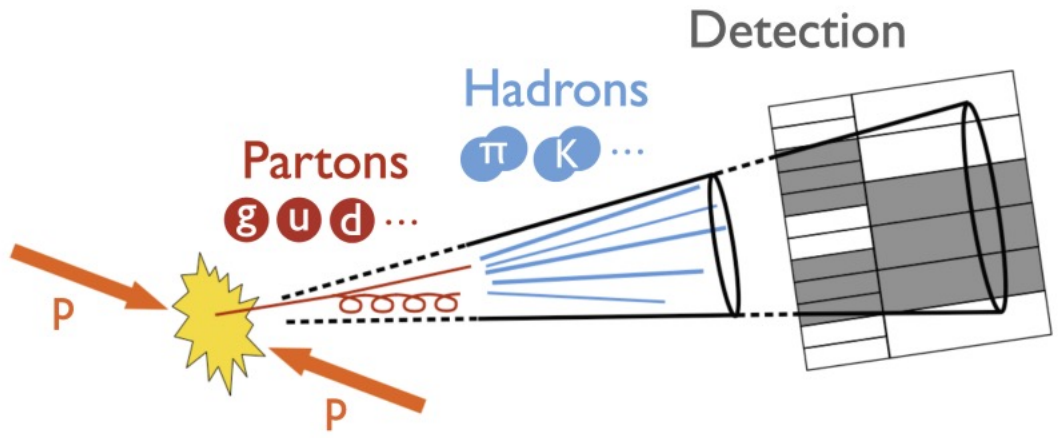
Improving ATLAS Hadronic Object Performance with ML/AI Algorithms

Xiang Chen

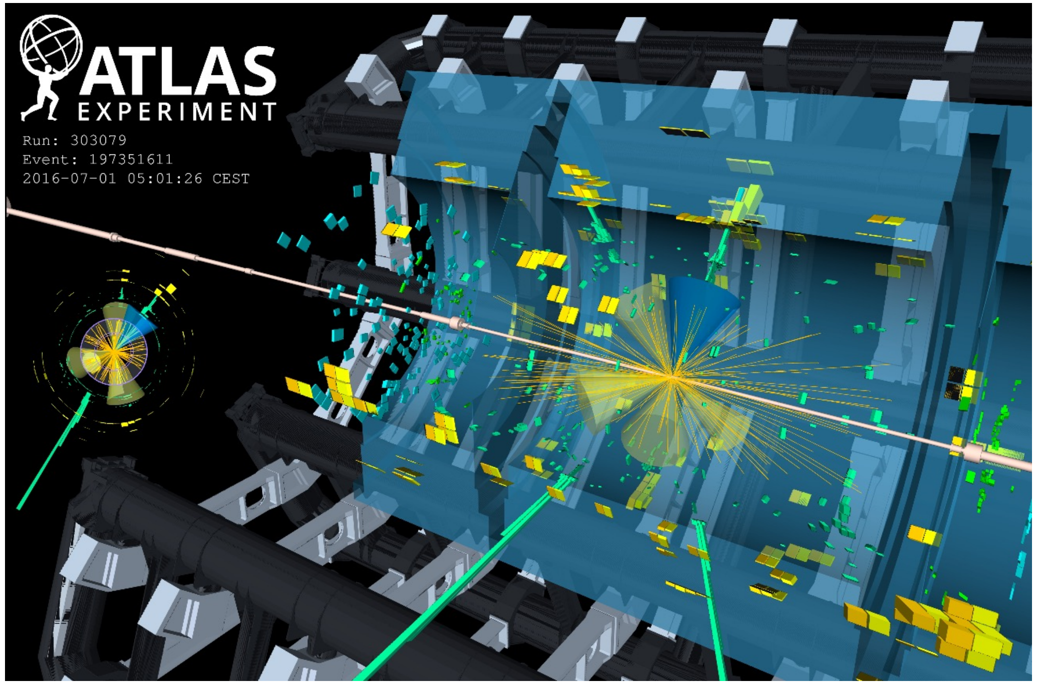
On behalf of ATLAS Collaboration

Introduction - Jets in High Energy Physics

- Jets are formed by showers of particles originating from the hadronization of quarks and gluons
- In ATLAS physics analyses, jets are typically reconstructed using anti-kT algorithm, using the combined information of tracks and energy deposits in calorimeters as inputs
- Jets are important for physics analyses for BSM searches and SM measurements in collider experiments
- Jets contains complex information and it is very suitable for the application of machine learning methods



https://cms.cern/news/jets-cms-and-determination-their-energy-scale



https://atlas.cern/updates/press-statement/atlas-observes-tth-production

Introduction – Hadronic Objection in ATLAS

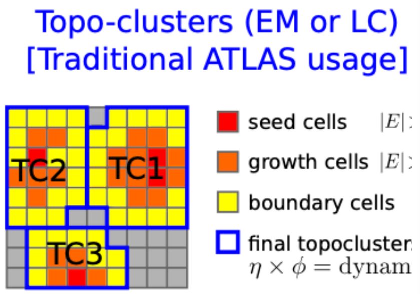
- There are several steps for building jets:
  - **Inputs finding**: to get the constituents as the 4-vector inputs from the tracker and/or calorimeters
  - **Reconstruction**: to group the constituents to form the ‘cone’, applying the grooming method and pile-ups removal at the same time
  - **Calibration**: to correct the jet energy scale
  - **Tagging**: to identify which particle the jet is coming from (such as W/Z, top quark, or gluon)
- **Missing transverse momentum:** to derive the missing momenta from calibrated hadronic objects in events

Highlighted for today:

1. Jet reconstructions:
   - calorimeter signals reconstruction/ classification – ATL-PHYS-PUB-2022-040
2. Boost jet tagging in UFO jets
   - W/Z tagging - ATL-PHYS-PUB-2021-029
   - Top tagging - ATL-PHYS-PUB-2021-028, ATL-PHYS-PUB-2022-039
3. Jet Calibration:
   - METNet - ATL-PHYS-PUB-2021-025 (More information from Peter’s talk)
   - The Global Neural Network Calibration – arXiv:2303.17312

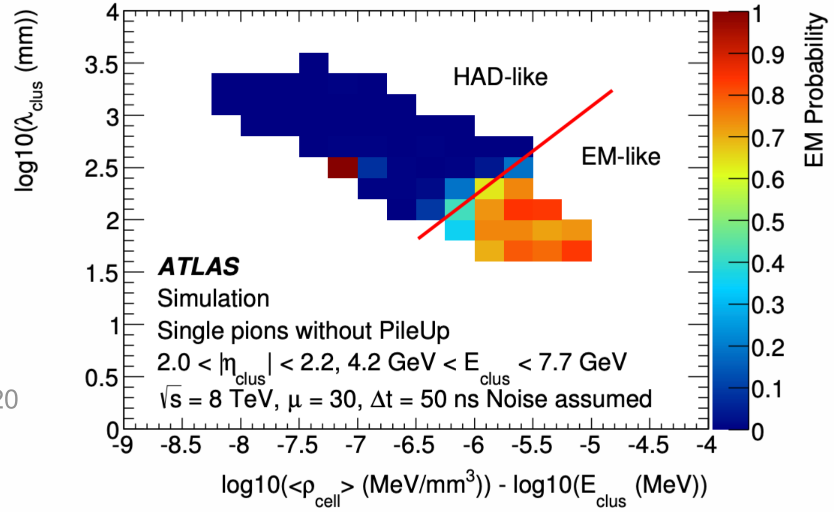
Pion Reconstruction

- The reconstruction and classification of π^0 and $\pi^{+/-}$ are important in reconstruction of hadronic signals
- Three-dimensional clusters of topologically-connected calorimeter cells called **topo-clusters** are the starting point of hadronic reconstructions
- Current approach:
  - Cluster classification
  - Local Cell Weighting(LCW)



From Steven Schramm, BOOST 2020

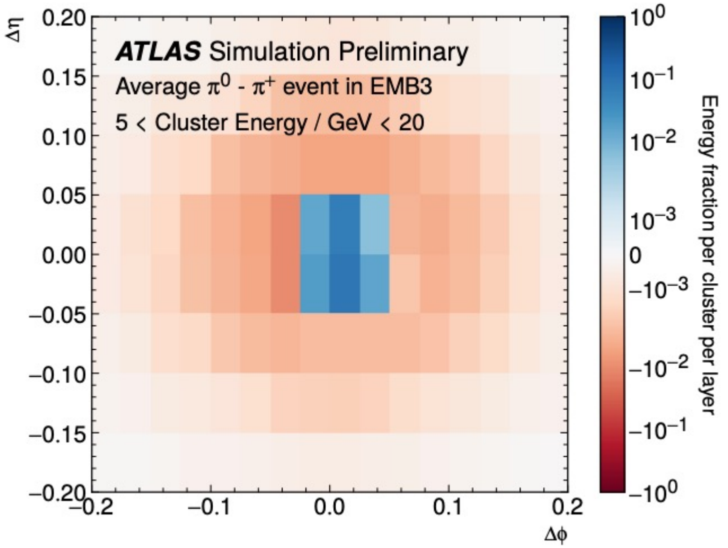
Arxiv:1603.02934

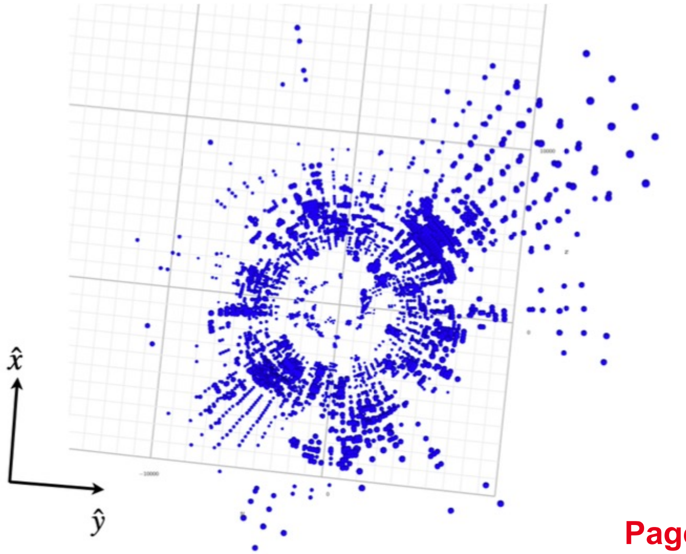


- Use **ML methods**:
  - To classify π^0 and $\pi^{+/-}$ and reconstruct its energy
  - As an important step towards the upgraded low-level hadronic reconstruction scheme

ML methods:
- Cell as pixels: **CNN**
- Cell as points: **DeepSets(PFN)**, **Transformer**
- Cell as nodes: **GNN**

ATL-PHYS-PUB-2022-040

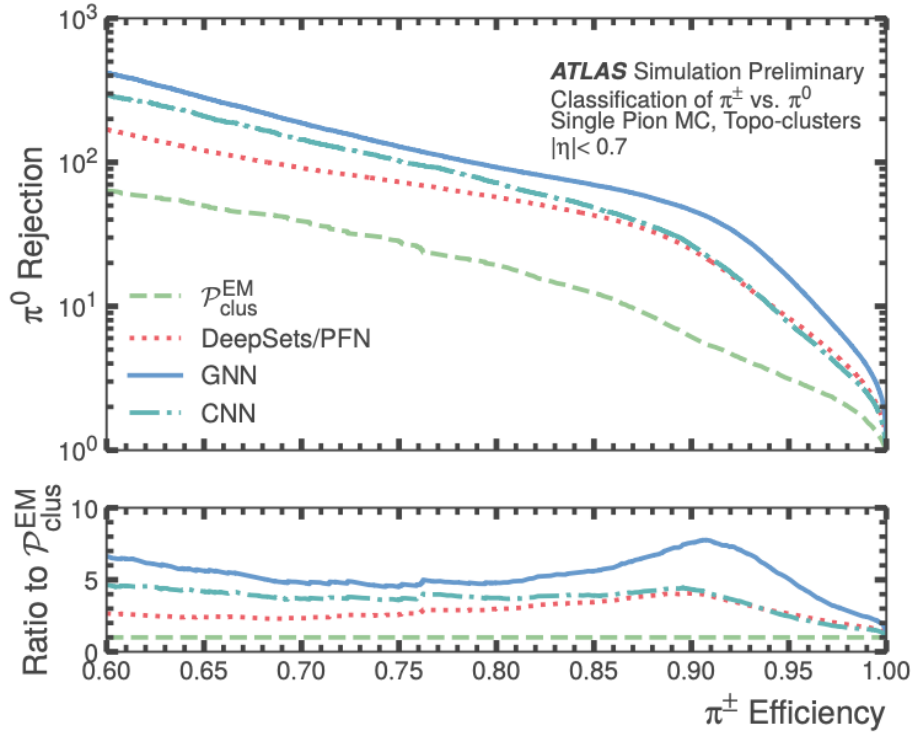


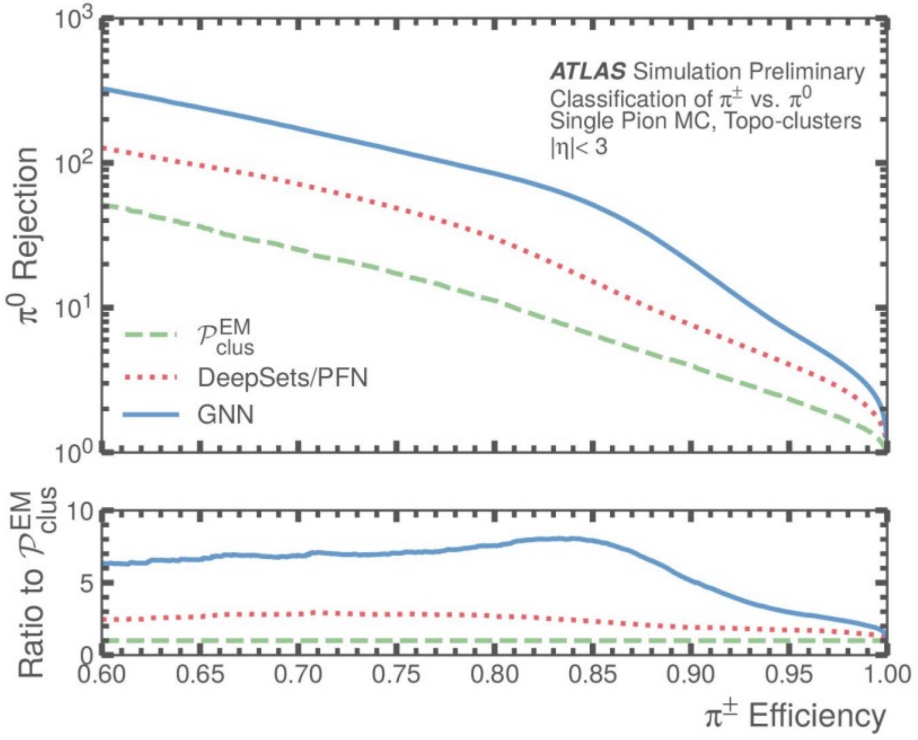


Pion Classification

ATL-PHYS-PUB-2022-040

- Performance evaluated by comparing the rejection rate of π^0 with fixing of $\pi^{+/-}$ efficiency
- ML methods (PFN, CNN, GNN) all outperform the baseline method ($\mathcal{P}_{clus}^{EM}$)
- **GNN** is the best in overall: improving ~5 times rejection in the full pseudo-rapidity range





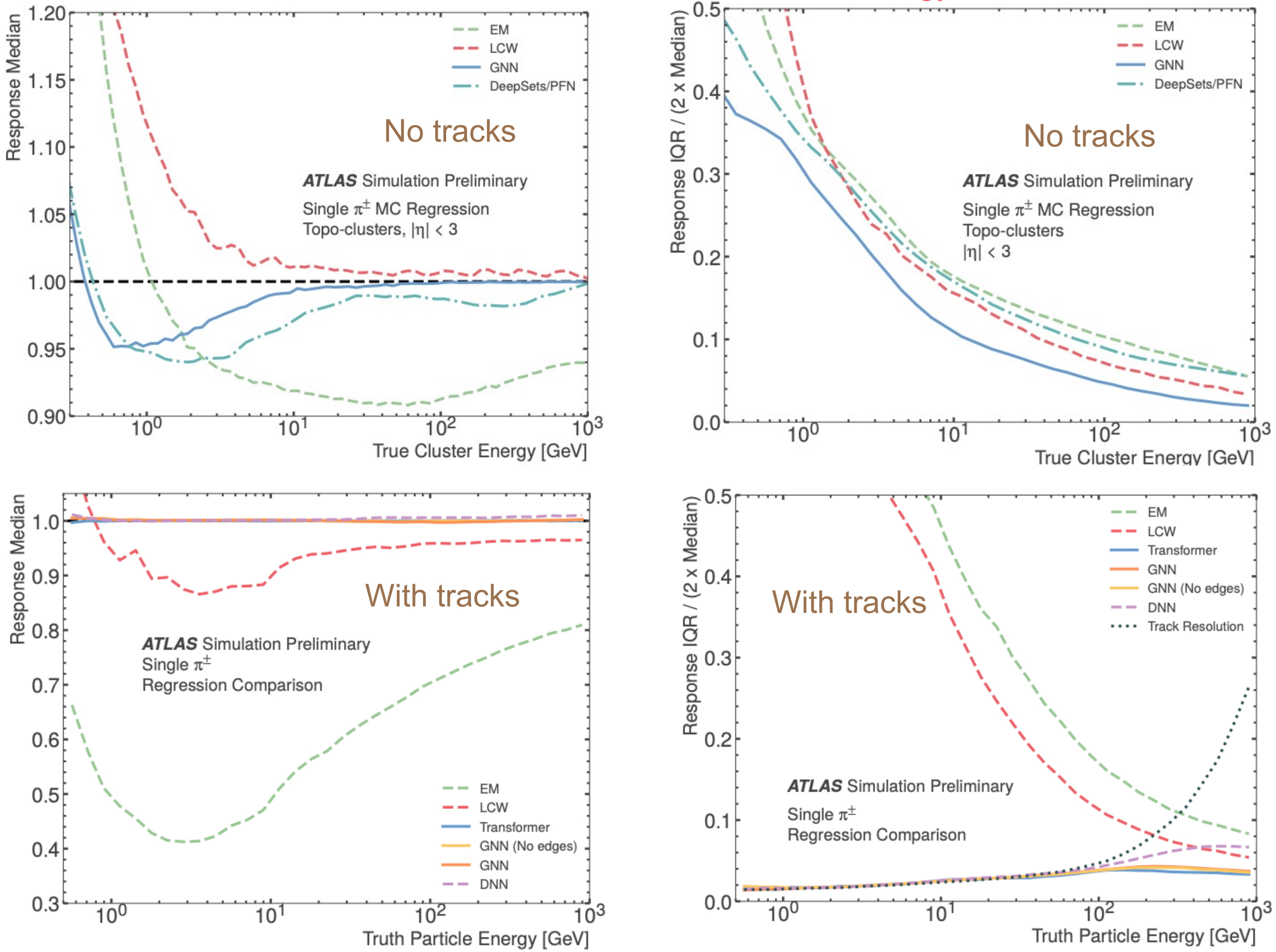
Pion Calibration Performance

ATL-PHYS-PUB-2022-040

- The performance of the models can be quantified by measuring the energy response
- Compared to LCW(calibrated) and EM(uncalibrated):
  - **All machine learning-based methods perform better** in energy calibration accuracy and resolution of energy measurement
- The model including track information (bottom plots) performs significantly better

Energy response

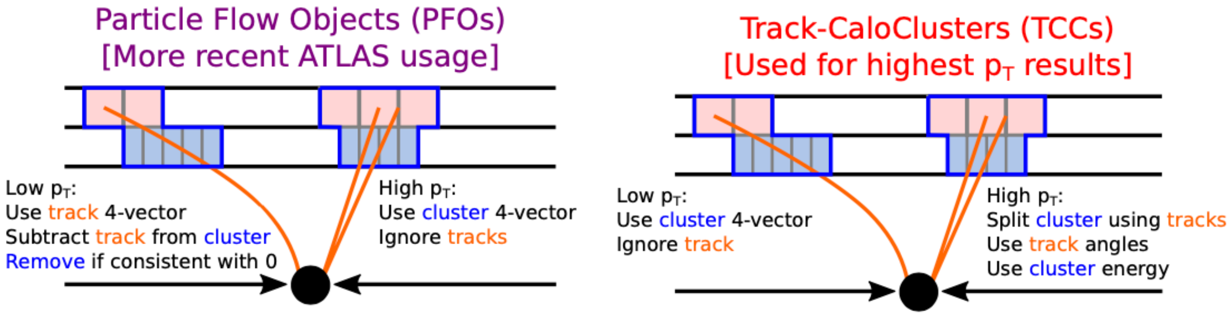
Energy resolution

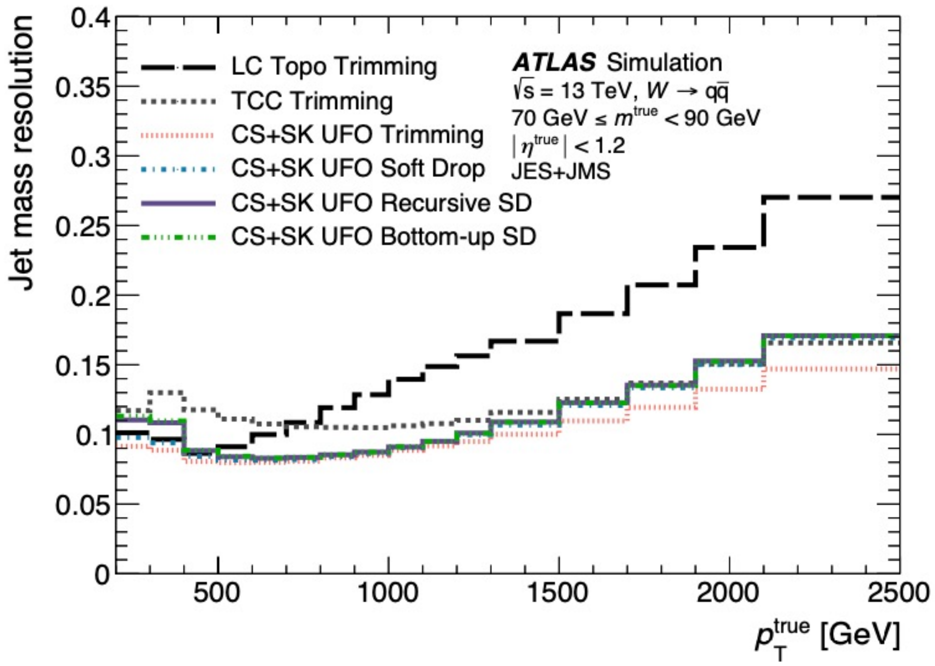


Boost Jet Tagging: UFO Jets

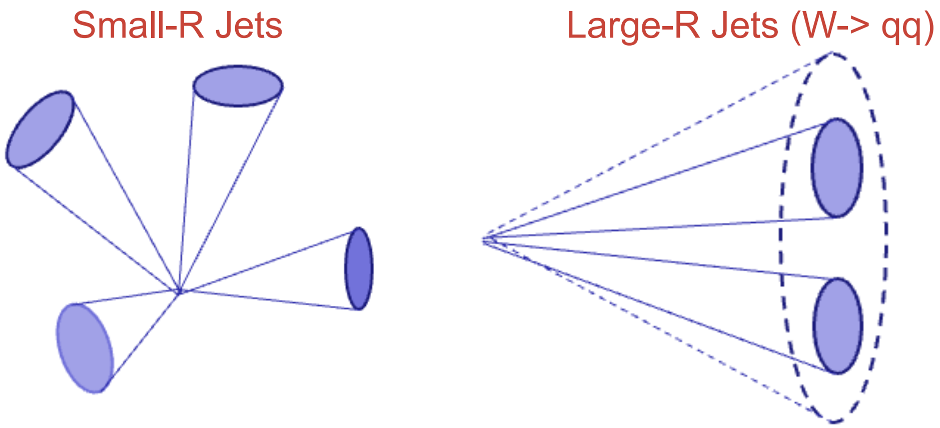
- For better reconstruction of jets in Run2, new jet reconstructions algorithms are introduced to build **Unified Flow Objects(UFO) -** *Eur. Phys. J. C* **81**, 334 (2021)
- UFO combines good tracker resolution/low-pt resolution with calorimeter information
- UFOs can perform better on jet mass reconstruction and removing pile-up

From Steven Schramm, BOOST 2020





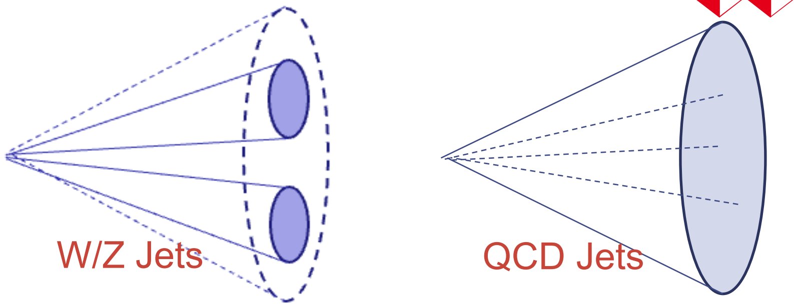
- Decay products from heavy particle decay can collimate into a single jet at high pT called **Large-R jet** (R = 2*pT/m)
- ML methods are applied for taggers identifying large–R jets from heavy particles (top, W/Z) with light QCD jets

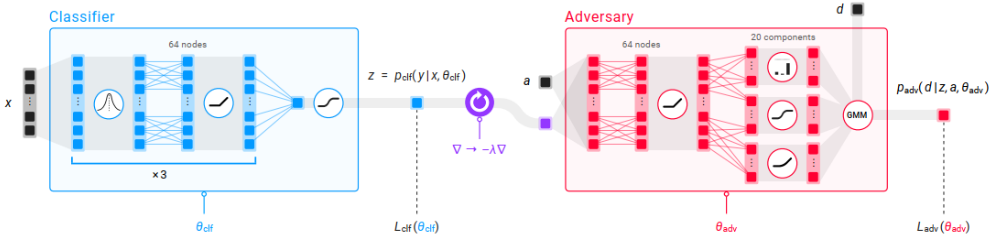


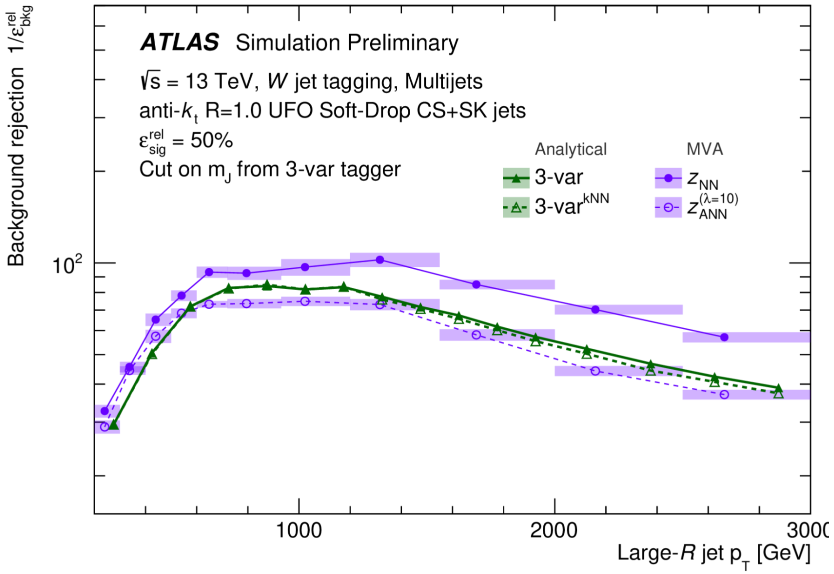
arXiv:2009.04986

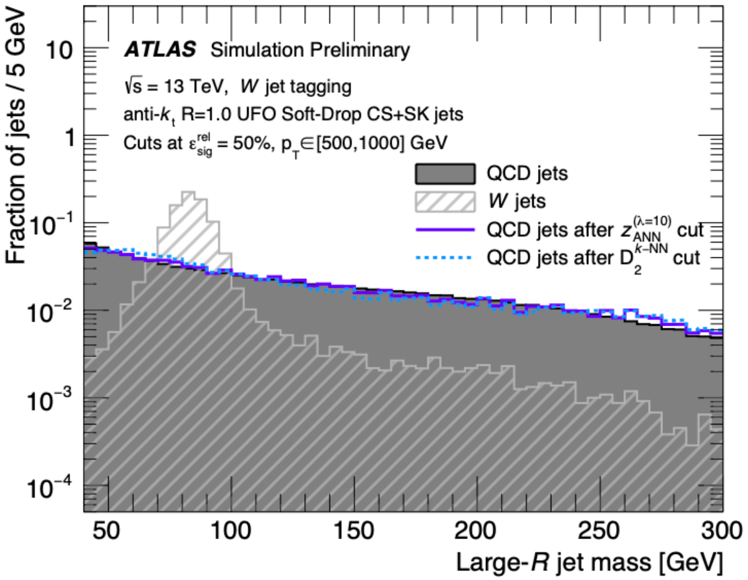
Boost Jet Tagging: W/Z tagging

- Using UFO jets to construct taggers: two taggers are available
- 3 variable tagger: using 3 jet substructures for cuts
- DNN tagger: using substructure variables better performance (~3 times of 3-var tagger) but the unwanted mass shape
- ANN mass-decorrelated tagger: design adversarial NN based on DNN tagger for better background estimation



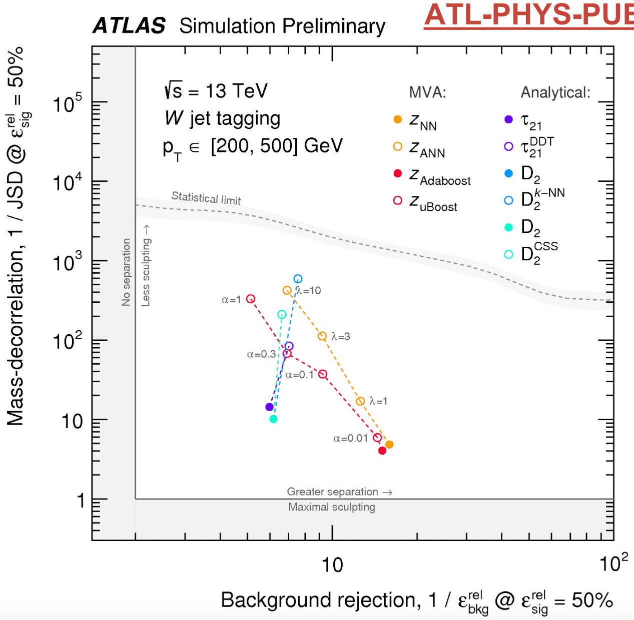






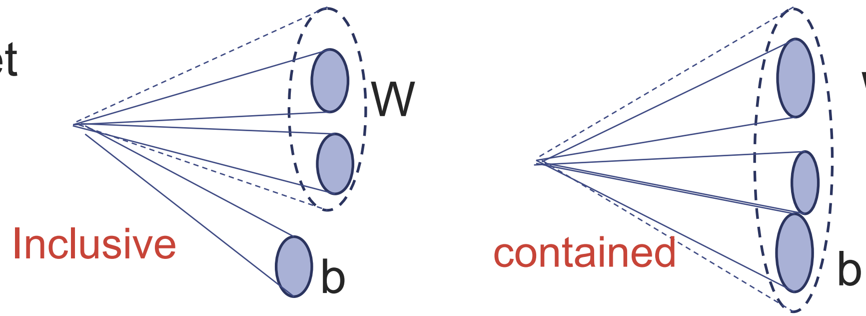
ATL-PHYS-PUB-2021-029

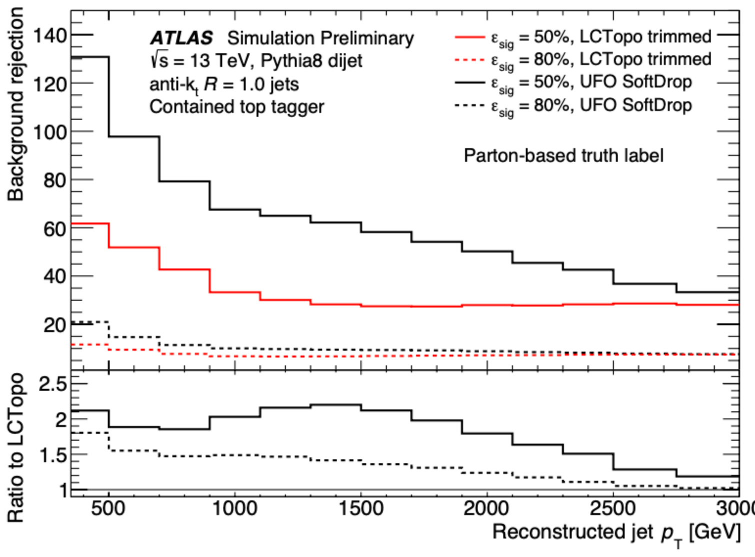
ATL-PHYS-PUB-2018-014

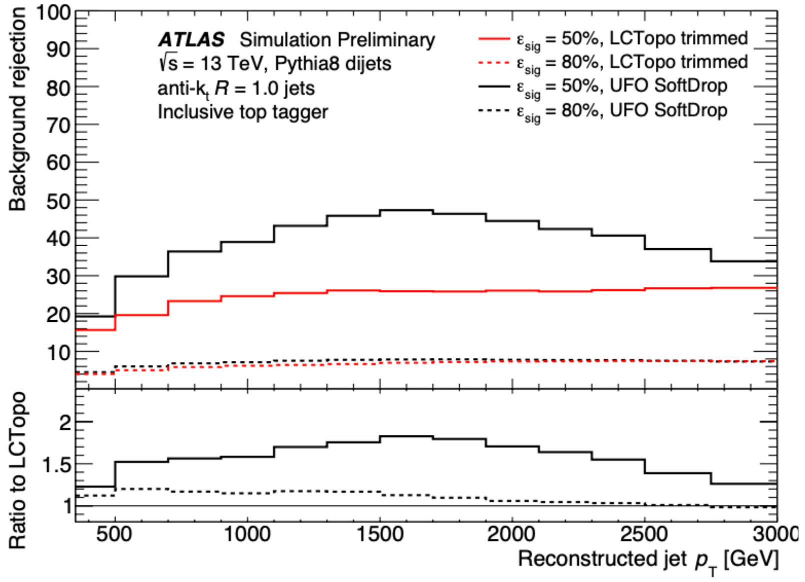


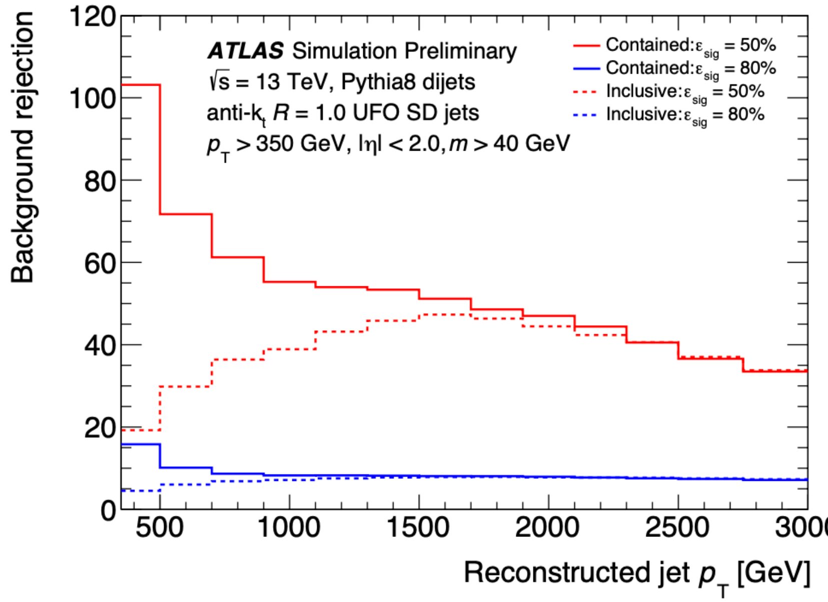
Boost Jet Tagging: Top tagging

- Using UFO jets to construct taggers: two DNN taggers using jet substructure variables are available
- Two different scenarios: **contained** and **inclusive** tops
- Using 15 different variables as inputs
- Compared to previous results (previous ATLAS jet constituent algorithm), DNN tagger performance at 50%(80%) signal efficiency is better overall









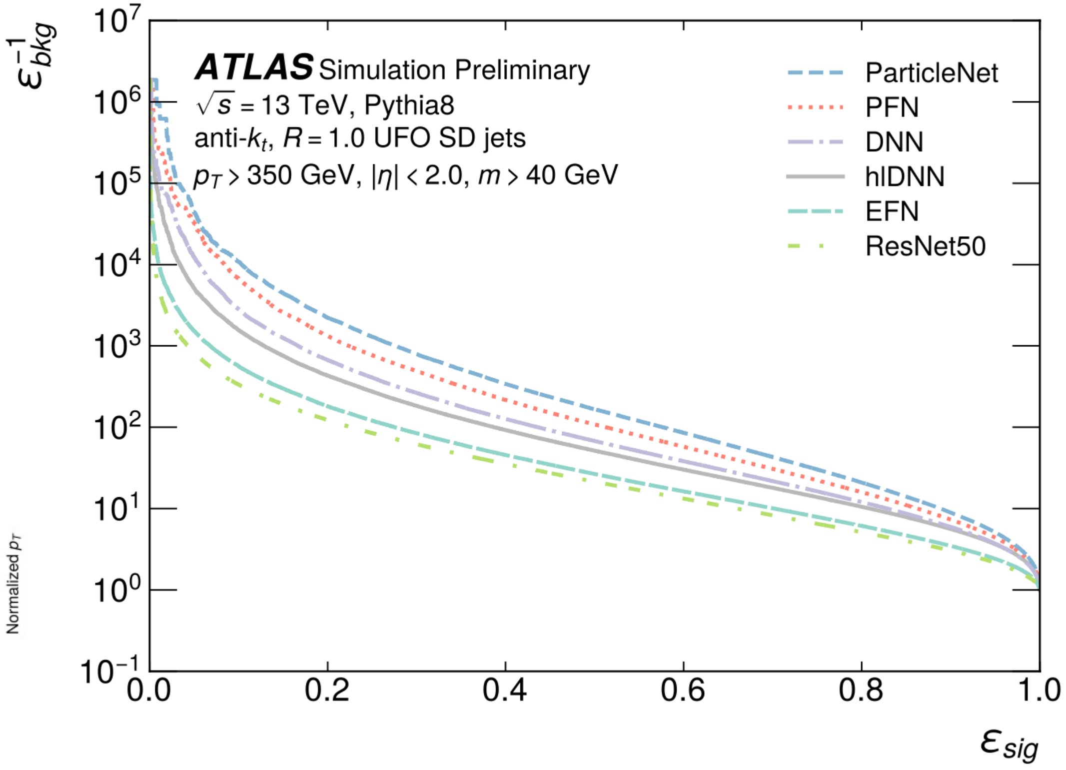
ATL-PHYS-PUB-2021-028

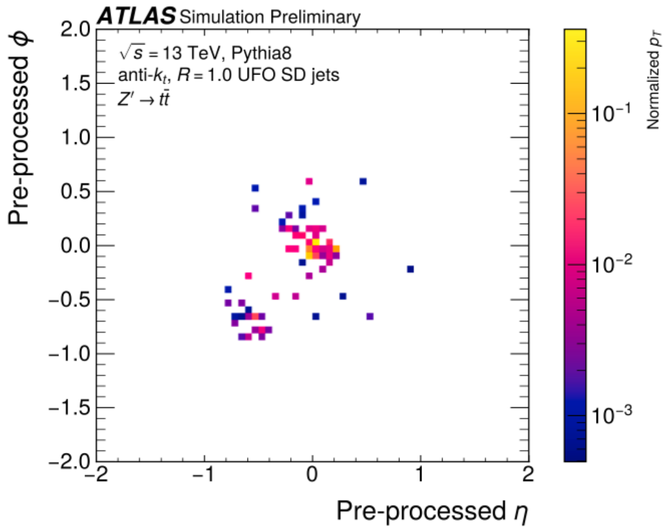
Boost Jet Tagging: Constituent Level Top Tagging

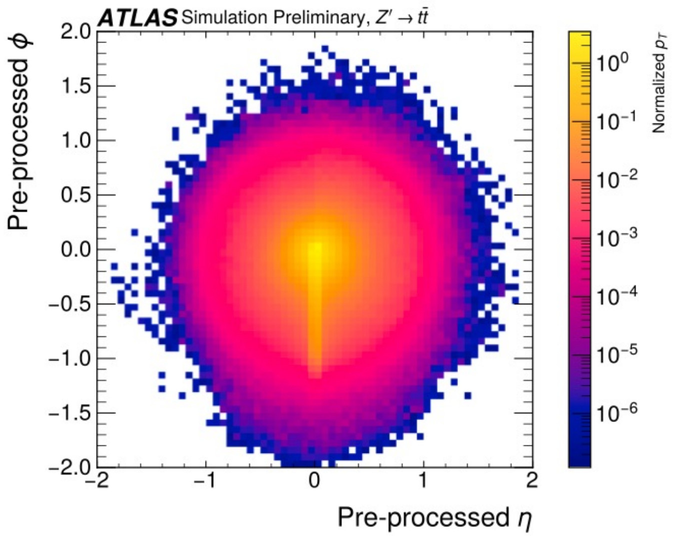
ATL-PHYS-PUB-2022-039

- Using 4-vector jet constituents as inputs: using lower level features for complex machine learning models

1. **Baseline DNN (hlDNN)** : trained with jet substructure variables
2. **Constituent level DNN**: jet constituent algorithm jets
3. **Energy Flow Network**: DeepSet
4. **Particle Flow Network**: EFN-like
5. **ResNet50**: large-scale CNN network
6. **ParticleNet**: graph neural network (GNN) which represents jets as a graph





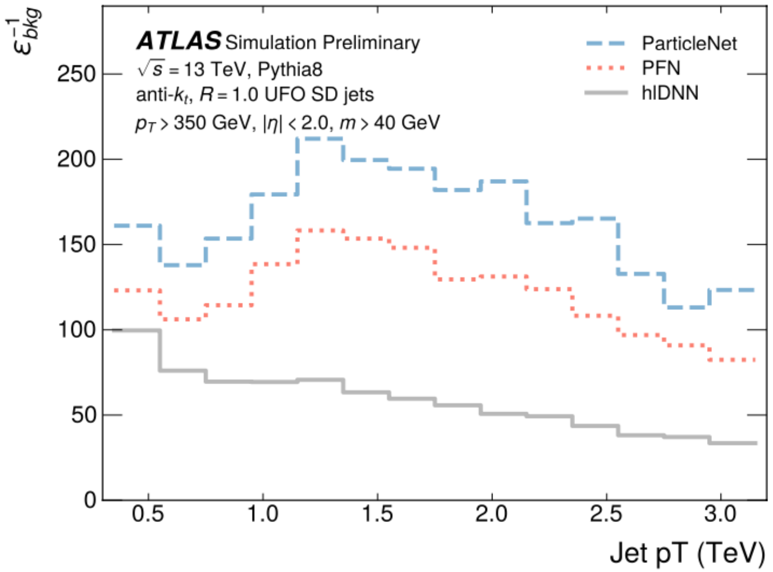


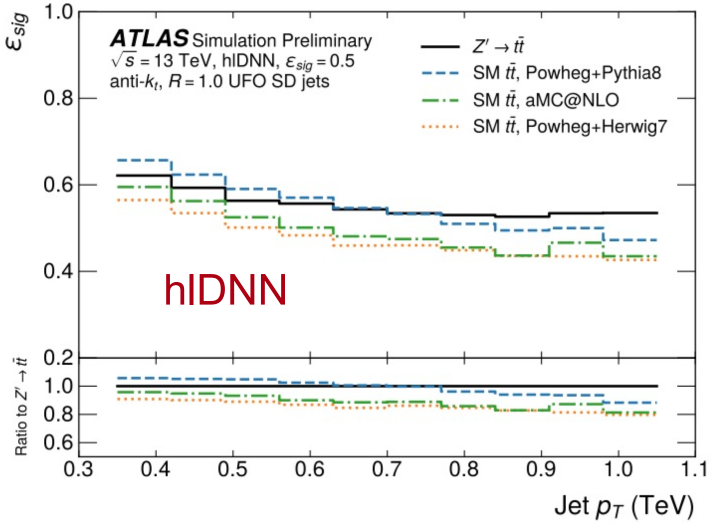
Three of the constituent-based taggers (**ParticleNet, PFN, DNN**) surpass the high-level-quantity-based tagger's (hlDNN) performance.

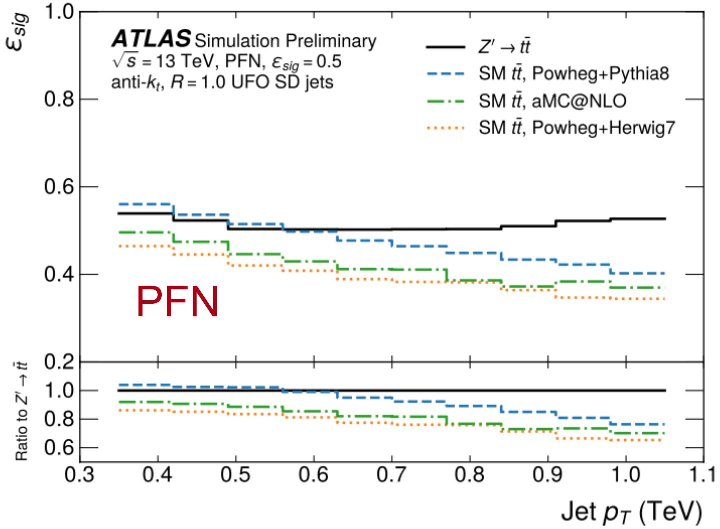
Boost Jet Tagging: Constituent Level Top tagging

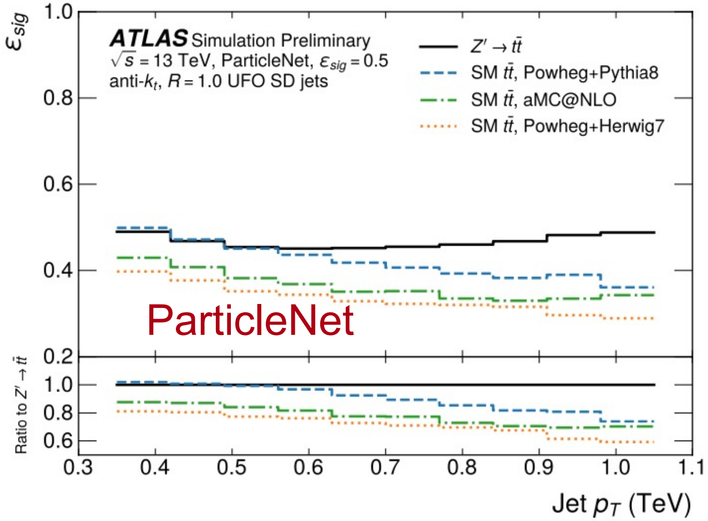
, ATL-PHYS-PUB-2022-039

- ParticleNet shows the best performance, ~2 times of hlDNN
- The constituent-based taggers' performance are better in the mid pT range of about 1-2 TeV compared to hlDNN
- Model dependence is also studied:
  - PFN and ParticleNet have sizeable dependence on QCD modelling
- Further study the impacts on modelling uncertainties for constituent level taggers planned



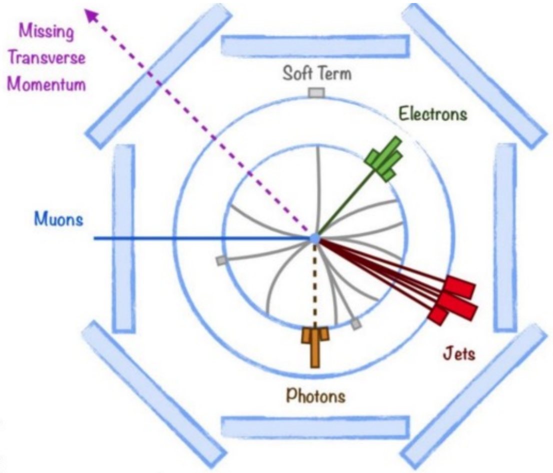






p_T^{miss} in ATLAS

ATL-PHYS-PUB-2021-025



- Missing transverse momentum (p_T^{miss}) is a key observable in many physics analyses using data recorded by the ATLAS detector
  - $p_T^{miss\ true}$: the missing momentum caused by Standard Model (SM) neutrinos dark matter
  - Fake p_T^{miss} :the missing transverse momentum comes from reconstruction
- Traditionally, p_T^{miss} is calculated by negative sum of objects: providing some working points to meet analysis requirements but not flexible

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

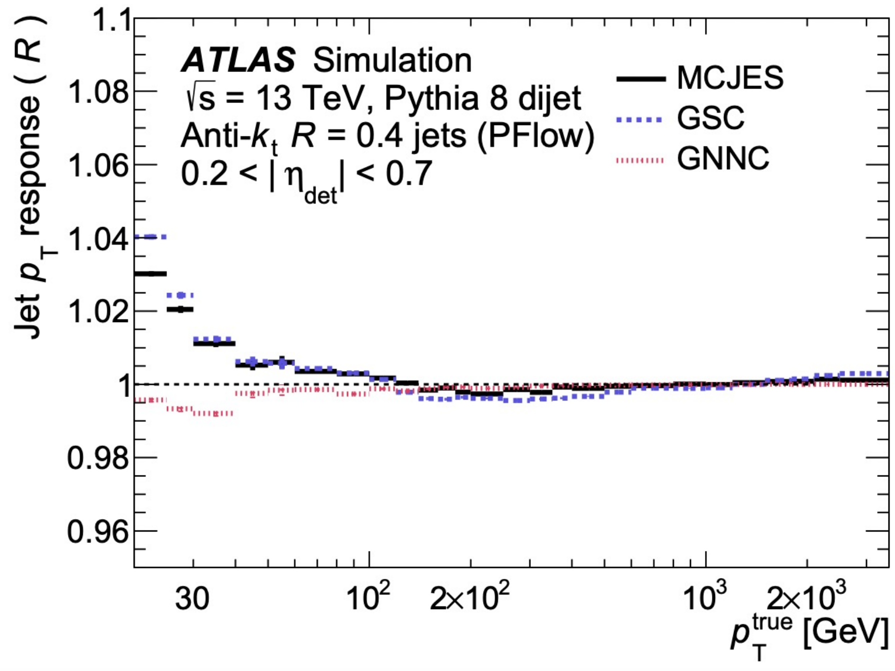
- **METNet:** the neural network to combine different working points and adjust them according to each event
  - To regress the $p_T^{miss\ true}$. Inputs are p_T^{miss} variables for each working point and variables which characterise the pile-up and topology of each event
  - Trained on ttbar and di-boson events such as WW, ZZ processes

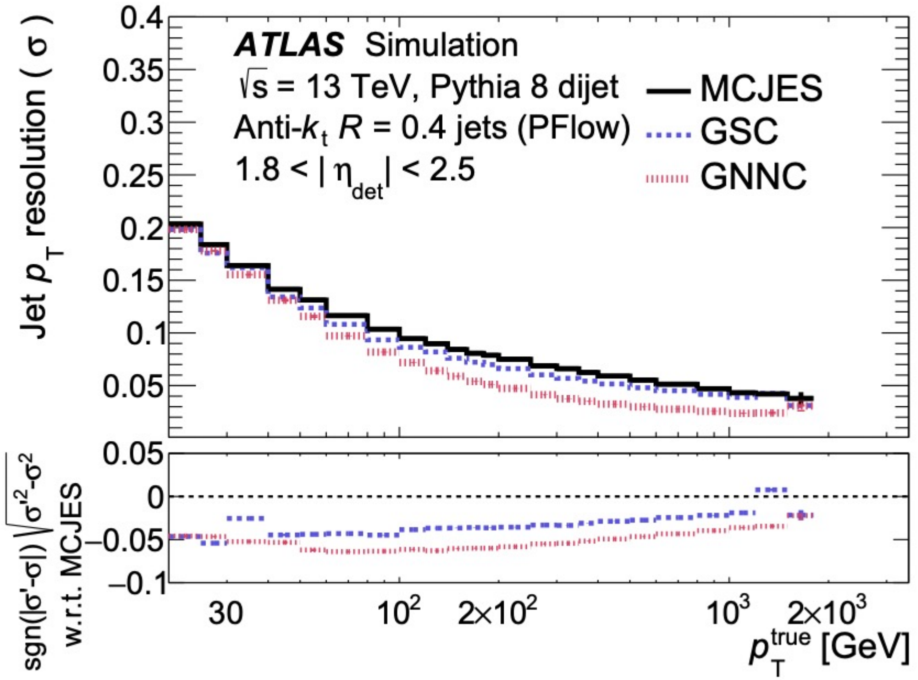
(More information from Peter’s talk)

The Global Neural Network Calibration

arXiv:2303.17312

- The jet energy is only partially observed in the detector so it needs to apply jet calibration to correct the jet energy response (more information about jet calibration in Peter's talk), typically using E and eta (MCJES)
- The global jet property calibration applies further corrections to jets
  - The global sequential calibration (GSC): using 6 observables to improve jet energy response
  - **The global neural network calibration (GNNC): DNN** trained with more variables
- Compared to GSC, GNNC has better jet pT closure and has over 15% improvements in jet pT resolution

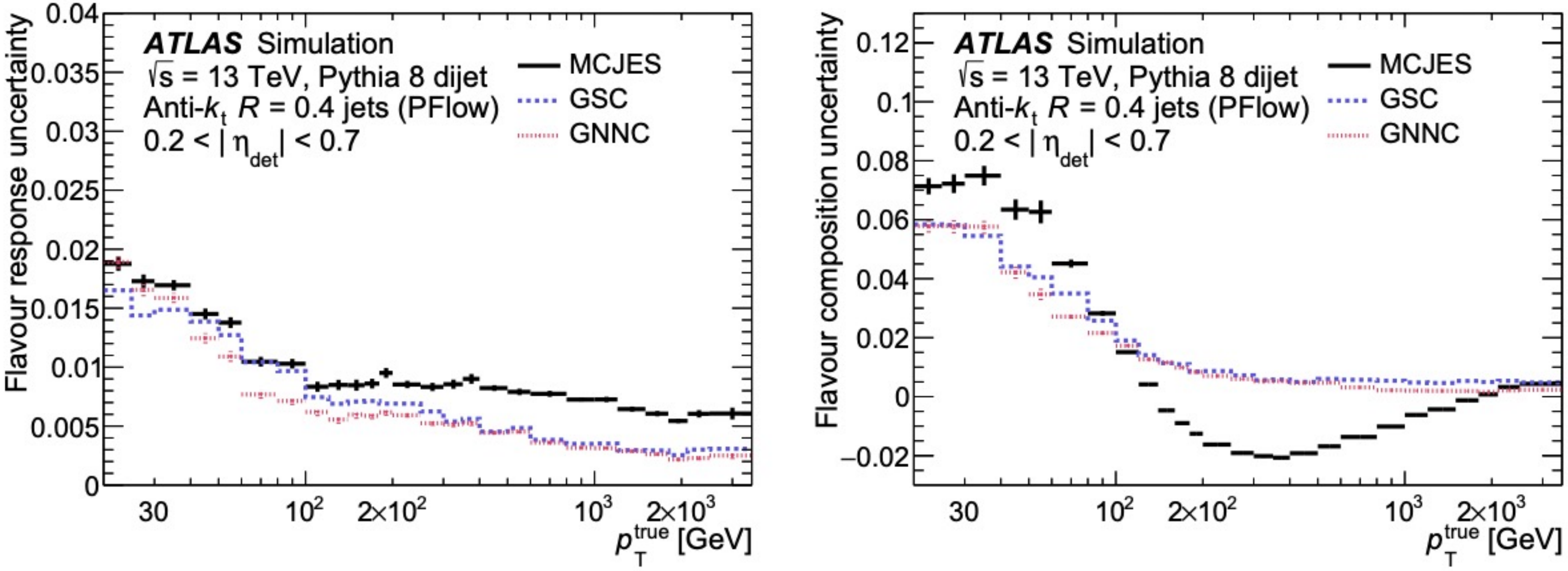




The Global Neural Network Calibration

arXiv:2303.17312

- The two flavour-dependence uncertainties in the JES are derived from simulation and account for relative flavour fractions and differing responses to quark- and gluon-initiated jets
- Comparing the flavour composition and flavour response uncertainties for the MCJES, GNNC and GSC.
- Both the GSC and GNNC can reduce these uncertainties
- GNNC provides a greater reduction, improvement ~15%-25% compared to GSC



Summary

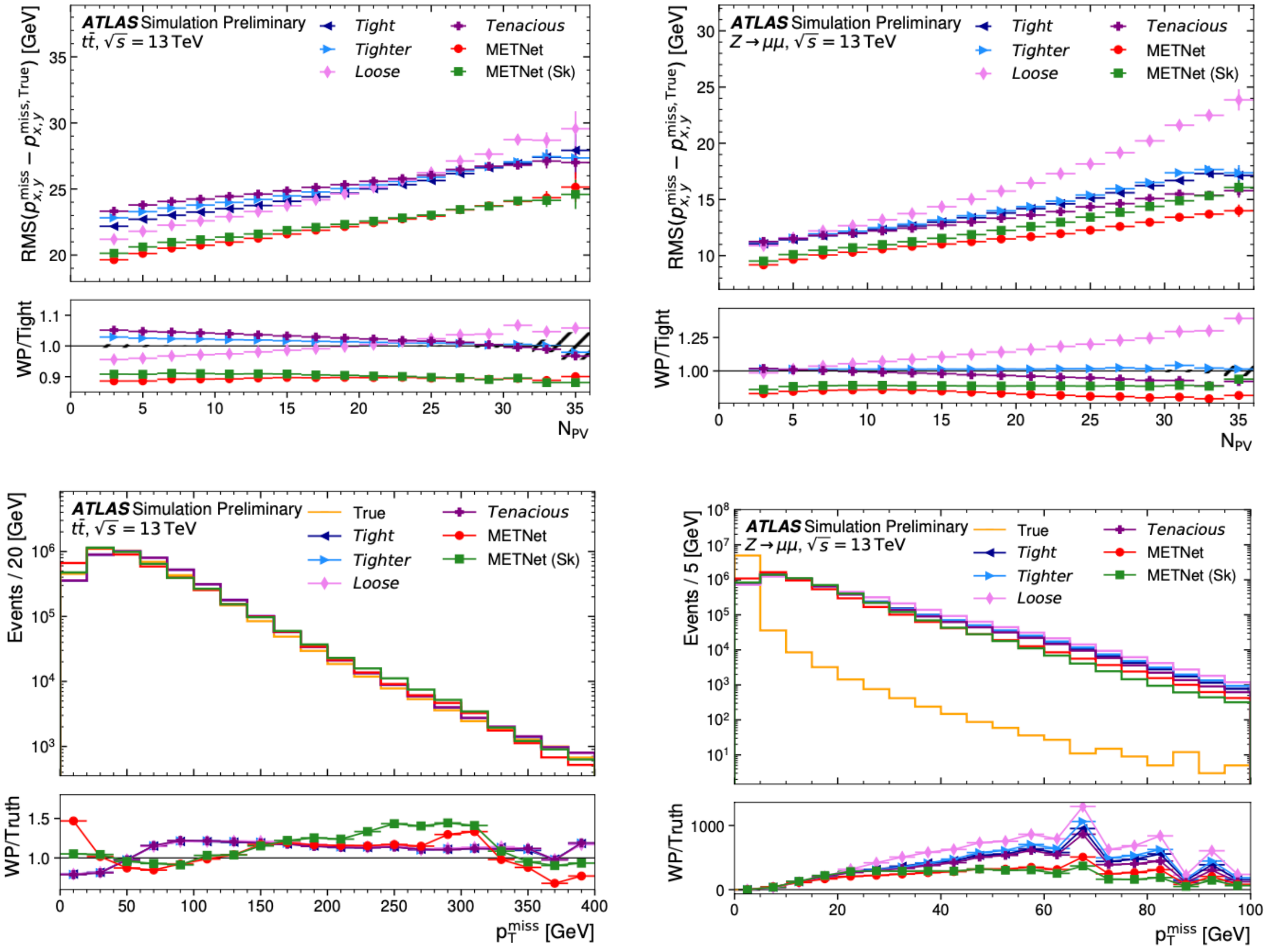
- Machine learning techniques are widely used in hadronic object studies in the ATLAS collaboration
  - Pion classification and calibration
  - Calibration and reconstruction:
    - METNet
    - The Global Neural Network Calibration
  - Tagging:
    - boost top and W/Z tagging
    - Constituent-level top tagging
- The progress in hadronic object study with machine learning/AI can greatly improve the performance for BSM searching and SM measurements
- Run3 has started and there is more chance to develop new machine learning techniques and test their performance in the ATLAS analyses

Backups

METNet Performance

ATL-PHYS-PUB-2021-025

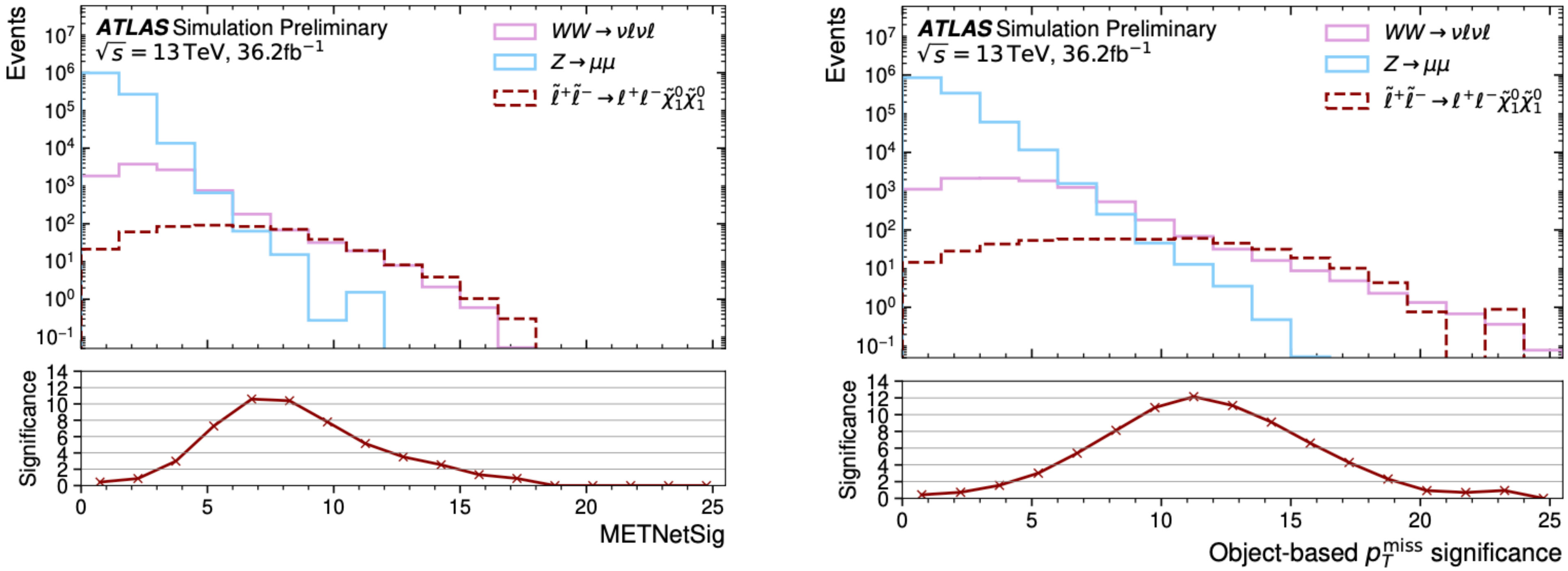
- Compared to previous working points, the resolution is **improved** by METNet
- MET Net can handle **different topologies**: Good performance on Z->mu mu events which is not included in the training
- Good distribution bias: using a special loss function, Sinkorn loss can reduce the bias effect
- It shows the potential to improve p_T^{miss} resolution with ML methods



METNetSig Performance

ATL-PHYS-PUB-2021-025

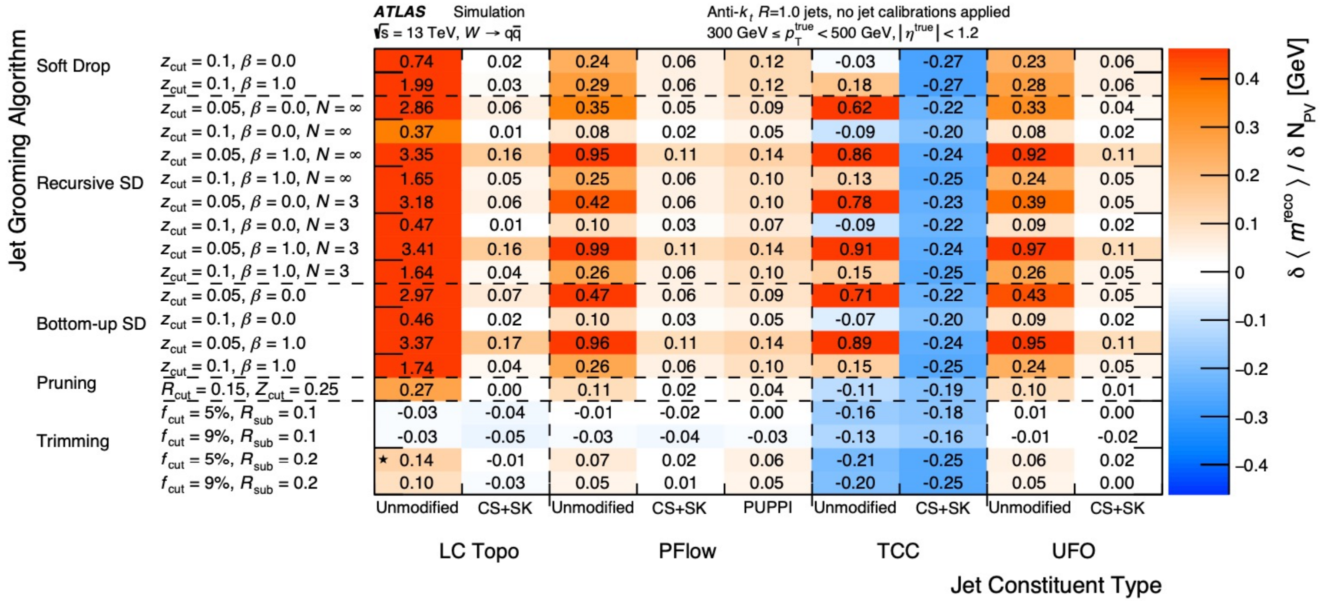
- METNetSig: extending METNet to calculate p_T^{miss} significance in machine learning method and this is not limited by separate measurements of each object's resolution
- It can be used to discriminate real and fake p_T^{miss}
- Similar performance to ALTAS object-based p_T^{miss} significance



Jet Reconstruction at ATLAS

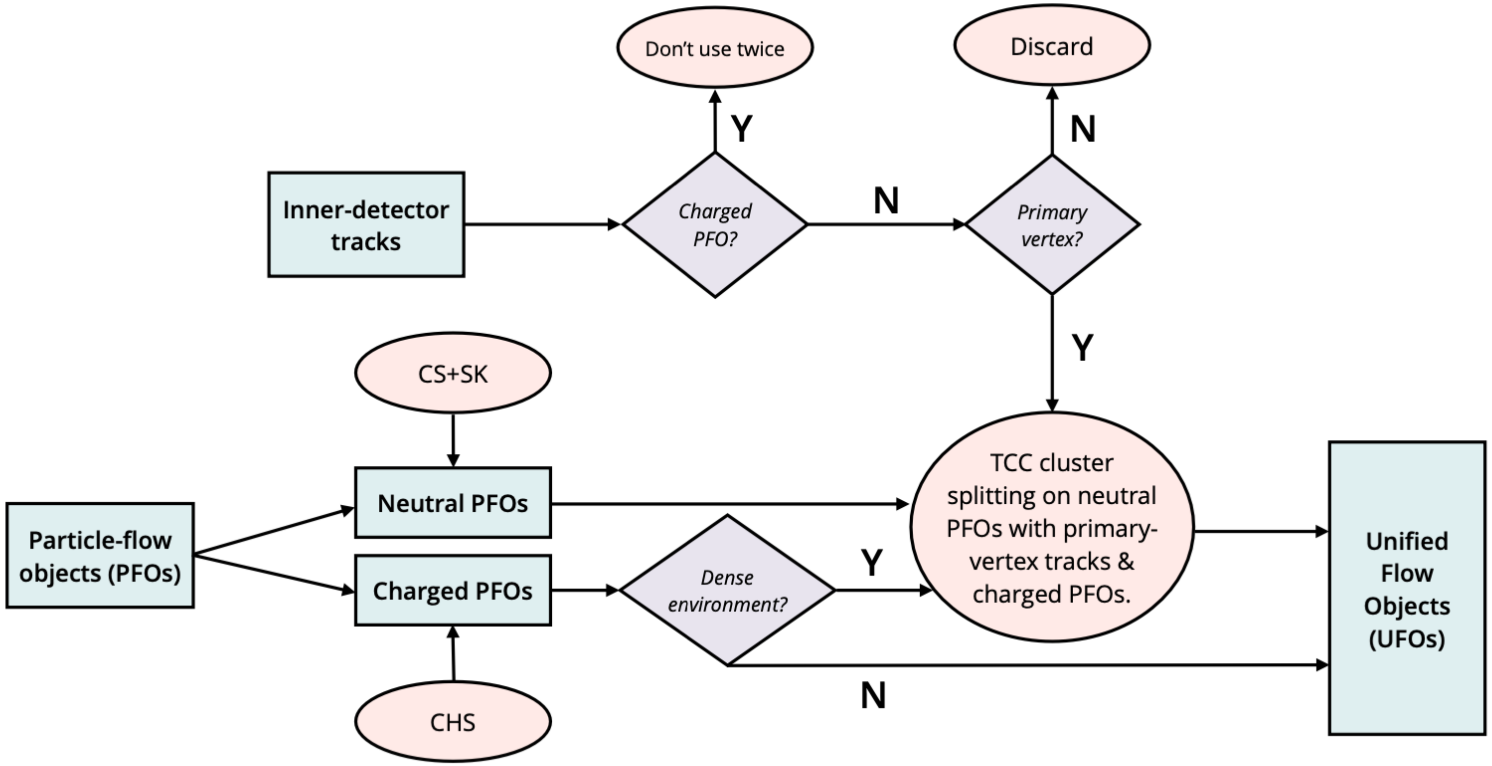
arXiv:2009.04986

	Algorithm	Abbreviation	Settings
Jet input objects	Topological Clusters	Topoclusters	N/A
	Particle-Flow	PFlow	N/A
	Track-CaloClusters	TCCs	N/A
	Unified Flow Objects	UFOs	N/A
Pile-up mitigation algorithms	Constituent Subtraction	CS	$A_g = 0.01$ $\Delta R_{\max} = 0.25$ $\alpha = 0$
	Voronoi Subtraction (*)	VS	N/A
	SoftKiller	SK	$\ell = 0.6$
	Pile-up Per Particle Identification	PUPPI	$R_{\min} = 0.001$ $R_0 = 0.3$ $a = 200$ MeV $b = 14$ MeV
Jet grooming algorithms	Soft-Drop	SD	$z_{\mathrm{cut}} = 0.1$ $\beta = 0, 1, 2(*)$
	Bottom-up Soft-Drop	BUSD	$z_{\mathrm{cut}} = 0.05, 0.1$ $\beta = 0, 1, 2(*)$
	Recursive Soft-Drop	RSD	$z_{\mathrm{cut}} = 0.05, 0.1$ $\beta = 0, 1, 2(*)$ $N = 3, 5(*), \infty$
	Pruning	N/A	$z_{\mathrm{cut}} = 0.15$ $R_{\mathrm{cut}} = 0.25$
	Trimming	N/A	$f_{\mathrm{cut}} = 5\%, 9\%$ $R_{\mathrm{sub}} = 0.1, 0.2$



Unified Flow Object Reconstruction Algorithm

Arxiv:2009.04986



Substructure Variables

ATL-PHYS-PUB-2021-028

ATL-PHYS-PUB-2021-029

- Variables used in training for W/Z and top taggers in UFO jets

W/Z tagger (NN/ANN)		**Top tagger (DNN)**	
$D_2,\ C_2$	Energy correlation ratios	$\tau_1,\ \tau_2,\ \tau_3,\ \tau_4$	N-subjettiness
τ_{21}	N-subjettiness	$\sqrt{d_{12}},\ \sqrt{d_{23}}$	Splitting scales
R_2^{FW}	Fox-Wolfram moment	$\mathrm{ECF}_1,\ \mathrm{ECF}_2,\ \mathrm{ECF}_3$	Energy correlation (EC) functions
$\mathcal{P}$	Planar flow	$C_2,\ D_2$	EC ratios
a_3	Angularity	$L_2,\ L_3$	Generalised EC ratios
A	Aplanarity	Q_{W}	Invariant mass / virtuality
Z_{cut}	$Z-$Splitting scales	T_{M}	Thrust major
$\sqrt{d_{12}}$	$d-$Splitting scales		
$Kt\Delta R$	k_t-subjet ΔR		
n_{trk}	number of tracks		

Constituent Level Top Tagger Tagging

ATL-PHYS-PUB-2022-039

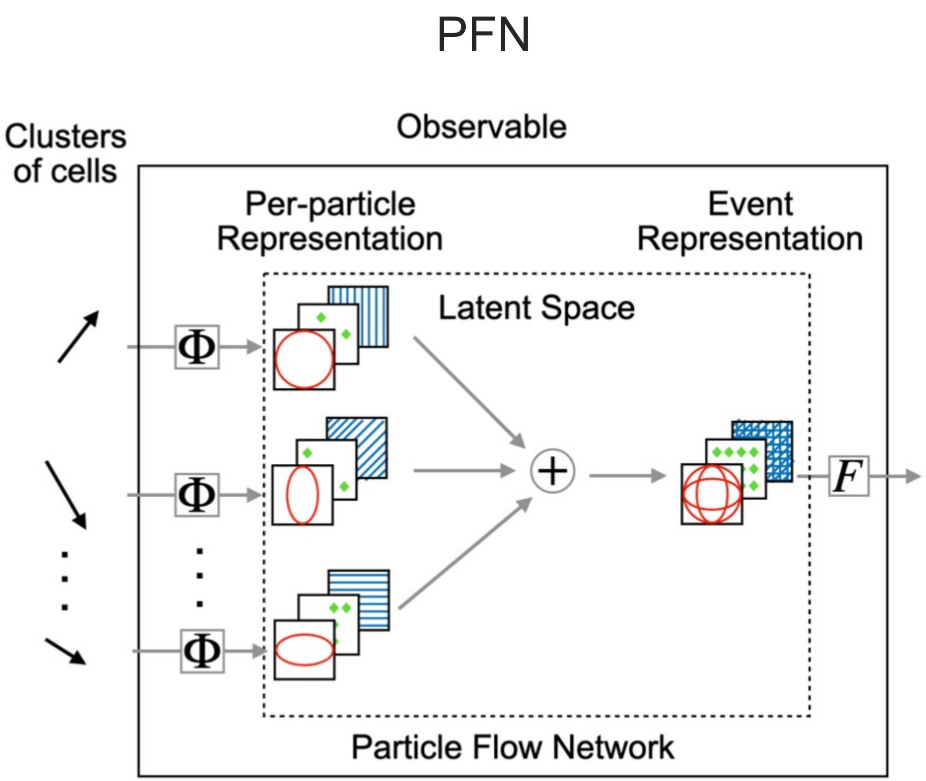
Model	Hyper-parameters
hlDNN	Hidden Layers: 5 Nodes per Layer: 180 *Activation Functions: ReLU* *Kernel Initialization: glorot uniform* Learning Rate: 4×10^{-5} Batch Size: 250 Batch Normalization: not used
DNN	Hidden Layers: 5 Nodes per Layer: 400 *Activation Functions: ReLU* *Kernel Initialization: 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

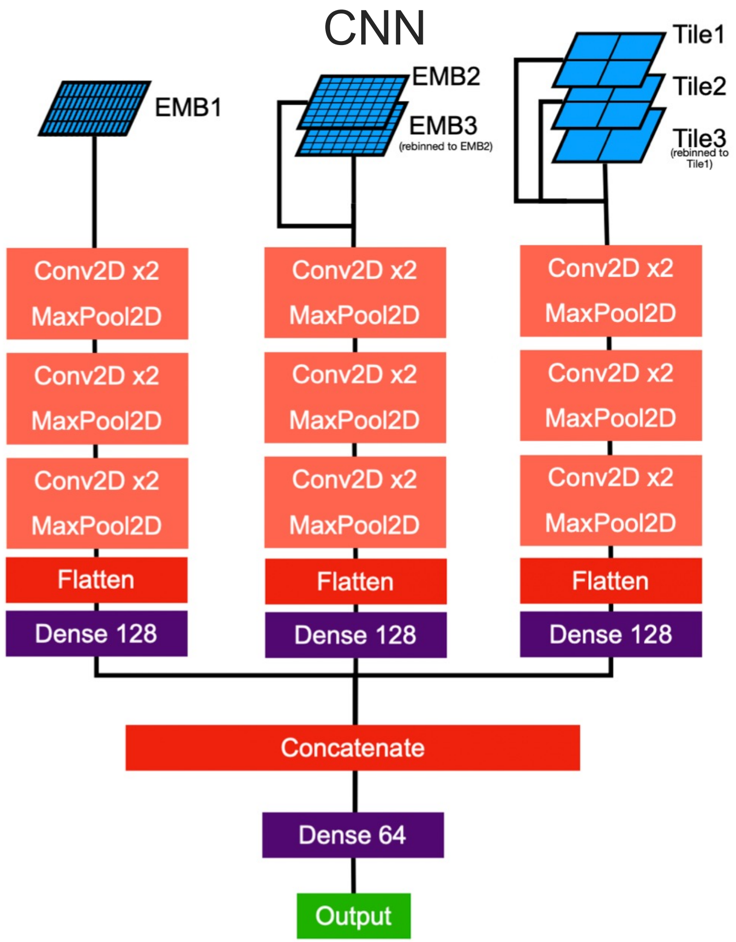
Model	AUC	ACC	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.5$	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.8$	# Params	Inference Time
ResNet 50	0.885	0.803	21.4	5.13	1,486,209	9 ms
EFN	0.901	0.819	26.6	6.12	1,670,451	4 ms
hlDNN	0.938	0.863	51.5	10.5	93,151	3 ms
DNN	0.942	0.868	67.7	12.0	876,641	3 ms
PFN	0.954	0.882	108.0	15.9	689,801	4 ms
ParticleNet	0.961	0.894	153.7	20.4	764,887	38 ms

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

ML Algorithm in Jet Tagging

ATL-PHYS-PUB-2022-039

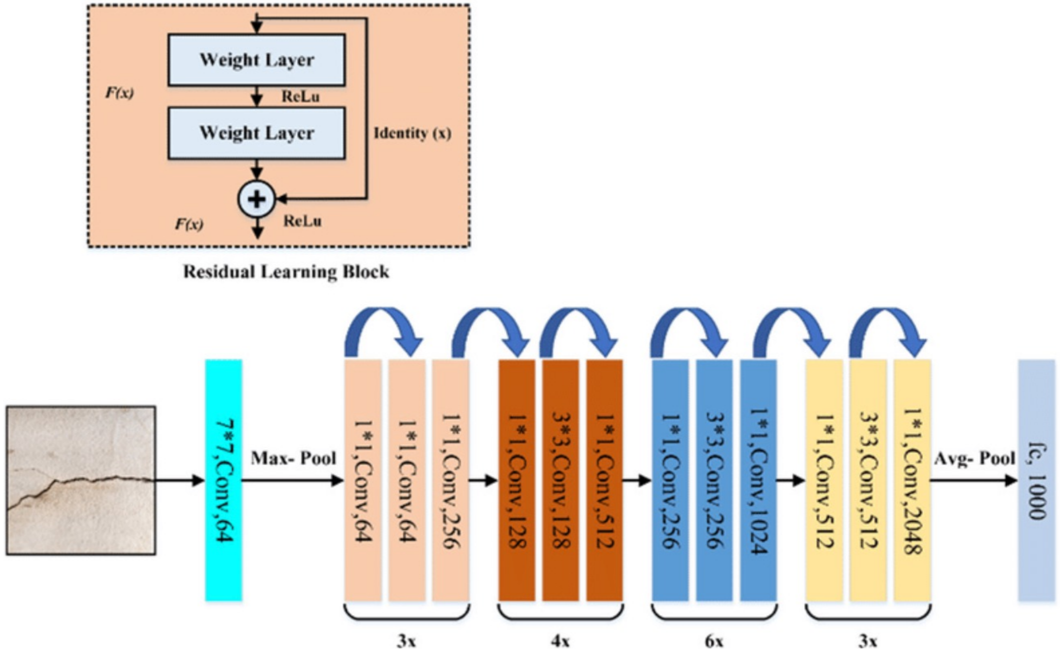




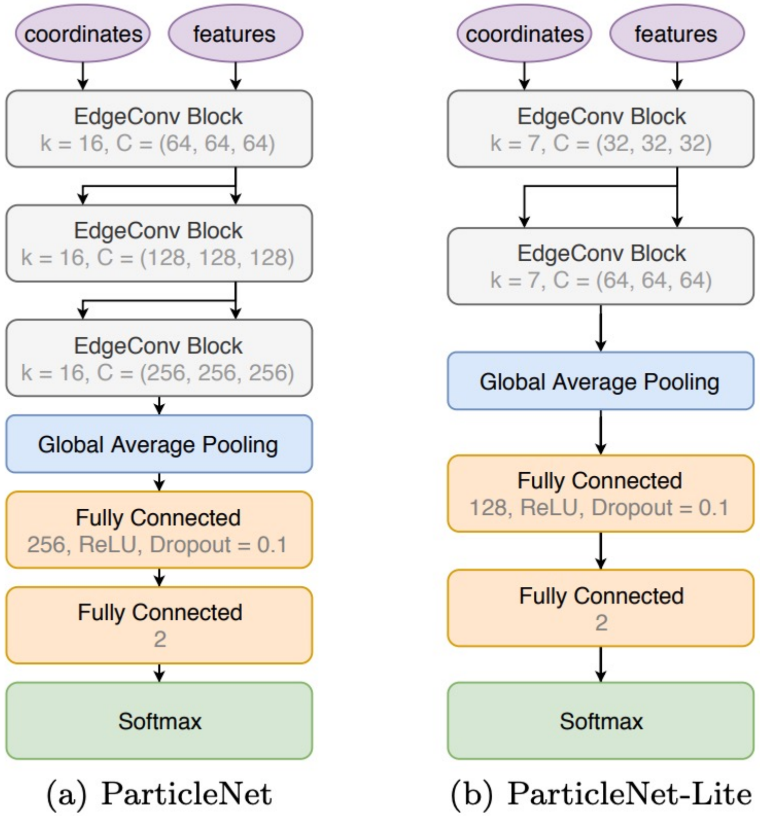
ML Algorithm in Jet Tagging

ATL-PHYS-PUB-2022-039

ResNet



Particle Net



The Global Neural Network Calibration

JETM-2022-01

- Variables for training network

Calorimeter	$f_{\text{LAr}0-3*}$	The E_{frac} measured in the 0th-3rd layer of the EM LAr calorimeter
	$f_{\text{Tile}0*-2}$	The E_{frac} measured in the 0th-2nd layer of the hadronic tile calorimeter
	$f_{\text{HEC},0-3}$	The E_{frac} measured in the 0th-3rd layer of the hadronic end cap calorimeter
	$f_{\text{FCAL},0-2}$	The E_{frac} measured in the 0th-2nd layer of the forward calorimeter
	$N_{90\%}$	The minimum number of clusters containing 90% of the jet energy
Jet kinematics	$p_{\text{T}}^{\text{JES}}$ *	The jet p_{T} after the MCJES calibration
	η^{det}	The detector η
Tracking	w_{track}*	The average p_{T}-weighted transverse distance in the η-ϕ plane between the jet axis and all tracks of $p_{\text{T}} > 1$ GeV ghost-associated with the jet
	N_{track}*	The number of tracks with $p_{\text{T}} > 1$ GeV ghost-associated with the jet
	f_{charged}*	The fraction of the jet p_{T} measured from ghost-associated tracks
Muon segments	N_{segments}*	The number of muon track segments ghost-associated with the jet
Pile-up	μ	The average number of interactions per bunch crossing
	N_{PV}	The number of reconstructed primary vertices

Table 1: List of variables used as input to the GNNC. Variables with a * correspond to those that are also used by the GSC.