Improving ATLAS Hadronic Object Performance with ML/AI Algorithms

DIS2023

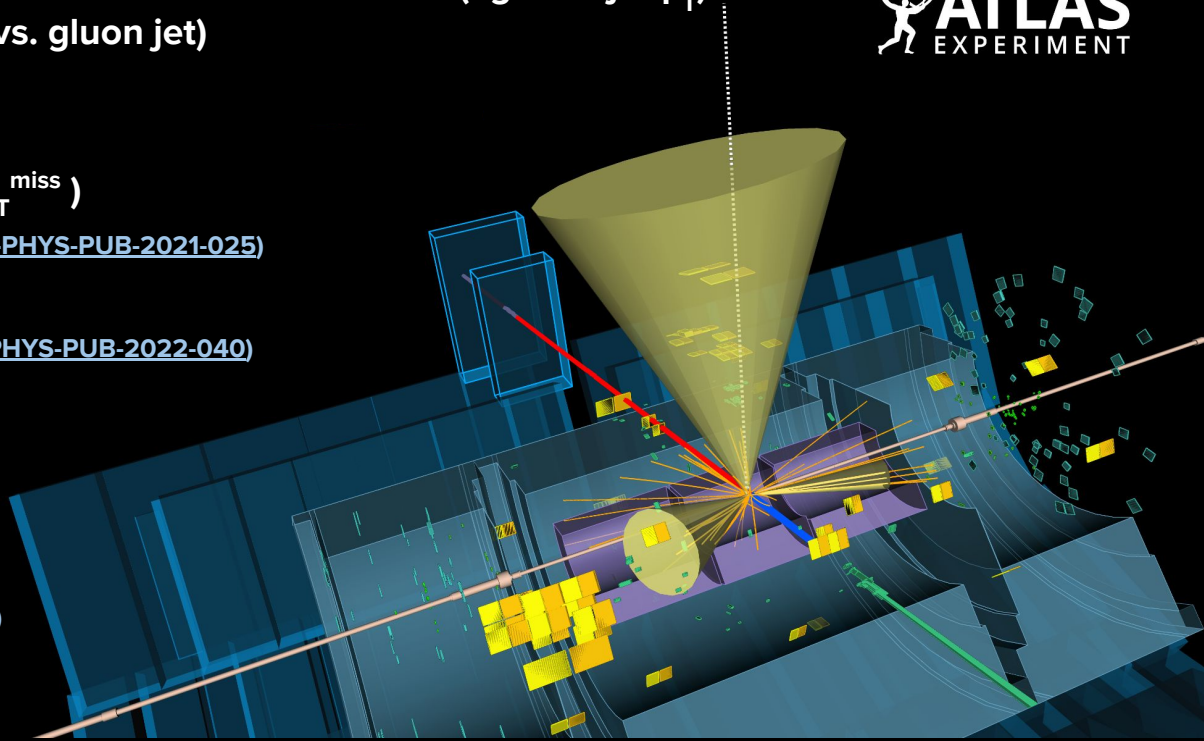
Ben Hodkinson
University of Cambridge

On behalf of the ATLAS Collaboration

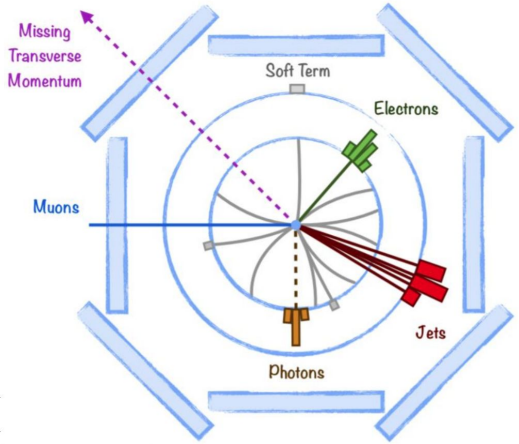
March 29, 2023

Introduction

- Jets and p_T^{miss} are complex objects
  - ➢ Promising setting for ML applications
- ML applications:
  - ➢ Regress truth-level quantities from detector-level information (eg. true jet p_T)
  - ➢ Classify type of object (eg. top jet vs. gluon jet)
- Highlights covered here:
  - ➢ Missing transverse momentum (p_T^{miss})
    - → Regressing truth-level p_T^{miss} (ATL-PHYS-PUB-2021-025)
  - ➢ Pion reconstruction
    - → Classification & calibration (ATL-PHYS-PUB-2022-040)
  - ➢ Boosted jet taggers:
    - → Top jets (ATL-PHYS-PUB-2021-028, ATL-PHYS-PUB-2022-039)
    - → W/Z jets (ATL-PHYS-PUB-2021-029, JETM-2022-006)
  - ➢ Jet calibration
    - → (See Naseem Bouchhar's talk)



1) Missing transverse momentum (p_T^{miss}) performance



p_T^{miss} in ATLAS

Jet selection specified by working point (Loose/Tight/Tighter/Tenacious)

User-defined object selection

$$\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\text{selected electrons}} \mathbf{p}_{\mathrm{T}}^{e} - \sum_{\text{accepted photons}} \mathbf{p}_{\mathrm{T}}^{\gamma} - \sum_{\text{accepted } \tau\text{-leptons}} \mathbf{p}_{\mathrm{T}}^{\tau_{\mathrm{had}}} - \sum_{\text{selected muons}} \mathbf{p}_{\mathrm{T}}^{\mu} - \sum_{\text{accepted jets}} \mathbf{p}_{\mathrm{T}}^{\mathrm{jet}} - \sum_{\text{unused tracks}} \mathbf{p}_{\mathrm{T}}^{\mathrm{track}}$$

$$\underbrace{\underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},e}}_{\text{selected electrons}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\gamma}}_{\text{accepted photons}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\tau_{\mathrm{had}}}}_{\text{accepted } \tau\text{-leptons}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss},\mu}}_{\text{selected muons}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,jet}}}_{\text{accepted jets}}}_{\text{hard term}} + \underbrace{\mathbf{E}_{\mathrm{T}}^{\mathrm{miss,soft}}}_{\text{unused tracks}}$$

soft term

(+ internal overlap removal)

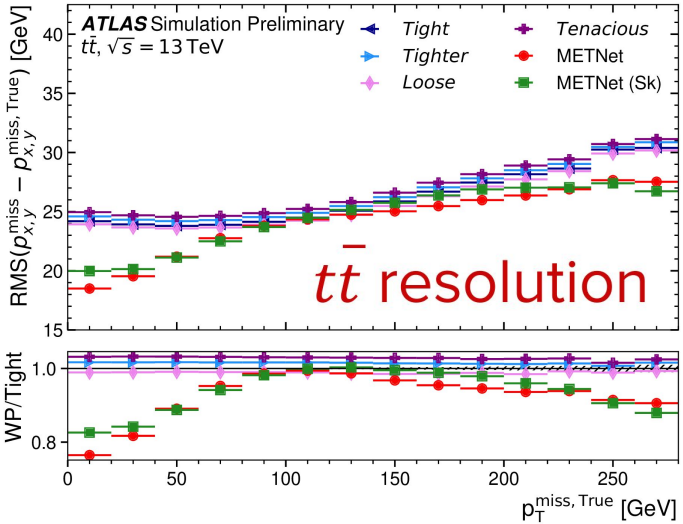
Optimal working point for a given event depends on the topology and pile-up conditions

- Tighter working points reject more pile-up but risk rejecting hard-scatter jets
- **METNet** = Neural network to pick and combine working points into a single $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$ estimate
  - Regresses **truth-level** $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$
  - Trained on **ttbar** and **di-boson** MC events.

ATL-PHYS-PUB-2021-025

METNet performance

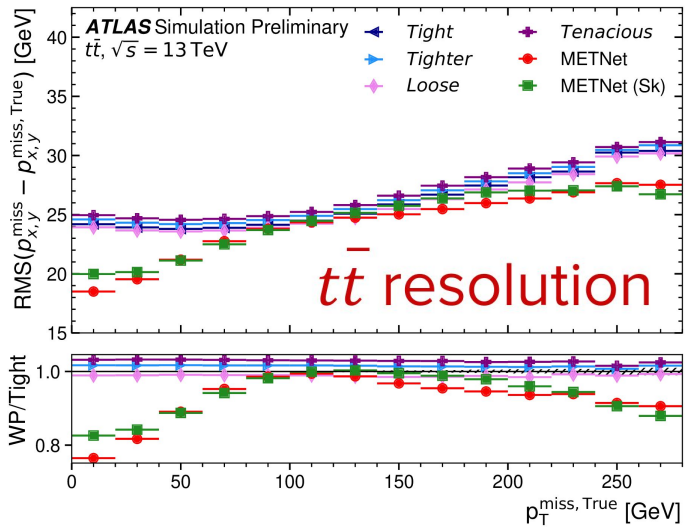
- Improved resolution compared to individual working points.

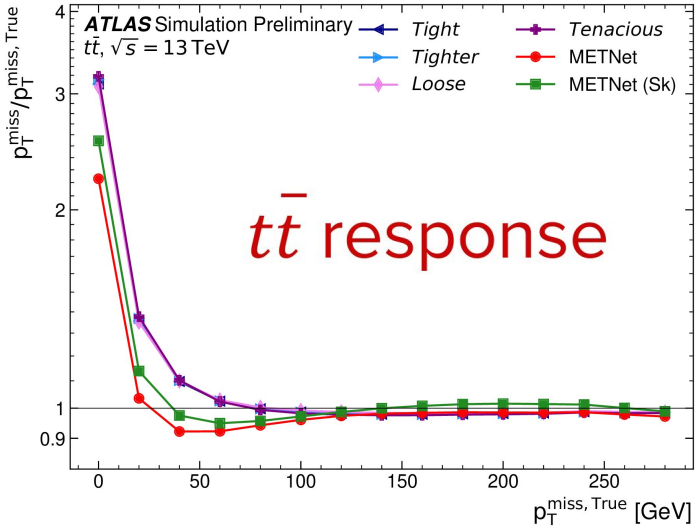


[ATL-PHYS-PUB-2021-025]

METNet performance

- Improved resolution compared to individual working points.
- Negative bias
  - ➢ Reduced using sample weights and sinkhorn contribution to loss function

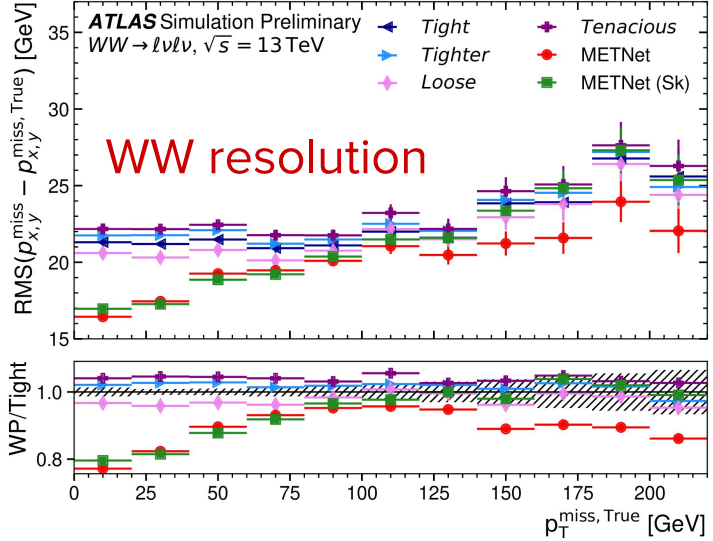


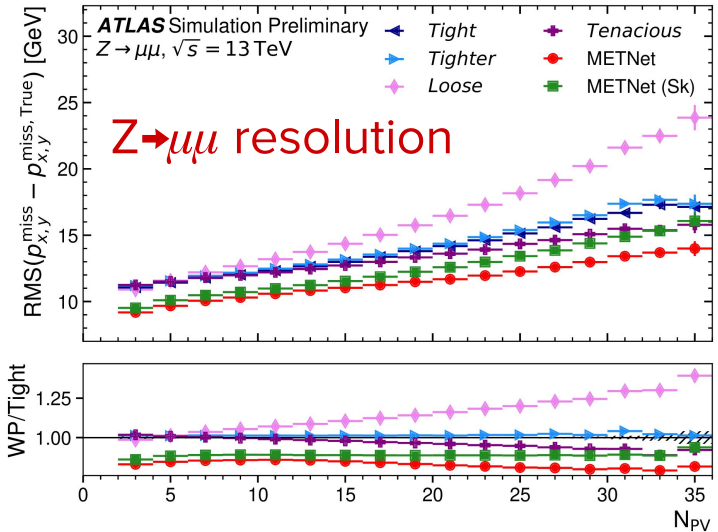


[ATL-PHYS-PUB-2021-025]

METNet performance

- Improved resolution compared to individual working points.
- Negative bias
  - ➢ Reduced using sample weights and sinkhorn contribution to loss function
- METNet can generalize to a number of event topologies
- Autotunes to event conditions





[ATL-PHYS-PUB-2021-025]

METNetSig

Mismeasured objects + finite acceptance/resolution

- p_T^{miss} significance: separates real and fake p_T^{miss}

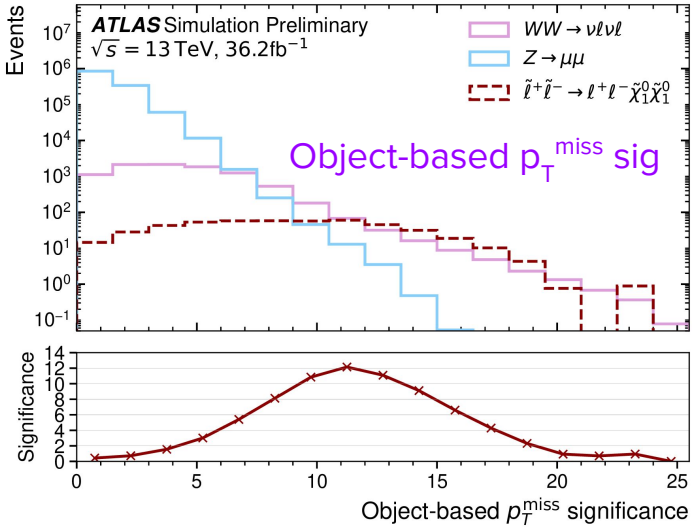
Neutrinos / bSM

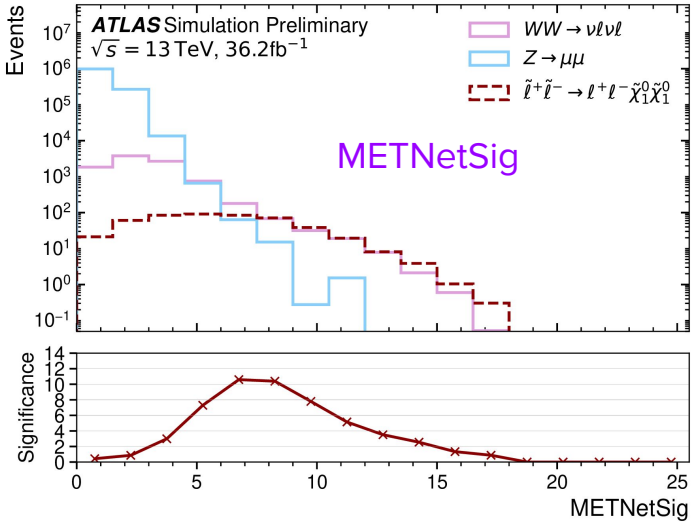
- Use Gaussian Negative Log Likelihood loss to produce 'confidence' σ as well as central prediction:

$$\mathcal{L}_{\mathrm{GNLL}} = \log\sigma + 0.5\left(\frac{y-\hat{y}}{\sigma}\right)^2$$

$$\mathrm{METNetSig} = p_{\mathrm{T}}^{\mathrm{miss,\,NN}}/\sigma$$

- Similar behaviour and performance to ATLAS object-based p_T^{miss} significance
  - ➢ Despite being constructed from entirely different inputs





[ATL-PHYS-PUB-2021-025]

2) Pion reconstruction

Pion reconstruction

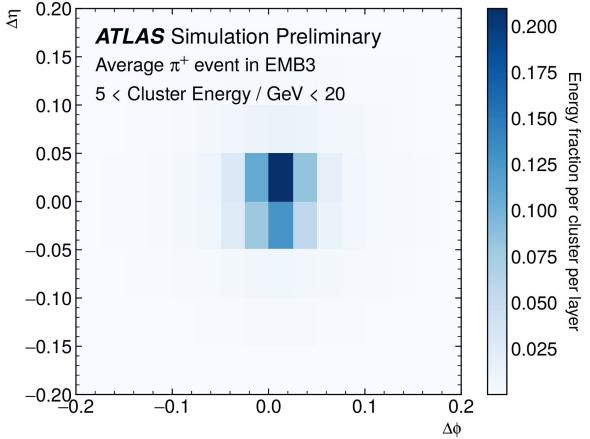
ATL-PHYS-PUB-2022-040

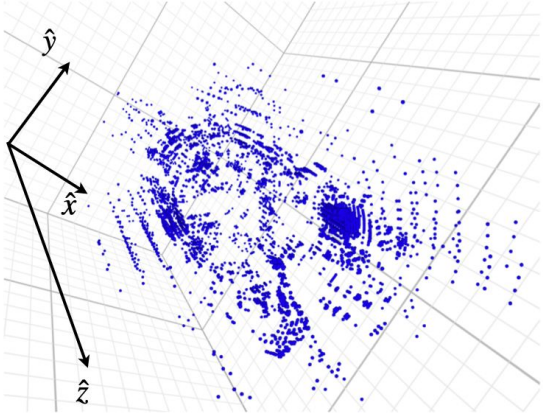
- Goal:
  - ➢ Classify pions as charged (π^{+-}) or neutral (π^0)
  - ➢ Regress true pion energy
- Why?
  - ➢ ATLAS has non-compensating calorimetry[1]
    - ➔ Deposits from charged and neutral pions need to be restored to different scales
  - ➢ First step towards larger goal of using ML in jet energy reconstruction (eg. for Particle Flow)
- ML methods:
  - ➢ **Convolutional Neural Network (CNN)**
  - ➢ **Deep Neural Network (DNN)** - regression only
  - ➢ **DeepSets**
  - ➢ **Graph Neural Network (GNN)**
  - ➢ **Transformer** - regression only

[1] arXiv.1603.02934

Input representations:

Images of calorimeter data



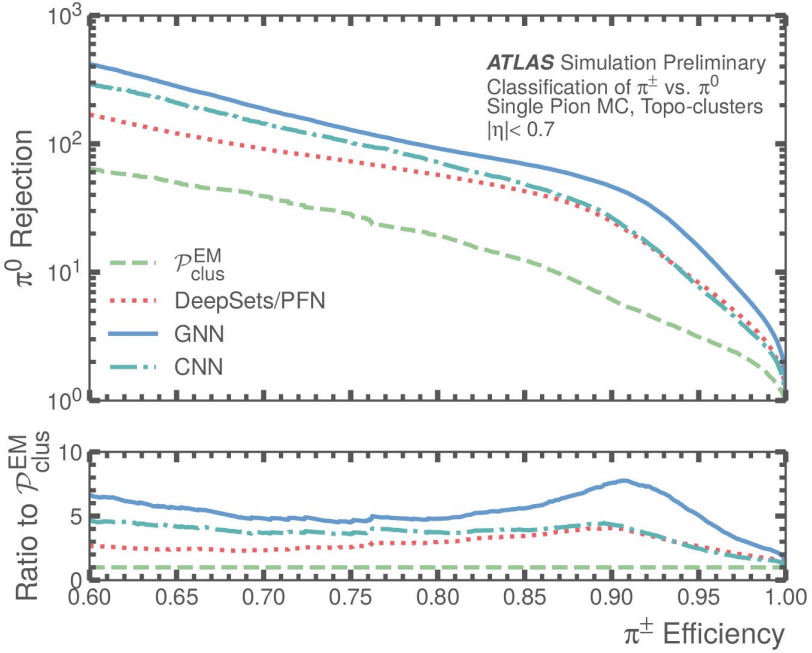


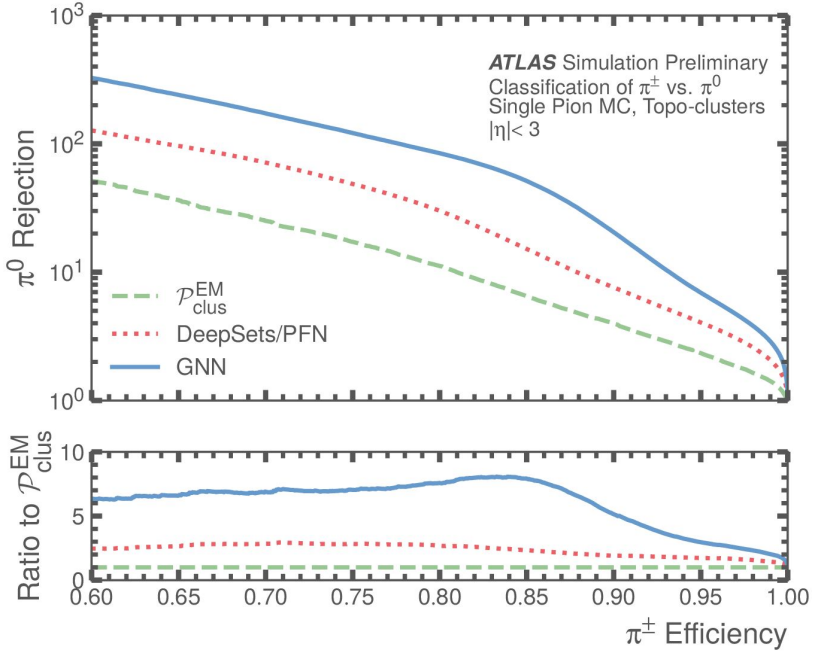
Point clouds
➔ 3D vectors of individual topocluster cells and tracks
➔ Target truth particle energy instead of truth cluster energy

Pion classification

ATL-PHYS-PUB-2022-040

- All methods outperform **baseline**
- Excellent performance for **GNN** in particular

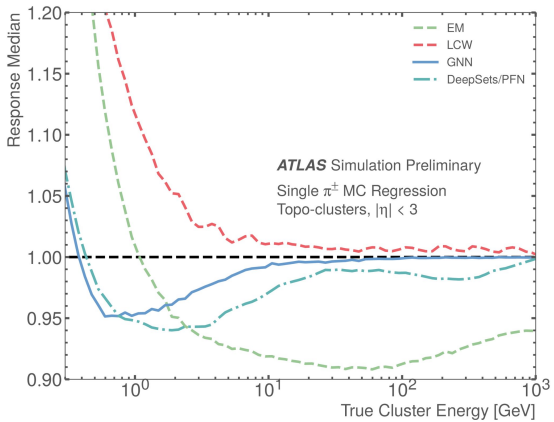


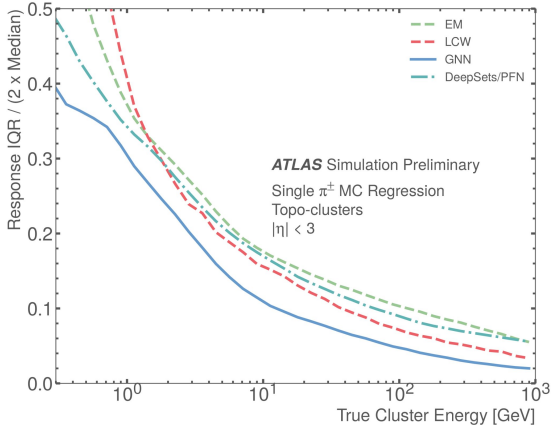


Pion energy calibration

ATL-PHYS-PUB-2022-040

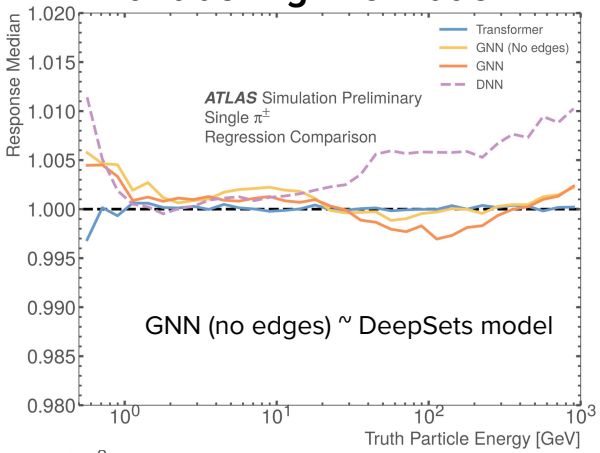
Without tracking information

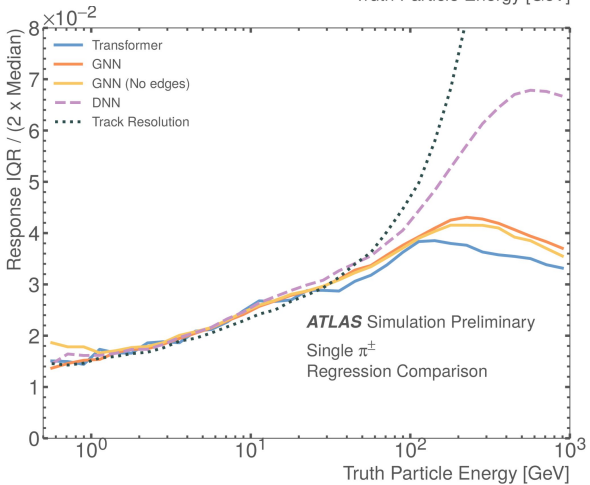




- All ML models significantly **outperform** EM and LCW **baselines**
- Including **tracking** information significantly improves performance
- Up to ~50 GeV, the results approximate **tracker energy resolution**
  - ➢ Beyond this, calorimetry energy resolution dominates
  - ➢ ML giving the **best of both**
  - ➢ Advantage of **point-cloud methods**

With tracking information

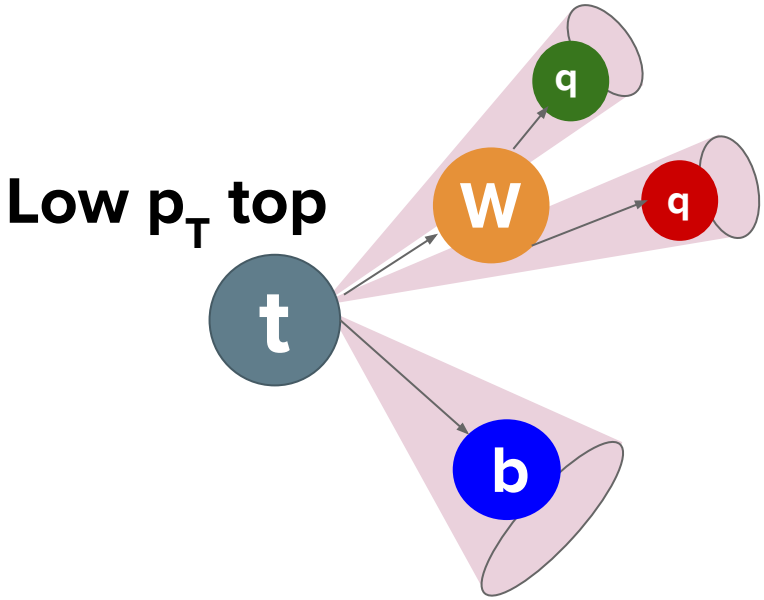


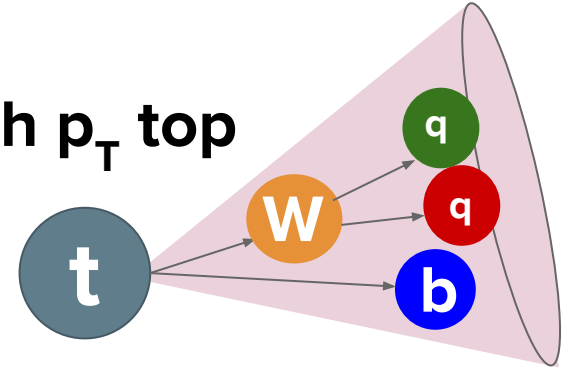


3) Boosted jet tagging

Boosted jet tagging

- Taggers distinguish large-radius jets from massive particles (W/Z/top) from light quark/gluon-initiated jets
  - ➢ Use jet substructure information
  - ➢ Enhances performance of bSM searches and precision SM measurements
- Latest taggers use jets reconstructed from Unified Flow Objects[1]
  - ➢ Combine topocluster and tracking information
  - ➢ Improved pile-up resilience and jet mass resolution
- ML classification techniques can improve on previous cut-based taggers





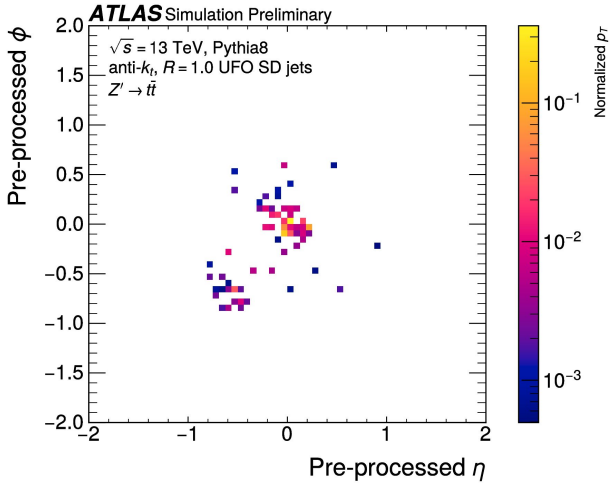
[1] arXiv:2009.04986

Top taggers

ATL-PHYS-PUB-2022-039

Latest taggers use kinematic properties of **jet constituents** rather than **high-level** jet substructure quantities

- ➢ Impressive performance on simplified Delphes simulated data-sets
- ➢ How do they perform with realistic GEANT4-simulated samples?

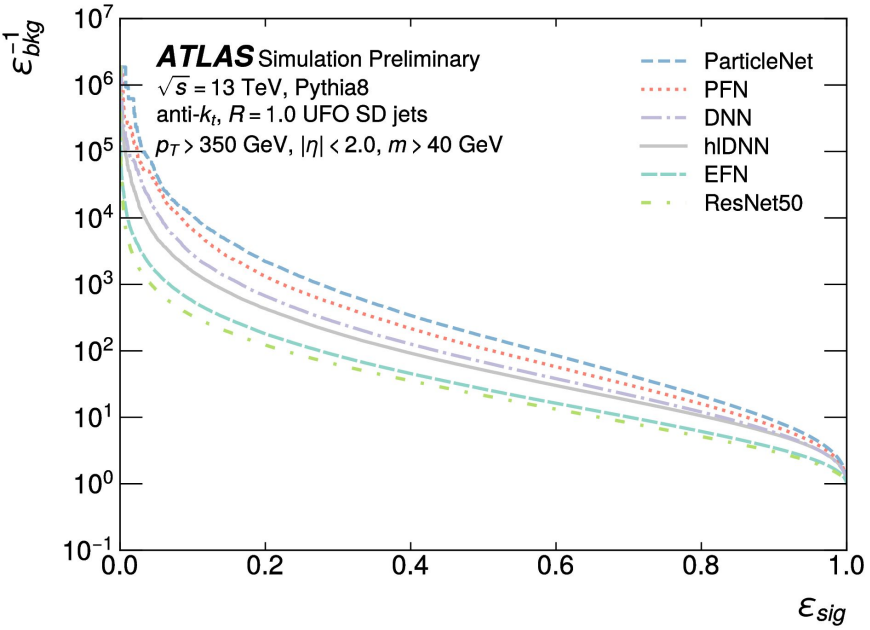


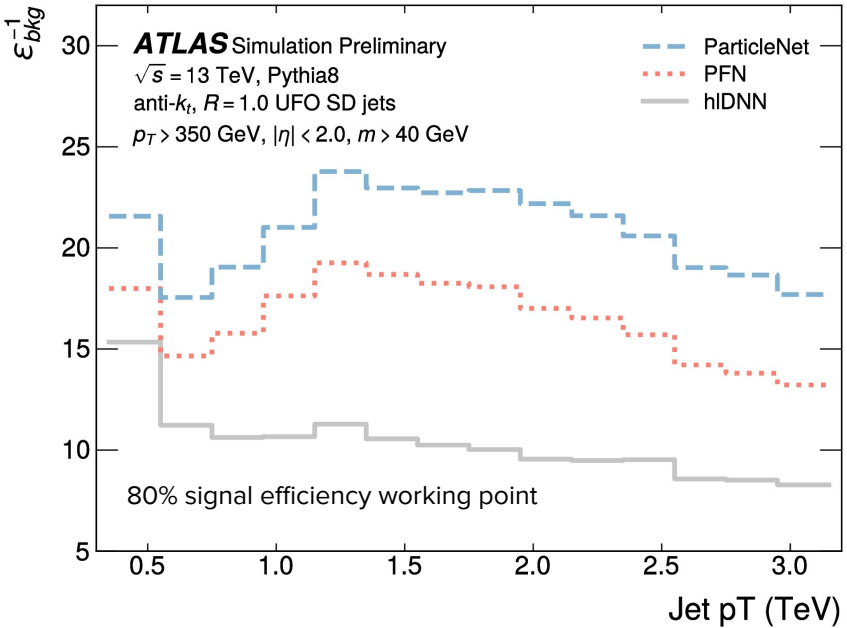
1) Baseline **DNN** trained on **high-level** input features
   (from ATL-PHYS-PUB-2021-028)
2) **DNN** trained on **constituent-level** input features
3) **ParticleNet** (arXiv:1902.08570)
   - ➢ Graph neural network (GNN)
   - ➢ Jets represented as graph (nodes = constituents)
4) **Energy Flow Network (EFN)** (arXiv:1810.05165)
   - ➢ DeepSets structure
   - ➢ Can only consider constituent-level quantities linear in p_T
5) **Particle Flow Network (PFN)** (arXiv:1810.05165)
   - ➢ Similar to EFN but permits all constituent-level quantities
6) **ResNet50** (arXiv:1512.03385)
   - ➢ Large-scale convolutional neural network (CNN)
   - ➢ Trained on "images" of constituent p_T in (η,ϕ) plane

Top taggers

ATL-PHYS-PUB-2022-039

- **3 / 5** constituent-based taggers **outperform** **hlDNN** baseline
- **EFN** and **ResNet50** **underperform** relative to **Delphes**-based studies
  - ➢ Highlights the need to develop taggers in a realistic context
- Data-set publicly available for future development: https://gitlab.cern.ch/atlas/ATLAS-top-tagging-open-data
- Work to assess systematic uncertainties ongoing

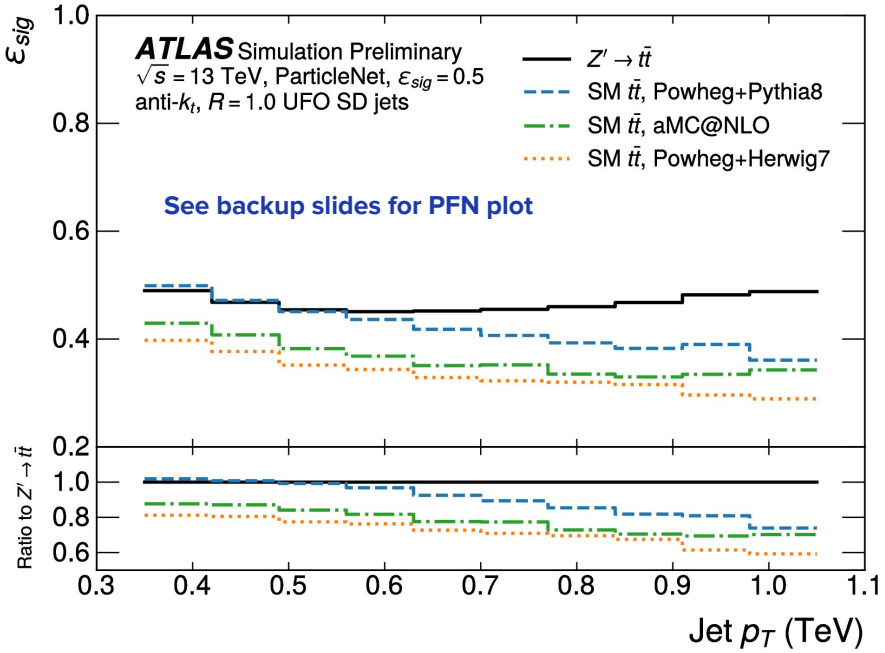


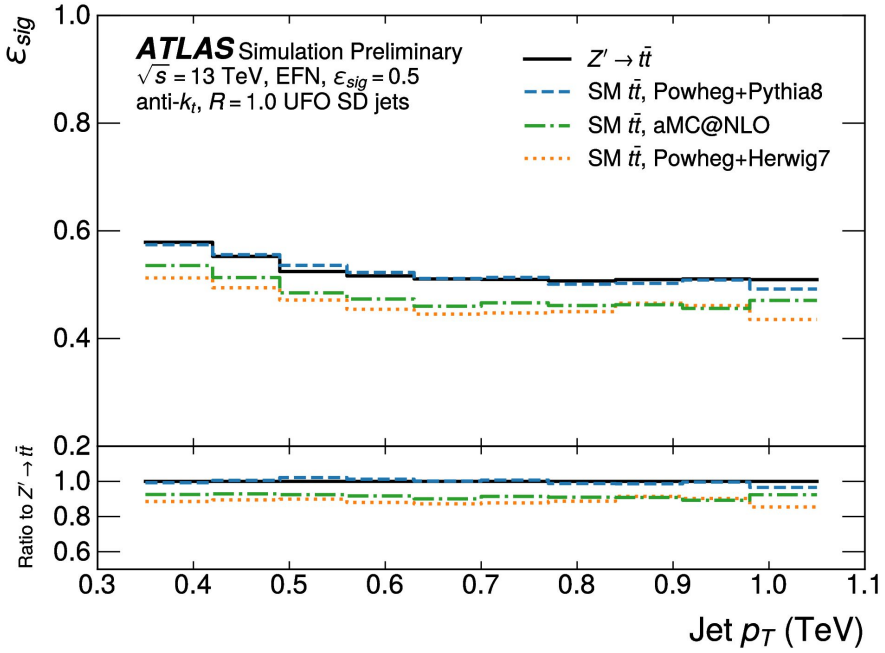


Top Taggers: Model-dependence

ATL-PHYS-PUB-2022-039

- **PFN** and **ParticleNet** show increased model dependence
  - ➢ More dependent on QCD modelling than high-level tagger
- **EFN** shows lowest model dependence due to IRC safety constraint

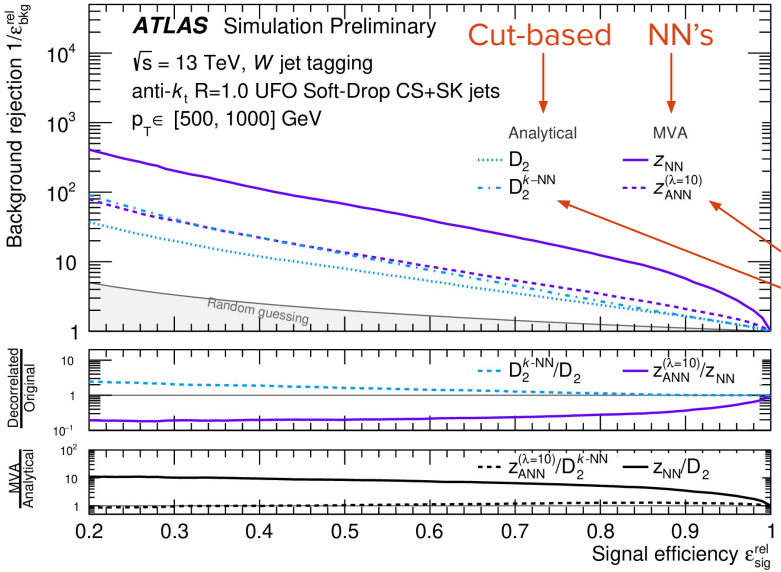




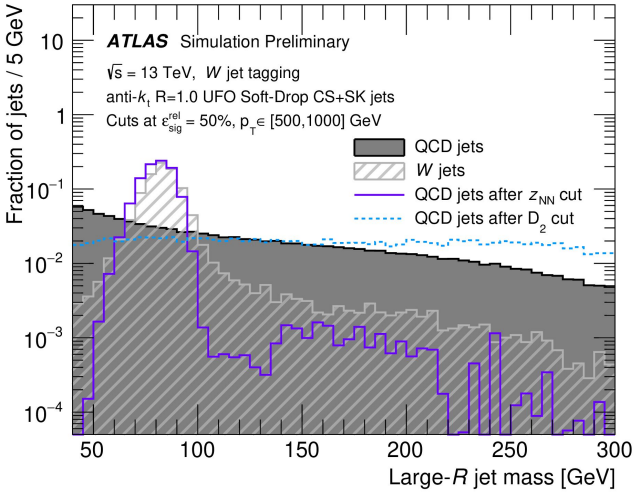
W/Z taggers

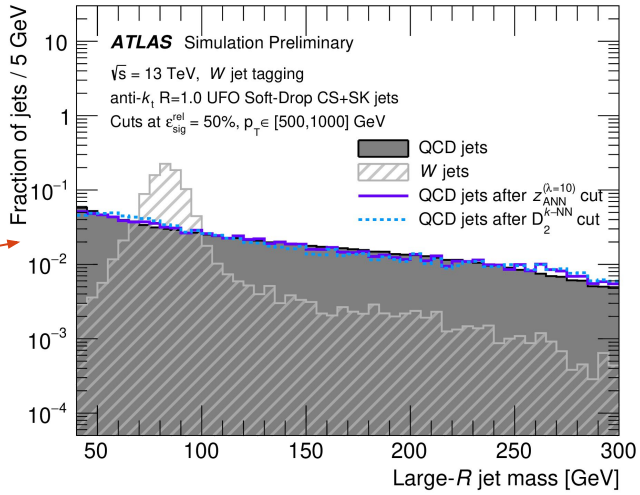
ATL-PHYS-PUB-2021-029
JETM-2022-006

- DNN has improved performance but high mass correlation
  - ➢ Complicates background estimation strategies
- Adversarial NN used to de-correlate jet mass
  - ➢ Corresponding decrease in performance could be partially recovered with analysis-specific mass-window cuts
  - ➢ Allows use of better side-band regions



With mass decorrelation





Summary

- Hadronic object reconstruction is ripe for ML applications
- Calibration (predicting truth-level energy/pT with detector-level quantities)
  - ➢ METNet
  - ➢ Pion energy calibration
  - ➢ Jet calibration (See Naseem Bouchhar's talk)
- Classification
  - ➢ Boosted top and W/Z taggers
  - ➢ π^{+-} vs. π^{0}
- Development of all of these applications is ongoing and promises to enhance the performance of bSM searches and precision SM measurements

BACKUP SLIDES

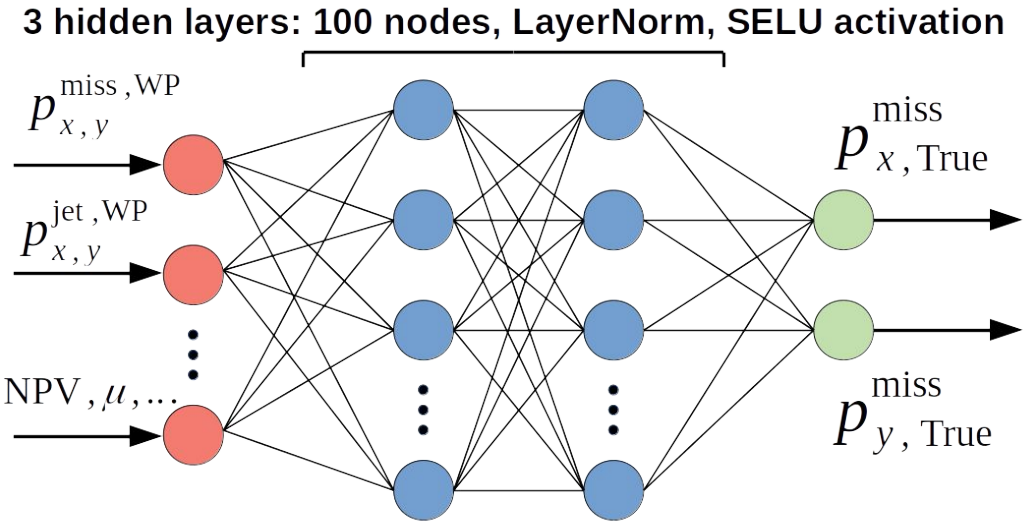
METNet set-up

- Dense neural network **regresses truth-level p_T^{miss}**
  - ➢ Trained on **ttbar** and **di-boson** MC events.
- **Sample weights** flatten the p_T^{miss} distribution of the training data ➔ reduces bias
- Loss functions: Huber + Sinkhorn

Pre-processing

Rotate to common p_T^{miss} direction

Standardize inputs



[ATL-PHYS-PUB-2021-025]

p_T^{miss} working points

Working point	Selections: p_T [GeV] for jets with: $\lvert\eta\rvert < 2.4$	$2.4 < \lvert\eta\rvert < 4.5$	JVT for jets with $\lvert\eta\rvert < 2.4$	fJVT for jets with $2.5 < \lvert\eta\rvert < 4.5$ and $p_T < 120$ GeV
Loose	> 20	> 20	> 0.5 for $p_T < 60$ GeV jets	-
Tight	> 20	> 30	> 0.5 for $p_T < 60$ GeV jets	< 0.4
Tighter	> 20	> 35	> 0.5 for $p_T < 60$ GeV jets	-
Tenacious	> 20	> 35	> 0.91 for $20 < p_T < 40$ GeV jets > 0.59 for $40 < p_T < 60$ GeV jets > 0.11 for $60 < p_T < 120$ GeV jets	< 0.5

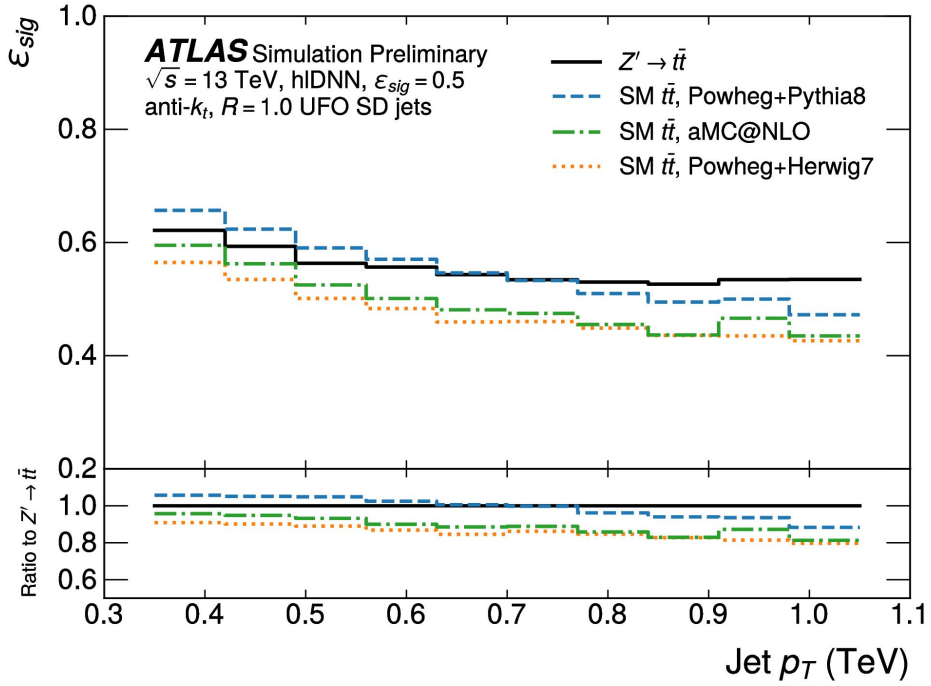
METNet set-up

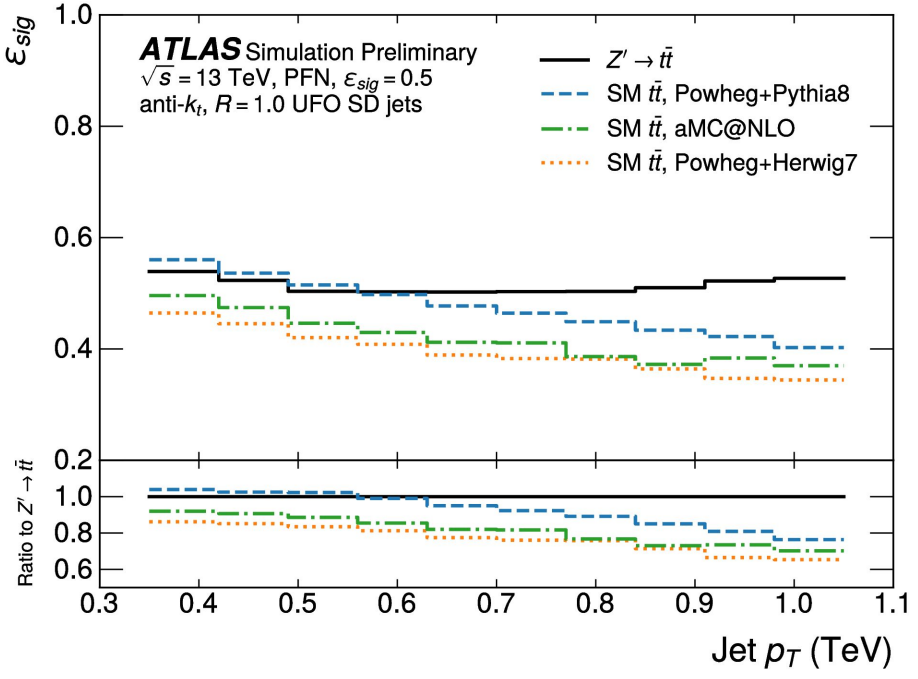
Layer	Nodes	Activation	Parameters
Input	60	None	0
Hidden 1	100	SELU	6100
LayerNorm 1	100	SELU	200
Hidden 2	100	SELU	10100
LayerNorm 2	100	SELU	200
Hidden 3	100	SELU	10100
LayerNorm 3	100	SELU	200
Output	2	Linear	202
Total			27,102

Hyperparameter	Value
Optimiser	Adam [67]
Weight initialization	Kaiming He [68]
Learning rate	0.001
Batch size	256
Huber loss δ	1.5

Top Taggers: Model-dependence

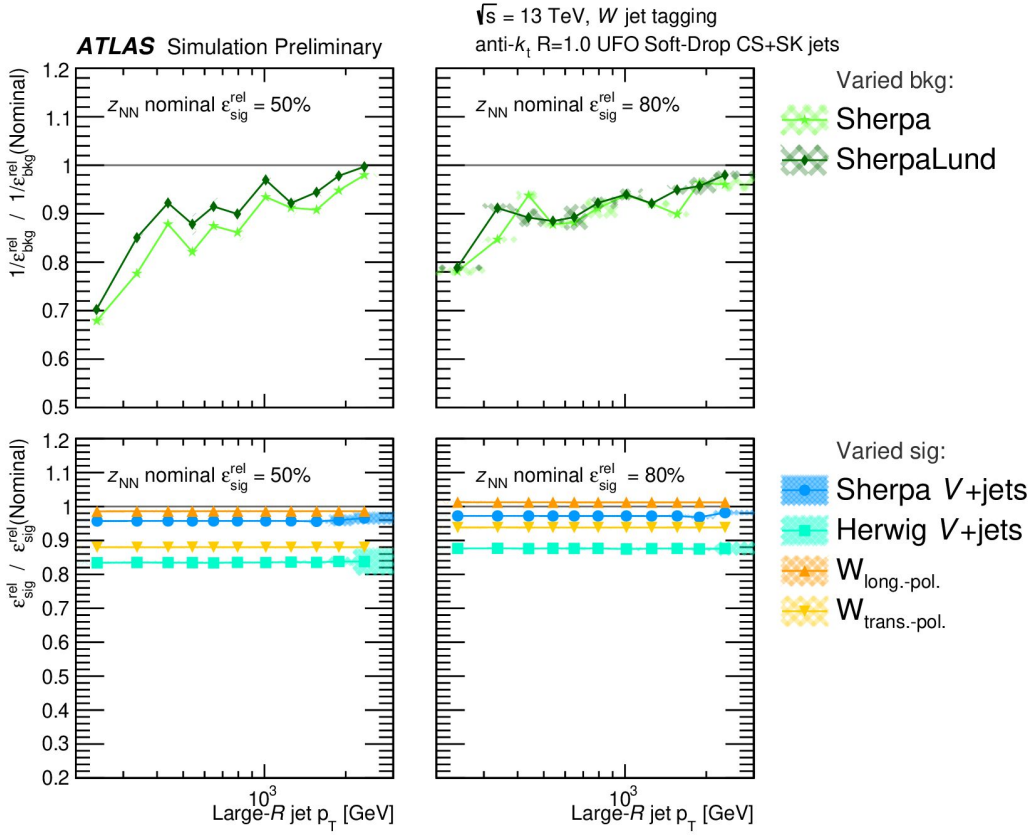
ATL-PHYS-PUB-2022-039





W/Z taggers: Model-dependence

- Sensitive to modelling differences between MC generators
- Also sensitive to W boson polarization
- Further studies and careful calibration required to avoid large systematic uncertainties



Top tagger: ML set-up

Model	Hyper-parameters
ResNet 50	Bottom Layer: 7x7 2D convolution with strides (2, 2) and zero padding Number of Stages: 4 Blocks per Stage: (3, 4, 6, 3) *Block Type: bottleneck* Block Output Filters: (64, 128, 256, 512) *Activation Functions: ReLU* *Kernel Initialization: he uniform* Batch Normalization Momentum: 0.1 *Global Pooling: average* Initial Learning Rate: 1×10^{-2} *Scheduler: decrease learning rate by factor of 0.1 every 10 epochs* Batch Size: 256
ParticleNet	Φ Number of Stages: 3 Blocks per Stage: (3, 3, 3) Block Output Features: (64, 224, 384) k Nearest Neighbors: 18 Top Layer Nodes: 125 *Activation Functions: ReLU* *Kernel Initialization: glorot normal* Batch Normalization Momentum: 0.7 Global Pooling: max Learning Rate: 4.2×10^{-4} Batch Size: 250

Model	Hyper-parameters
hlDNN	Hidden Layers: 5 Nodes per Layer: 180 *Activation Functions: ReLU* *Kernel Initialization: glorot uniform* Learning Rate: 4×10^{-5} Batch Size: 250 Batch Normalization: not used
DNN	Hidden Layers: 5 Nodes per Layer: 400 *Activation Functions: ReLU* *Kernel Initialization: glorot uniform* L1 Regularization: 2×10^{-4}, applied to all layers Learning Rate: 1.2×10^{-5} Batch Size: 250 Batch Normalization: applied before activation function for all layers except output layer
EFN	Φ Hidden Layers: 5 Φ Nodes per Layer: 350 Latent Dropout: 0.084 F Hidden Layers: 5 F Nodes per Layer: 300 F Dropout: 0.036 *Activation Functions: ReLU* *Kernel Initialization: glorot normal* Learning Rate: 6.3×10^{-5} Batch Size: 350
PFN	Φ Hidden Layers: 5 Φ Nodes per Layer: 250 Latent Dropout: 0.072 F Hidden Layers: 5 F Nodes per Layer: 500 F Dropout: 0.022 *Acitvation Functions: ReLU* *Kernel Initialization: glorot normal* Learning Rate: 7.9×10^{-5} Batch Size: 250

Top-tagger: Baseline DNN inputs

Quantity Type	Symbols	References
N-subjettiness	τ_1, τ_2, τ_3, τ_4	[66] [67]
k_t Splitting Scales	$\sqrt{d_{12}}$, $\sqrt{d_{23}}$	[15]
Generalized Energy Correlation Functions	ECF_1, ECF_2, ECF_3, C_2, D_2, L_2, L_3	[68] [69] [70]
Minimum Pair-wise Invariant Mass	Q_w	[15]
Thrust Major	T_m	[71]

Table 1: A listing of the 15 quantities used to train the baseline high level quantity tagger