

Boosted W/Z Boson and Top Tagging in ATLAS

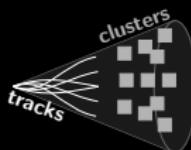
BOOST 2022, Hamburg

Tobias Fitschen

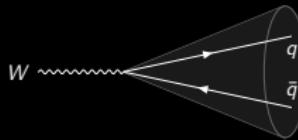
17 August 2022

University of Manchester

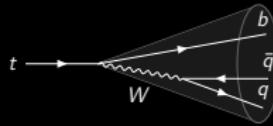




UFO Jets



W/Z Tagging

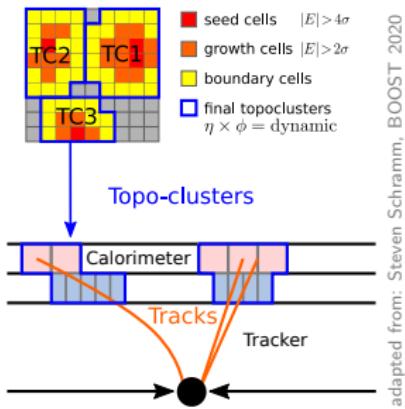


Top Tagging

Why UFO Jets?

Calorimeter only:

- **LCTopo**: Topological calorimeter clusters



Combined with tracking:

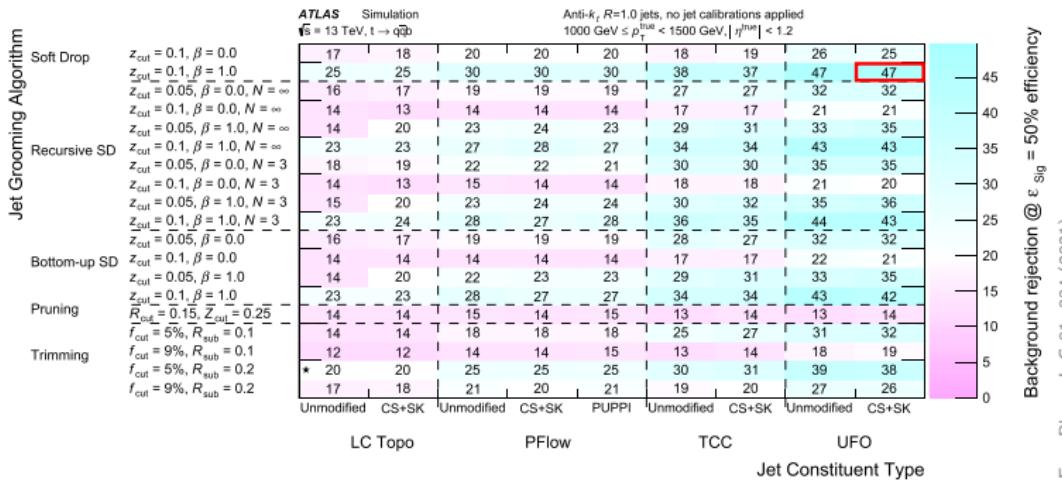
- **PFlow**: Particle Flow Objects
 - Low p_T : Use **track** 4-vector for charged particles, subtract energy from **cluster** 4-vectors
 - High p_T : Use **cluster** 4-vectors, ignore **tracks**
- **TCC**: Track Calo Clusters
 - Low p_T : Use **cluster** 4-vectors, ignore **tracks**
 - High p_T : Split **clusters** using **tracks**, get energy from **clusters** but angles from **tracks**

Combining PFlow and TCC:

- **UFO** combines **TCC** and **PFlow** to achieve optimal performance over a broad kinematic (p_T) range

Background rejection for various pileup mitigations and groomings:

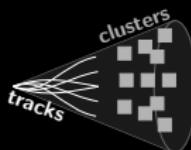
Here: 2-variable top tagger, high- p_T range
(plots for W and low- p_T in backup)



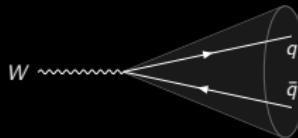
Best background rejection with:

- $R = 1.0$ anti- k_T **UFO** jets
- Pileup Mitigation: Constituent Subtraction + SoftKiller (**CS+SK**)
- Grooming: Soft Drop (**SD**) with $\beta = 1.0$ $z_{\text{cut}} = 0.1$

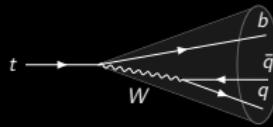
Other factors: Good pileup stability, mass resolution, ...



UFO Jets



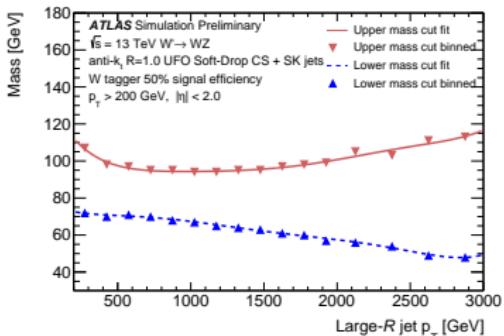
W/Z Tagging



Top Tagging

3-Variable Tagger

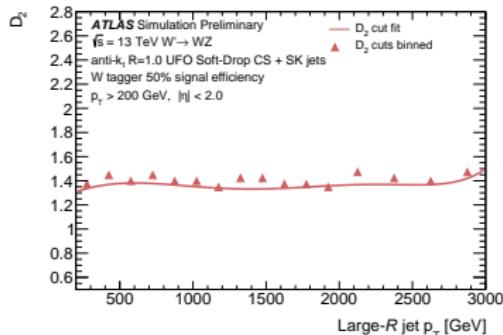
Mass window



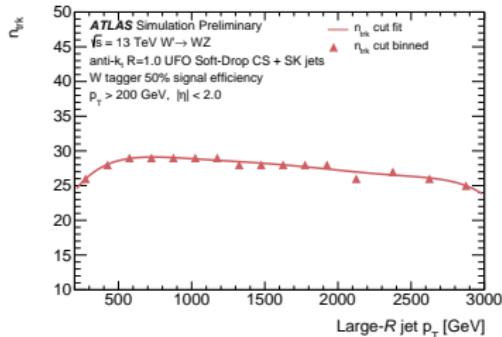
3-variable tagger: W

- p_T -dependent cuts on 3 features
- Maximise background rejection while keeping 50% signal efficiency per bin
- D_2 nearly flat in p_T
 - Thanks to angular resolution of UFO constituents
 - Fixed D_2 cut possible

Energy correlation ratio D_2

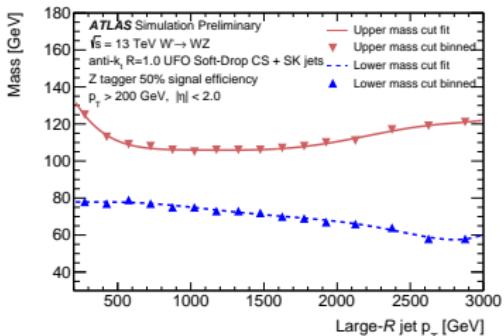


Number of tracks n_{trk}



3-Variable Tagger

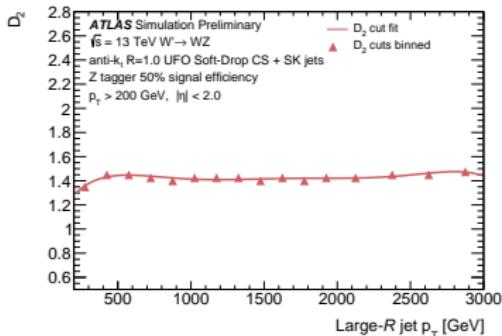
Mass window



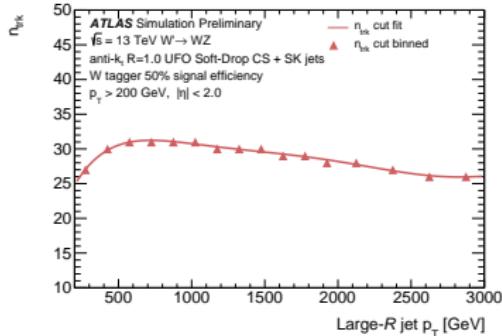
3-variable tagger: Z

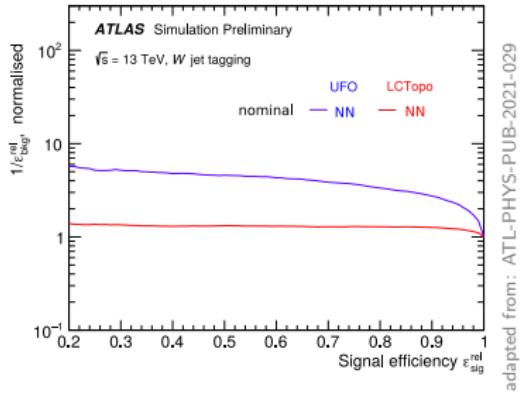
- p_T -dependent cuts on 3 features
- Maximise background rejection while keeping 50% signal efficiency per bin
- D_2 nearly flat in p_T
 - Thanks to angular resolution of UFO constituents
 - Fixed D_2 cut possible

Energy correlation ratio D_2



Number of tracks n_{trk}





Machine learning based (NN):

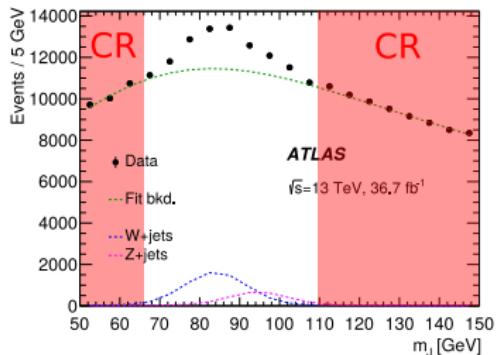
- Various substructure variables

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}	Z -Splitting scales
$\sqrt{d_{12}}$	d -Splitting scales
$Kt\Delta R$	k_t -subjett ΔR

(see backup for definitions)

→ ϵ^{bkg} improved by factor 2-3 in NN **UFO tagger** w.r.t. **LCTopo tagger**

Data-driven background estimates:



adapted from : arxiv.org/abs/1708.04445

- Define mass side-bands as control regions (CR)
- Fit smooth function to data from left to right side-band
- Estimate background in signal region from fit

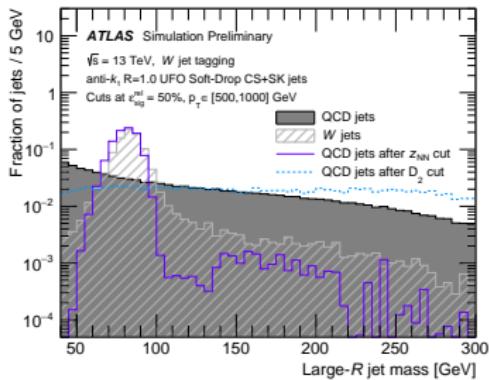
Problem: Tagger may introduce unwanted shaping of background, de-populating the sideband regions

Solution: Decorrelate tagger decision from m_j :

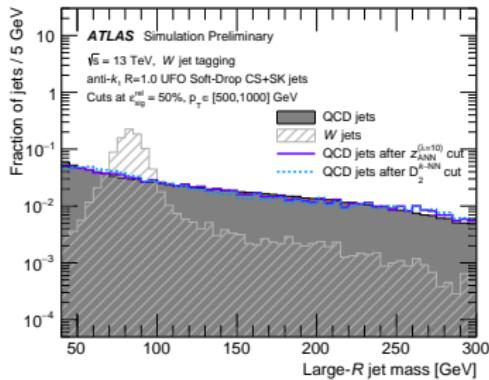
- Adversarial neural networks (ANN) for NN tagger

Mass Decorrelation with Adversarial Training

NN: Correlated to m_J



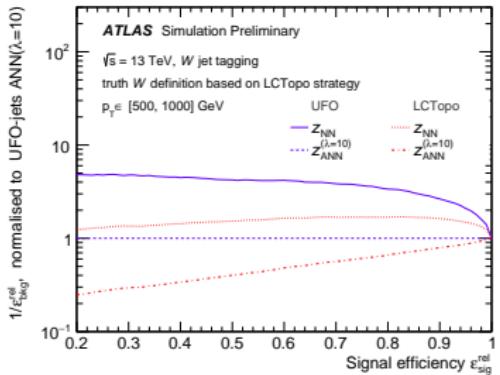
ANN: Active decorrelation



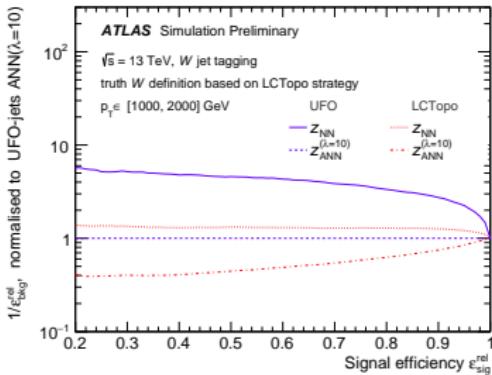
- Background mass distribution shaped according to signal by NN
- Adversarial Neural Network (ANN) successfully decorrelates

Tagger Performance: UFO vs LCTopo

Low p_T



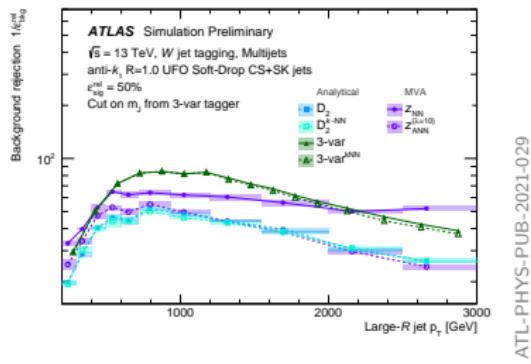
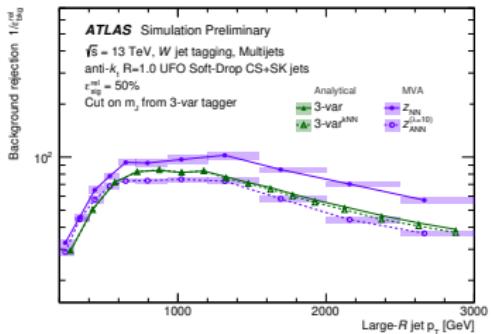
High p_T



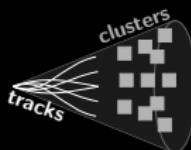
- Bkg rejection with **UFO** improved by a factor of 2-3 w.r.t. **LCTopo**!
- Mass decorrelation (ANN) comes with tradeoff of reduced efficiency
 - But may be better option for data-driven background estimates on m_j distribution

Previously:

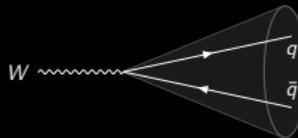
ATL-PHYS-PUB-2021-029

**Now:**With n_{trk} as additional feature

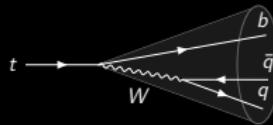
- Previously: 3-variable tagger showed better performance than NN
- Now: NN much better after including n_{trk} as additional feature!
 - ANN comparable with 3-variable tagger, but with decorrelation!
- Reason for such strong improvement:
 - Most other features exploit 2-prongedness of W/Z decay
 - n_{trk} is good quark/gluon discriminator



UFO Jets



W/Z Tagging



Top Tagging

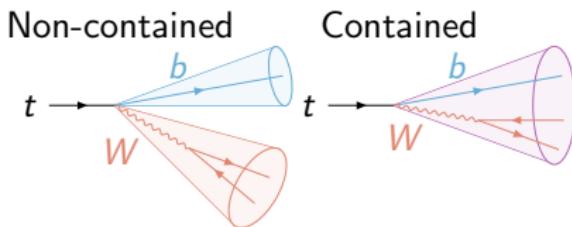
Two types of DNN-based top taggers defined:

Inclusive: Purely defined by ΔR matching:

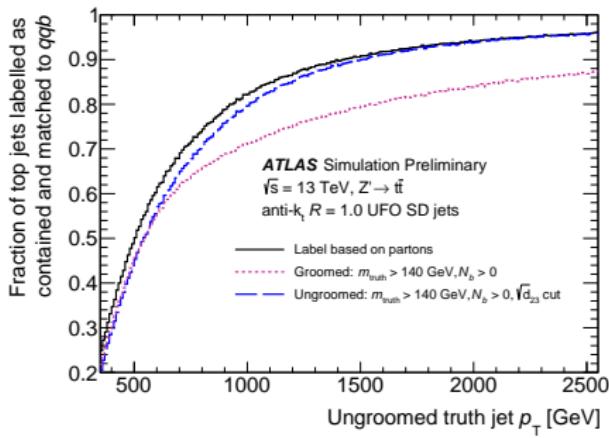
- $\Delta R(j^{\text{reco}}, j^{\text{truth}}) < 0.75$ and $\Delta R(j^{\text{truth}}, t^{\text{truth}}) < 0.75$

Contained: Extra cuts to ensure $t \rightarrow qqb$ fully contained within jet:

- Same ΔR requirement
- $N_B \geq 1$
- $m_{\text{ungroomed}}^{\text{truth}} > 140 \text{ GeV}$
- $\sqrt{d_{23}} > 27 e^{-\frac{p_T}{1433 \text{ GeV}}} \text{ GeV}$



Contained top labelling efficiency:



DNN Top Taggers: Definition

2 taggers: Inclusive and contained

- Fixed working points: 50% and 80%
 - Defined as function of p_T
- DNN features optimised for UFO jets:

$\tau_1, \tau_2, \tau_3, \tau_4$ N -subjettiness

$\sqrt{d_{12}}, \sqrt{d_{23}}$ Splitting scales

ECF_1, ECF_2, ECF_3 Energy correlation functions

C_2, D_2 Energy correlation ratios

L_2, L_3 Generalised energy correlation ratios

Q_W Invariant mass / virtuality

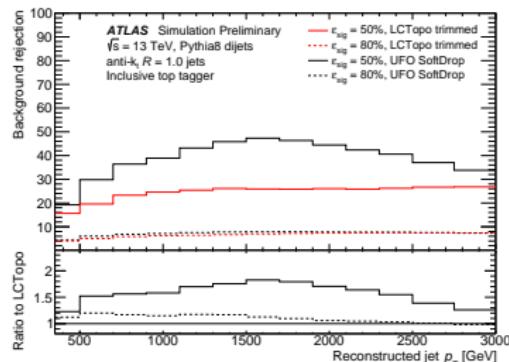
T_M Thrust major

(see backup for definitions)

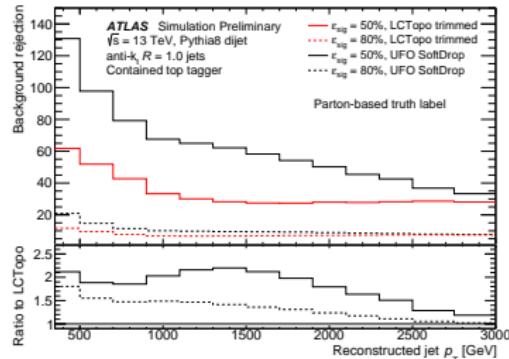
Observations: UFO vs LCTopo

- 80% working point:
 - Inclusive: $\approx 20\%$ better rejection at $p_T < 1.5$ TeV
 - Contained: Better over whole range
- 50% working point:
 - Clear improvement for both

Inclusive: UFO vs LCTopo



Contained: UFO vs LCTopo

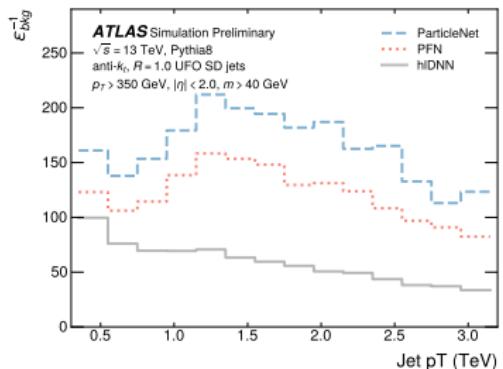


ATL-PHYS-PUB-2022-39: Constituent-Based Top-Quark Tagging

DNN top tagger (prev. slides):

- Set of high-level features (substructure variables)
 - Used as baseline (hIDNN)

More info in poster by Kevin Greif



ATL-PHYS-PUB-2022-39

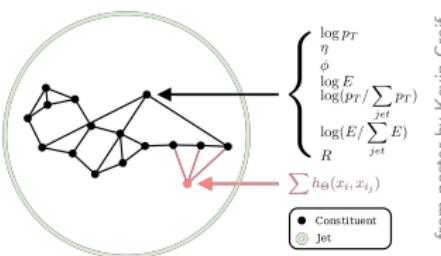
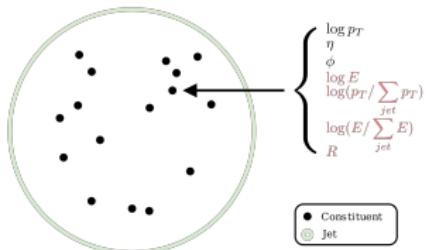
Constituent-based taggers:

- Low-level features based on 4-vectors of jet constituents

→ Up to $\times 2$ improvement over baseline (hIDNN)!

PFN: Particle Flow Network

ParticleNet: Dynamic Graph-CNN



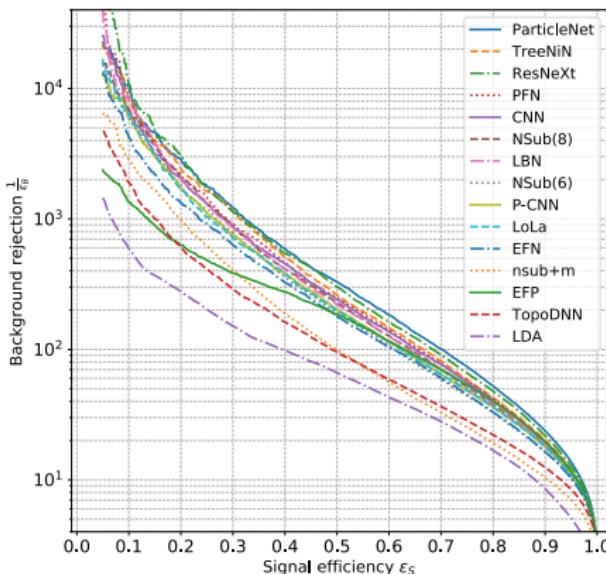
from poster by Kevin Greif

Constituent-Based Top Tagger

Based on: The Machine Learning landscape of top taggers

doi:10.21468/SciPostPhys.7.1.014

- Comparison of many modern ML techniques applied to the top tagging task
- Simplified detector simulation with Delphes + ATLAS card
- Calorimeter information only
→ No tracking as in UFO jets



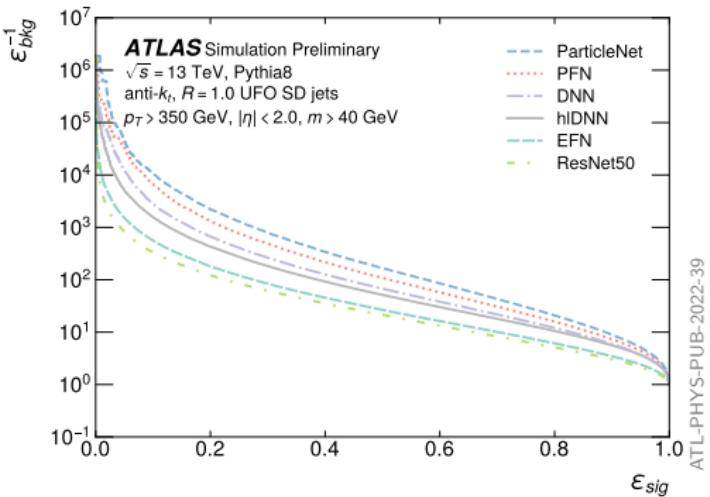
doi:10.21468/SciPostPhys.7.1.014

→ Best performance with **ParticleNet** and **ResNeXt**

ATL-PHYS-PUB-2022-39

More info in poster by Kevin Greif

New: How do these algorithms perform on ATLAS simulated UFO jets?



- **hDNN:** Baseline similar to DNN top tagger
 → [ATL-PHYS-PUB-2021-028](#)
- **DNN:** Using constituent 4-momenta
 → [arxiv.org/abs/1704.02124](#)
- **EFN/PFN:** Energy/Particle-flow networks
 → [arxiv.org/abs/1810.05165](#)
- **ResNet50:** CNN using jet images
 → [arxiv.org/abs/1512.03385](#)
- **ParticleNet:** Dynamic Graph-CNN
 → [arxiv.org/abs/1902.08570](#)

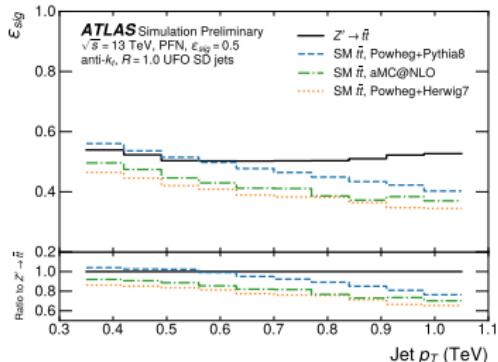
- **ParticleNet** and **PFN** show best performance
- **ResNet50** & **EFN** underperform → Do not translate well from Delphes study

Simulated data made public for ML experts along with PUB-note!

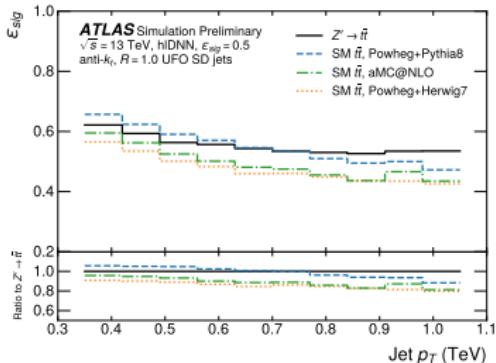
Model dependence

- Different parton shower and hadronisation models
- ϵ^{sig} measured at threshold for $\epsilon^{\text{sig}} = 50\%$ in nominal sample
- **PFN** and **ParticleNet**: Slightly more model dependent than baseline hIDNN

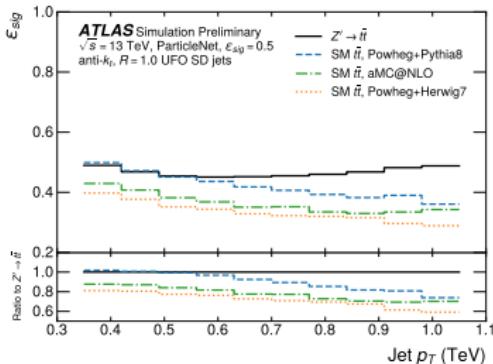
PFN (Particle Flow Network)



baseline: hIDNN



ParticleNet



Summary

W/Z taggers:

- 3-variable cut-based, NN-based with/without mass-decorrelation
- Clear advantage of NN tagger w.r.t. 3-variable when including n_{trk}
- Background rejection improved by factor 2-3 in UFO NN tagger w.r.t. previous NN tagger using LCTopo jets!

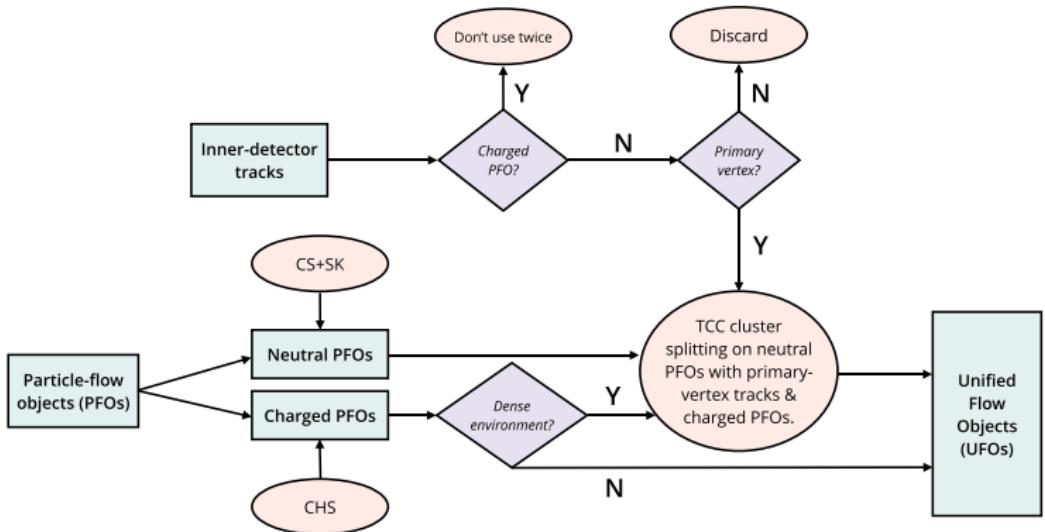
Top taggers:

- Inclusive and contained DNN tagger, new: Constituent-based taggers
- Clear improvement of DNN tagger using UFO w.r.t. LCTopo, especially at 50% working point
- Constituent-based ParticleNet tagger may add $\times 2$ improvement on top of that!

Appendix

UFO Jets

The UFO Algorithm

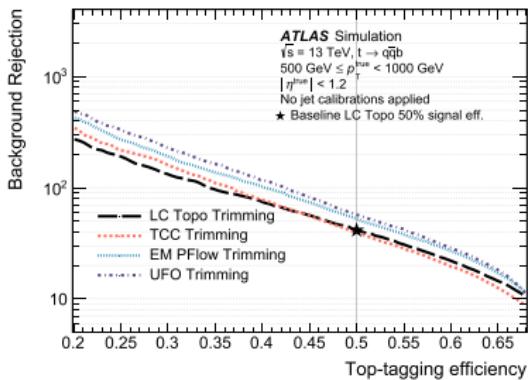


Extensive effort in ATLAS to find best jet definition for tagging:

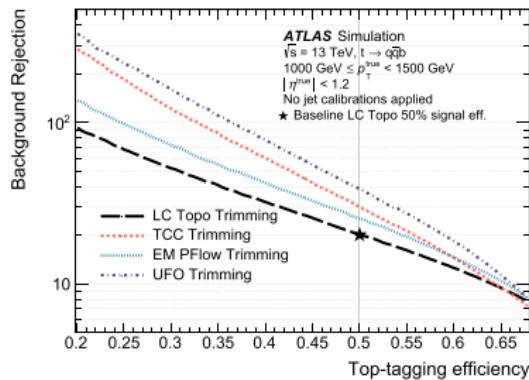
Eur. Phys. J. C 81, 334 (2021)

- Expected tagger performance evaluated for simple 2-variable cuts:
 - W/Z tagger: m, D_2
 - Top tagger: m, τ_{32}

UFO jets show best performance for simple top tagger:



$500 \text{ GeV} < p_T^{\text{true}} < 1000 \text{ GeV}$

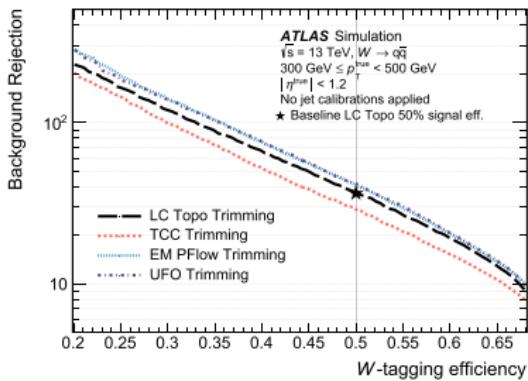


$1000 \text{ GeV} < p_T^{\text{true}} < 1500 \text{ GeV}$

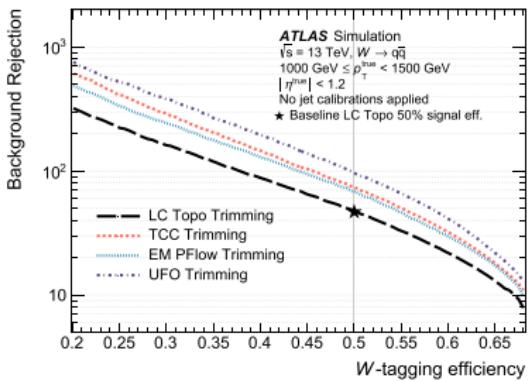
Extensive effort in ATLAS to find best jet definition for tagging: Eur. Phys. J. C 81, 334 (2021)

- Expected tagger performance evaluated for simple 2-variable cuts:
 - W/Z tagger: m, D_2
 - Top tagger: m, τ_{32}

...as well as simple W tagger:



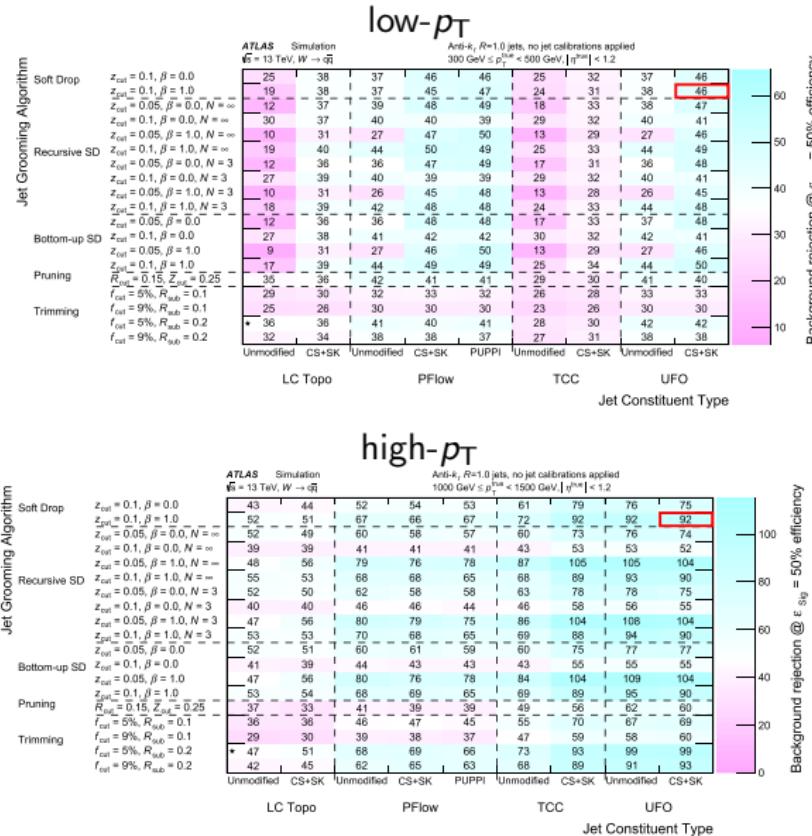
$500 \text{ GeV} < p_T^{\text{true}} < 1000 \text{ GeV}$



$1000 \text{ GeV} < p_T^{\text{true}} < 1500 \text{ GeV}$

Optimisation of Jet Definition

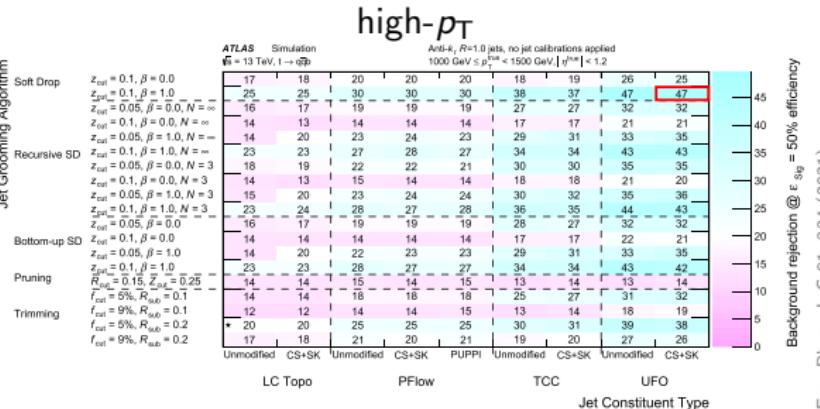
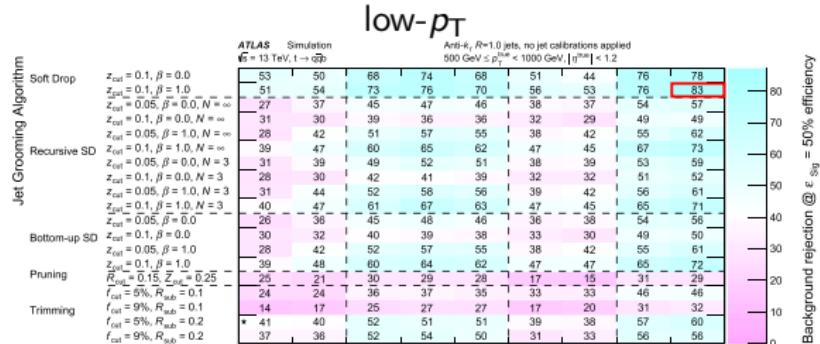
2-variable W tagger



Optimisation of Jet Definition

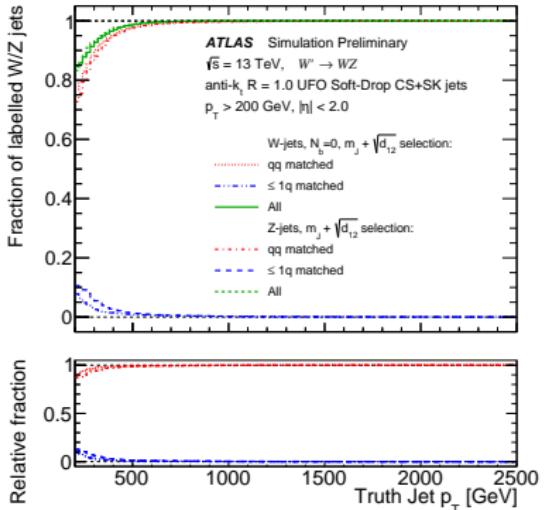


2-variable top tagger



W/Z Taggers

Truth Labelling



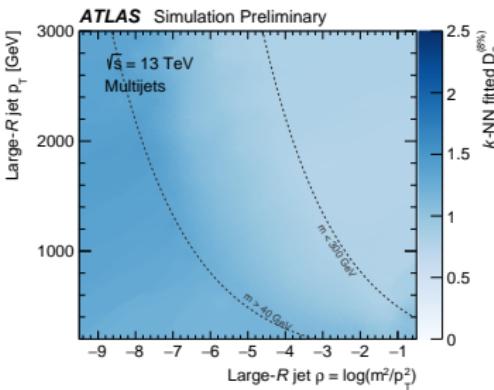
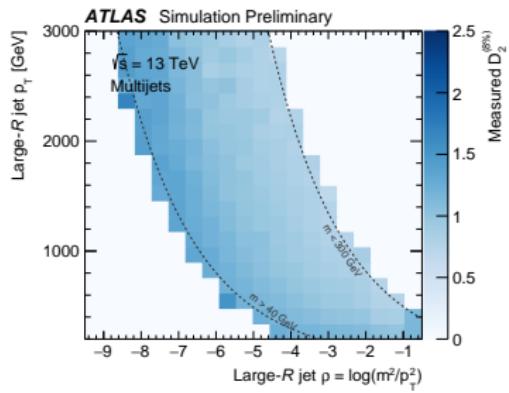
Truth jet definition:

- Parton level MC to label jets as:
 - Signal: Containing full $W \rightarrow q\bar{q}$
 - Background: From single q/g
- Truth jets: Reconstructed from stable particles, anti- k_T $R = 1.0$
 - No grooming applied to ensure independence of grooming algorithm
- Requirements for truth signal jets:
 - Truth W/Z within $\Delta R < 0.75$
 - $m_J > 50$ GeV
 - $\sqrt{d_{12}} > 55.25 + e^{-2.35 \frac{p_T}{\text{GeV}}} \text{ GeV}$
 - $N_B = 0$ for W jets to reduce top contamination
- Matched to UFO jets with $\Delta R < 0.75$
- Optimised for $\epsilon_{\text{sig}} = 85\%$ at $p_T = 200$ GeV, 100% at $p_T = 300$ GeV

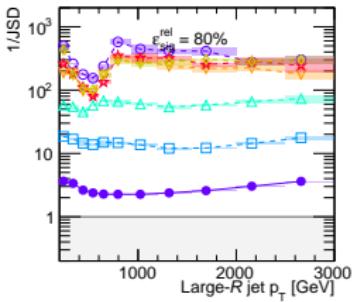
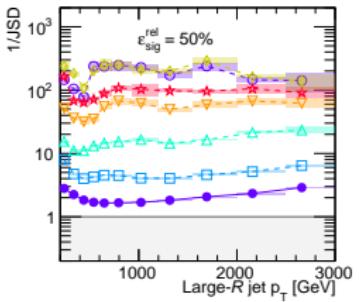
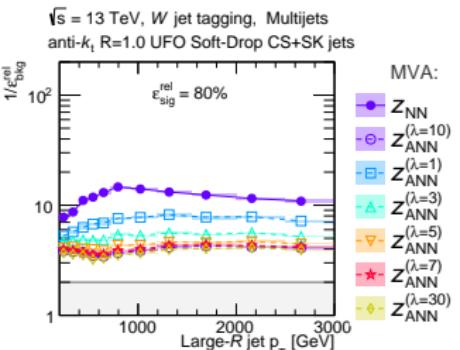
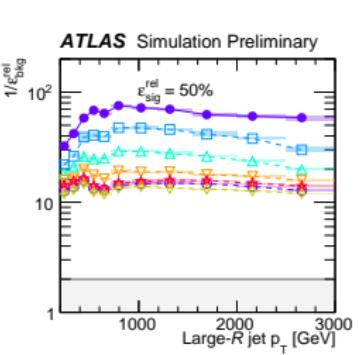
Matching to daughter quarks q, q' :

- Fraction of W (Z) containing both q, q' within $\Delta R < 0.75$:
 $> 98\%$ (96%) at $p_T > 300$ GeV
 (previously: 90%)

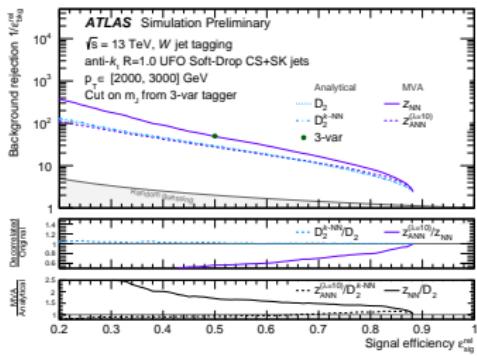
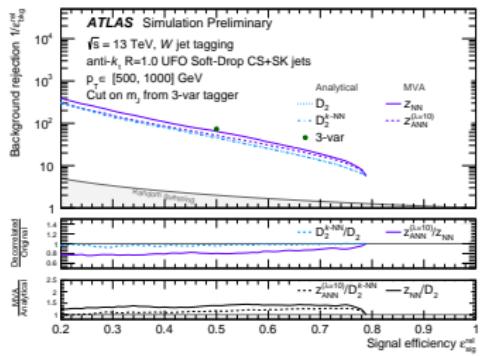
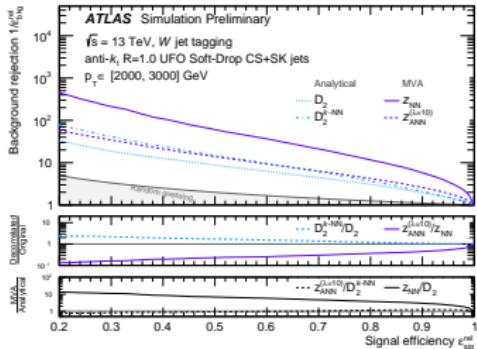
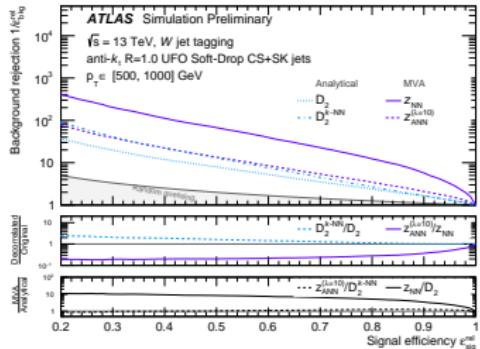
W/Z Tagger: k-NN Method



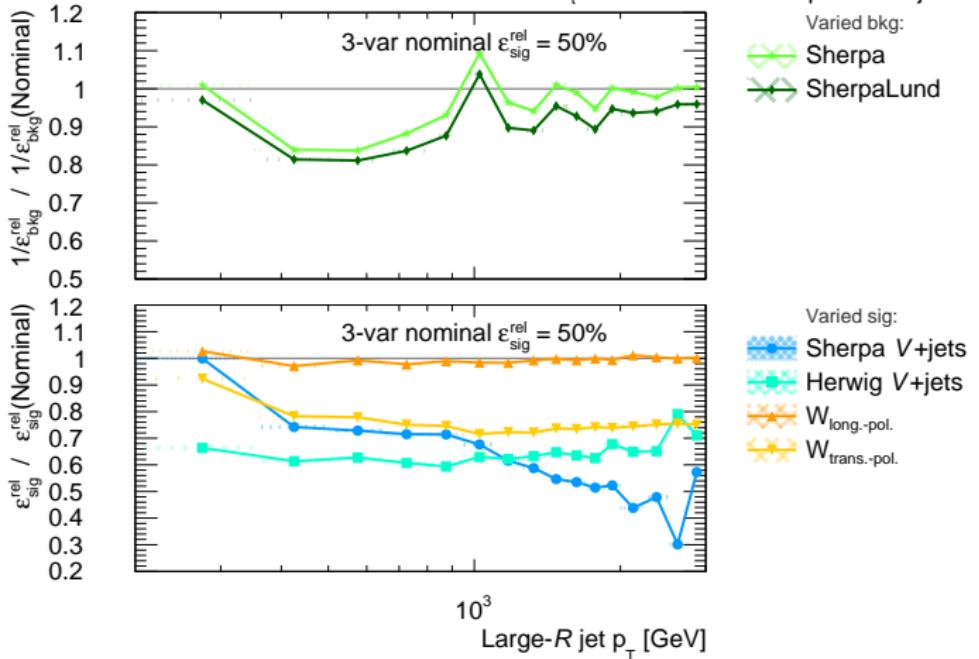
W/Z Tagger: Effect of λ



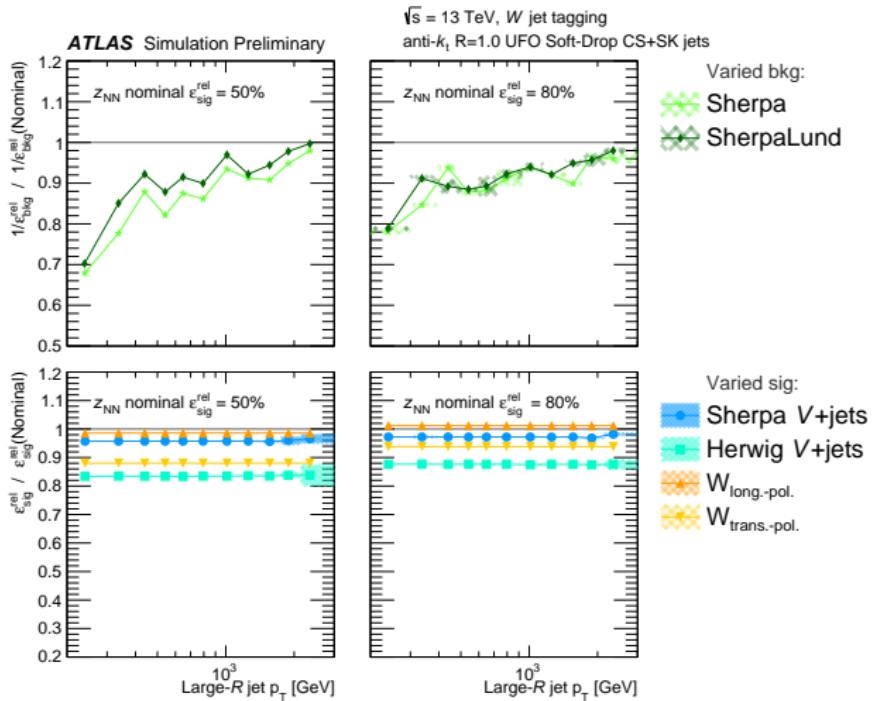
W/Z Tagger: ROC Curves



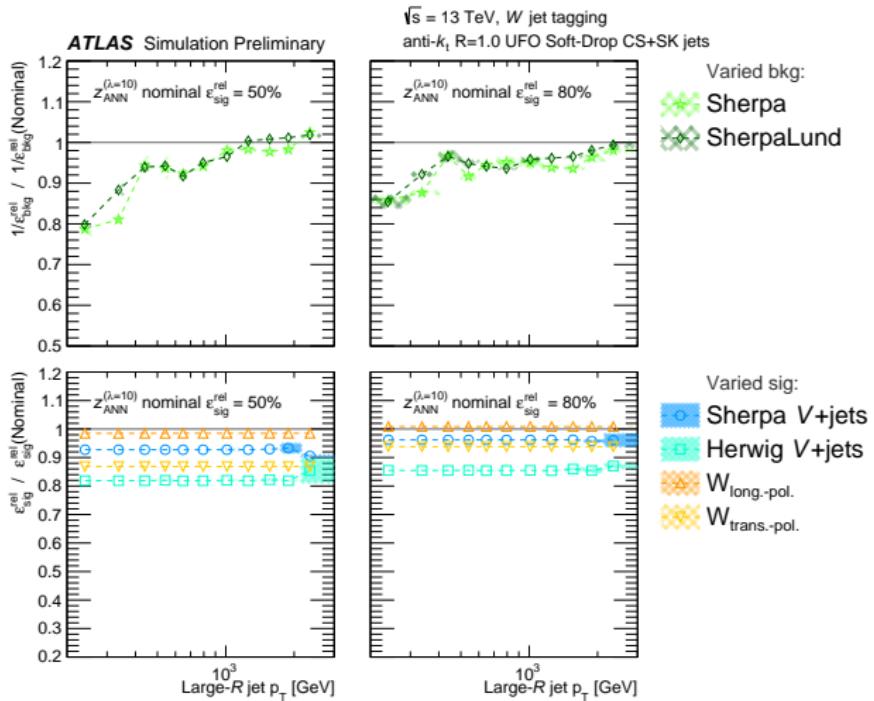
ATLAS Simulation Preliminary $\sqrt{s} = 13 \text{ TeV}$, W jet tagging
 anti- k_t R=1.0 UFO Soft-Drop CS+SK jets



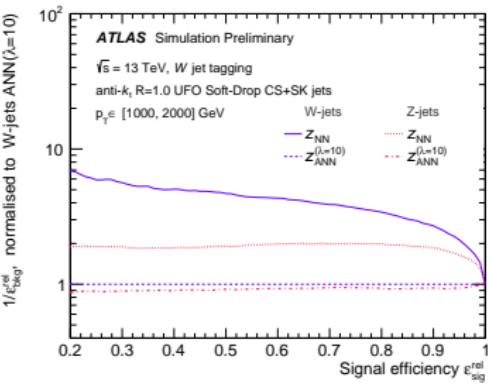
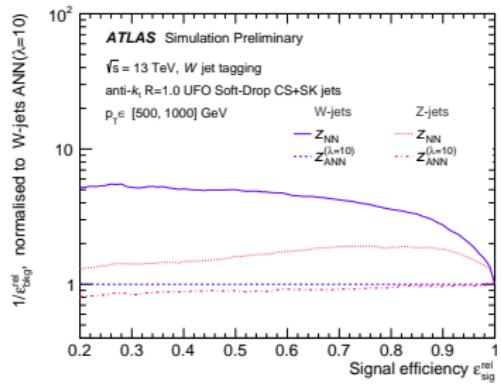
W/Z Tagger: Modelling



W/Z Tagger: Modelling

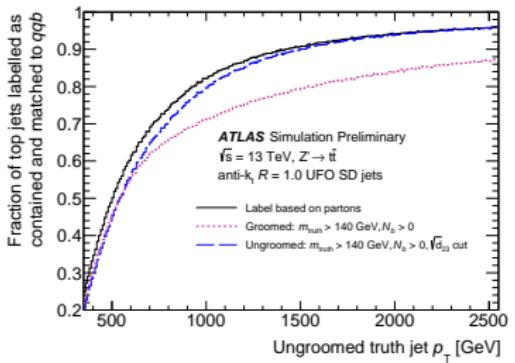
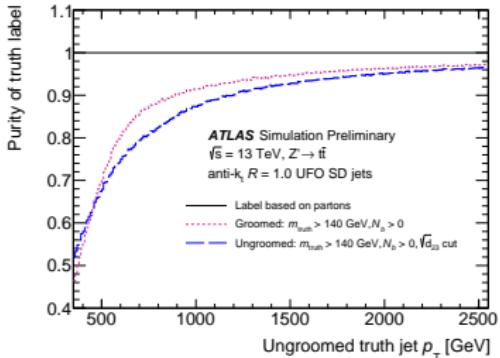
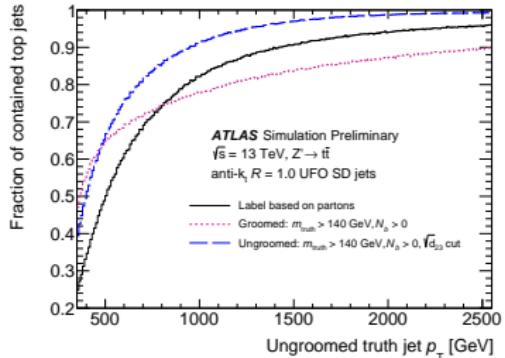


W/Z Tagger: Z vs W



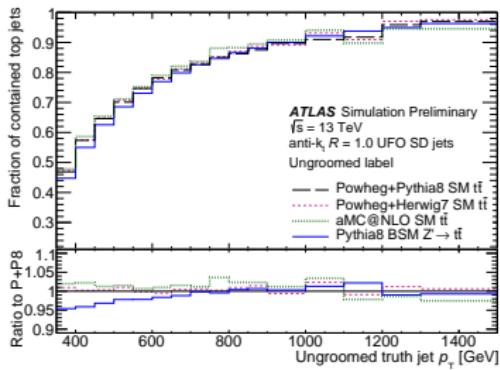
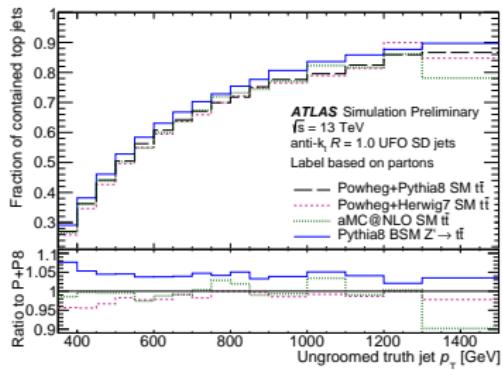
Top Taggers

DNN Top Tagger: Truth Labelling

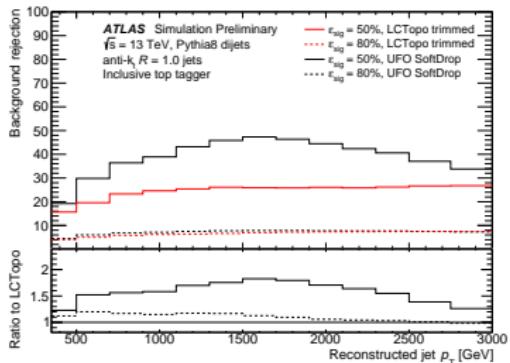
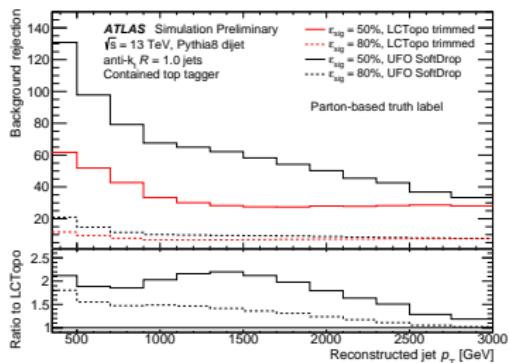
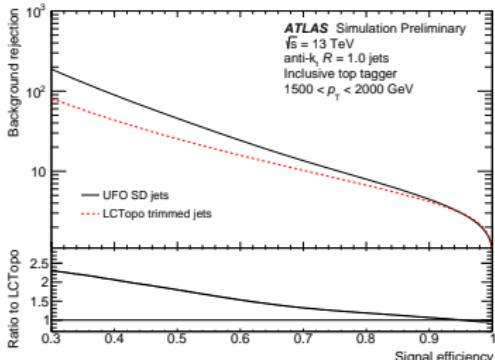
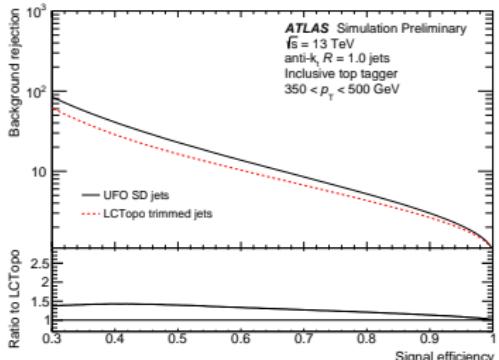


ATL-PHYS-PUB-2021-028

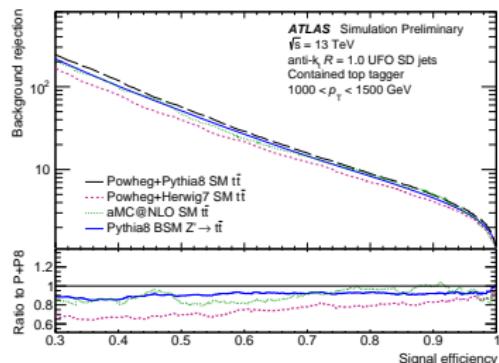
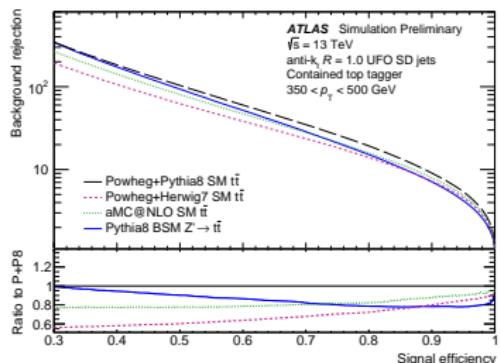
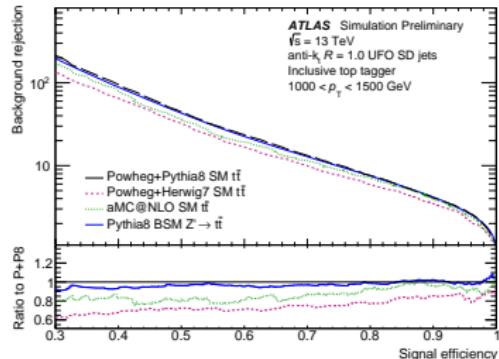
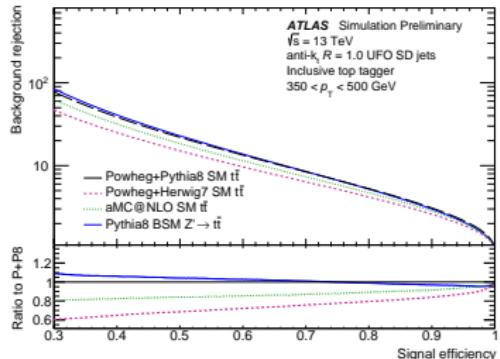
DNN Top Tagger: Truth Labelling



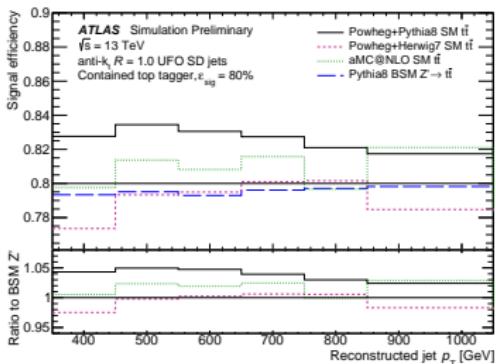
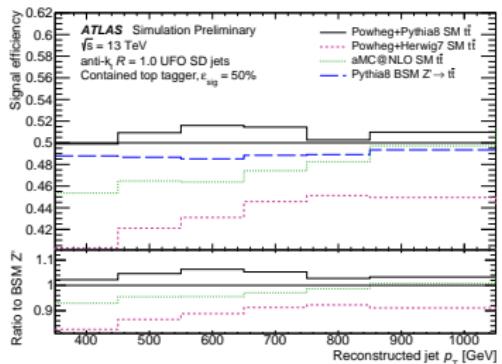
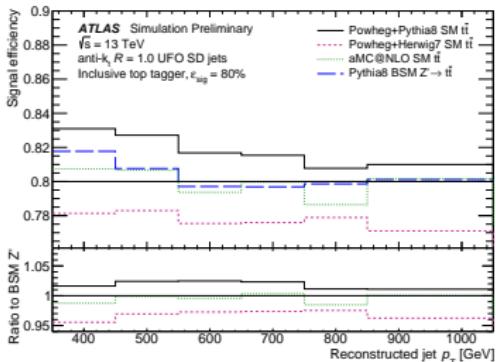
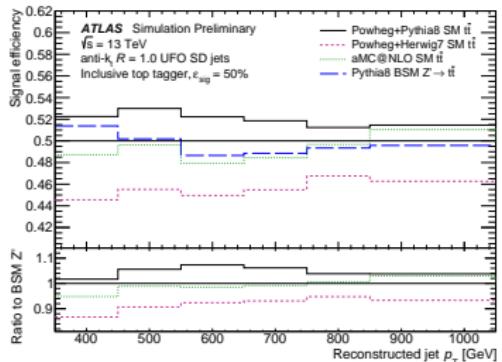
DNN Top Tagger: ROC Curves



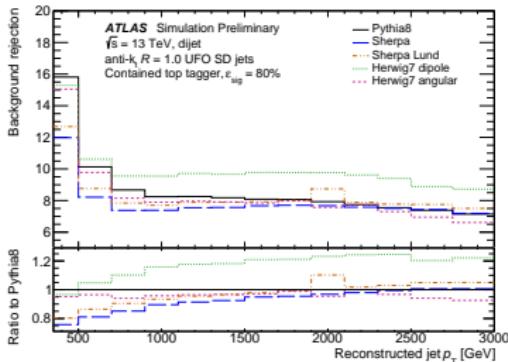
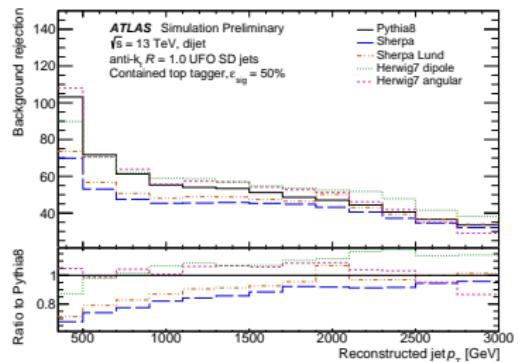
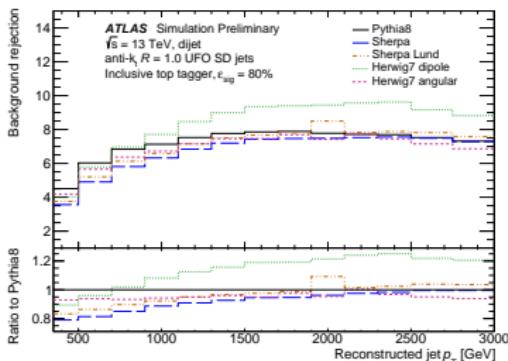
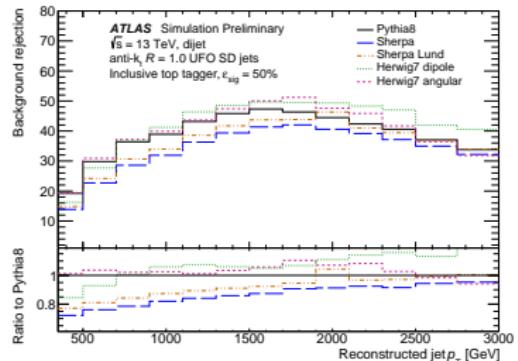
DNN Top Tagger: Modelling



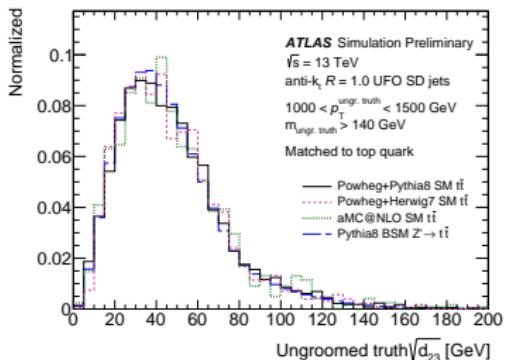
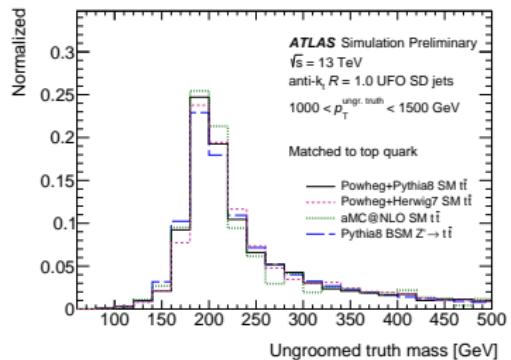
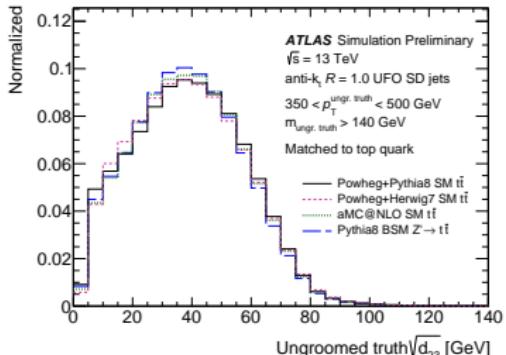
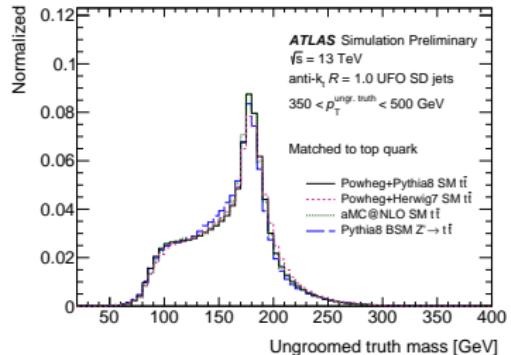
DNN Top Tagger: Modelling



DNN Top Tagger: Modelling



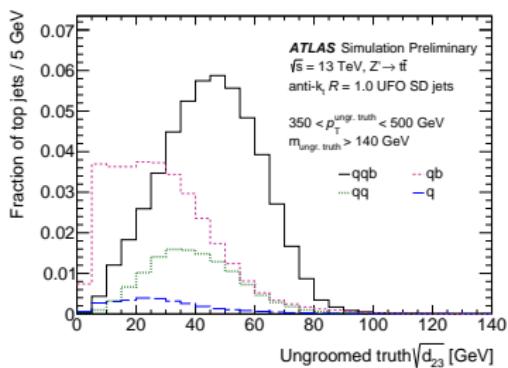
DNN Top Tagger: Modelling for Truth Labelling



DNN Top Tagger: Truth Labelling

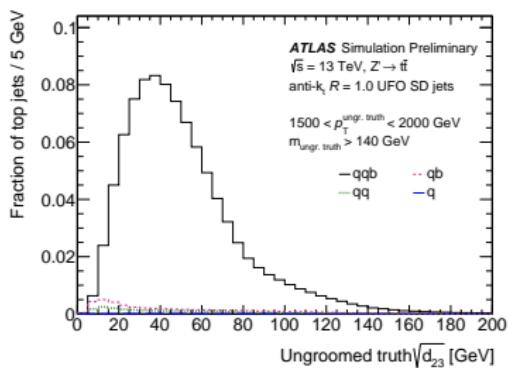
Low p_T :

→ optimal cut: $\sqrt{d_{23}} \approx 21$ GeV



High p_T :

→ optimal cut: $\sqrt{d_{23}} \approx 7$ GeV



Constituent-Based Top Tagger: Performance

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
hIDNN	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

Substructure Variables

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	$\text{ECF}_1, \text{ECF}_2, \text{ECF}_3$	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 -subjett ΔR		
n_{trk}	number of tracks		

arxiv.org/abs/1305.0007

$$\text{ECF}(N, \beta) = \sum_{i_1 < i_2 < \dots < i_N \in J} \left(\prod_{a=1}^N p_{T,i_a} \right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^N R_{i_b i_c} \right)^\beta$$

N constituents i of the jet J with Euclidian distance:

$$R_{i_b i_c} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

- IRC (infrared & collinear) safe $\forall \beta > 0$
 - Goes to $\rightarrow 0$ in infrared/collinear limit

Here: ECF₁, ECF₂, ECF₃;

Energy Correlation Ratios C_2 , D_2

[doi.org/10.1007/JHEP12\(2014\)009](https://doi.org/10.1007/JHEP12(2014)009)

Normalised ECFs e_n^β :

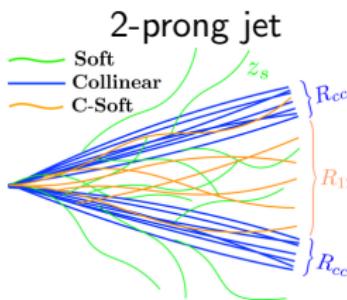
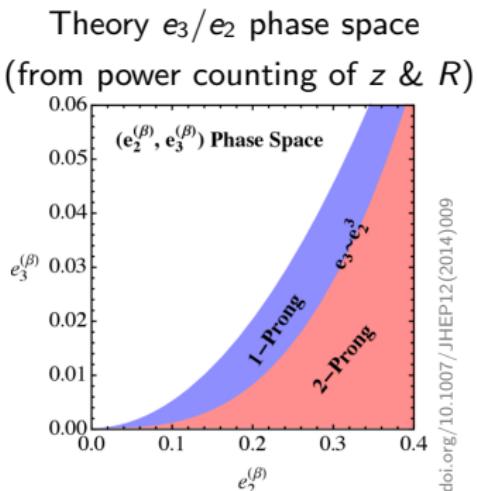
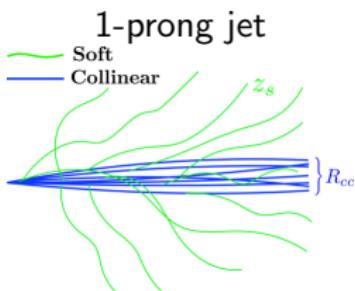
$$e_n^\beta = \frac{\text{ECF}(n, \beta)}{\text{ECF}(1, \beta)^n}; \quad z_i = \frac{p_{T_i}}{p_{T_J}}$$

$$\Rightarrow e_2^\beta = \sum_{1 \leq i \leq j \leq n_J} z_i z_j R_{ij}^\beta$$

$$\Rightarrow e_3^\beta = \sum_{1 \leq i \leq j \leq k \leq n_J} z_i z_j z_k R_{ij}^\beta R_{ik}^\beta R_{jk}^\beta$$

Ratios of e_n^β :

$$C_2 = \frac{e_3^\beta}{\left(e_2^\beta\right)^2}, \quad D_2 = \frac{e_3^\beta}{\left(e_2^\beta\right)^3}$$



→ C_2 and D_2 Separate 1- and 2-prong jets on the e_3/e_2 plane

N-Subjettiness τ_N

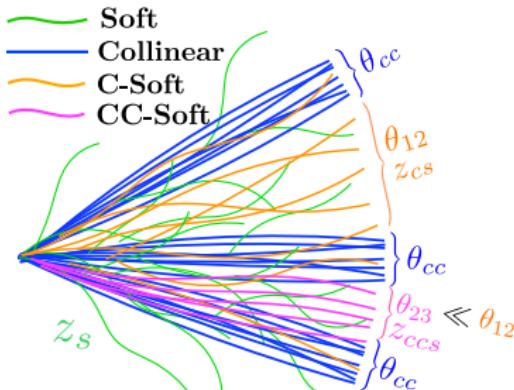
<https://arxiv.org/abs/1011.2268>

$$\tau_N^\beta = \sum_{1 \leq i \leq n_J} z_i \min \left\{ R_{i1}^\beta, \dots, R_{iN}^\beta \right\}$$

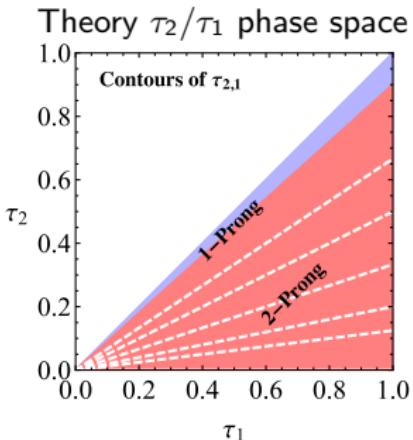
$$\text{with } z_i = \frac{p_{T_i}}{p_{T_J}}$$

$$\rightarrow \tau_{21}^\beta = \frac{\tau_2^\beta}{\tau_1^\beta}$$

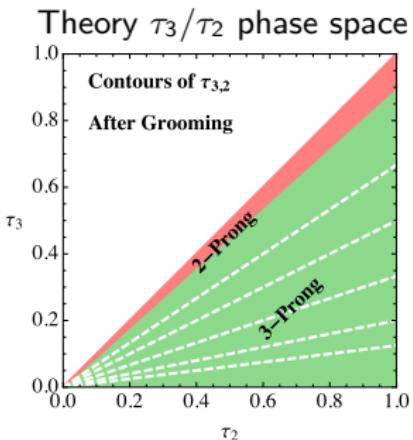
3-prong jet



arxiv.org/abs/1609.07483



arxiv.org/abs/1609.07483



arxiv.org/abs/1609.07483

arxiv.org/abs/1609.07483

$${}_\nu e_N^\beta = \sum_{i_1 < i_2 < \dots < i_N \in J} \left(\prod_{a=1}^N p_{T_{i_a}} \right) \left(\prod_{m=1}^{\nu} \min_{s < t \in \{i_1, i_2, \dots, i_N\}} R_{st} \right)^\beta$$

Where $\min_X^{(m)}$ denotes the m th smallest element in the set X

Reduces to nominal ECF in the case $\nu = \binom{N}{2}$:

$$\text{ECF}(N, \beta) = \sum_{i_1 < i_2 < \dots < i_N \in J} \left(\prod_{a=1}^N p_{T_{i_a}} \right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^N R_{i_b i_c} \right)^\beta$$

→ ${}_\nu e_N^\beta$ are sensitive to hierarchy of distinct angular (R) scales m in jet

- ECF average over them

Ratios to separate 2- & 3-prong jets: $L_2 = \frac{{}_3 e_3^{\beta=1}}{{}_1 e_2^{\beta=2}}^{\frac{3}{2}}$, $L_3 = \frac{{}_1 e_3^{\beta=1}}{{}_3 e_3^{\beta=1}}^{\frac{1}{3}}$

Number of Ghost-Associated Tracks n_{trk}



JHEP04(2008)005

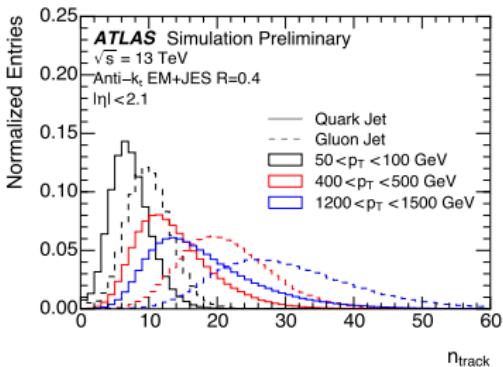
n_{trk} : number of tracks

- With $p_T > 500 \text{ MeV}$
 - Ghost-associated to jet
- Powerful q/g discriminant

Ghost-associated jet area

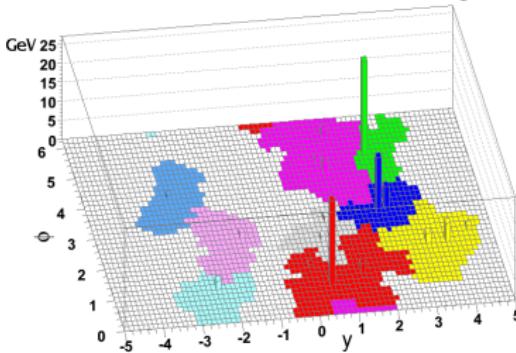
- Add dense coverage of 'infinitely' soft 'ghost' constituents
- Count how many are clustered within the jet

n_{trk} as q/g discriminant



ATL-PHYS-PUB-2017-009

Ghost associated areas of k_t jets



JHEP04(2008)005