19 (2024) P08018

Summary

- Presented recent advances in **quark/gluon, boosted W boson and top quark tagging** with **constituent-based** ML algorithms in ATLAS.

- State-of-the-art architectures and access to low-level information lead to **significant improvements** in performance, comparing the traditional methods using high-level variable.

- With greater performance, comes **greater uncertainty and MC modeling dependence**.

- Understanding and addressing the uncertainty and model dependence is **crucial** to generalize the performance of these taggers.

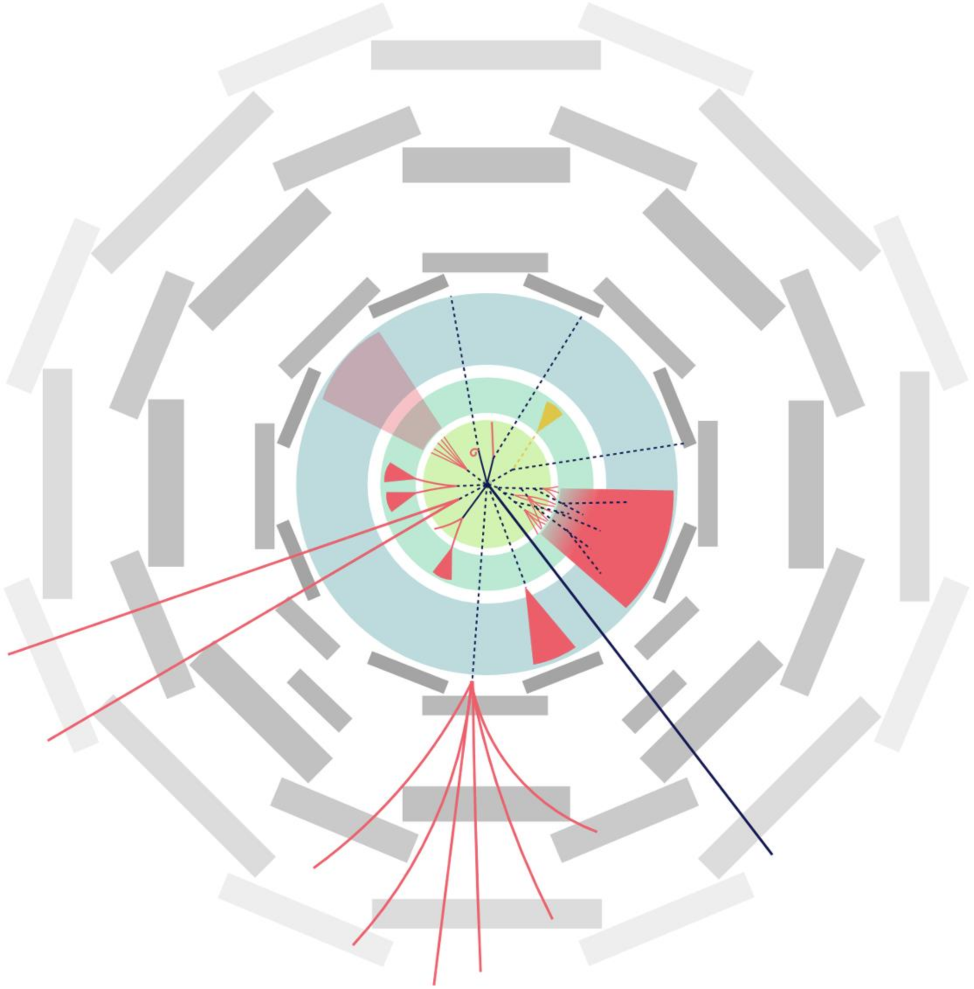


image credit

Backup

References of the ML algorithms

[1] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[2] H. Qu and L. Gouskos, ParticleNet: Jet Tagging via Particle Clouds, Phys. Rev. D 101 (2020) 056019, arXiv: 1902.08570 [hep-ph]

[3] F. A. Dreyer and H. Qu, Jet tagging in the Lund plane with graph networks, JHEP 03 (2021) 052, arXiv: 2012.08526 [hep-ph]

[4] P. T. Komiske, E. M. Metodiev and J. Thaler, Energy Flow Networks: Deep Sets for Particle Jets, JHEP 01 (2019) 121, arXiv: 1810.05165 [hep-ph]

[5] H. Qu, C. Li and S. Qian, Proceedings of the 39th International Conference on Machine Learning, PMLR 162:18281-18292, 2022.

Quark/gluon tagging

Model	AUC	$\varepsilon_g^{-1}@\varepsilon_q = 0.5$	# Params [10^6]	Inference Time [ms]	GPU Memory [MB]
DeParT	0.8489	15.4242	2.62	266.51	1684
ParT	0.8479	15.2457	2.62	233.84	1730
ParticleNet	0.8476	15.4402	2.59	768.74	5410
PFN	0.8406	14.2387	2.64	136.93	393
FC	0.8280	13.5199	2.63	65.53	76
FC reduced	0.8038	10.3639	2.63	84.84	47
EFN	0.7761	7.7222	2.60	101.53	337

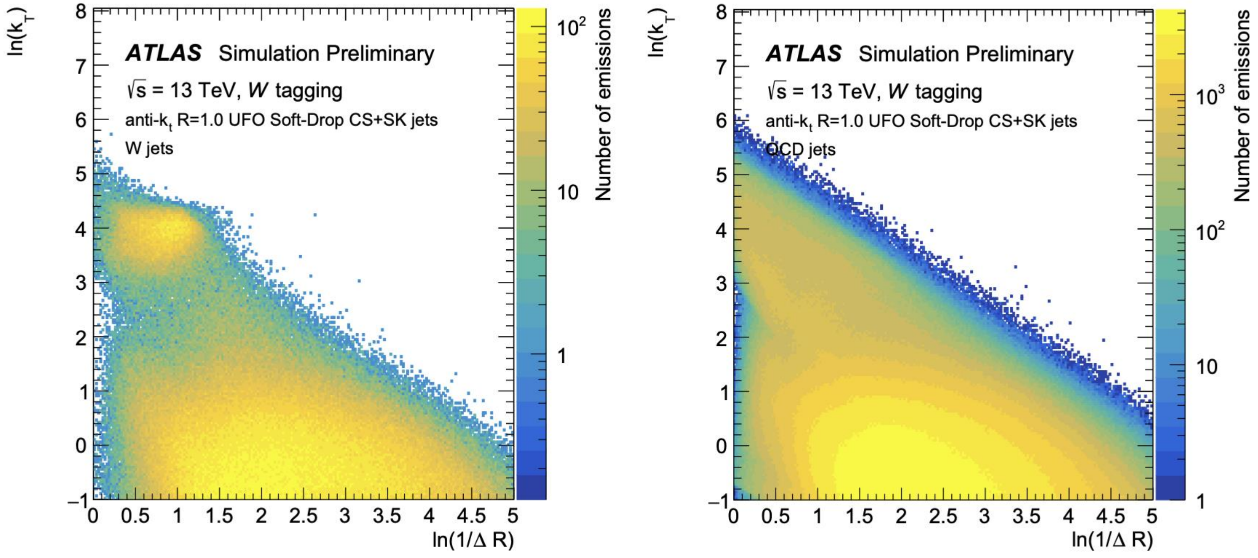
Results and technical information of the different taggers. The total number of parameters is given in millions. The time and memory are measured on two NVIDIA Tesla V100 GPU 32GB. The inference time corresponds to one batch of size 512. Memory is the maximum GPU memory usage during inference of batch with size 512

W tagging

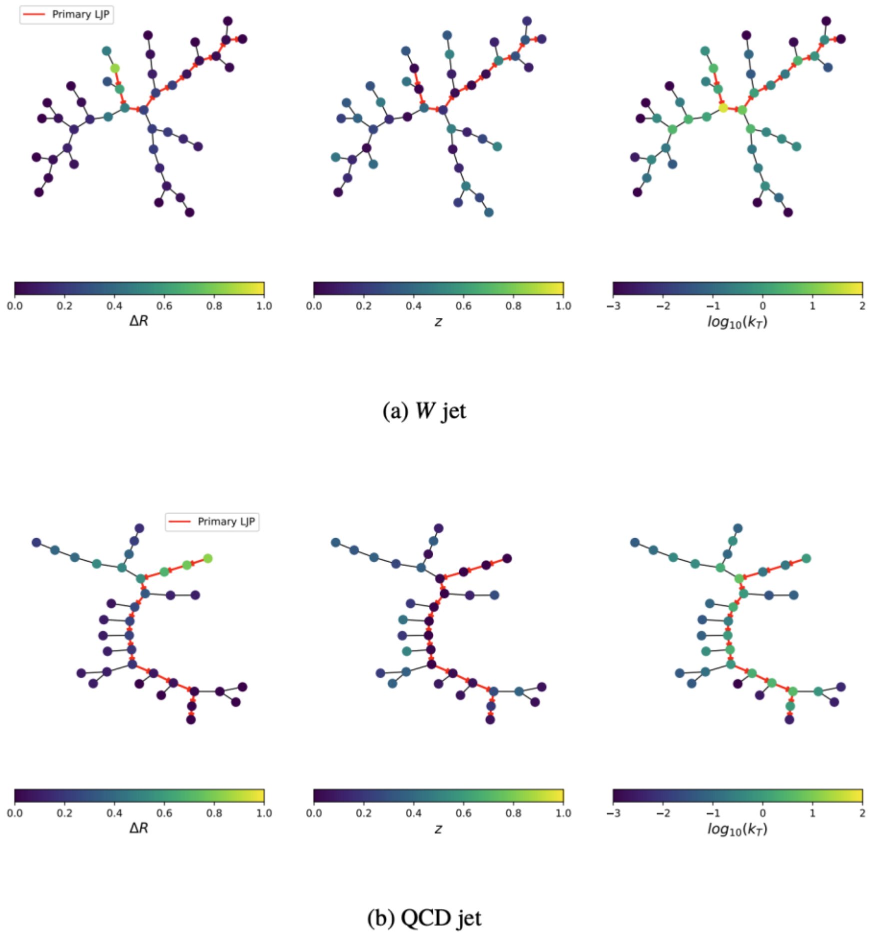
Model	AUC	ACC	ε_{bkg}^{-1} @ ε_{sig} = 0.5	ε_{bkg}^{-1} @ ε_{sig} = 0.8	# Params	Inference Time
EFN	0.920	0.835	35.1	7.95	56.73k	0.065 ms
PFN	0.931	0.853	44.7	9.50	57.13k	0.11 ms
ParticleNet	0.933	0.826	46.2	9.76	366.16k	0.36 ms
ParticleTransformer	0.951	0.880	77.9	14.6	2.14M	0.28 ms

The performance metrics of the constituent-based taggers evaluated on the testing set, along with the number of trainable parameters and the inference time. Inference time is measured using an Nvidia Tesla V100-SXM2-32GB GPU, taking the average time need by processing one jet.

W tagging

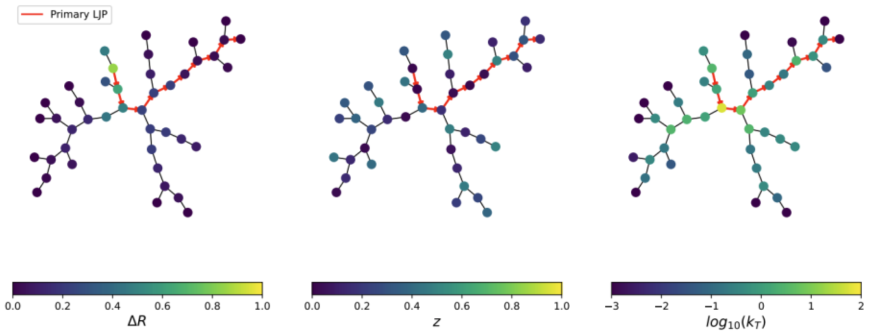


Lund jet plane emission density for hadronically decaying boosted W bosons and QCD background jets

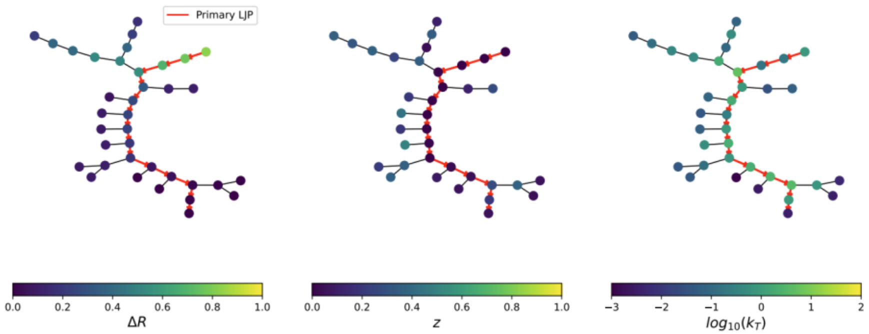


Graphs showing the full jet de-clustering and the primary Lund jet plane is shown in red. Different panels are shown with the color axis showing the values of the Lund variables.

W tagging

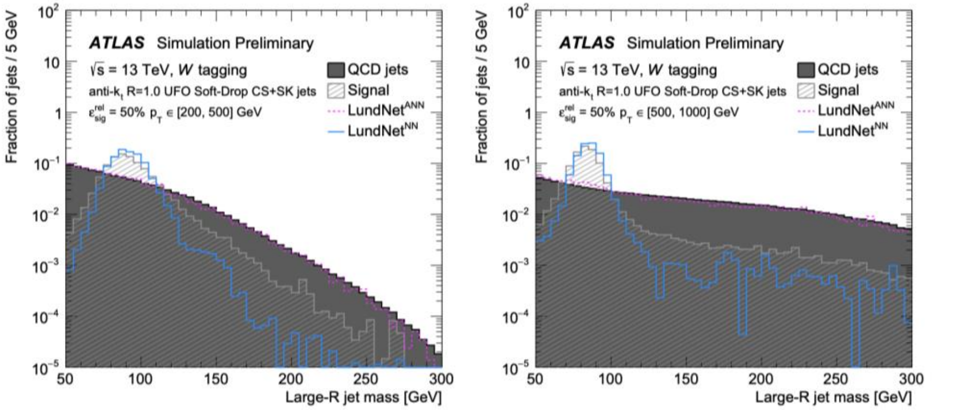


(a) W jet



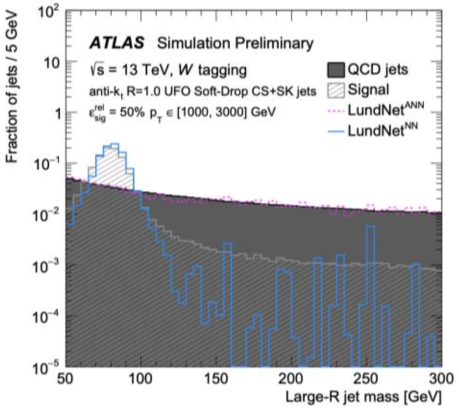
(b) QCD jet

Graphs showing the full jet de-clustering and the primary Lund jet plane is shown in red. Different panels are shown with the color axis showing the values of the Lund variables.



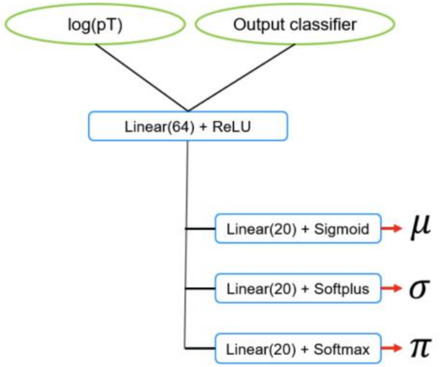
(a) $p_T \in [200, 500]$ GeV

(b) $p_T \in [500, 1000]$ GeV



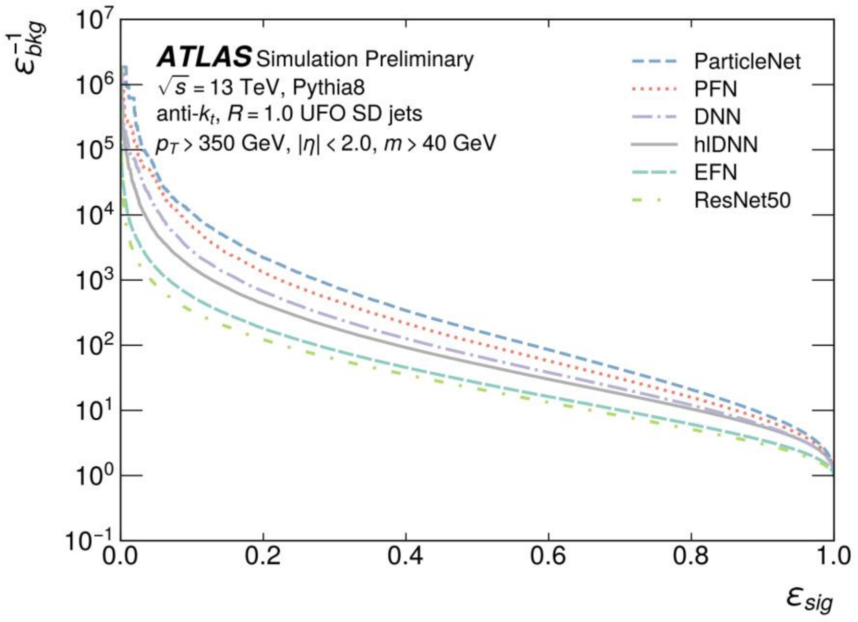
(c) $p_T \in [1000, 3000]$ GeV

Normalized jet mass distribution for background jets before tagging (black line and dark gray shading), compared with the same distributions after tagging to achieve 50% signal efficiency.



Adversary network architecture used for training mass-decorrelated LundNet, LundNet^{ANN}.

Top tagging



Background rejection versus signal efficiency curves for constituent-based top taggers, along with curve for high-level tagger (hlDNN) for comparison.

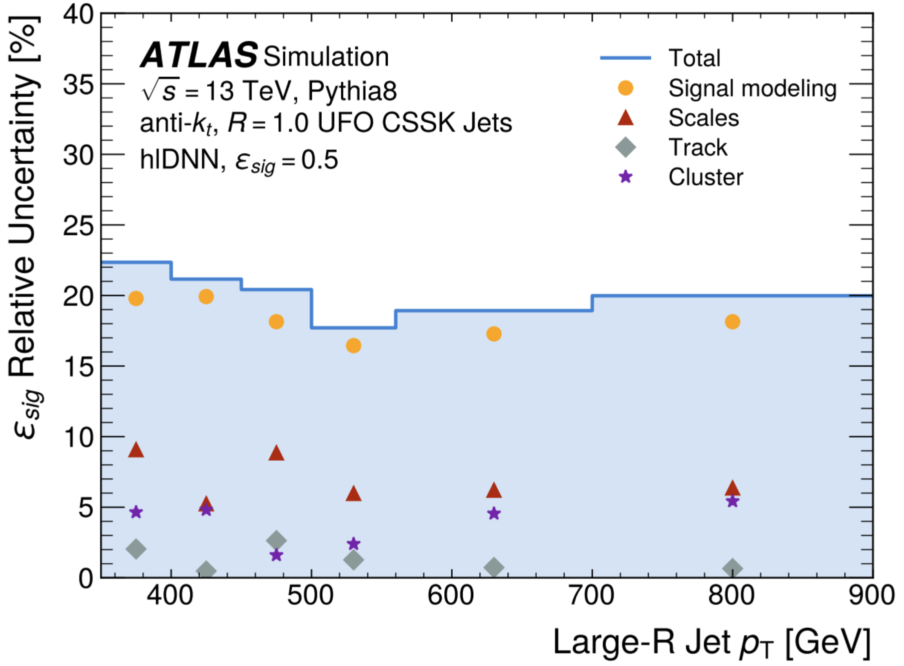
Tagger	Number of parameters	Inference time
hlDNN	133,381	3 ms
DNN	876,641	3 ms
EFN	959,251	4 ms
PFN	754,501	3 ms
ResNet 50	1,499,585	20 ms
ParticleNet	764,887	143 ms

The number of trainable parameters and inference time for each tagger considered in this study. Inference time is defined as the amount of time required to run inference for a batch of 256 jets on an NVIDIA Tesla V100 GPU.

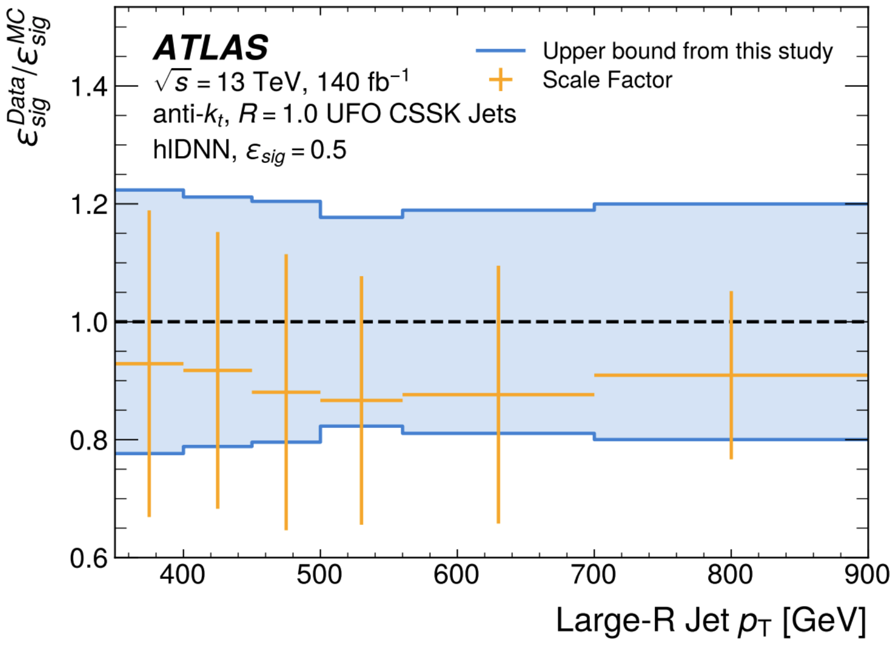
Tagger	AUC	ACC	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.5$	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.8$
ResNet 50	0.872 ± 0.006	0.787 ± 0.006	18.4 ± 1.1	4.63 ± 0.2
EFN	0.894 ± 0.001	0.810 ± 0.001	23.8 ± 0.5	5.74 ± 0.07
hlDNN	0.9374 ± 0.0001	0.8628 ± 0.0002	47.2 ± 0.4	10.36 ± 0.03
DNN	0.9447 ± 0.0004	0.8715 ± 0.0008	73.0 ± 1.3	12.5 ± 0.1
PFN	0.9502 ± 0.0004	0.878 ± 0.001	92.7 ± 1.8	14.6 ± 0.2
ParticleNet	0.9614 ± 0.0005	0.895 ± 0.001	155.8 ± 3.8	20.6 ± 0.4

The performance of each top quark tagger is measured with several metrics evaluated on the testing set.

Top tagging



The uncertainty in the signal efficiency for the hlDNN tagger



A comparison between the total uncertainty derived and the scale factor and its uncertainty for the hlDNN tagger.