Boosted W/Z boson and top tagging in ATLAS

Qibin LIU

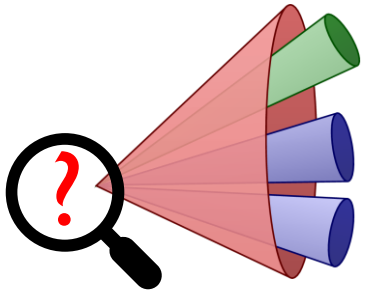
Tsung-Dao Lee Institute
Shanghai Jiao Tong Univ.

On behalf of the ATLAS collaboration

ICNFP2022

Introduction

- Differentiating between hadronically decaying massive particles is increasingly important for HEP
  - Extensive interest from di-boson, di-top resonance, SUSY and especially more on highly boosted topology
- Variety of boosted tagging algorithms on large-radius jet are actively developed and studied in ATLAS
  - Exploit all the information of jet kinematics and substructure with combinational cut and machine learning
- Latest work of boosted W/Z and Top tagging on Unified-Flow-Objects (UFO) large-*R* jet to be presented
  - Mass decorrelation, model dependence and comparison with previous also discussed



Eur. Phys. J. C 81 (2021) 4, 334

Large-Radius Jet Reconstruction

- Highly boosted and collimated jets from W/Z/Top reconstructed with large-*R* jet (R=1.0)
- LCTopo large-*R* jet: **L**ocal **C**ell reweighting **Topo**logical cluster → deployed in ATLAS during Run-2
  - Use information of energy deposition in calorimeter solely
- UFO large-*R* jet: **U**nified-**F**low-**O**bjects : New! Talk from Romain, 7th Sept.
  - Particle-Flow (PFO): calo. cluster + expected energy deposition from track
  - Track-CaloCluster (TCC): calo. cluster + angular information from inner tracker
  - Optimized pileup-mitigation and jet grooming [backup]
  - Better performance over entire p_T: e.g. good recon. of mass and jet substructure variables

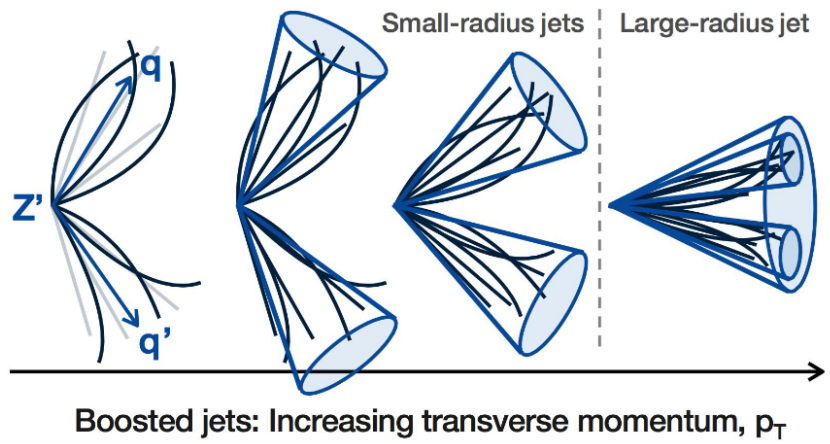
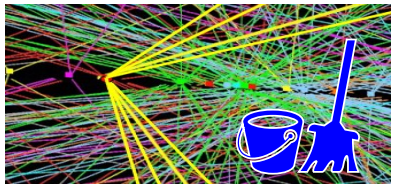


Image credit: Juan Varela

Grooming:
clean soft part in jet
(not from hard scatter)



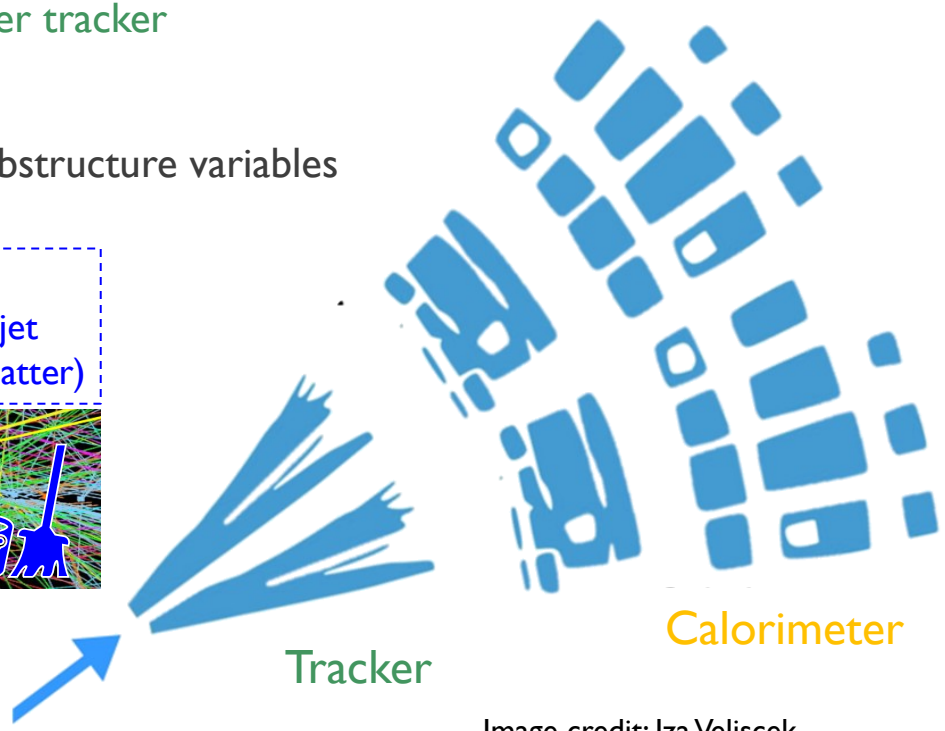
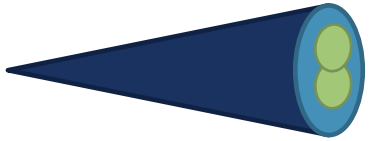


Image credit: Iza Veliscek

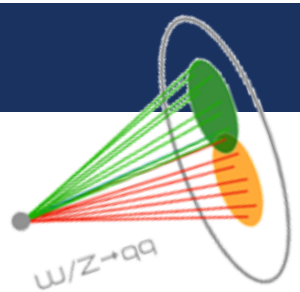
W/Z TAGGING ON UFO JET

ATL-PHYS-PUB-2021-029



$\tau_i, \sqrt{d_{ij}}, ECF, \ldots$

W/Z Jet Definition

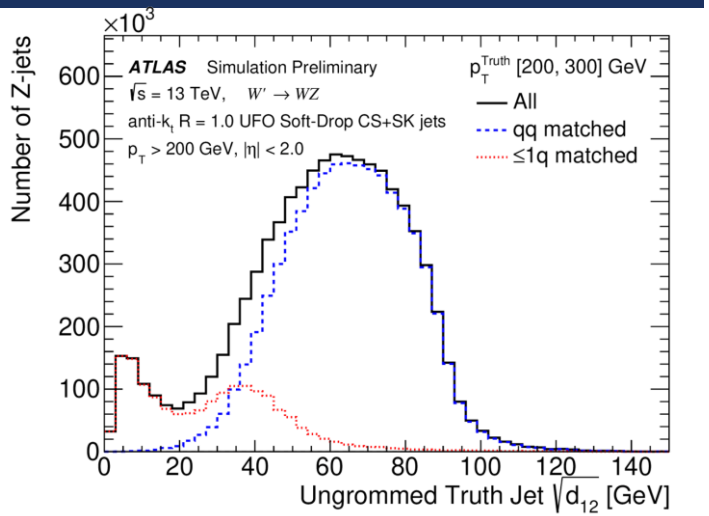


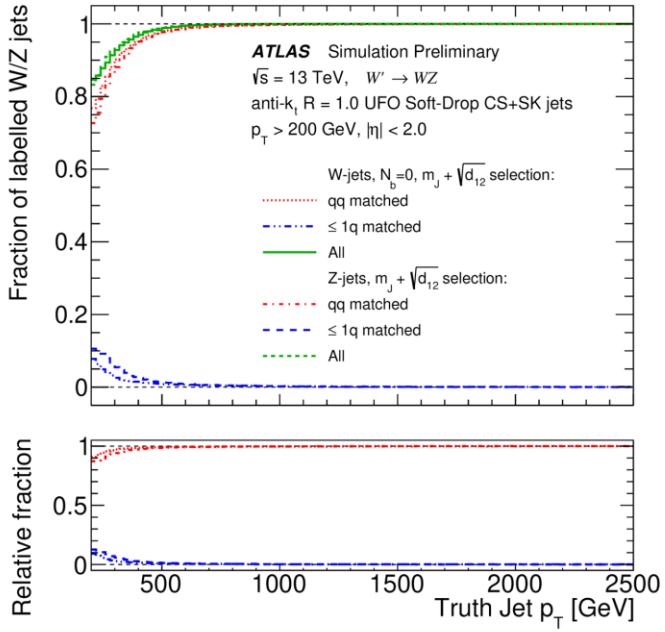
- Truth W/Z jets:
  - Recon. from stable truth particle w/ Anti-*Kt* R=1.0 (~~grooming~~)
  - Match to a truth W/Z boson with $\Delta R < 0.75$
  - 0 truth b-hadron associated (for truth W)
  - $m_J > 50GeV$
  - p_T depedent $\sqrt{d_{12}}$ requirement [backup]

Angular distance:
$\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$

d-Splitting scale $\sqrt{d_{ij}}$:
pT weighted distance (i^{th} / j^{th} constituents)

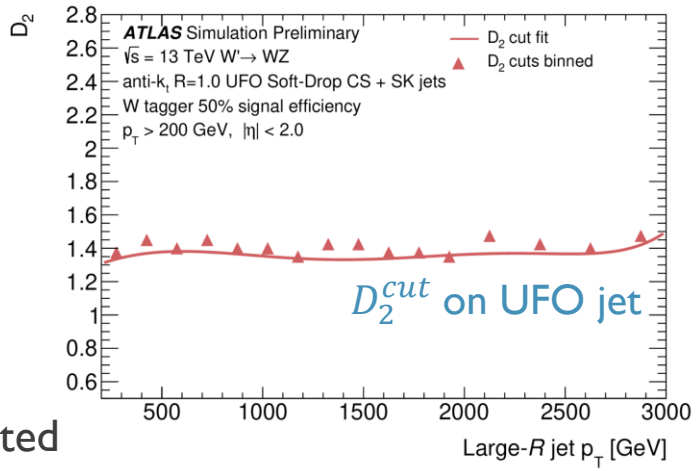
- Signal (W/Z jet): truth W/Z jet matched UFO jet in $W' \rightarrow WZ$
- Background (QCD jet): UFO jet from multijet (not match)
- High contamination of two quarks:
  - More than 98% for W jet and 96% for Z jet at high p_T
- Robust labeling for different topology and generator

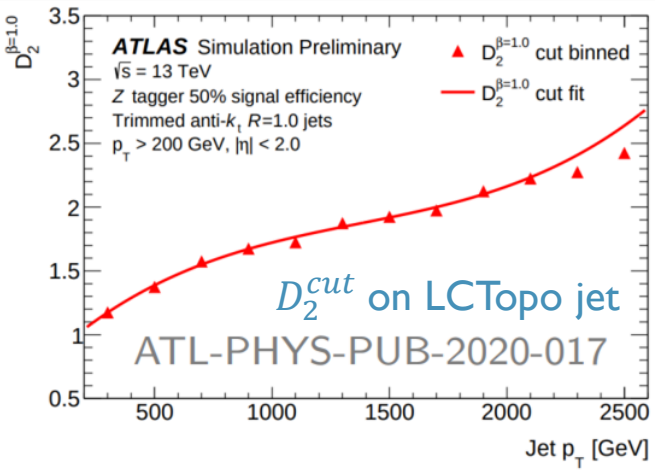




Cut-based: 3-var/D_2 Tagger

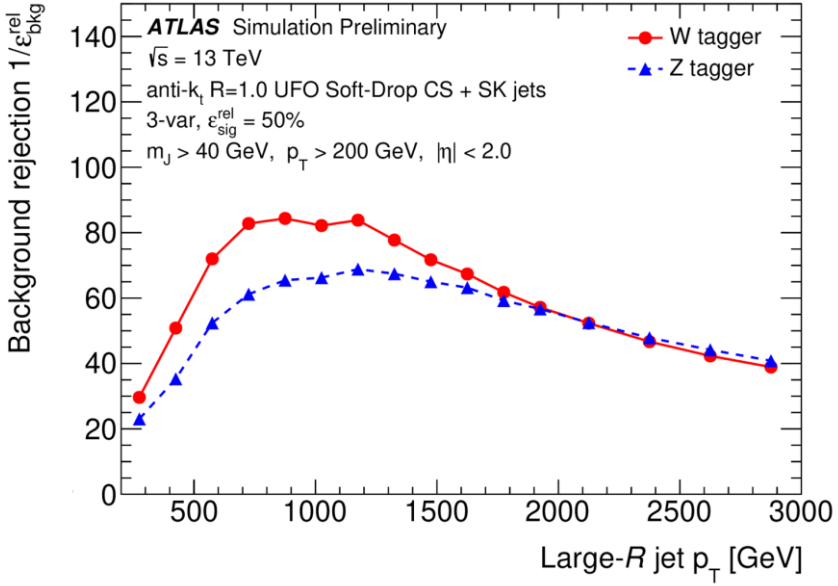
- Rectangular cut on substructures variables:
  - $m_J \in [m_J^{low}, m_J^{high}]$: window of jet recon. mass
  - $D_2 < D_2^{cut}$: energy correlation ratio [ref][backup]
  - $N_{trk} < N_{trk}^{cut}$: No. of inner-detector tracks associated





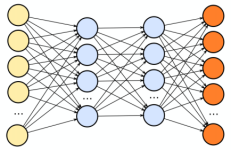
- Cut value optimized for best $1/\varepsilon_{bkg}$ at flat $\varepsilon_{sig} = 50\%$ or 80% ("50WP"/"80WP")
- D_2^{cut} shows less dependence of p_T compared with LCTopo

  → good angular resolution of UFO jet
- Up to 98% rejection of QCD jet for W/Z tagger at $\varepsilon_{sig} = 50\%$

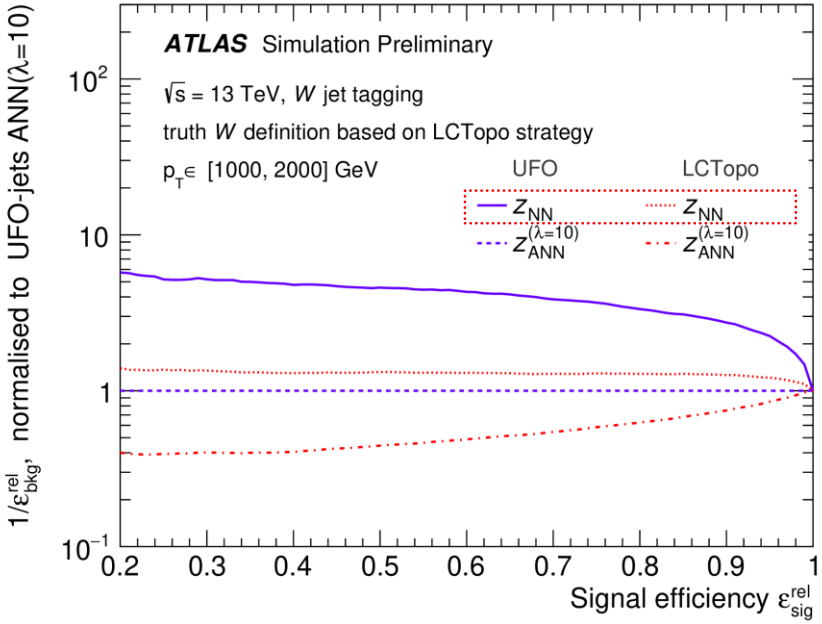


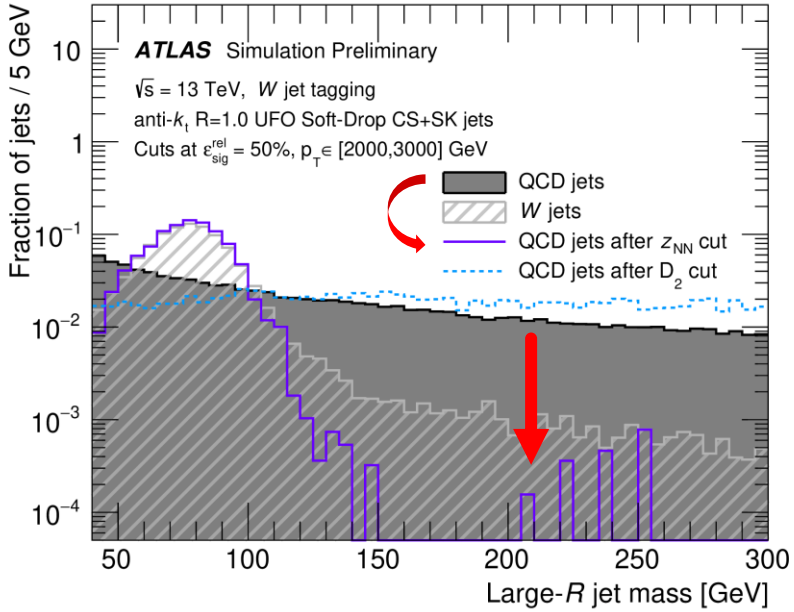
Deep Neural Network: DNN Tagger

- Deep neural network (DNN) trained to tag W-jet
  - 3x32 fully connected architecture combine 10 inputs
- Bkg. rejection increased by 2~4 times w.r.t the DNN tagger on LCTopo
  - Improved reconstruction of substructure variables with UFO jets
- Tagging efficiency of DNN tagger highly depend on mass
  - Modified the pass-tag background shape → **mass sculpting effect**

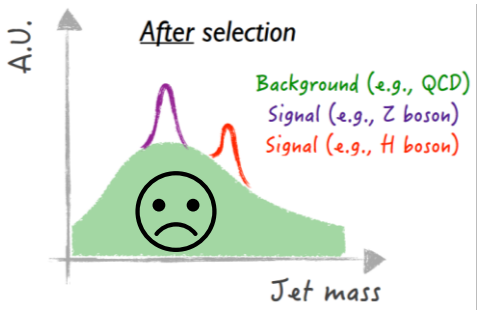


Variable	Description
D_2, C_2	Energy correlation ratios
τ_{21}	N-subjettiness
R_2^{FW}	Fox-Wolfram moment
$\mathcal{P}$	Planar flow
a_3	Angularity
A	Aplanarity
Z_{cut}, $\sqrt{d_{12}}$	Splitting scales
$Kt\Delta R$	k_t-subjet ΔR



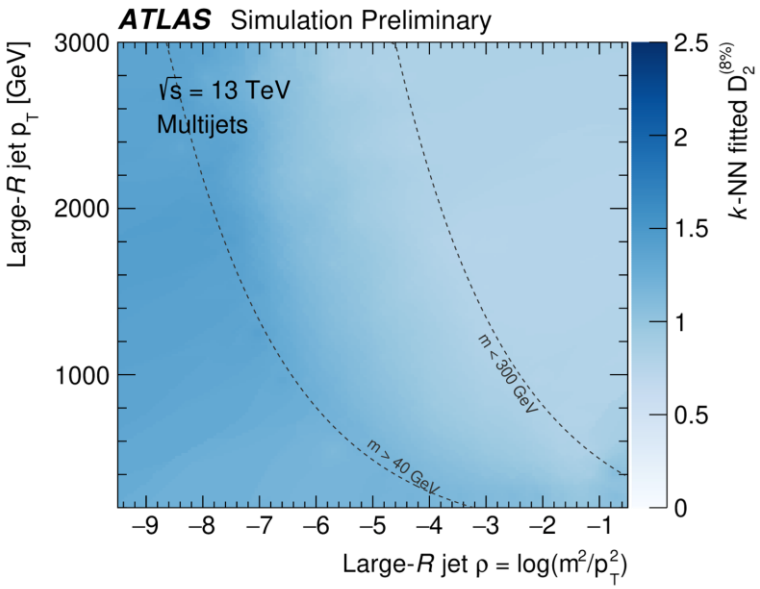


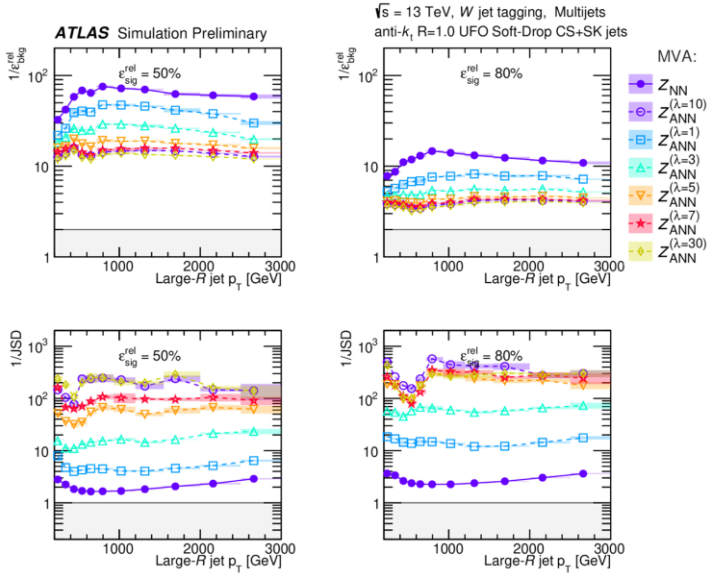
Mass Decorrelation

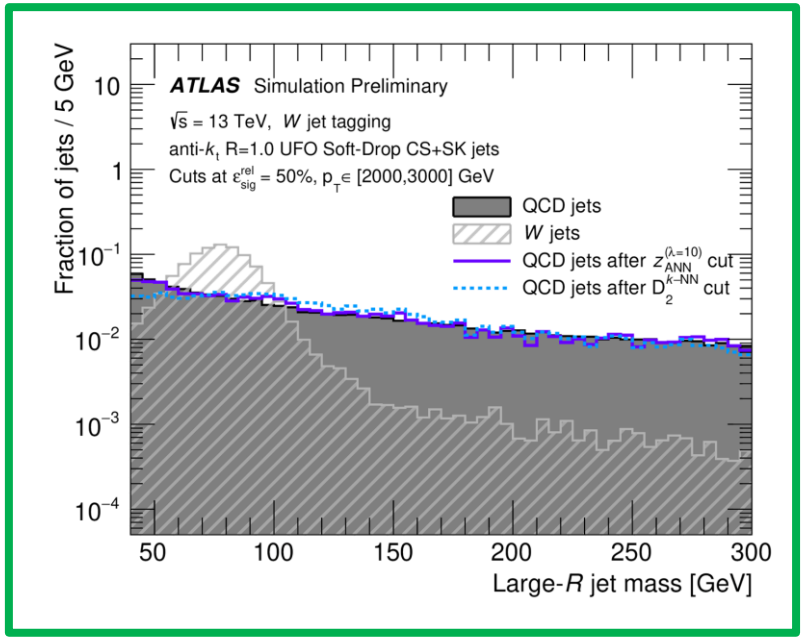


From H.Qu

- Mass sculpting effect caused by m_J dependence of tagging efficiency esp. for DNN tagger
  - Difficulty for analyses using m_J sideband for background estimation or bump-hunting
- Analytical method based on k-nearest neighbor regression (**k-NN**) to decorrelate $D_2^{cut} \rightarrow D_2^{kNN-cut}$
  - **Subtract** the fully correlated part : $D_2^{8\%}(m_J, p_T)$ regressed at fixed ε_{bkg}=8% (when ε_{sig}=50%)
- Adversary network (**ANN**) introduced to train mass decorrelated DNN tagger → ANN tagger
  - Make **adversary** network harder to infer the m_J from DNN output with the cost of weaker tagging

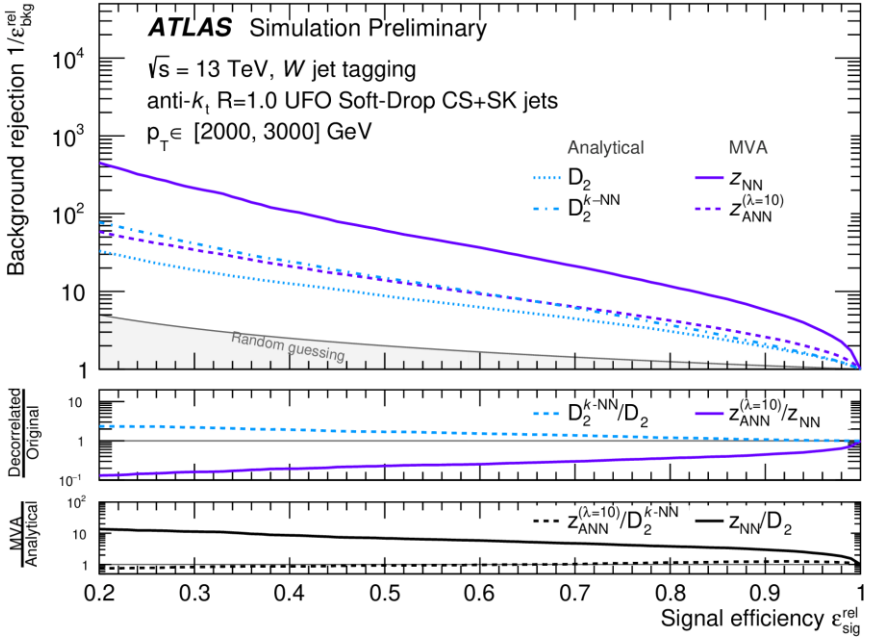


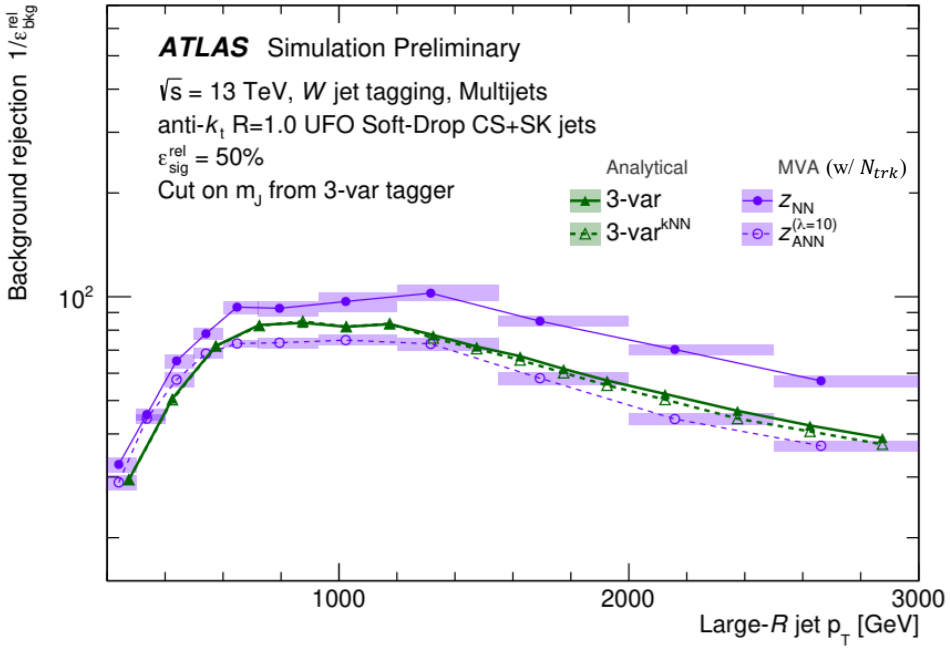




Performance

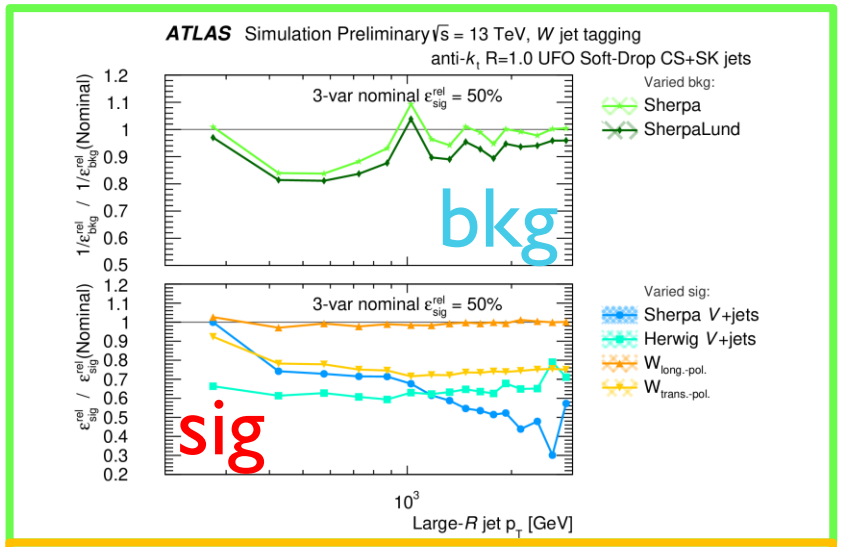
- DNN tagger shows up to 10 times better bkg. rejection than cut-based D_2 tagger
- Mass decorrelated DNN (ANN) shows similar performance as cut-based D_2 -kNN
  - DNN benefits just from the information (correlation) of mass
- At around $m_{W/Z}$, the performance of different tagger gets closer
  - N_{trk} found to be effective separating quark/gluon and included in the updated DNN tagger

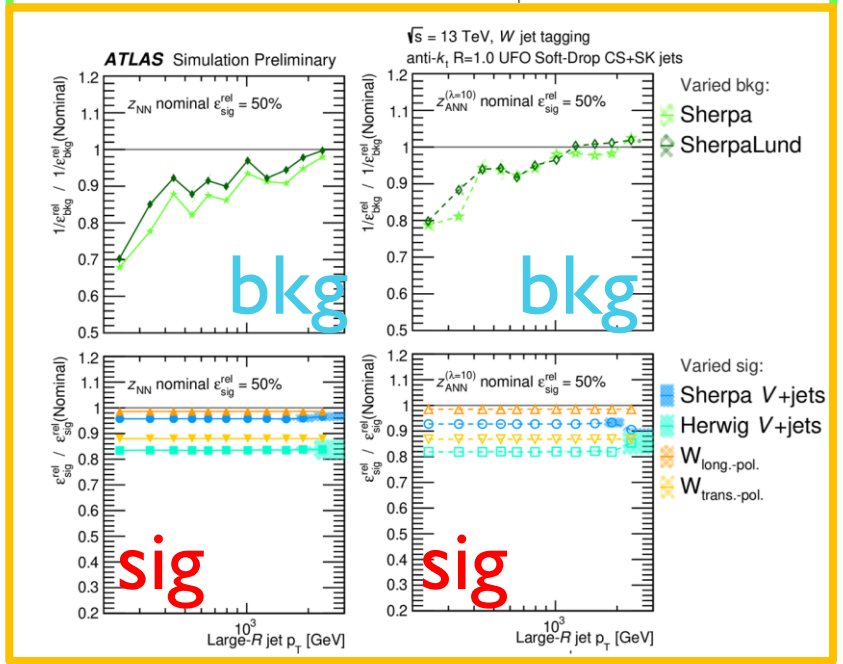




Model Dependence of W/Z Tagger

- Robustness against different physics modelling:
  - Especially the soft QCD contribution
  - Impact on the calibration and systematics for analysis
- Signal modelling:
  - 20%/15% drop for ***3-var***/**DNN** on trans. polarized W (SM V+jets)
  - ~10% lower for Herwig v.s. Sherpa due to LO/NLO diff. of m_J
- Background modelling:
  - <5% from hadronization
  - 20%~10% from generator and parton shower
- Overall 20% variation of both ε_{sig} and ε_{bkg}
  - Compatible with the taggers on LCTopo

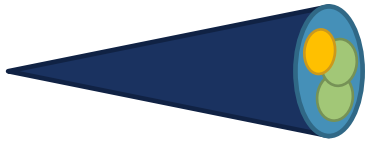




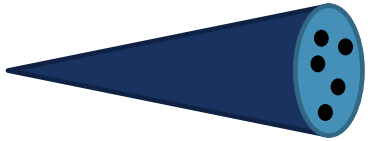
TOP TAGGING ON UFO JET

ATL-PHYS-PUB-2021-028

ATL-PHYS-PUB-2022-039



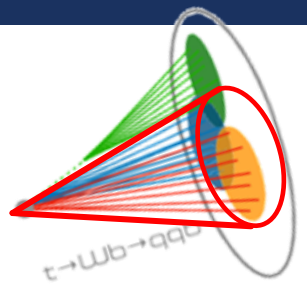
$\tau_i, \sqrt{d_{ij}}, ECF, \ldots$

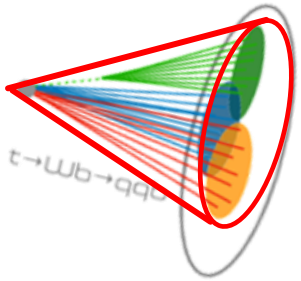


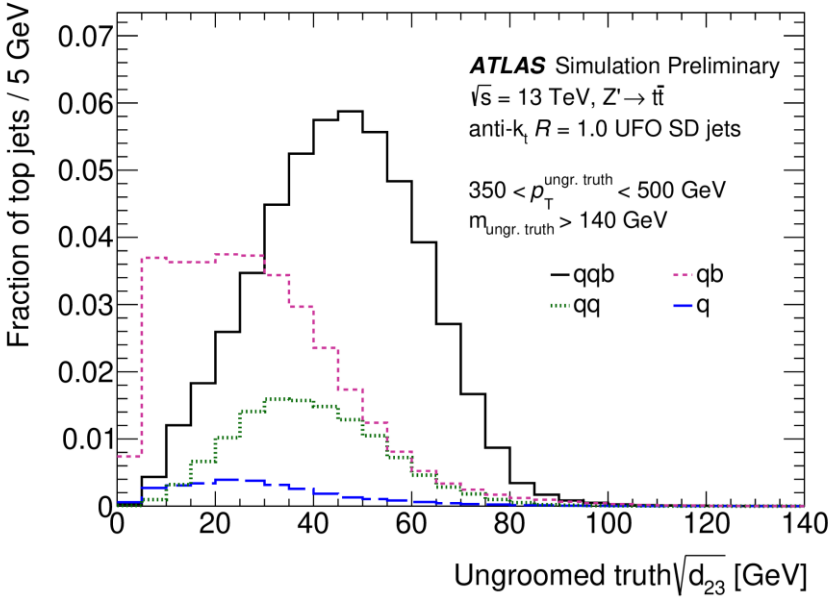
$\overrightarrow{p_i}$

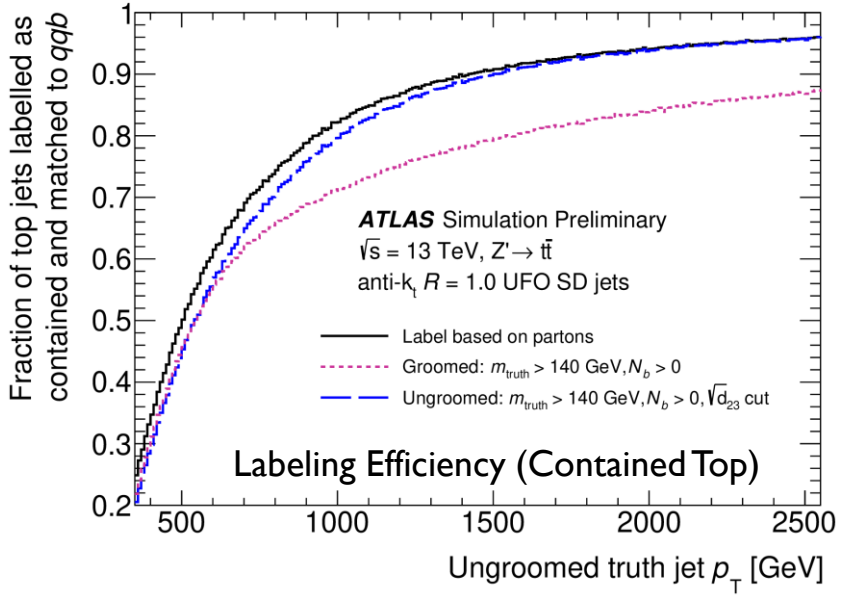
Top Jet Definition

- Selection: p_T>350GeV, $|\eta|$<2.0, m_J >40GeV and $N_{constituent}$ ≥3
- Signal: truth labelled top jet in $Z' \to t\bar{t}$
  - "**Inclusive Top**": contain any part of the hadronic top decay ($\Delta R < 0.75$ matched to top)
  - "***Contained Top***": "*Inclusive Top*" & contain all the energy from the hadronically decaying top
    - Truth **ungroomed** jet: m_J >140GeV & ≥1 b-hadron ghost-associated
    - $\sqrt{d_{23}}$ critira: robust substructure to seperate fully contained top ($qq'b$) [backup]
- Background: quark-/gluon-initiated jets in multijet





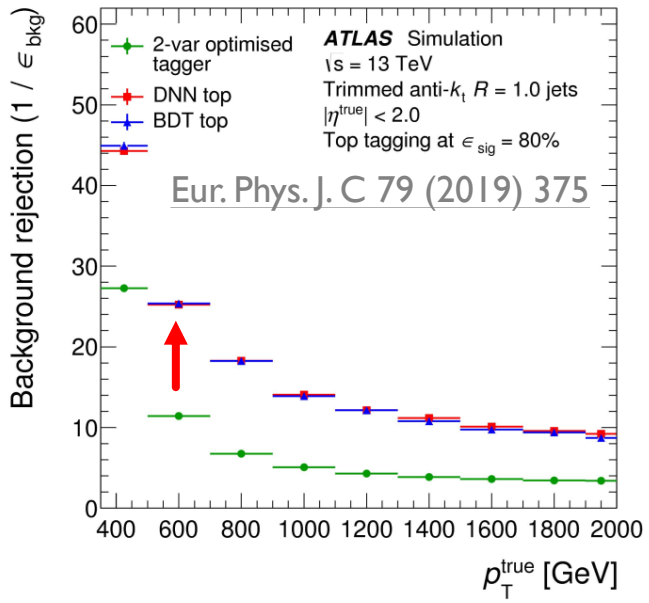




Baseline DNN Top Tagger

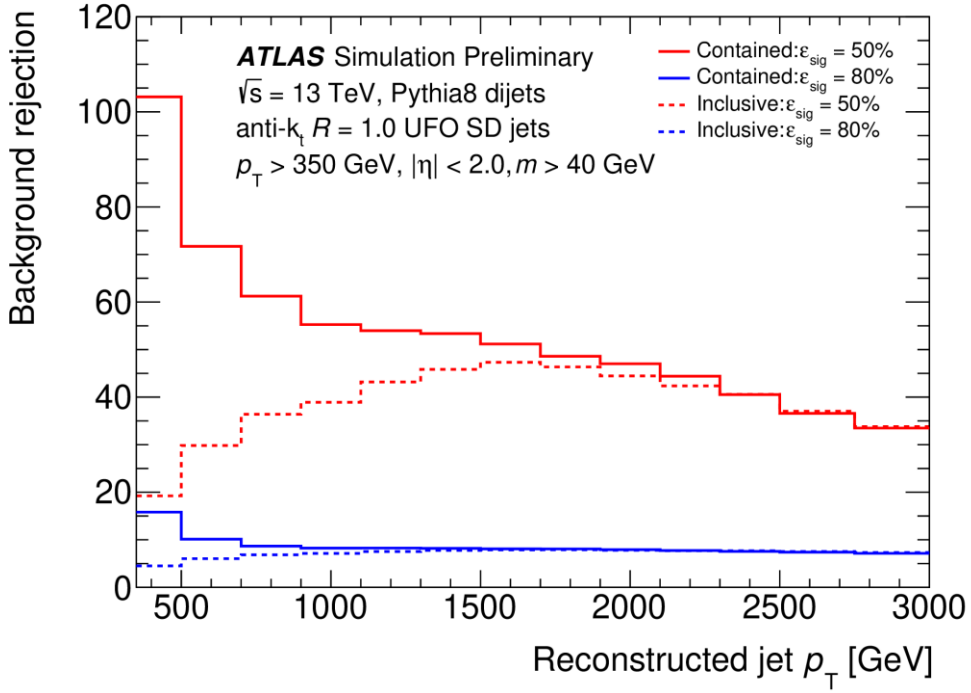
- DNN top tagger more powerful than cut-based refer to previous study [ref]
- 2 DNN taggers trained to tag **inclusive / *contained top*** jet on UFO jet
- 15 "high-level" substructure variables:
  - N-subjettiness: $\tau_1, \tau_2, \tau_3, \tau_4$ [ref]
  - K_t splitting scale: $\sqrt{d_{12}}, \sqrt{d_{23}}$ [ref]
  - Energy correlation: $ECF_1, ECF_2, ECF_3, C_2, D_2, L_2, L_3$ [ref]
  - Minimum pair-wise invariant mass: Q_W [ref]
  - Thrust major: T_M [ref]

$\tau_i, \sqrt{d_{ij}}, ECF, \ldots$



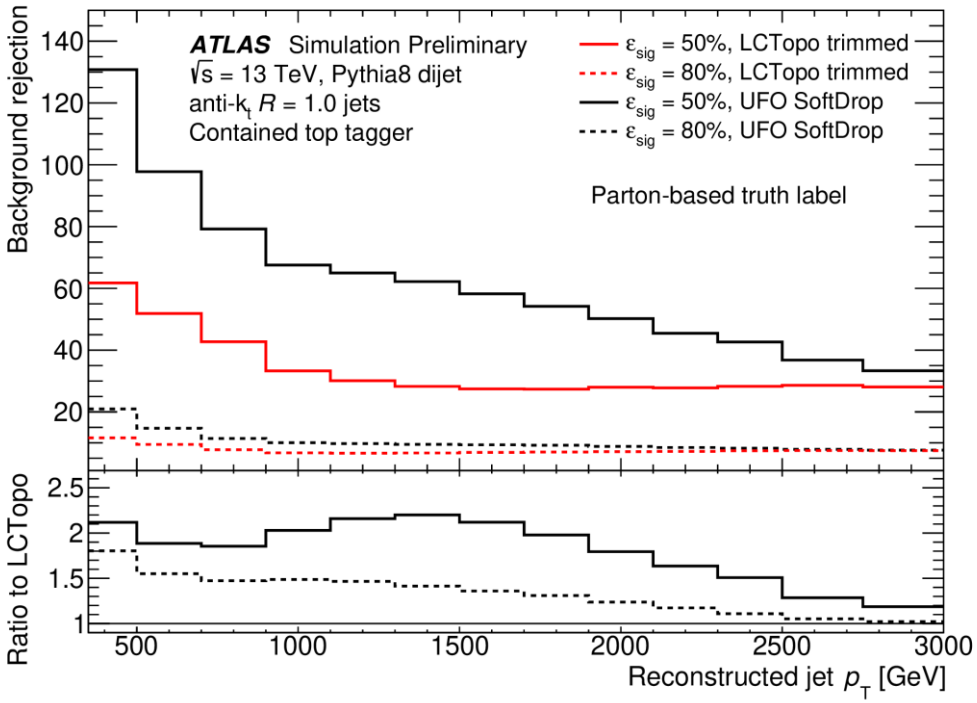
- Bkg. reject >99% at 50% ε_{sig} (50WP) for ***contained top*** at low p_T
- Flat $1/\varepsilon_{bkg}$ for both **inclusive** and ***contained top*** at 80% ε_{sig} (80WP)
- No diff. for **inclusive** or ***contained top*** tagging at high p_T

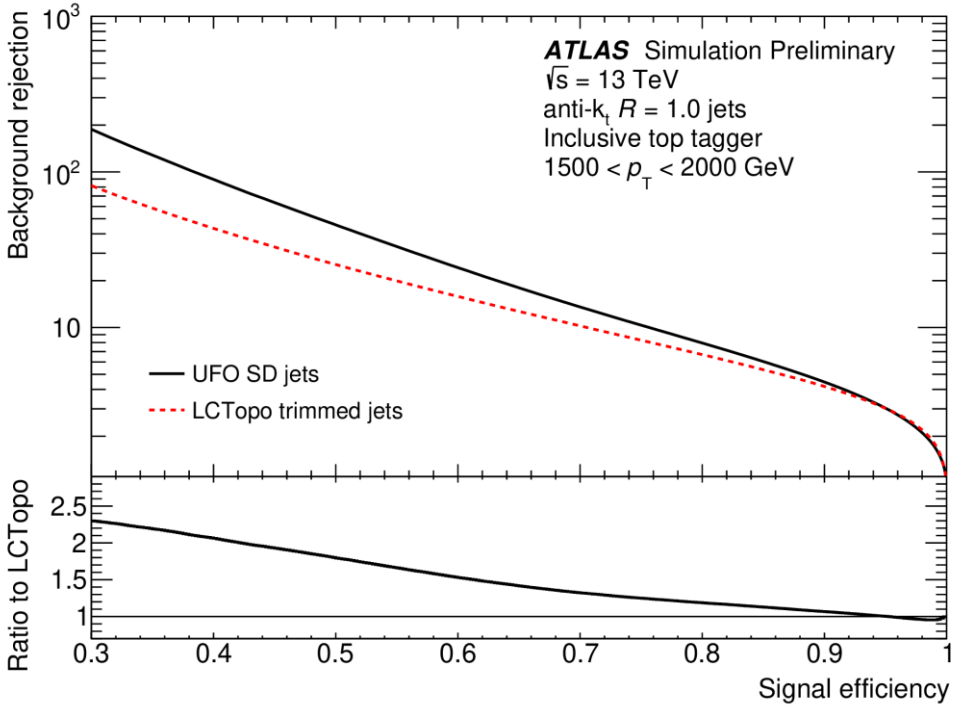
  → **inclusive top** contains full decay products at boosted case



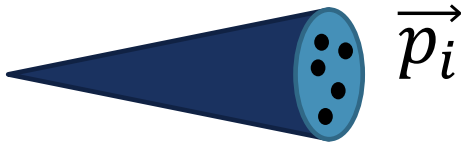
Performance

- DNN top tagger defined on UFO jet outperforms the one on LCTopo (with same labeling)
  - For both ***contained*** and **inclusive**, entire p_T and all working points
- At 50% eff.: rejection improved by a factor up to 2 for ***contained top***, 1.8 for **inclusive top** (p_T ~1.5TeV)
  - Better reconstruction of substructure variables with UFO jet



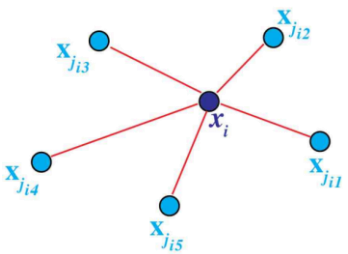


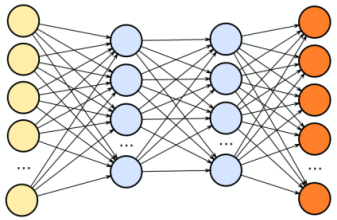
Constituent-based Top Tagger



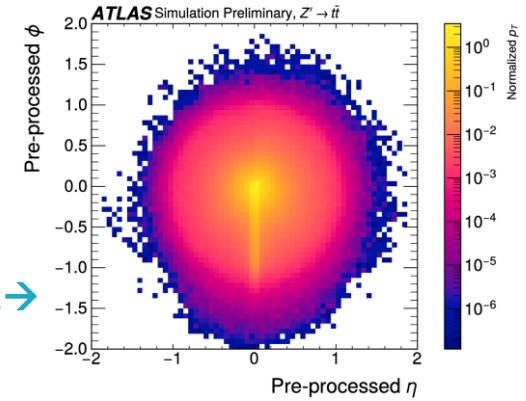
- Capture "low-level" information from jet constituents
- Using up to 80 constituents in each jet and 7 processed variables to use:
  - 4-vector(*) (p_T, η, ϕ, E), $R \equiv \sqrt{\eta^2 + \phi^2}$,
  - Exploit known physical symmetries by shift/rotation of jet axis
- Various advanced architectures adapted in ATLAS:
  - ➢ **ParticleNet (Graph Network) [ref] : ~2 times better**
  - ◆ **Particle Flow Network (PFN) (Set-based DNN) [ref]**
  - ◆ **Densely connected neural network (DNN) [ref]**
  - ◆ Energy Flow Network (EFN) [ref] : Infrared & collinear (IRC) safety
  - ● ResNet 50 (Convolutional Network) [ref]

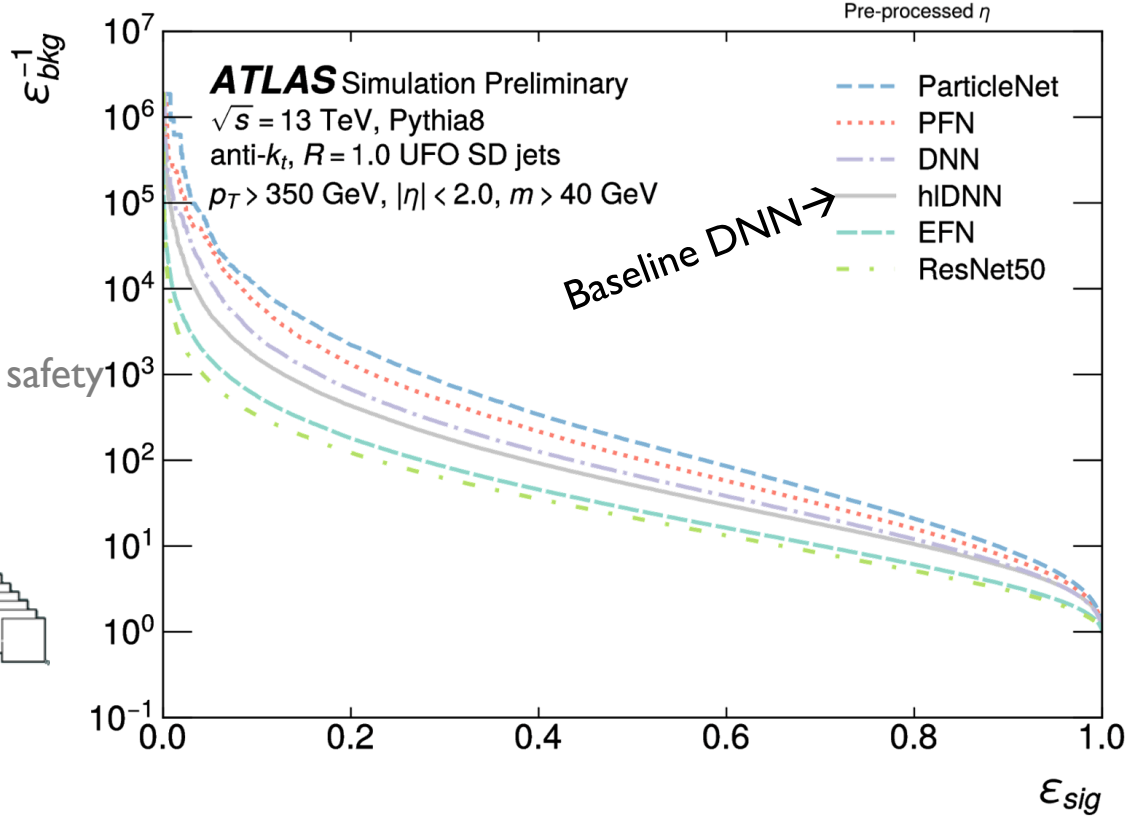
Better!





Processed Top Jet Avg.→ (Constituents)

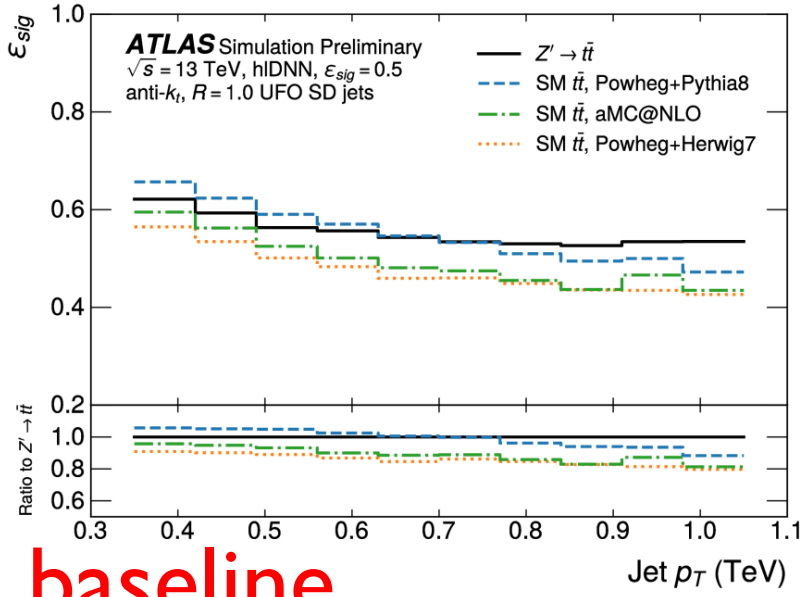




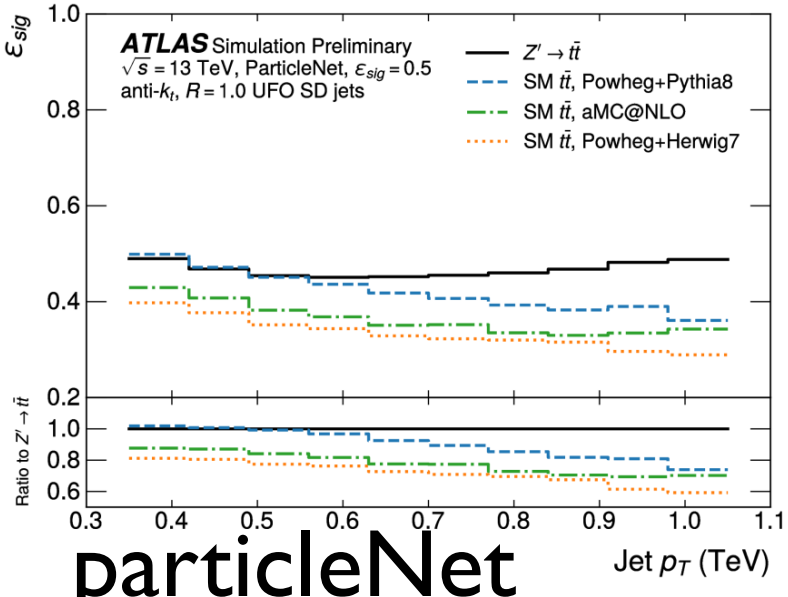
*: p_T and E in log scale

Model Dependence of Top tagger

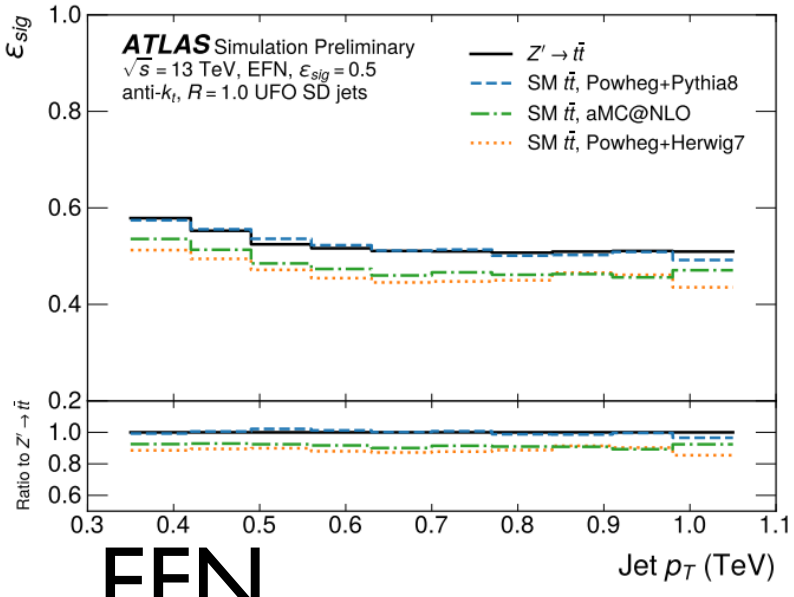
- Generator, parton showering (PS) variation and extrapolation from $Z' \to t\bar{t}$ to SM $t\bar{t}$
- High-level tagger(baseline DNN):
  - Small(~5%) variation from extrapolation and large(10~20%) from generator and PS
- Low-level tagger(constituent-based):
  - ParticleNet, PFN : larger dependence since sensitive to low level jet properties and ISR/FSR
  - EFN: highly reduced dependence due to IRC-safe transformation of 4-vector



baseline



particleNet



EFN

Summary

- Boosted W/Z and Top tagging algorithms, based on new UFO large-*R* jet in ATLAS are presented
  - Both cut-based and DNN-based tagger studied and various choices provided
  - Balance between tagging performance, mass decorrelation and complexity of understanding the systematics
- Performance is studied in detail and compared with previous tagger defined on LCTopo jet
  - Better performance mainly from good reconstruction of jet variable of UFO over entire p_T
  - Systematics effect at similar level compared with those of taggers on LCTopo
- Constituent-based top taggers studied in a realistic context of ATLAS at first time exploring new possibility
- Application in the future physics analysis is exciting and promising

BACKUPS

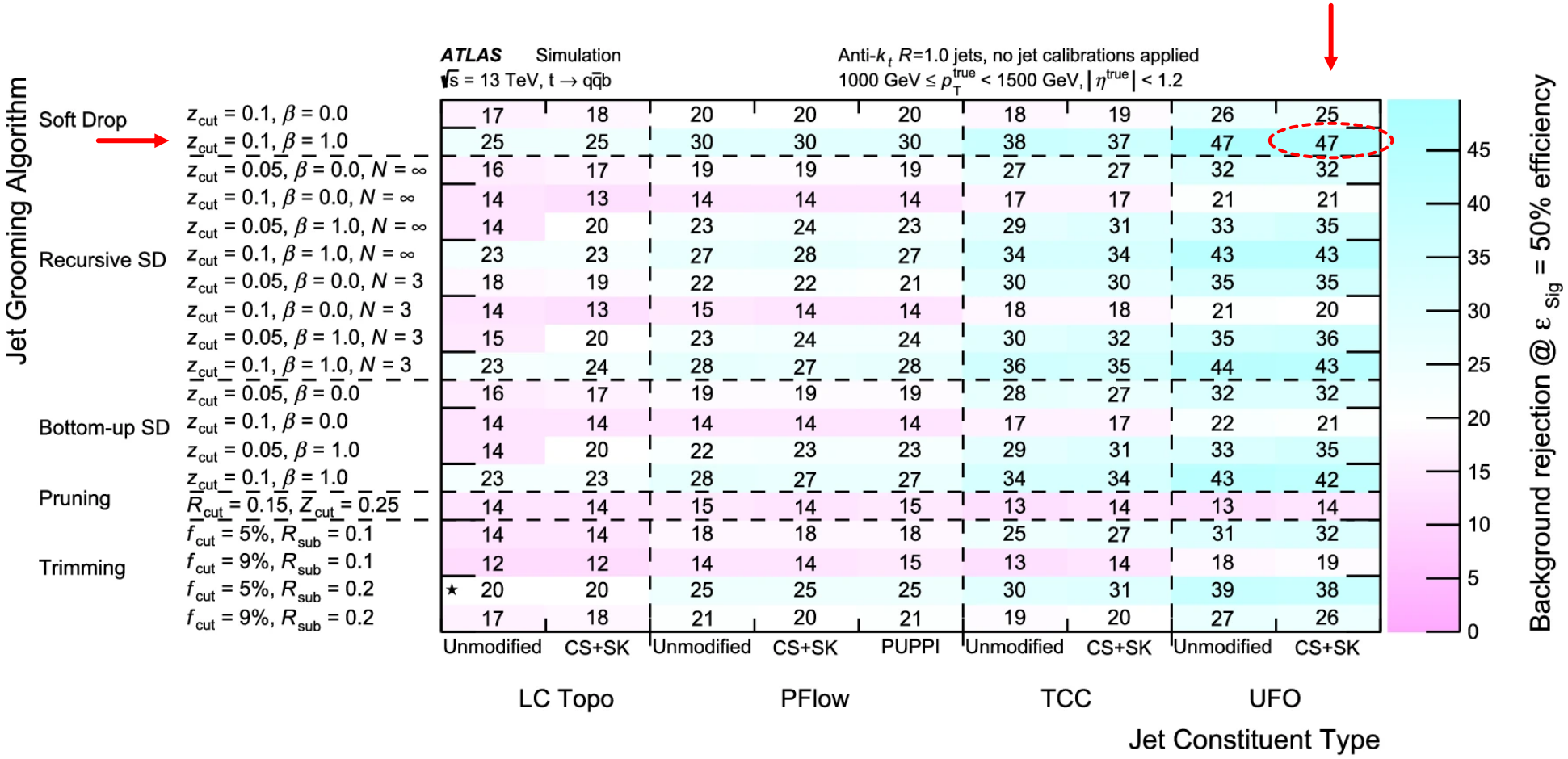
Eur. Phys. J. C 81 (2021) 4, 334

Detailed definition of UFO jet : UFO CS+SK SD jet

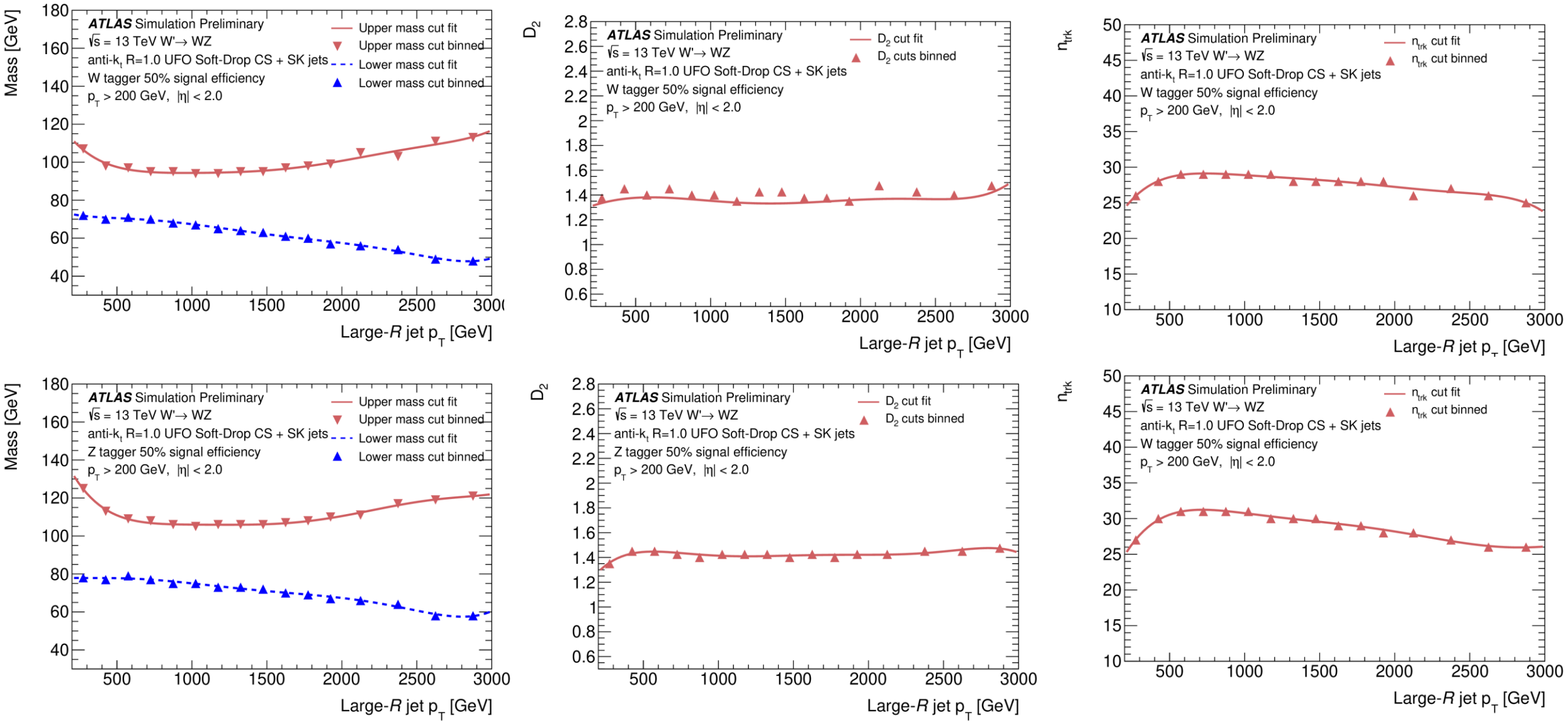
	Algorithm	Abbreviation	Settings
Jet input objects	Topological Clusters	Topoclusters	N/A
	Particle-Flow	PFlow	N/A
	Track-CaloClusters	TCCs	N/A
	Unified Flow Objects	UFOs	N/A
Pile-up mitigation algorithms	Constituent Subtraction	CS	$A_g = 0.01$ $\Delta R_{max} = 0.25$ $\alpha = 0$
	Voronoi Subtraction (*)	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	$z_{cut} = 0.1$ $\beta = 0, 1, 2(*)$
	Bottom-up Soft-Drop	BUSD	$z_{cut} = 0.05, 0.1$ $\beta = 0, 1, 2(*)$
	Recursive Soft-Drop	RSD	$z_{cut} = 0.05, 0.1$ $\beta = 0, 1, 2(*)$ $N = 3, 5(*), \infty$
	Pruning	N/A	$z_{cut} = 0.15$ $R_{cut} = 0.25$
	Trimming	N/A	$f_{cut} = 5\%, 9\%$ $R_{sub} = 0.1, 0.2$

Optimization of UFO jet

- Different pile-up mitigation and grooming algorithms tested and UFO CS+SK and SD chosen

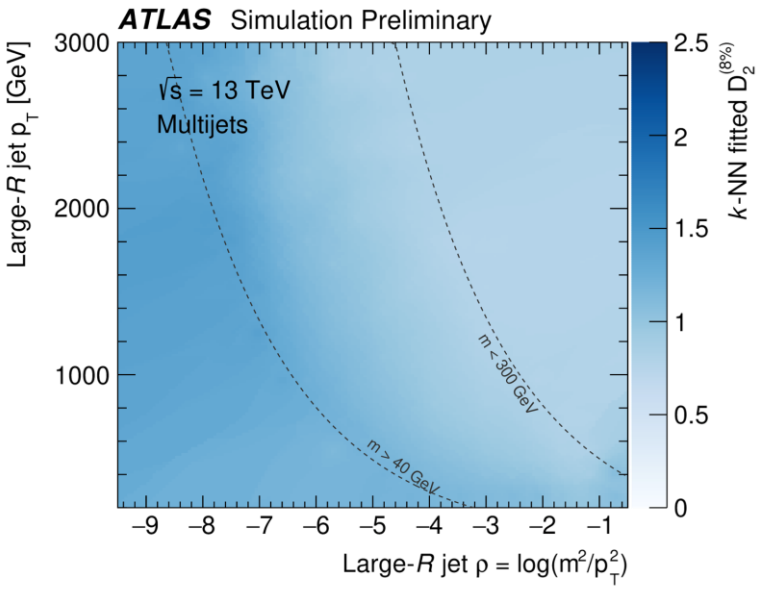


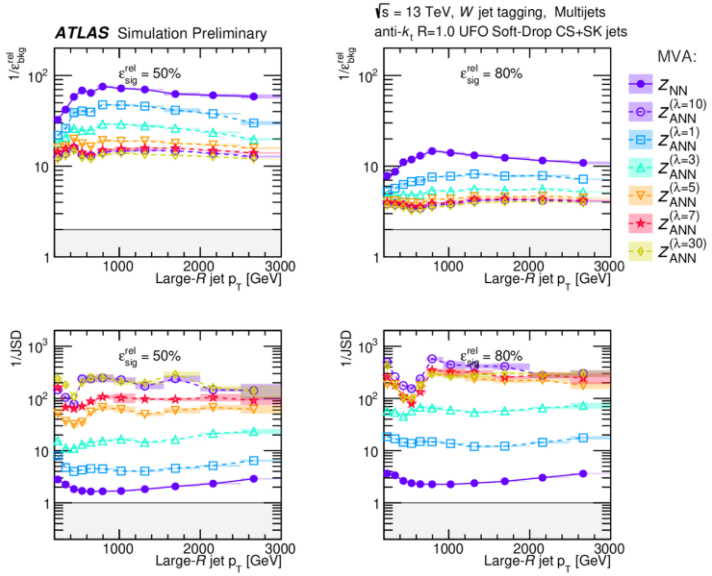
3-var cut value for W/Z tagging

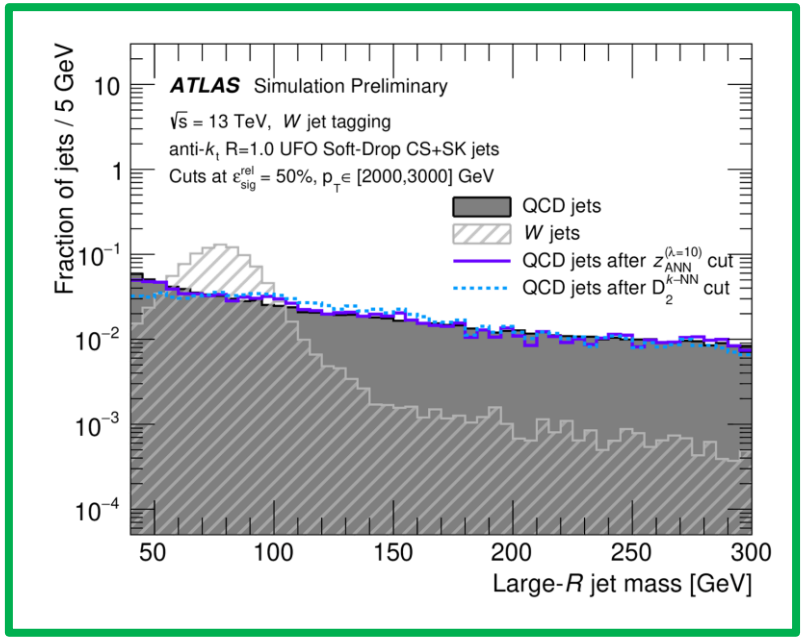


Mass Decorrelation

- Mass sculpting effect caused by m_J dependence of tagging efficiency esp. for DNN tagger
  - Difficulty for analyses using m_J sideband for background estimation
- Analytical method based on k-nearest neighbor regression (**k-NN**) to decorrelate D_2^{cut} from m_J → $D_2^{kNN-cut}$
  - Subtract the fully correlated part : $D_2^{8\%}(m_J, p_T)$ determined and regressed at fixed ε_{bkg}=8% (ε_{sig}=50%)
- Adversary network (**ANN**) introduced to train mass decorrelated DNN tagger → ANN tagger
  - Make adversary network harder to infer the m_J from DNN output with the cost of weaker tagging

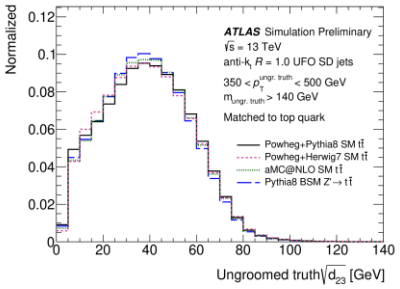


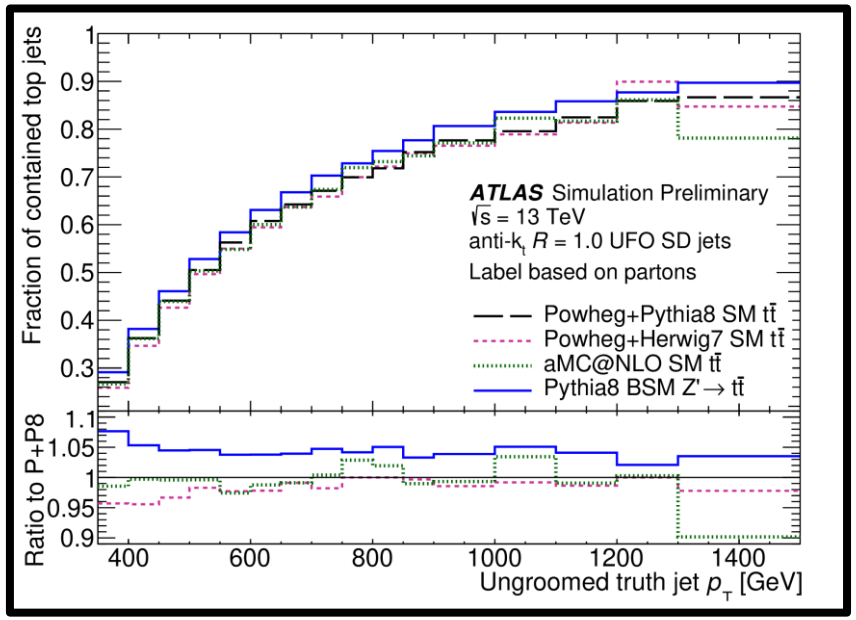


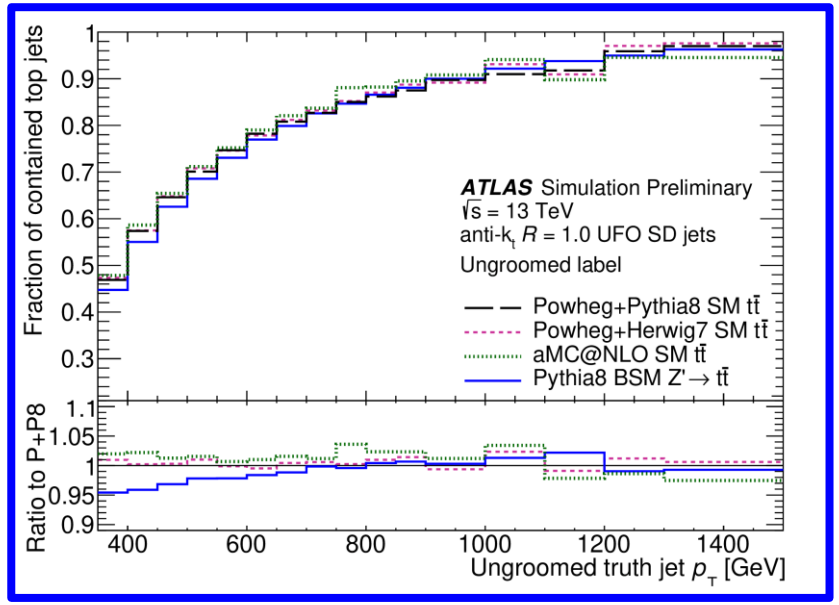


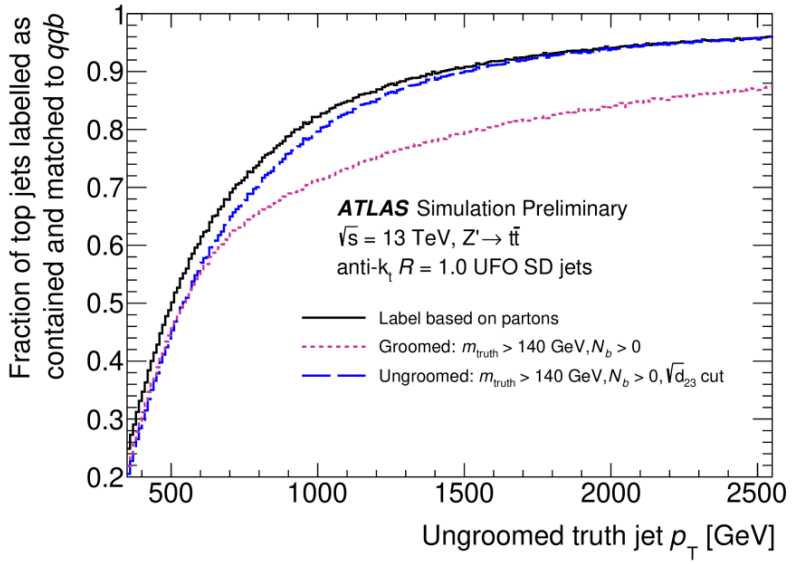
Optimization of Top Jet Definition

- "**Ungroomed**" labeling of contained top optimized for top tagging on UFO jet:
  - Improve the containment of top and $qq'b$ fraction in a robust way
- Better than **parton-based** or **groomed** labeling used for LCTopo jet:
  - **Parton-based**: ΔR matched to all the decay chain → natural but large impact from parton modelling
  - **Groomed**: based on groomd truth jet → optmized only for trimmed jet and less top ($qq'b$) with UFO jet



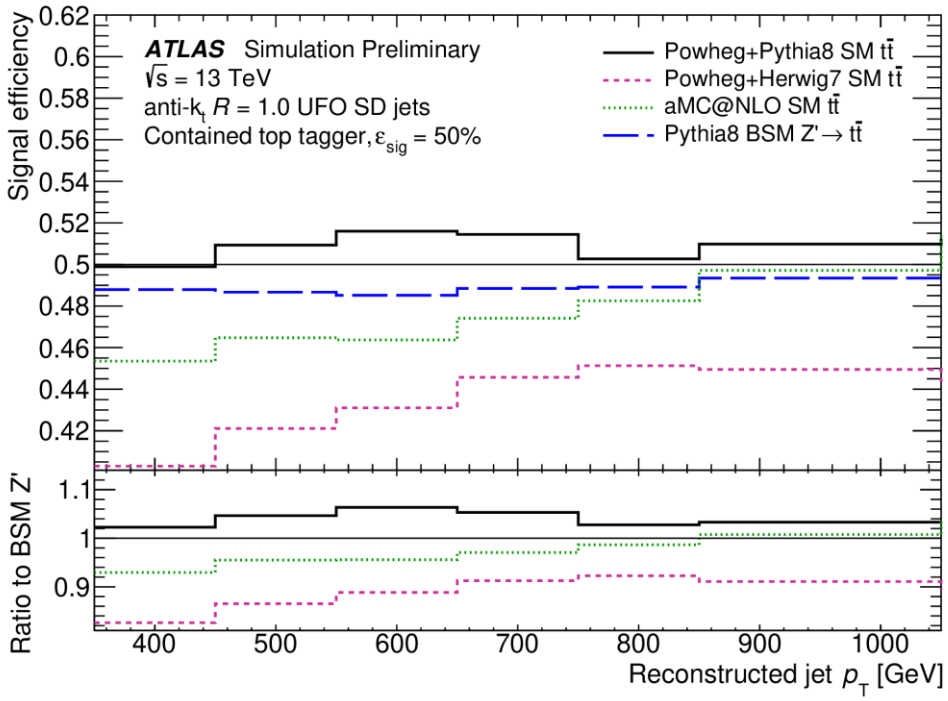


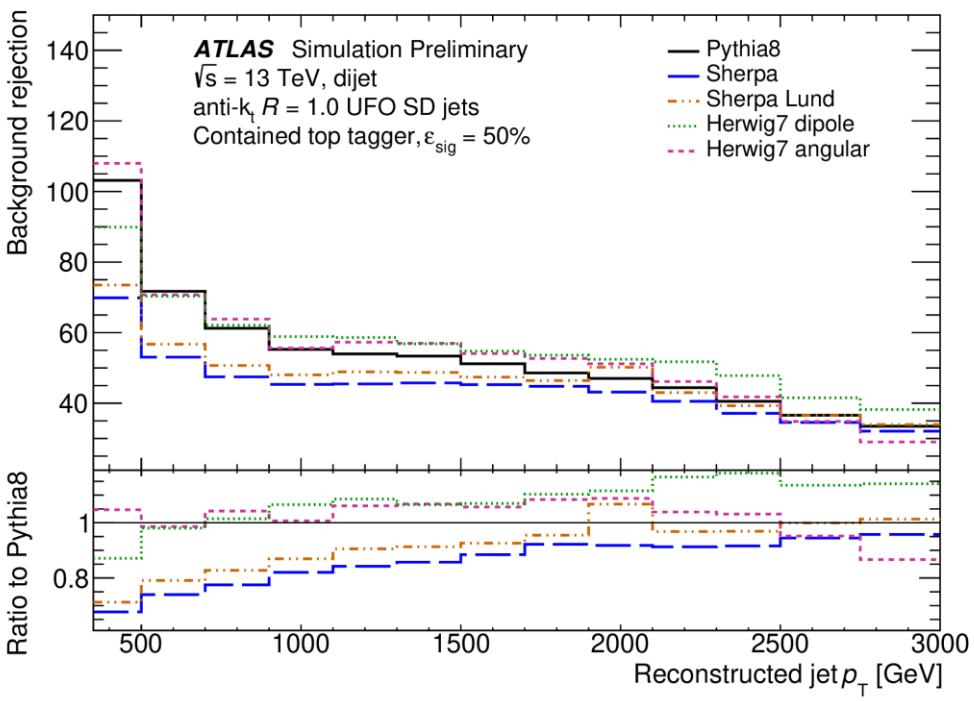




Modelling Dependence of baseline DNN Top tagger

- Small difference of ε_{sig} from generator variation and extrapolation from $Z' \rightarrow t\bar{t}$ training sample to SM $t\bar{t}$ (~5%)
- Large variation of ε_{sig} (10%) and ε_{bkg} (20%) from parton showering esp. for low p_T and tight (50%) WP
- Overall smaller impact for loose (80%) WP
- Similar order of modelling dependence for ***contained top*** and **inclusive top** tagger

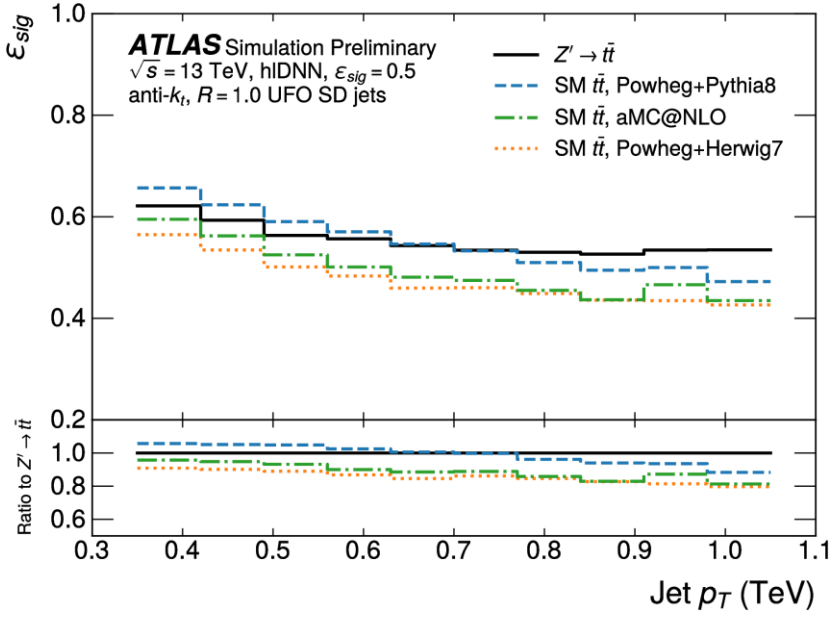


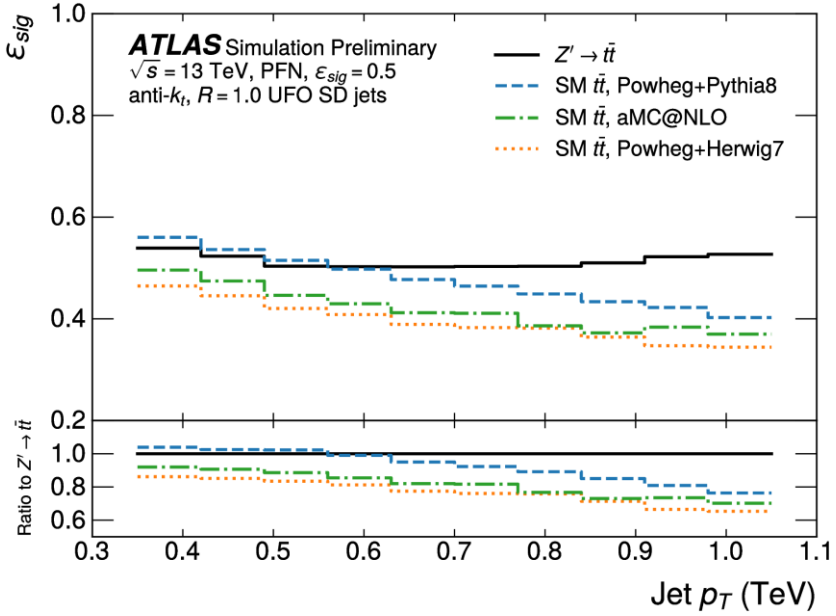


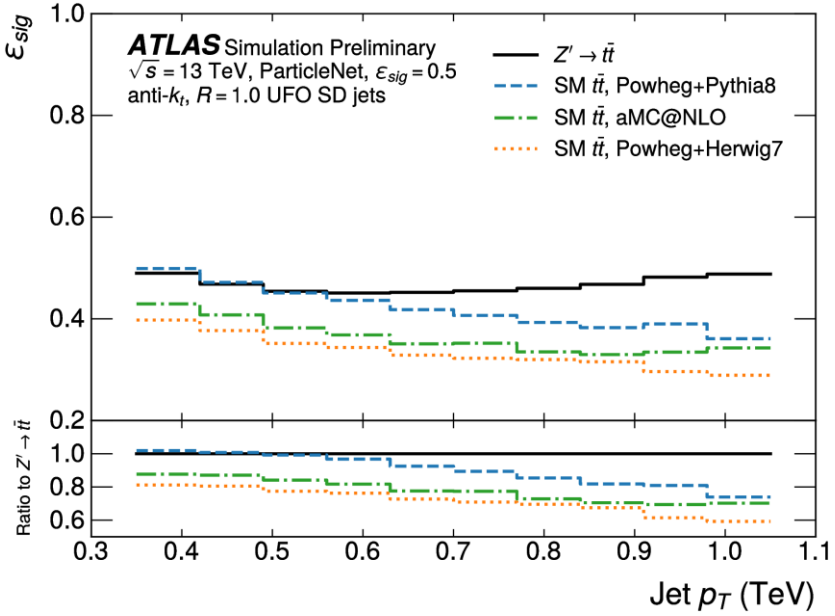
Modelling Dependence of constituent-based Top tagger

- Slightly large model dependence of PFN and ParticleNet than baseline DNN top tagger
  - Exploit more low level information (like constituent jet 4-vector) which depended more on modelling , like parton showering

baseline



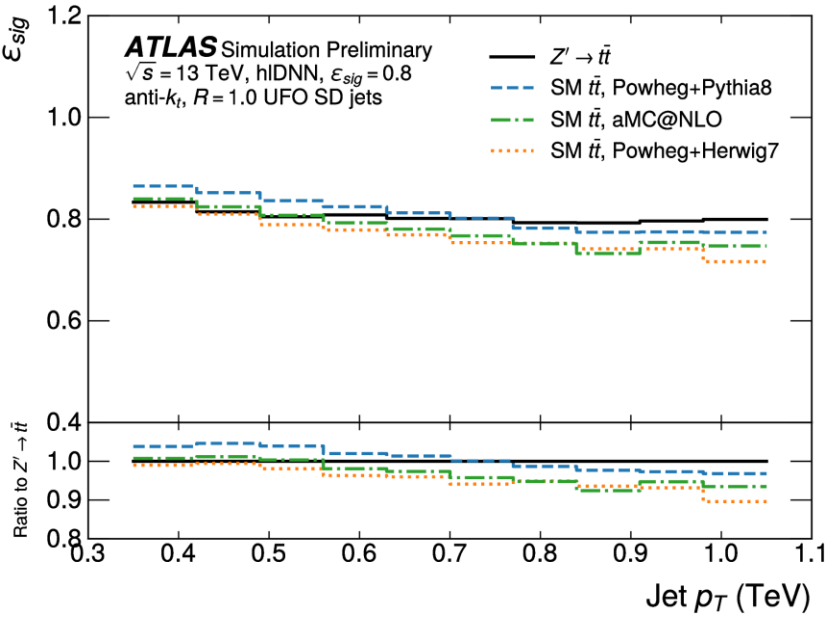


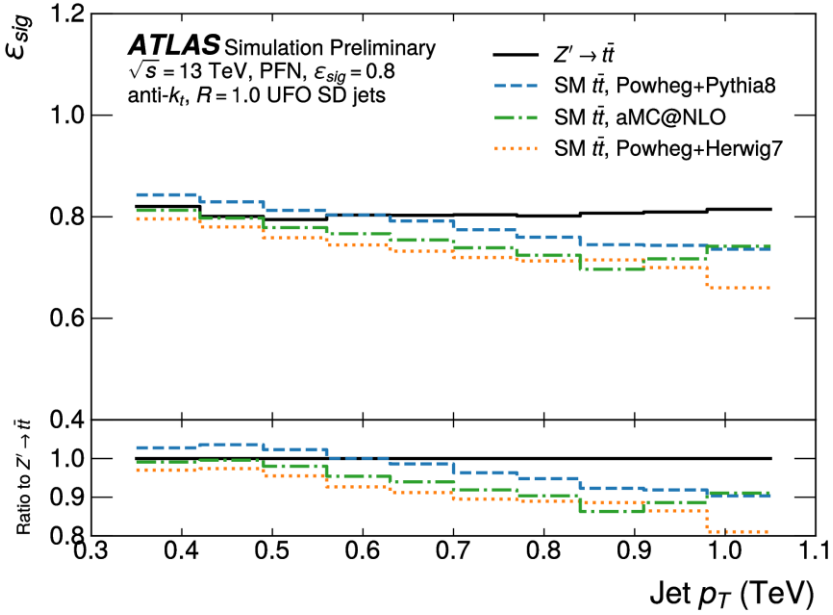


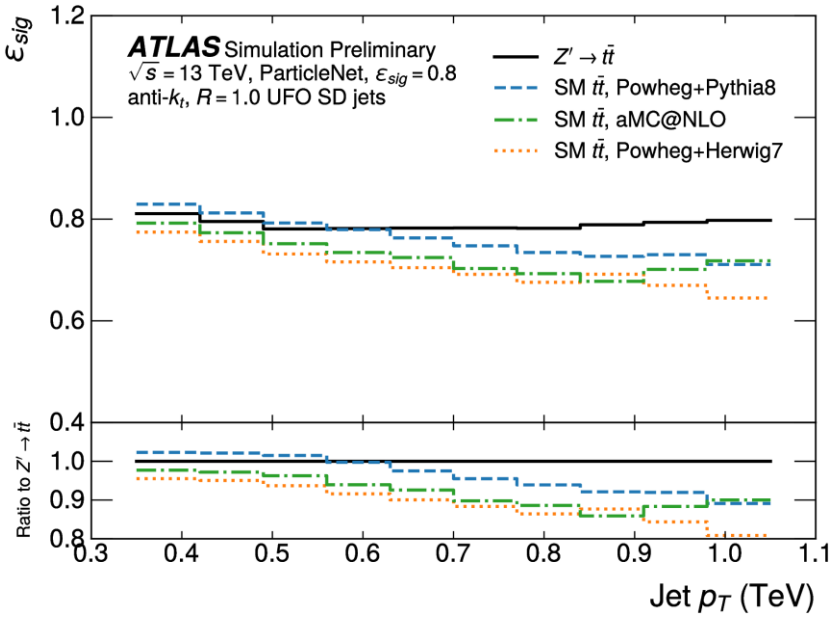
Modelling Dependence of constituent-based Top tagger

- Slightly large model dependence of PFN and ParticleNet than baseline DNN top tagger
  - Exploit more low level information (like constituent jet 4-vector) which depended more on modelling , like parton showering

baseline

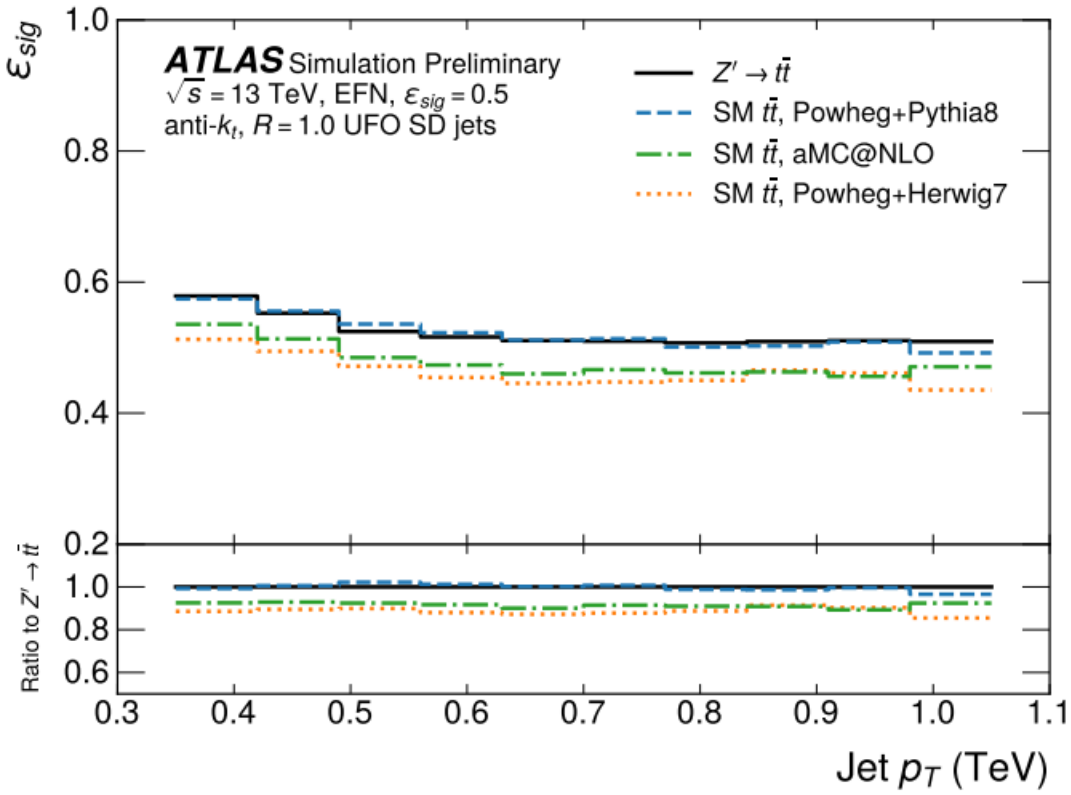


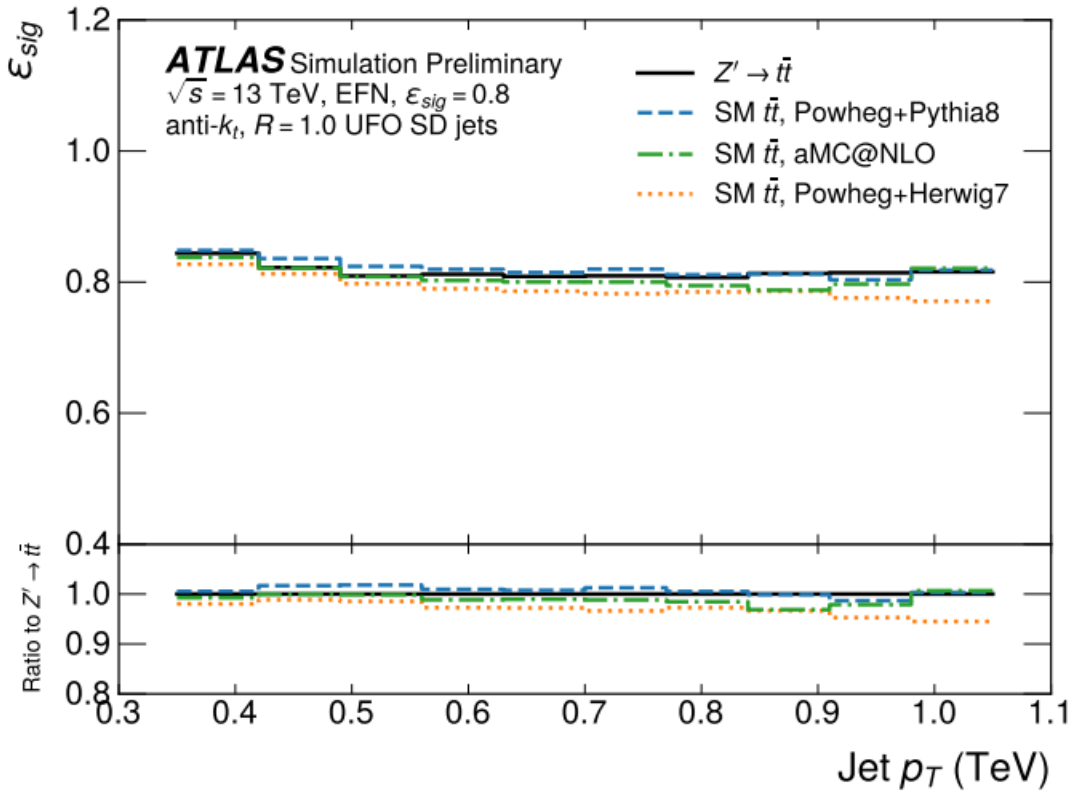




Modelling Dependence of constituent-based Top tagger

- EFN have very small model dependence due to IRC-safety



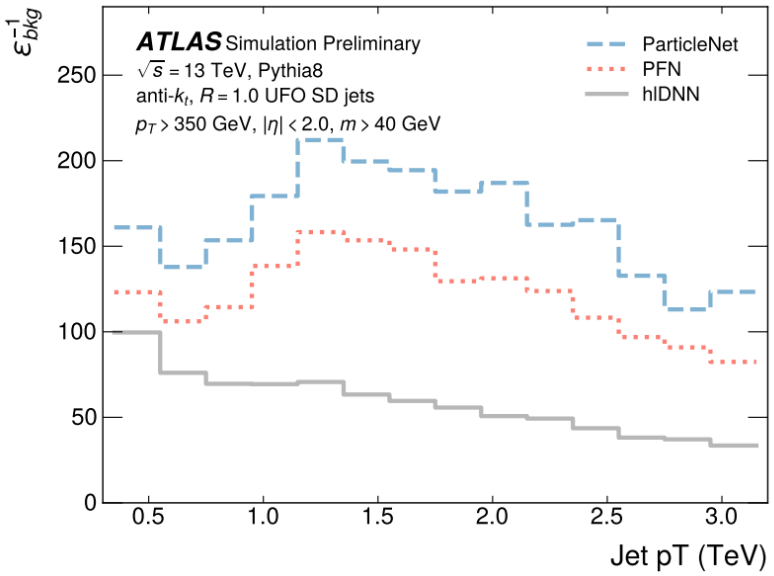


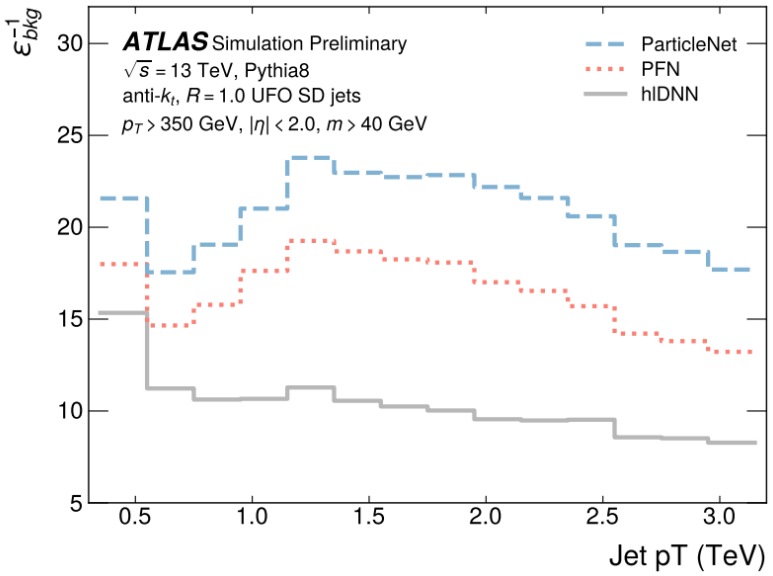
Details of constituent-based Top tagger architecture

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	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

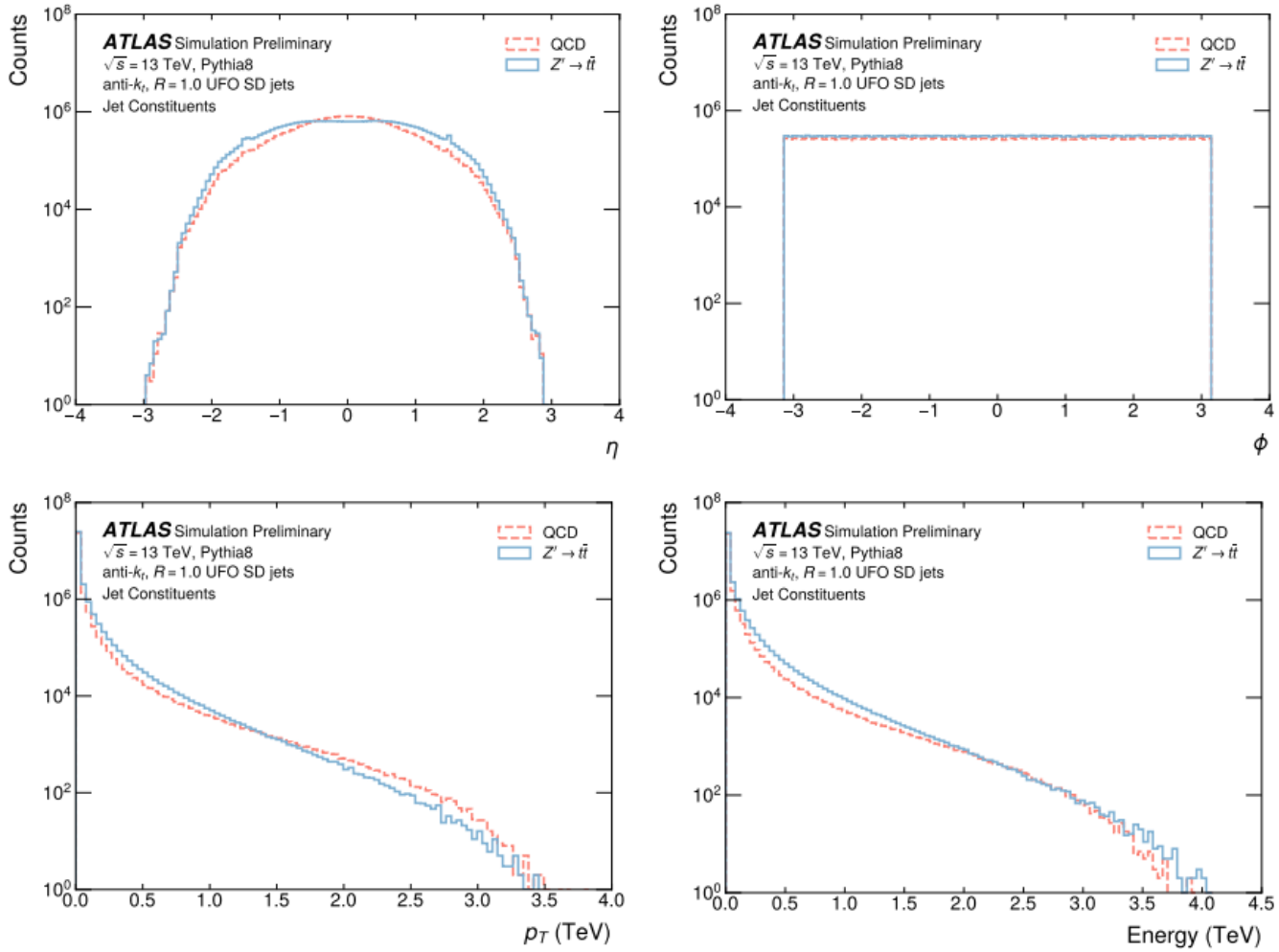
Performance of constituent-based Top tagger architecture



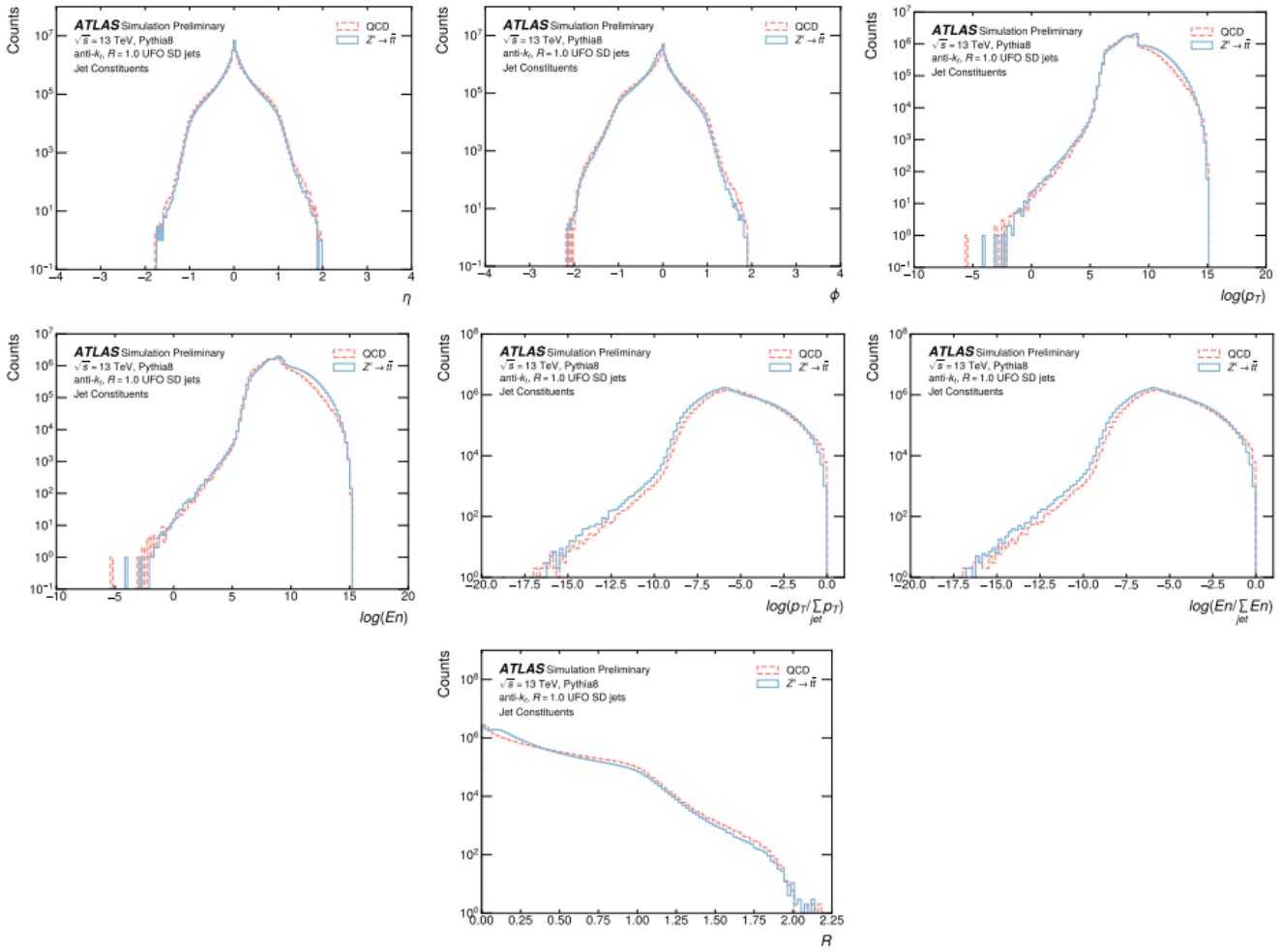


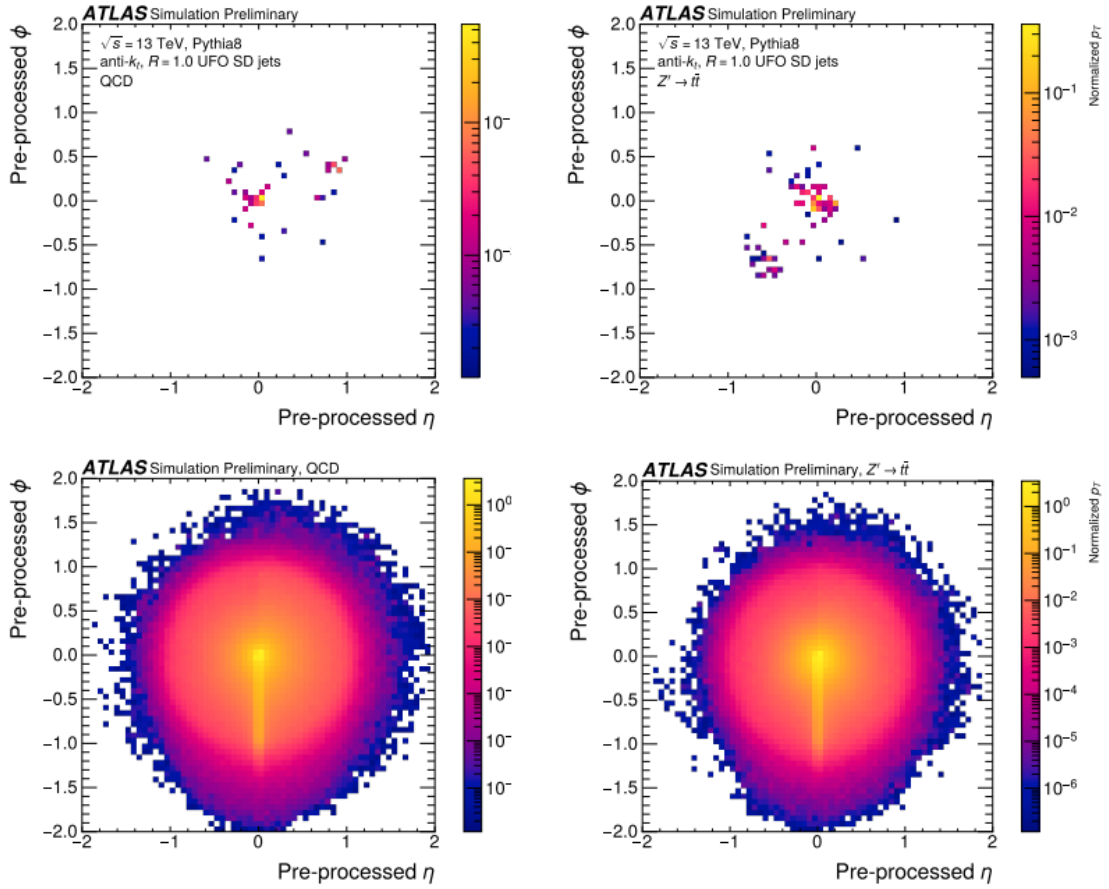
Model	AUC	ACC	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.5$	ε_{bkg}^{-1} @ $\varepsilon_{sig} = 0.8$	# Params	Inference Time
ResNet 50	0.885	0.803	21.4	5.13	1,486,209	9 ms
EFN	0.901	0.819	26.6	6.12	1,670,451	4 ms
hlDNN	0.938	0.863	51.5	10.5	93,151	3 ms
DNN	0.942	0.868	67.7	12.0	876,641	3 ms
PFN	0.954	0.882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

Input variables (low level constituents)

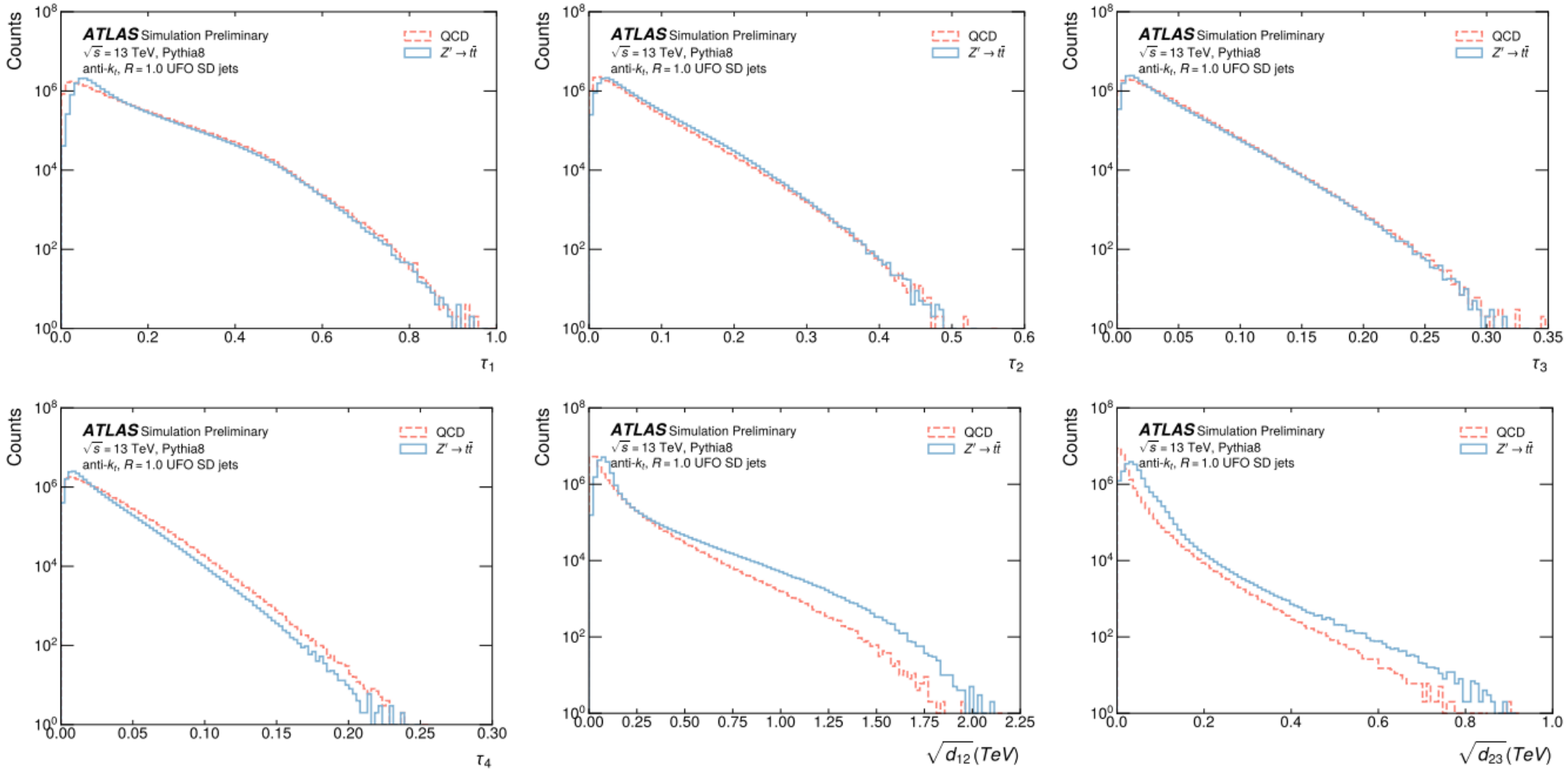


Input variables (low level constituents, after processing)

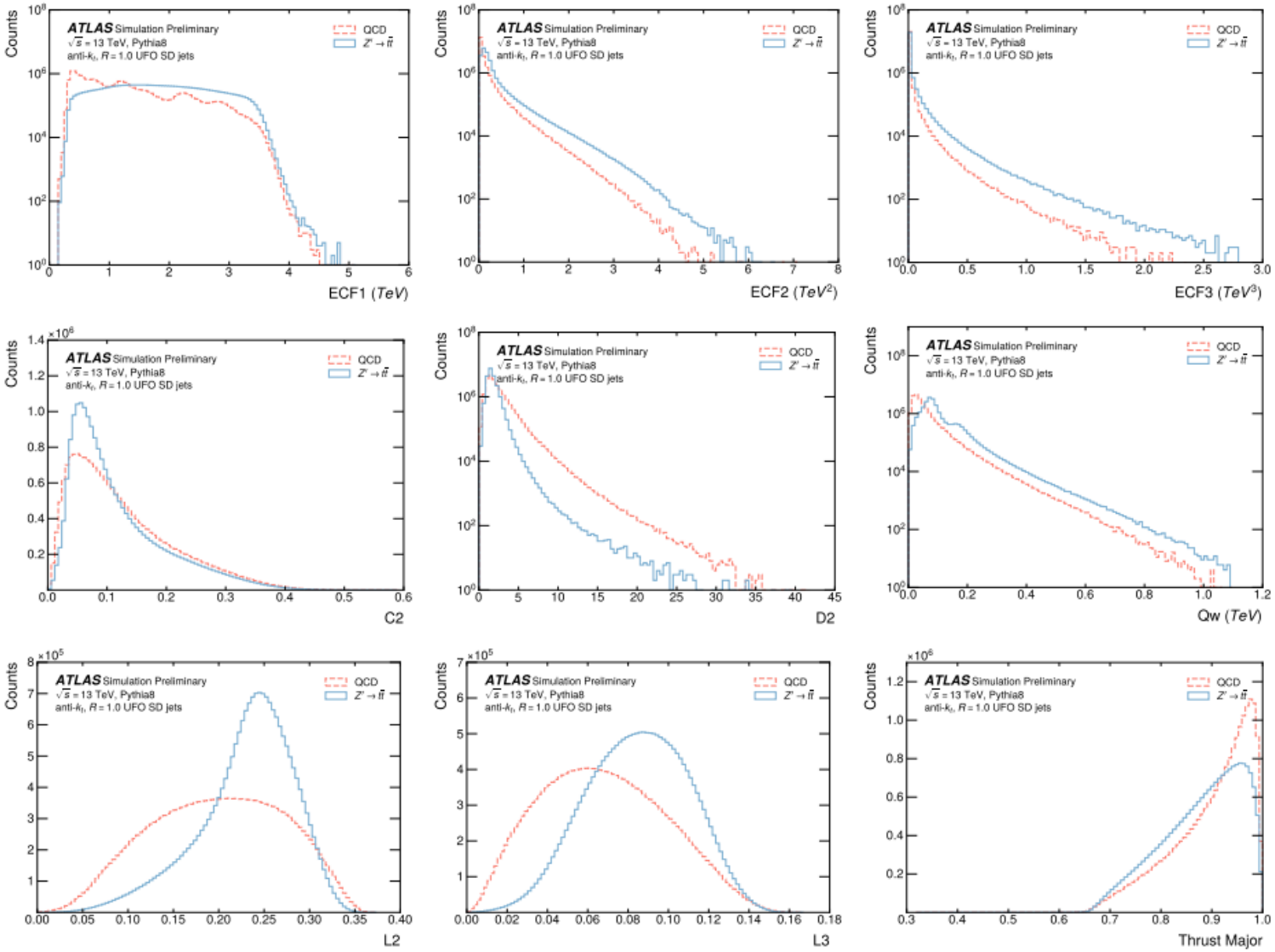




Input variables (high level, substructure)



Input variables (high level, substructure)



$\sqrt{d_{12}}$ and $\sqrt{d_{23}}$

- Energy scale of n-th Kt-declustering

$$d_{ij} = \min(p_{\mathrm{T}i}^2, p_{\mathrm{T}j}^2)\frac{\Delta R_{ij}^2}{R^2}, \ \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2,$$
$$d_{iB} = p_{\mathrm{T}i}^2, \qquad (1)$$

- ~ distance between n-th and j-th component/moment in Kt clustering algorithm
- $\sqrt{d_{12}}$: sensitive to two-prong topology
  - Selection in W/Z tagger to improve qq' fraction :

$$\sqrt{d_{12}} > 55.25 \cdot \exp\left(\frac{-2.34 \times 10^{-3}}{\mathrm{GeV}} p_{\mathrm{T}}\right) \mathrm{GeV},$$

- $\sqrt{d_{23}}$: sensitive to three-prong topology
  - Selection in W/Z tagger to improve qq'b fraction :

    Cut at 21GeV at p_T =350GeV and exponentially decrease with p_T to 7GeV at p_T =1500GeV

D_2 and energy correlation function

$$\theta \equiv R = \sqrt{\Delta y^2 + \Delta\phi^2}$$

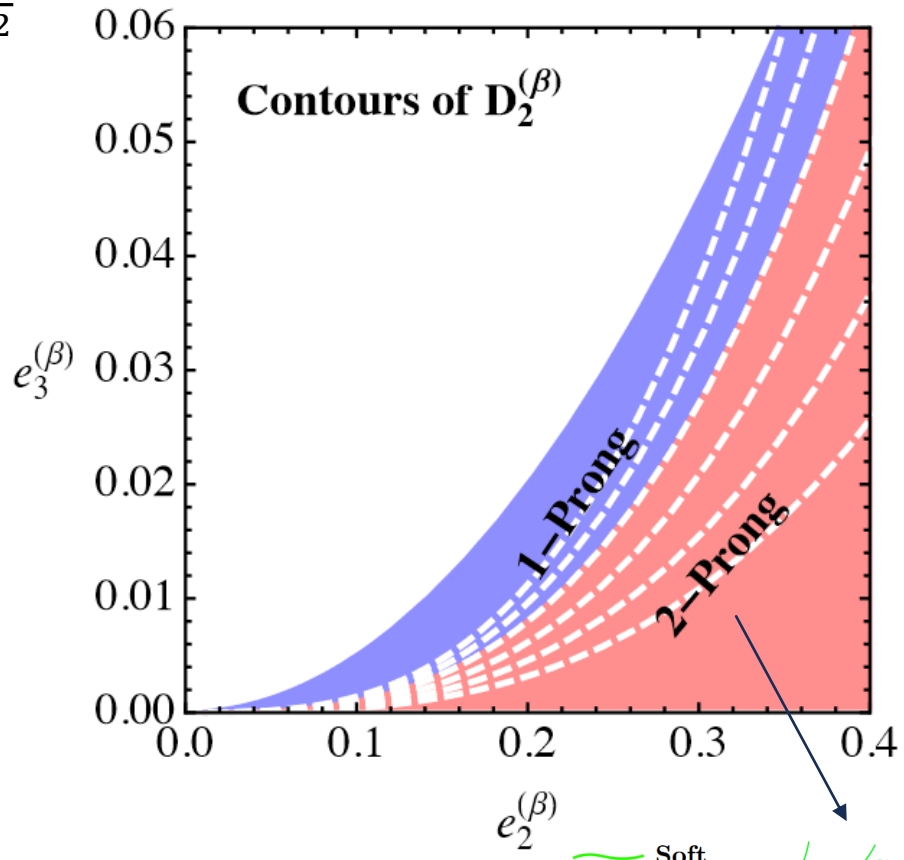
$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} E_{i_a}\right)\left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} \theta_{i_b i_c}\right)^{\beta}$$

$$e_n^{(\beta)} = \frac{\mathrm{ECF}(n,\beta)}{(\mathrm{ECF}(1,\beta))^n}$$

$$e_2^{(\beta)} = \frac{1}{p_{TJ}^2}\sum_{1\le i<j\le n_J} p_{Ti}p_{Tj}R_{ij}^{\beta}\,,$$

$$e_3^{(\beta)} = \frac{1}{p_{TJ}^3}\sum_{1\le i<j<k\le n_J} p_{Ti}p_{Tj}p_{Tk}R_{ij}^{\beta}R_{ik}^{\beta}R_{jk}^{\beta}$$

$$D_2^{(\beta)} \equiv \frac{e_3^{(\beta)}}{\left(e_2^{(\beta)}\right)^3}$$



$$e_2^{(\beta)} \sim R_{cc}^{\beta} + z_s\,,$$

$$e_3^{(\beta)} \sim R_{cc}^{3\beta} + z_s^2 + R_{cc}^{\beta} z_s$$

$$z_s \equiv \frac{p_{Ts}}{p_{TJ}} \ll 1\,, \qquad R_{sj} \sim 1$$

$$\frac{p_{Tc}}{p_{TJ}} \sim 1 \quad R_{cc} \ll 1\,, \qquad R_{cs} \sim 1$$