Classifying hadronic objects in ATLAS with ML/AI algorithms

Jad M. Sardain on behalf of the ATLAS Collaboration

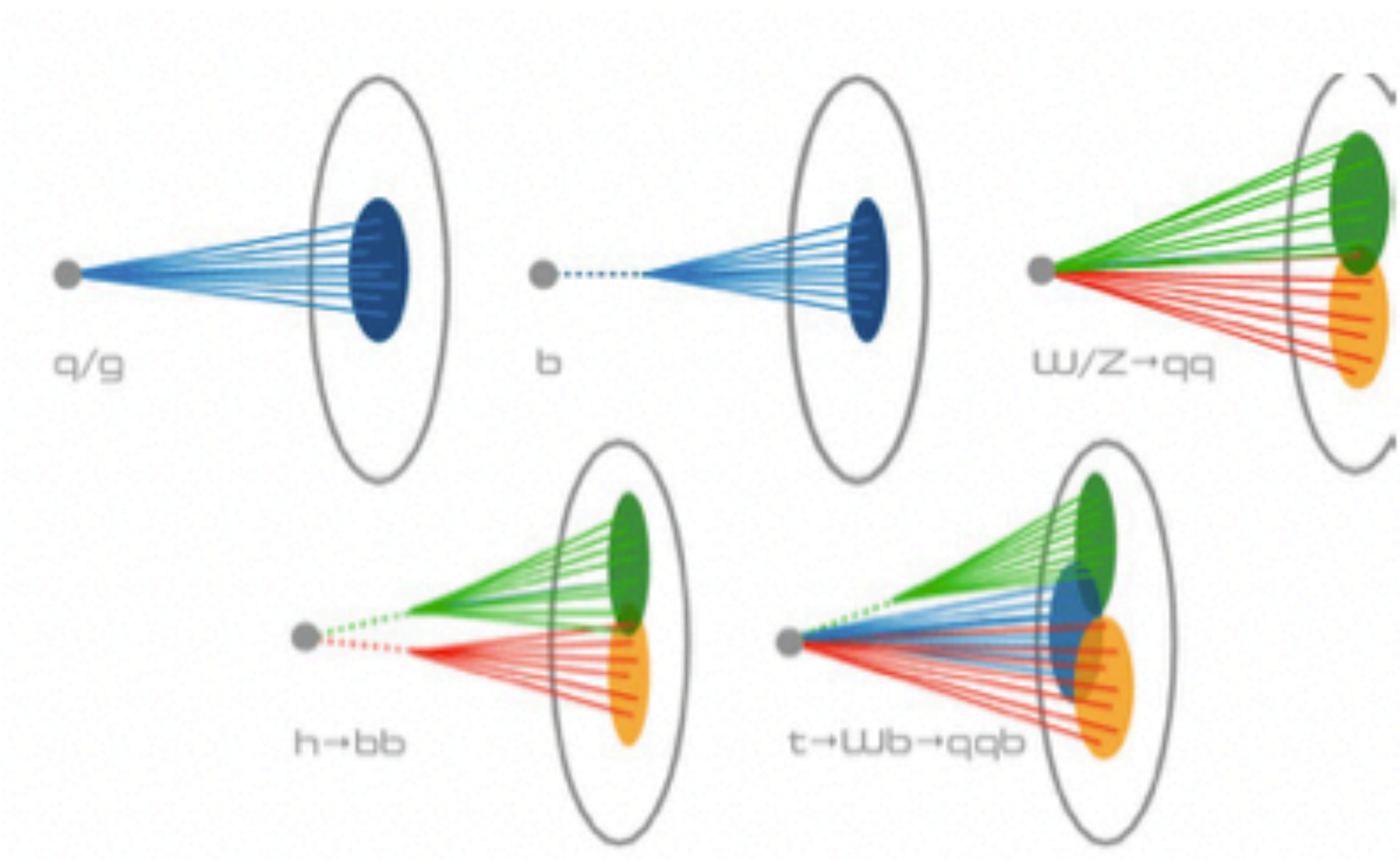
BOOST2023, July 31 - August 4 2023

Hadronic jet tagging: overview

- Taggers can improve SM measurements and the sensitivity of searches for new resonances in BSM

- Various approaches include:
  - Combine various high-level variables
  - Use low-level constituents
  - Jet images
  - Jet clustering sequence

- This talk will focus on: **1)** W-tagging **2)** q/g tagging **3)** Hbb/cc tagging



W tagging

Hadronic jet tagging: overview

ATL-PHYS-PUB-2021-029

- Two taggers are considered as baseline taggers and are used for comparisons :
  - **The "so-called" 3-var tagger**, using cuts on number of tracks associated to the un-groomed large-R jet, jet mass and the energy correlation ratio (D_2)

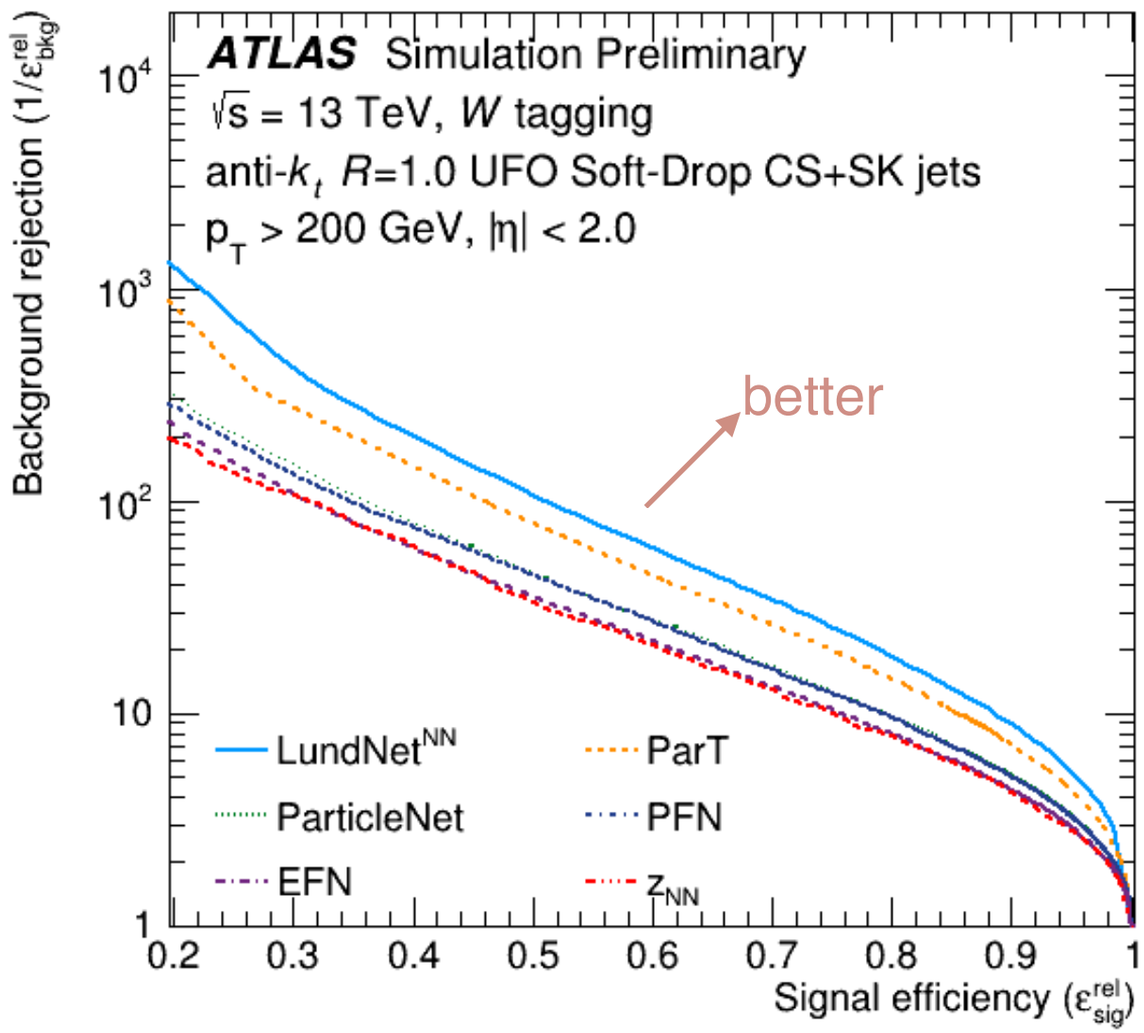
  - **The DNN/ANN tagger,** based on high level jet substructure observables

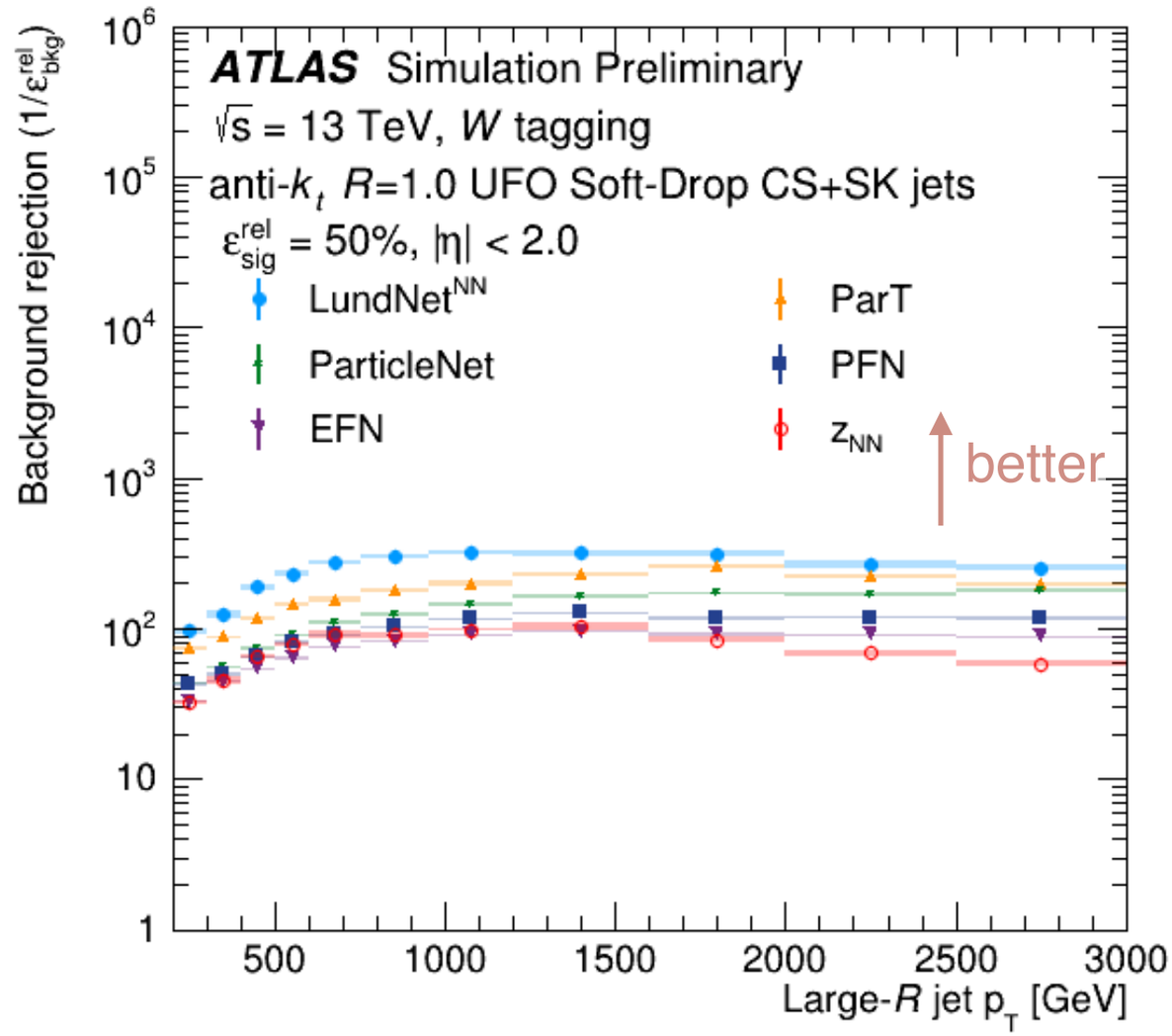
- All of these taggers:
  - **Signal:** W bosons from simulated $W' \to WZ(\to q\bar{q}q\bar{q})$ events with $m_{W'}$ = 2 TeV, Pythia8 + NNPDF2.3LO + A14 tune.
  - **Background:** QCD di-jet events @ LO, Pythia8 + NNPDF2.3LO + A14 tune.

- Large-R jets are reconstructed from Unified Flow Objects (UFO) with the radius parameter R = 1.0.

- Different taggers are considered with information coming from:
  1) jet constituents (ParticleTransformer, ParticleNet, PFN, EFN) or **2)** Lund jet plane (LundNet)

Different W taggers: performance





- The LundNet tagger shows the best performance, followed by constituent based taggers, then zNN

- For a $\varepsilon_{sig}^{rel} = 50\,\%$, the background rejection of **ParticleTransformer(LundNet)** is roughly **1.8–2.8(3)** times better than the baseline tagger.

- Let's go through these taggers!

Constituent based W tagger

ATL-PHYS-PUB-2023-020

- Try to maximize the use of the jet constituents' information using state-of-the-art ML/DL algorithms

Jet Requirements	W jet requirements
Jet $\lvert\eta\rvert < 2.0$	dR(truth jet, truth W) < 0.75
Jet $p_{\mathrm{T,truth}} > 200$ GeV	Ungroomed truth jet mass > 50 GeV
Number of constituents ≥ 2	Number ghost associated b-hadrons $= 0$
Jet mass > 40 GeV	Truth jet $\sqrt{d_{12}} > 55.25 \times \exp(-2.34 \times 10^{-3} \times$ truth jet $p_{\mathrm{T}})$
dR(jet, truth jet) < 0.75	

- Multiple models are used, each one taking different inputs

Models	Features
EFN	$\Delta\eta, \Delta\phi, \ln p_{\mathrm{T}}$
PFN	$\Delta\eta, \Delta\phi, \ln p_{\mathrm{T}}, \ln E, \ln\frac{p_{\mathrm{T}}}{\sum_{jet} p_{\mathrm{T}}}, \ln\frac{E}{\sum_{jet} E}, \Delta R$
ParticleNet	$\Delta\eta, \Delta\phi, \ln p_{\mathrm{T}}, \ln E, \ln\frac{p_{\mathrm{T}}}{\sum_{jet} p_{\mathrm{T}}}, \ln\frac{E}{\sum_{jet} E}, \Delta R$
ParticleTransformer	$\Delta\eta, \Delta\phi, \ln p_{\mathrm{T}}, \ln E, \ln\frac{p_{\mathrm{T}}}{\sum_{jet} p_{\mathrm{T}}}, \ln\frac{E}{\sum_{jet} E}, \Delta R$ (E, p_x, p_y, p_z)

JHEP 01(2019) 121

Phys.Rev.D 101 (2020) 5, 056019

arxiv: 2202.03772

- ParticleTransformer achieves the best performance, followed by the ParticleNet, PFN, and EFN

- All the constituent-based taggers outperform the high-level-feature-based tagger (noted as z_{NN} in the figure), since benefitting from additional information containing jet constituents

Jet tagging using Lund plane and ML

ATL-PHYS-PUB-2023-017

- An abstract representation of the jet formation, initially developed by theorists to better understand it

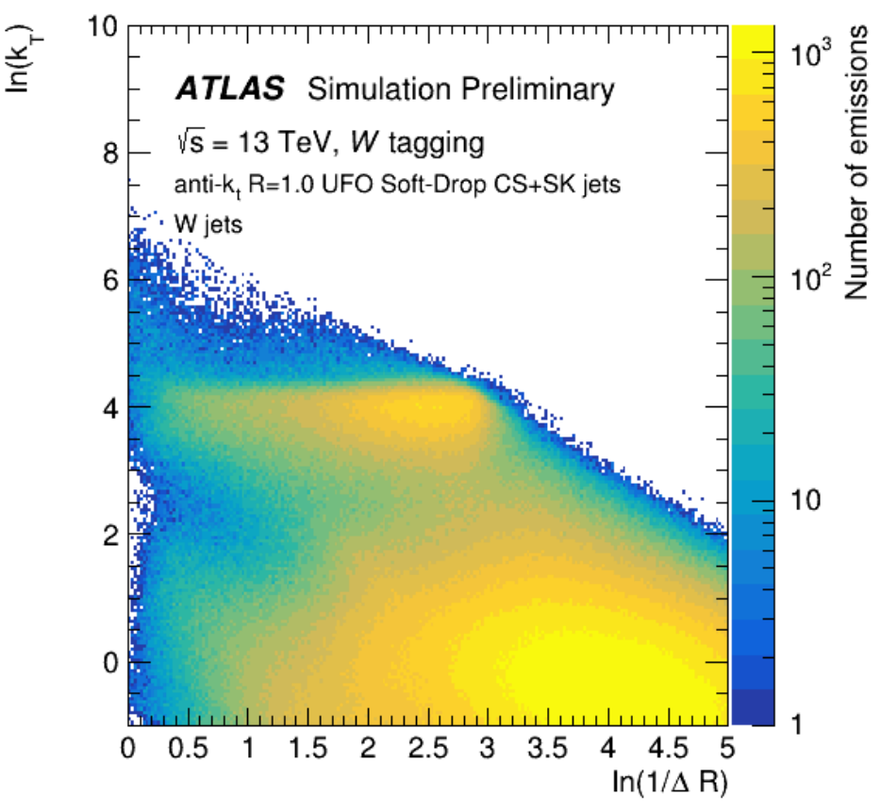
- Each emission represented by a point in the kT-emission angle plane (log scale), where different processes populate different regions of the plane

- Experimentally, we can have an approximate reconstruction of the Lund Plane by running back the C/A jet clustering and using the jet merging information

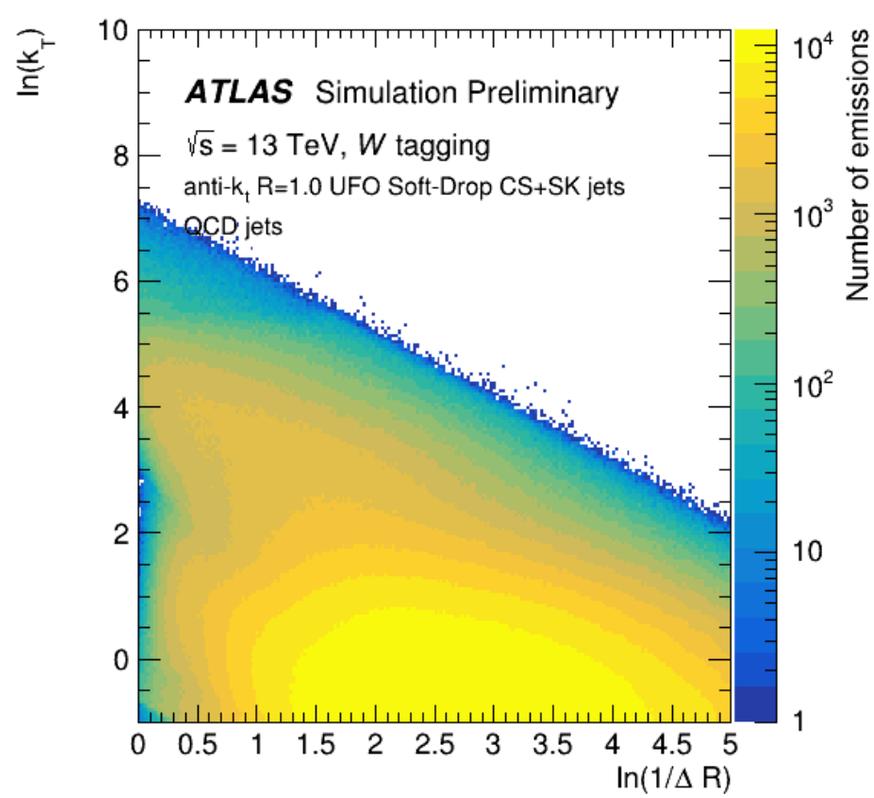
- One can apply a CNN but **a jet is more than an image**, the Lund plane has much more information coming from the sequence that produced it
    - Use GNN where each node has 3 vars: ***arXiv:2012.08526v2 [hep-ph]***
        z momentum fraction of the branching, k_t transverse momentum, Δ emission angle

Number of tracks per jet as a global feature to help classification

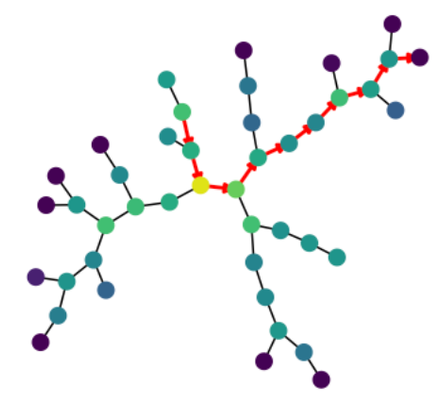
Signal



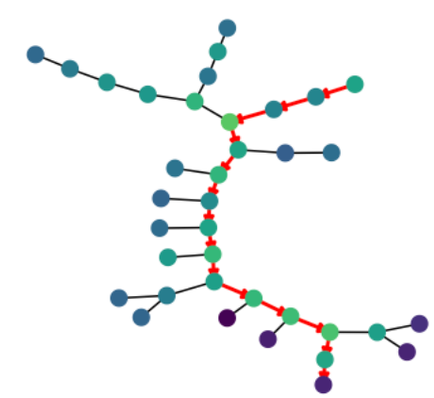
Background



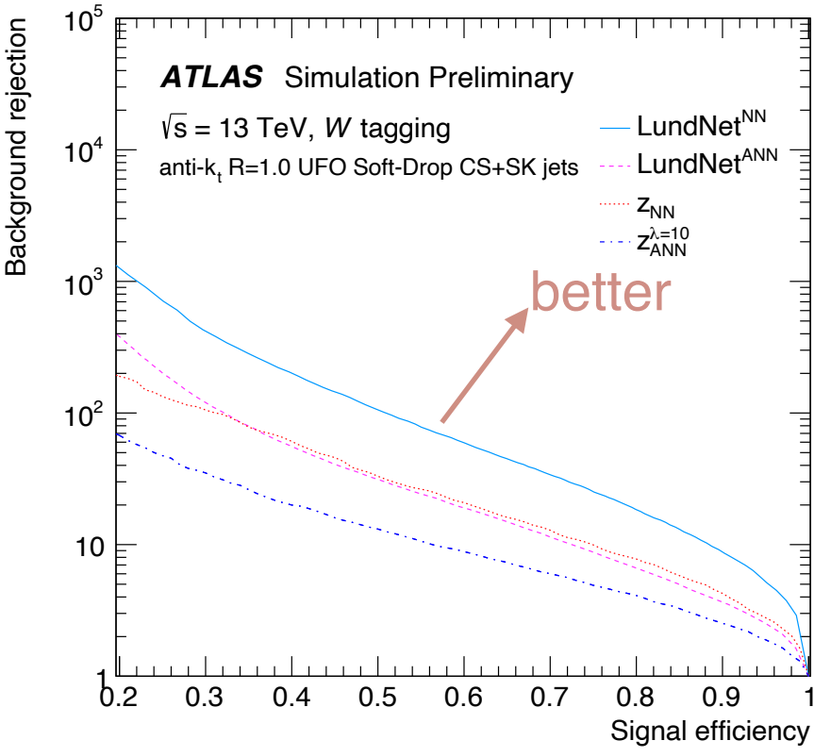
W-jet

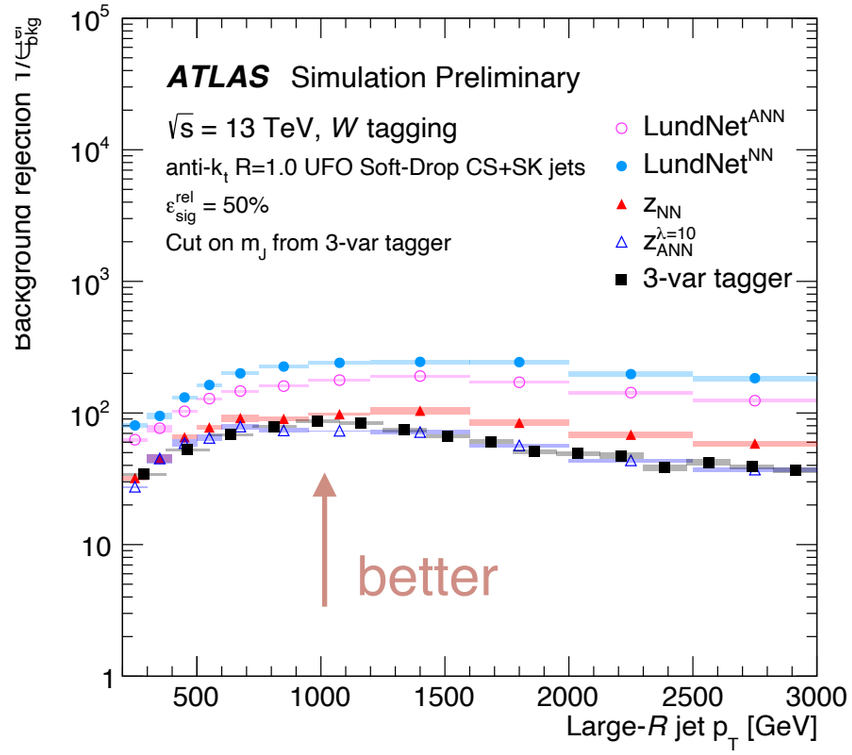


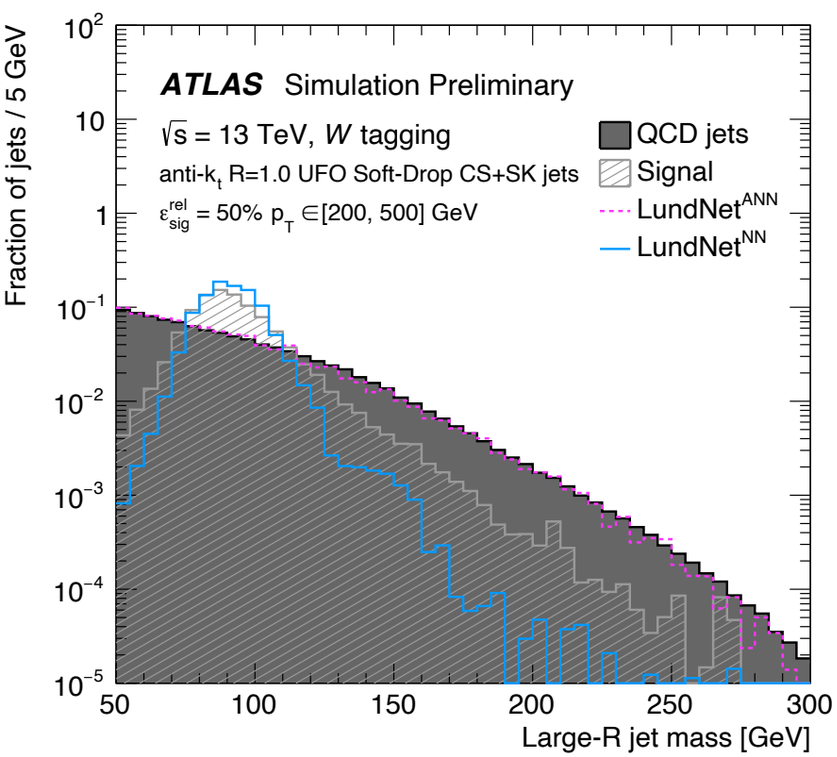
QCD

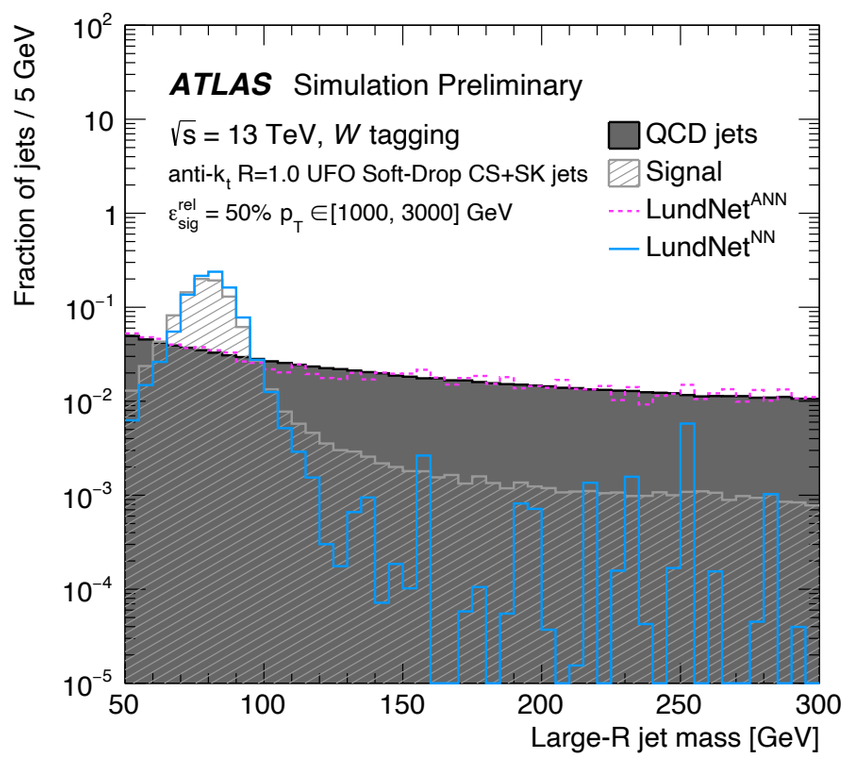


LundNet









- LundNet^{NN} achieves the best performance

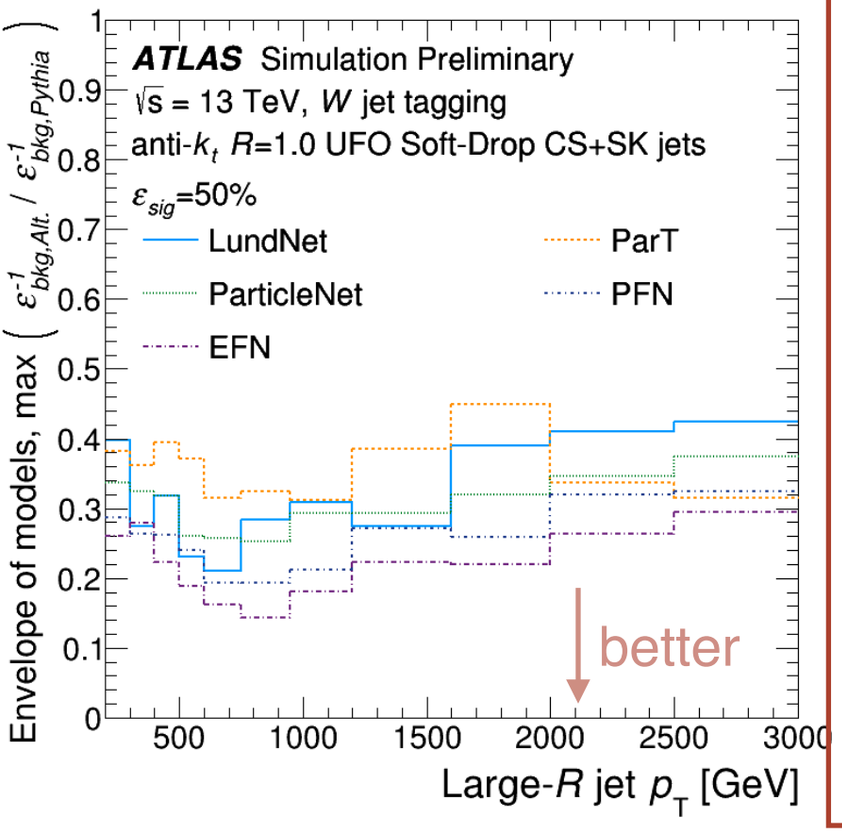
- The adversarial network (for mass decorrelation) significantly deteriorates performance (for both LundNet and DNN)

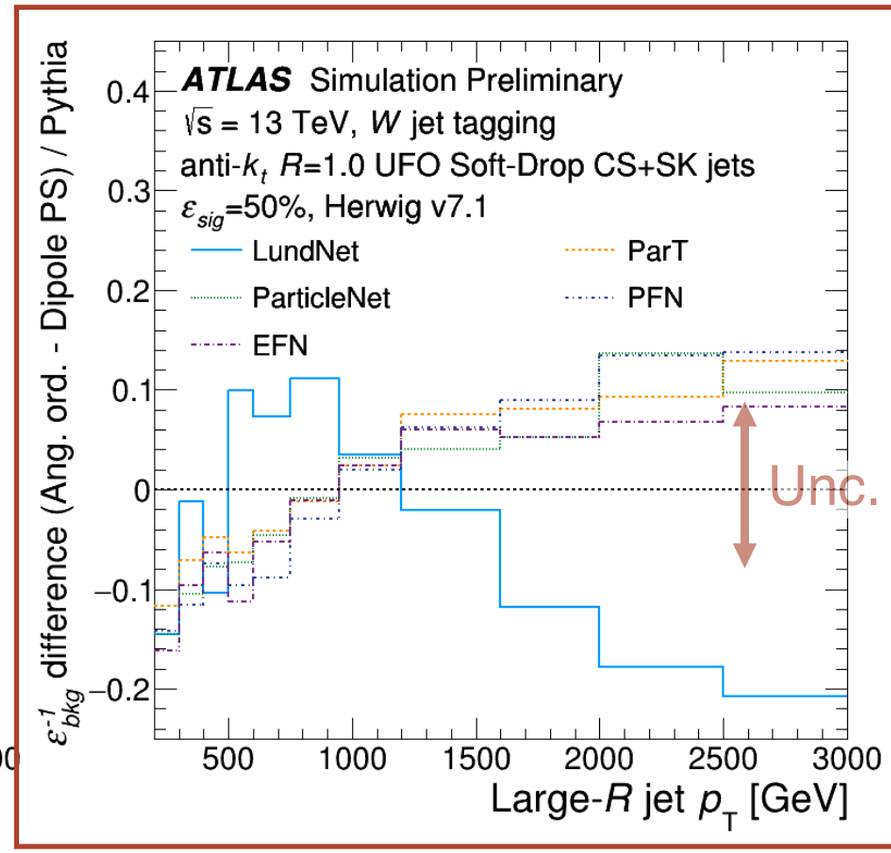
- At 50% signal efficiency and with the p_T-dependent 3-var tagger mass cut, the background rejection, after mass decorrelation, is **better** by a factor of **2.5(3)** with respect to the **3-var tagger (baseline ANN tagger)**

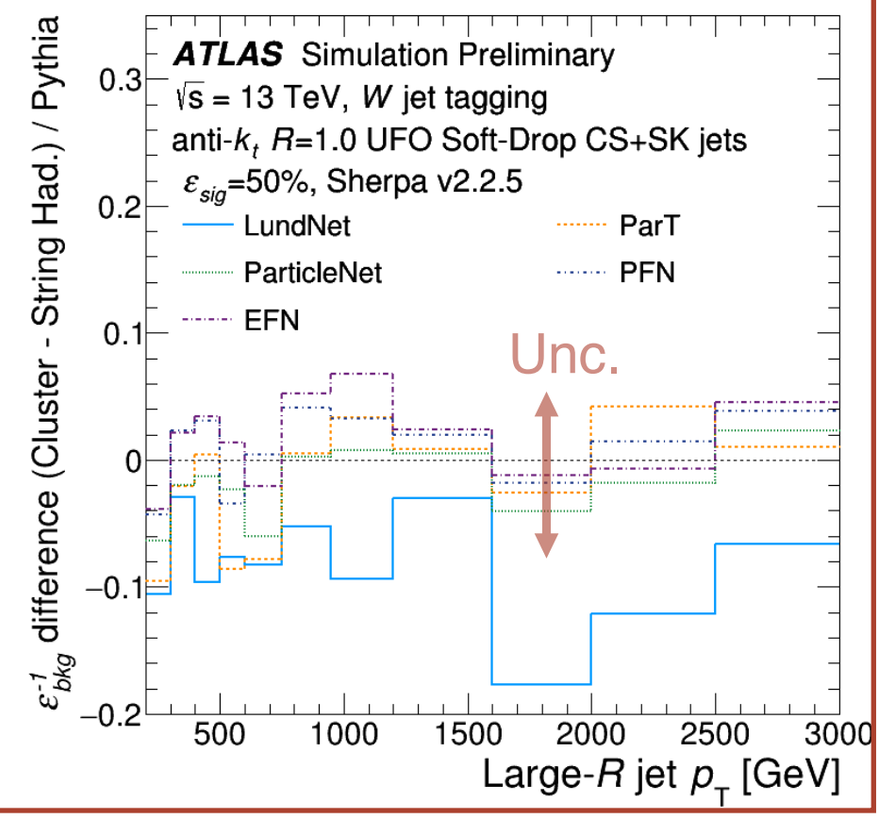
- Across all p_T ranges:
LundNet^{NN} retrieves W mass peak
LundNet^{ANN} retrieves QCD shape

- Kullback–Leibler (KL) divergence:
< 1% for both comparison:
LundNet^{NN} with signal
LundNet^{ANN} with QCD

Comparison among the W taggers: model dependence







- The **more complex the model** is/inputs are, the **more it is affected by modeling** uncertainties

- Envelope (high values = worse performance), for 50% working point, LundNet shows effects in background rejection of up to 40% whereas EFN shows a flat 30% effects across the whole p_T range

- Herwig with angle ordered parton shower has a higher contribution from soft collinear emission than Herwig with dipole parton shower.

- Higher contribution in the region factorizing the hard collinear emission for Sherpa with string model than Sherpa using the cluster model

- Understanding the source of this model-dependence is crucial in order to generalize the performance of these complex and new taggers

q/g tagging

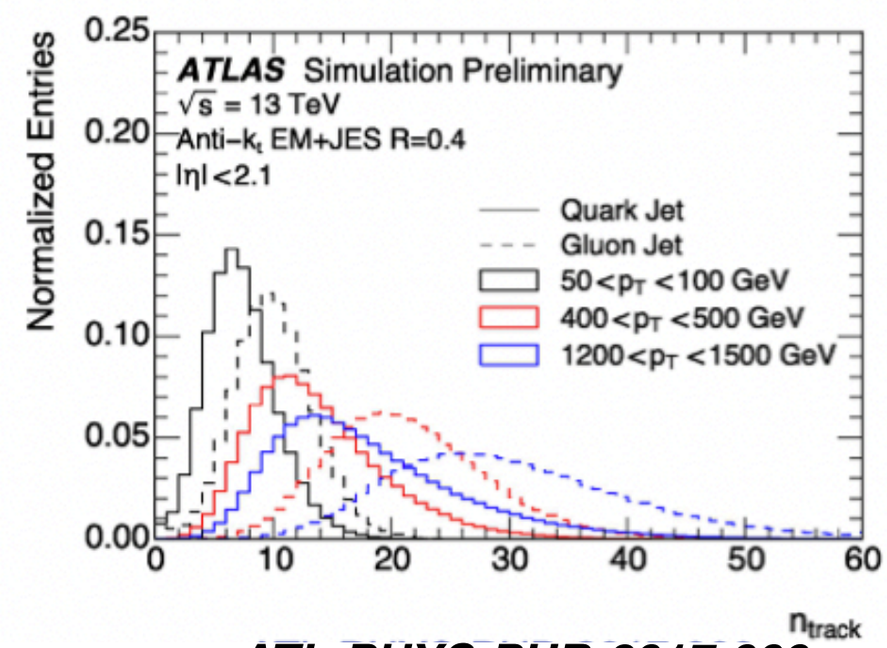
JETM-2020-02-002

arXiv coming soon

Strategy

- Quark and gluon jets are difficult to distinguish and many analyses need to know the origin of the jet

- Gluon jets tend to be wider and have more charged constituents than quark jets



ATL-PHYS-PUB-2017-009

- Using at least two R=0.4 PFlow jets with $p_T > 500$ GeV and $|\eta| < 2.1$

- Truth jets are matched with ΔR<0.4 requirement
- Truth flavor label of a jet is defined by the flavour of the highest-energy parton in the parton shower

- 2 taggers are defined:
  - using the number of tracks Ntrk
  - using a BDT with the following input features: Ntrk, w_{trk}, $C^{\beta=0.2}_{1,track}$, p_T

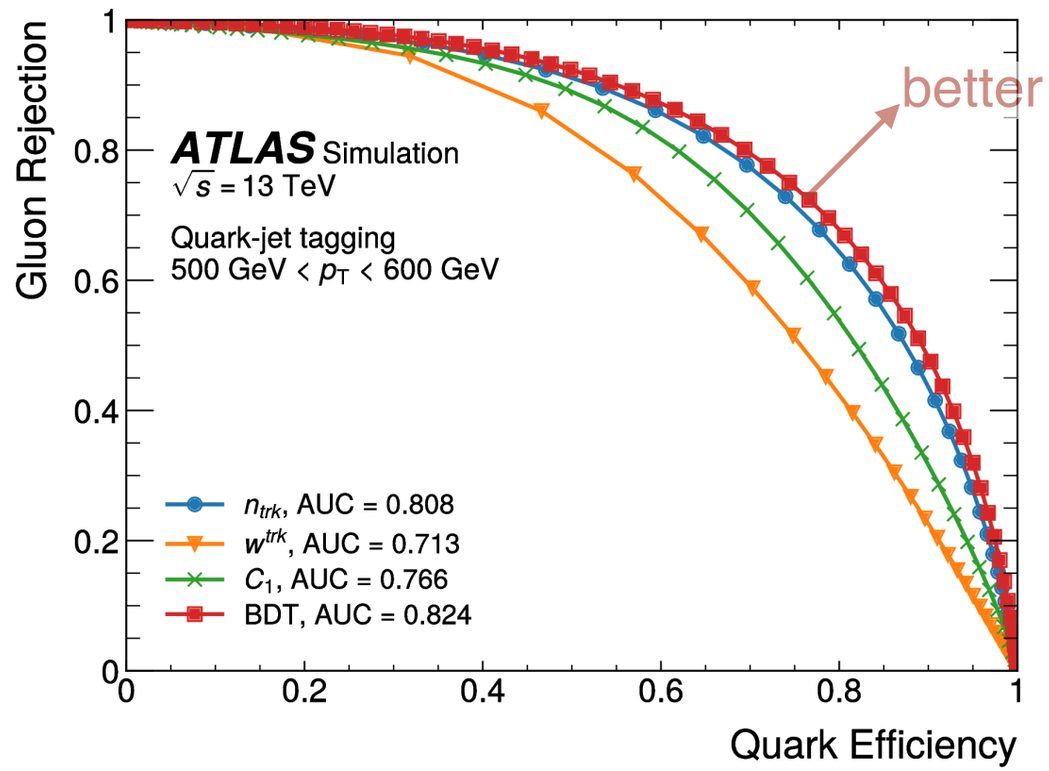
With

$$w_{\text{track}} = \frac{\sum_{i\in\text{Jet}} p_{\text{T},i}\Delta R_{i,\text{Jet}}}{\sum_{i\in\text{Jet}} p_{\text{T},i}}, \text{ tracks } i$$

$$C^{\beta=0.2}_{1,\text{ track}} = \frac{\sum_{i,j\in\text{Jet}} p_{\text{T},i}p_{\text{T,j}}(\Delta R_{i,j})^{\beta}}{\left(\sum_{i\in\text{Jet}} p_{\text{T},i}\right)^2}, \text{ tracks } i,j$$

Performance

- Ntrk, w_{trk}, $C_{1,track}^{\beta=0.2}$ show discrimination power but the BDT outputs the highest AUC and is better than Ntrk tagger across the whole jet p_T range



- A matrix method is then applied to calibrate such taggers. It takes as input
  - the distributions for the taggers in the forward/central region (from MC and data),
  - the fraction of quark/gluons in the forward/central region (from Pythia)

$$\begin{pmatrix} p_F(x) \\ p_C(x) \end{pmatrix} = \underbrace{\begin{pmatrix} f_{F,Q} & f_{F,G} \\ f_{C,Q} & f_{C,G} \end{pmatrix}}_{\equiv F} \begin{pmatrix} p_Q(x) \\ p_G(x) \end{pmatrix}.$$

- A reweighing procedure is applied as well to central jets since detector effects change the radiation pattern inside the jets

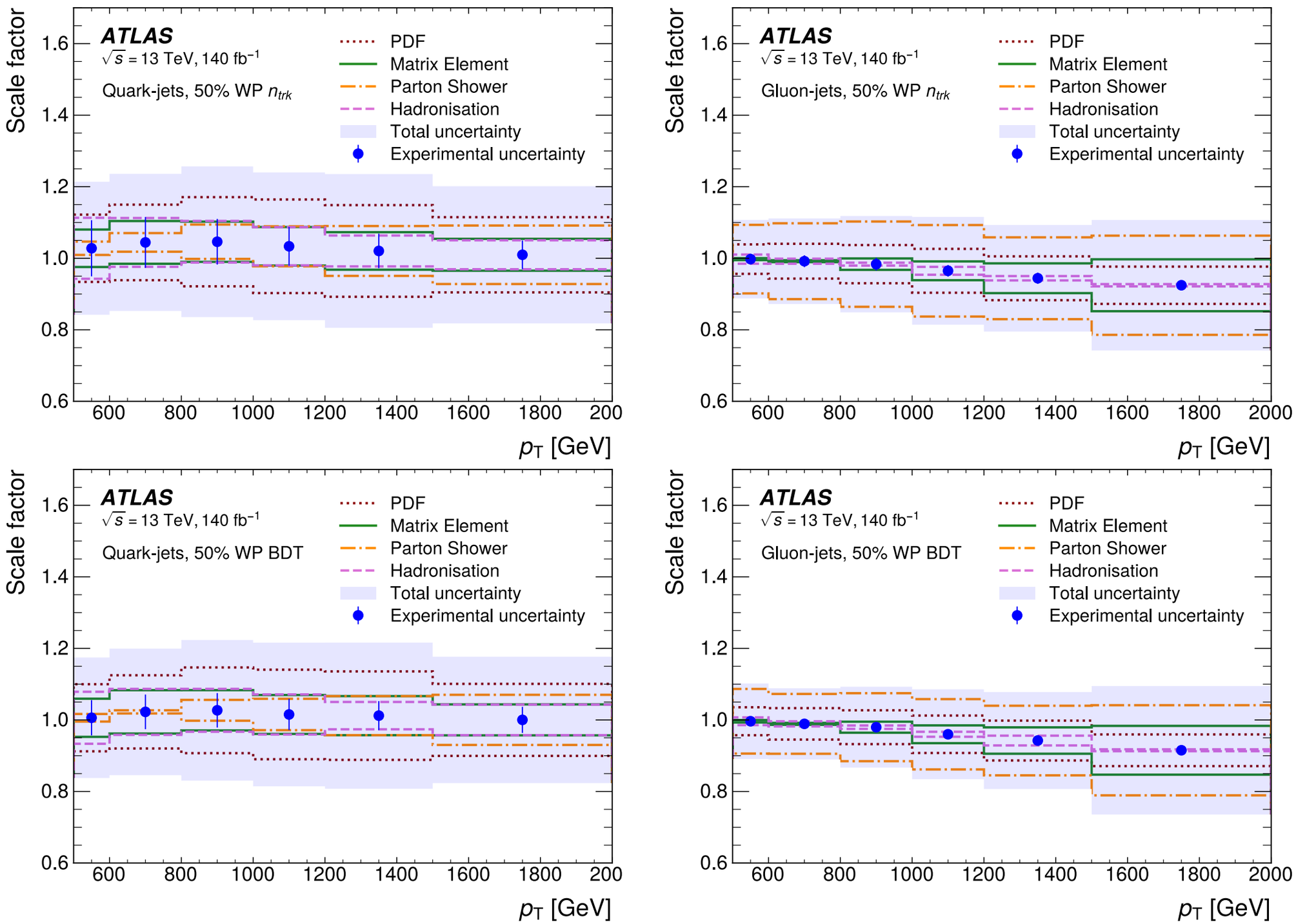
Results

$SF = \varepsilon^{data} / \varepsilon^{MC}$

Quark | Gluon

n_{trk}

BDT



- For WP@50%, σ_{total} **~ 20%** with SF from 0.92 to 1.02, with $\sigma_{th.}$ **most dominant** for both taggers

- Uncertainties from PDF for quark jets, and parton shower for gluon jets

Hbb/cc tagging

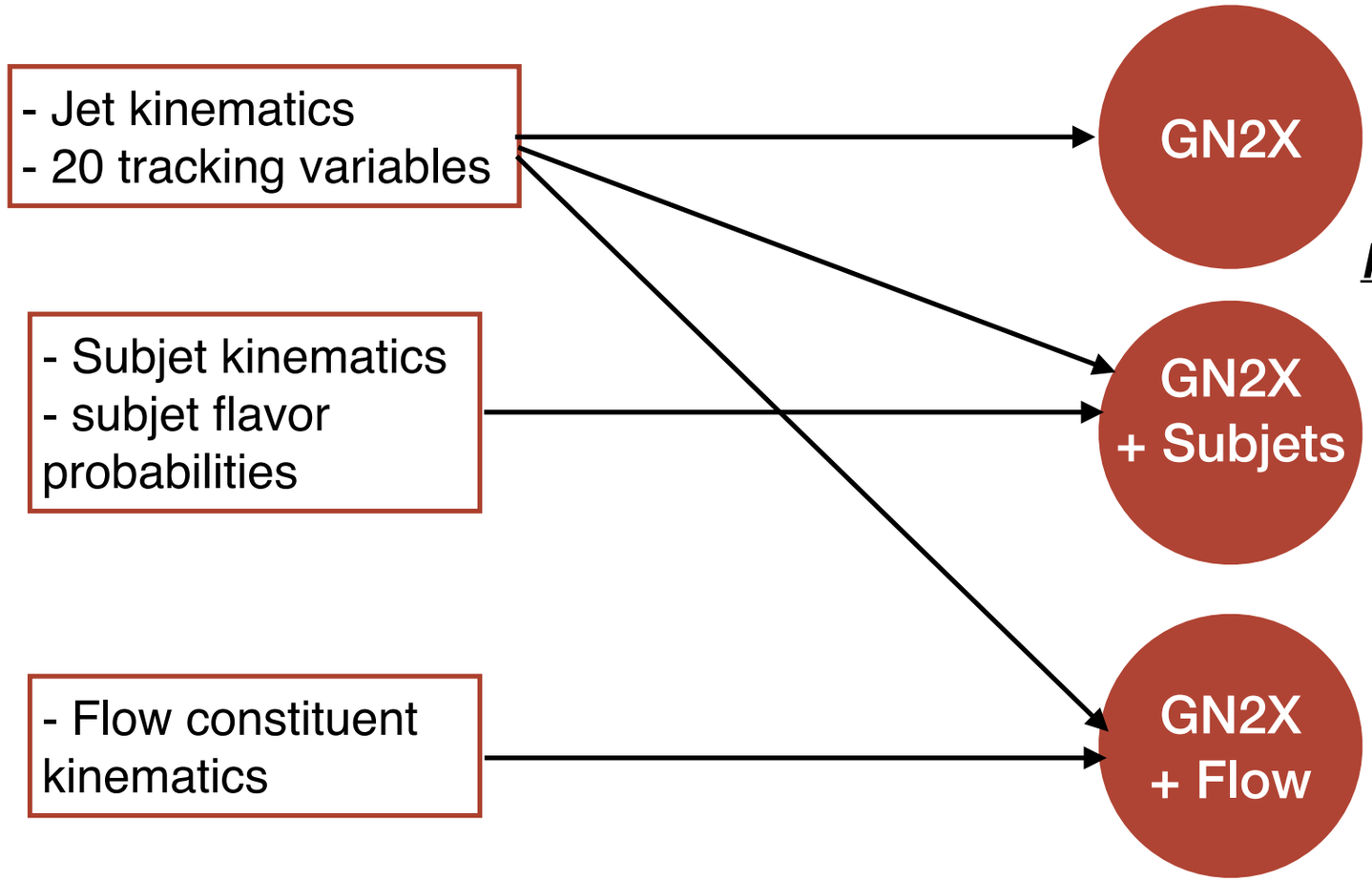
ATL-PUB-FTAG-2023-04

coming soon

Strategy

- $H \to b\bar{b}$ can improve the sensitivity of searches for new resonances in Beyond the Standard Model, and searches for $H \to c\bar{c}$ can benefit from the techniques used for $H \to b\bar{b}$

- Three new taggers have been developed to classify large-R jet based on origin:



More details on inputs

- Baseline taggers to be compared to:
  - 2 VR D_b^{GN2}: tagger using the same inputs as GN2X but training uses R=0.4 jets
  - D_{Xbb}: tagger based on DL1 discriminant

Selection

- R=1 SD+CSSK UFO jets with: $200 < p_T < 1500$ GeV, $|\eta| < 2.0$, $50 < m_J < 300$ GeV

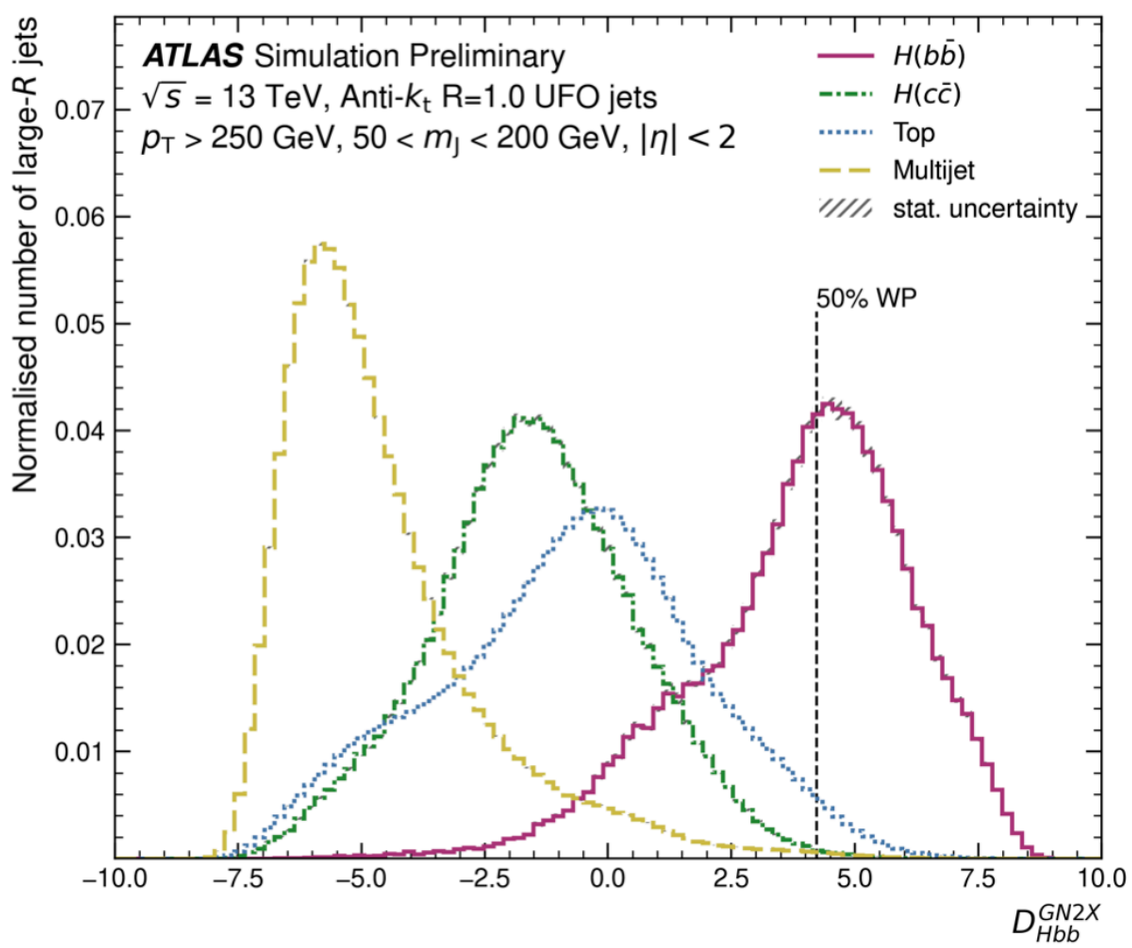
Performance

- The network generates probability scores that indicate the likelihood of a given jet being identified as Hcc, Hbb, top or multijet

- When assessing the Hbb tagging efficiency, these probability scores are combined into a 1D discriminant defined as

$$D_{\mathrm{Xbb}}^{\mathrm{GN2X}} = \ln\left(\frac{p_{\mathrm{Hbb}}}{f_{Hcc}\cdot p_{\mathrm{Hcc}} + f_{\mathrm{top}}\cdot p_{\mathrm{top}} + \left(1 - f_{\mathrm{Hcc}} - f_{\mathrm{top}}\right)\cdot p_{\mathrm{QCD}}}\right)$$

f_{Hcc} and f_{top} are free parameters that determine weights for a trade off among p_{Hcc}, p_{top} and p_{QCD}. For the performance studies, these values are found after an optimization study.



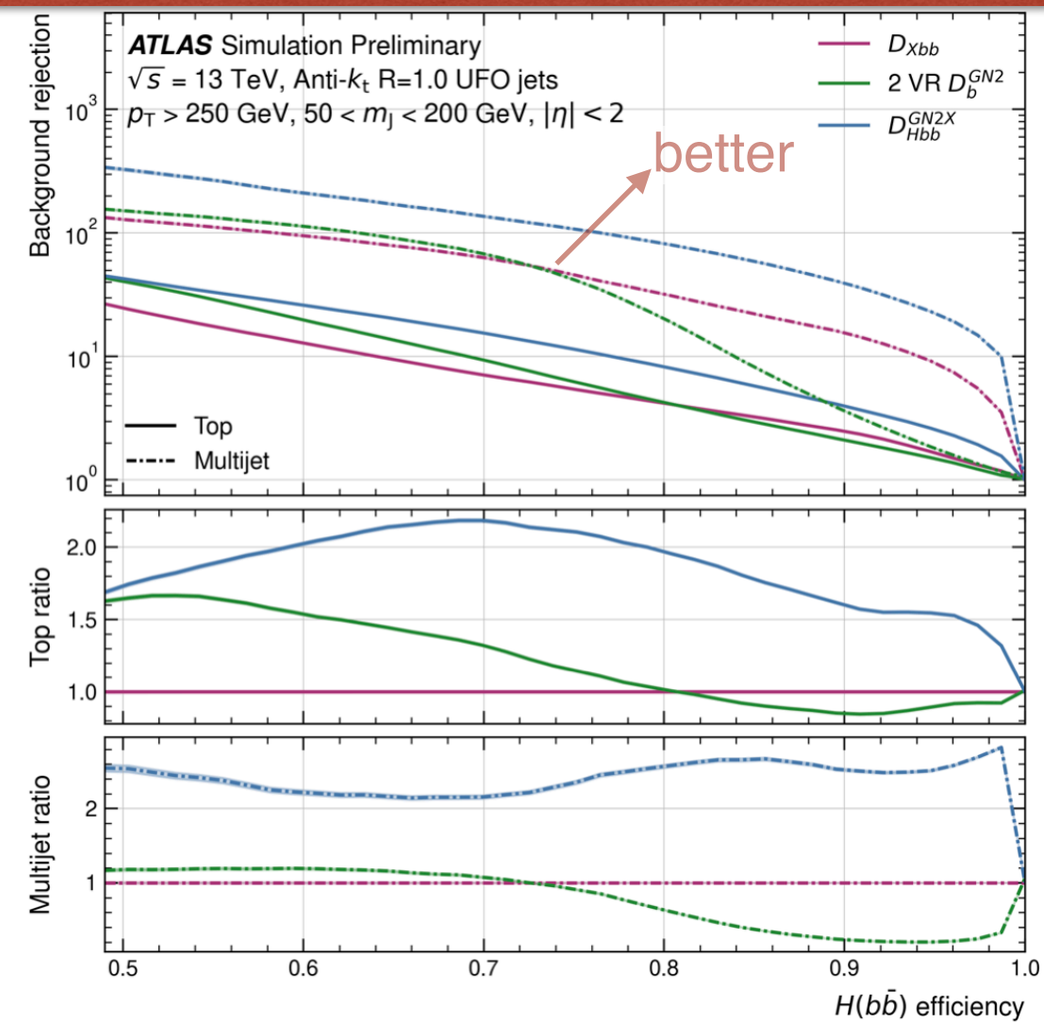
Performance

- Significant improvement from GN2X

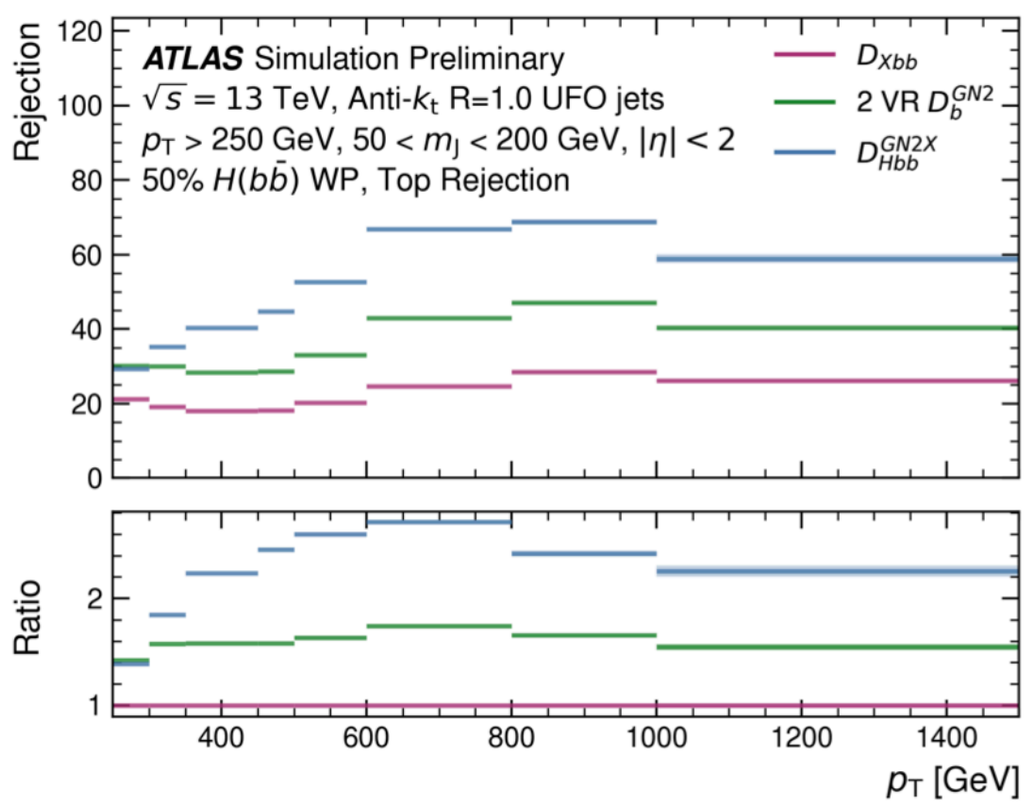
- At a 50% $Hb\bar{b}$ signal efficiency, GN2X provides an increase of **1.6(2.5)** in the **top jet (multijet)** rejection

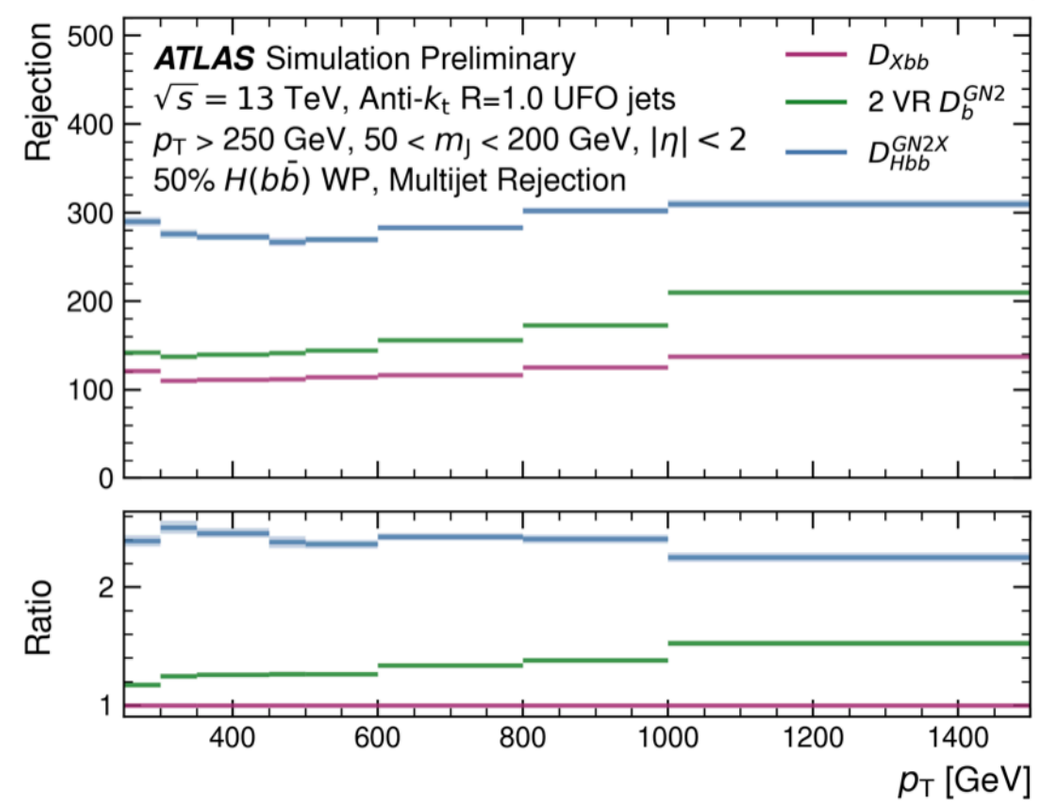
- GN2X also outperforms the 2-tag VR across all efficiencies

- For a 50% $Hb\bar{b}$ efficiency:
  - Top jet rejection improved by **1.4(2.2)** at **250(1500)GeV**
  - Multijet rejection improved by **>2** across **all pT** range



better

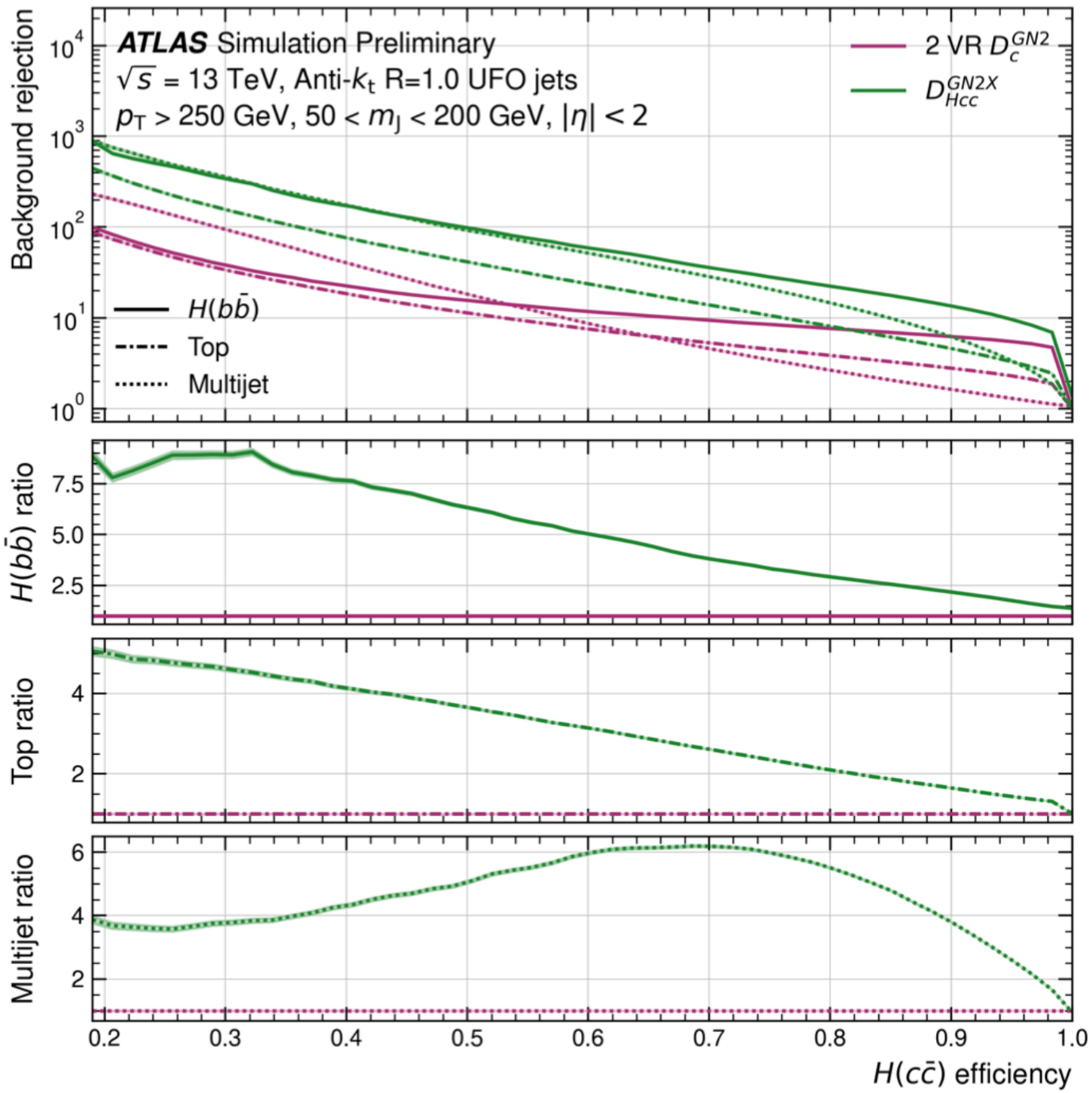




Performance

- Similarly, we can define:

$$D_{\mathrm{Xcc}}^{\mathrm{GN2X}} = \ln\left(\frac{p_{\mathrm{Hcc}}}{f_{Hbb}\cdot p_{\mathrm{Hbb}} + f_{\mathrm{top}}\cdot p_{\mathrm{top}} + \left(1 - f_{\mathrm{Hbb}} - f_{\mathrm{top}}\right)\cdot p_{\mathrm{QCD}}}\right)$$



D_{Xbb} tagger not used for Hcc^{-} identification

- For Hcc^{-} identification, Hbb^{-} events is non-negligible so the Hbb^{-} rejection is also shown

- At a 50% signal efficiency, GN2X provides an **improvement** by a factor of:
  - **3** for top jet rejection
  - **5** for the multijet rejection
  - **6** for the Hbb^{-} rejection.

Conclusions

Conclusions

- Multiple taggers using different techniques/models are applied

- New W taggers have been developed based on the constituents of the large R jet

- The more complex they are, the more dependent to modeling uncertainties they get

- New q/g taggers have been also developed using Ntrk and a BDT and applied to data!

- A new algorithm based on graph neural network was also developed to identify at Hbb/cc

Summary of new ATLAS results for BOOST 2023 *here*

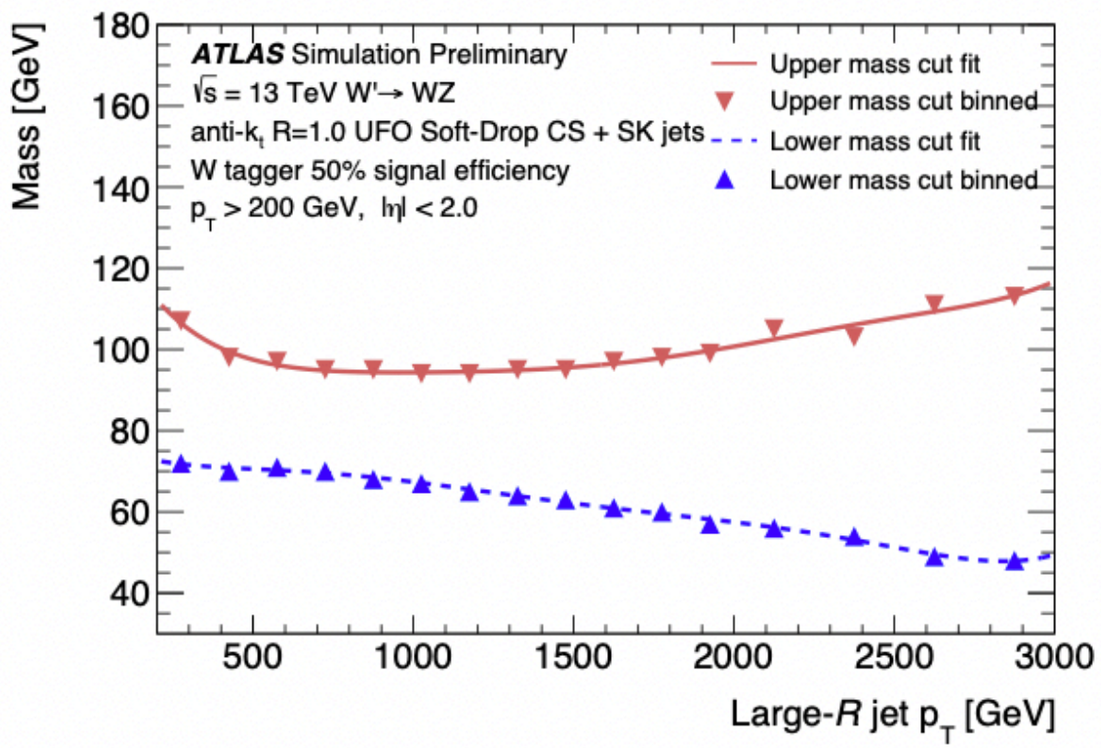
More info during the poster session!!!

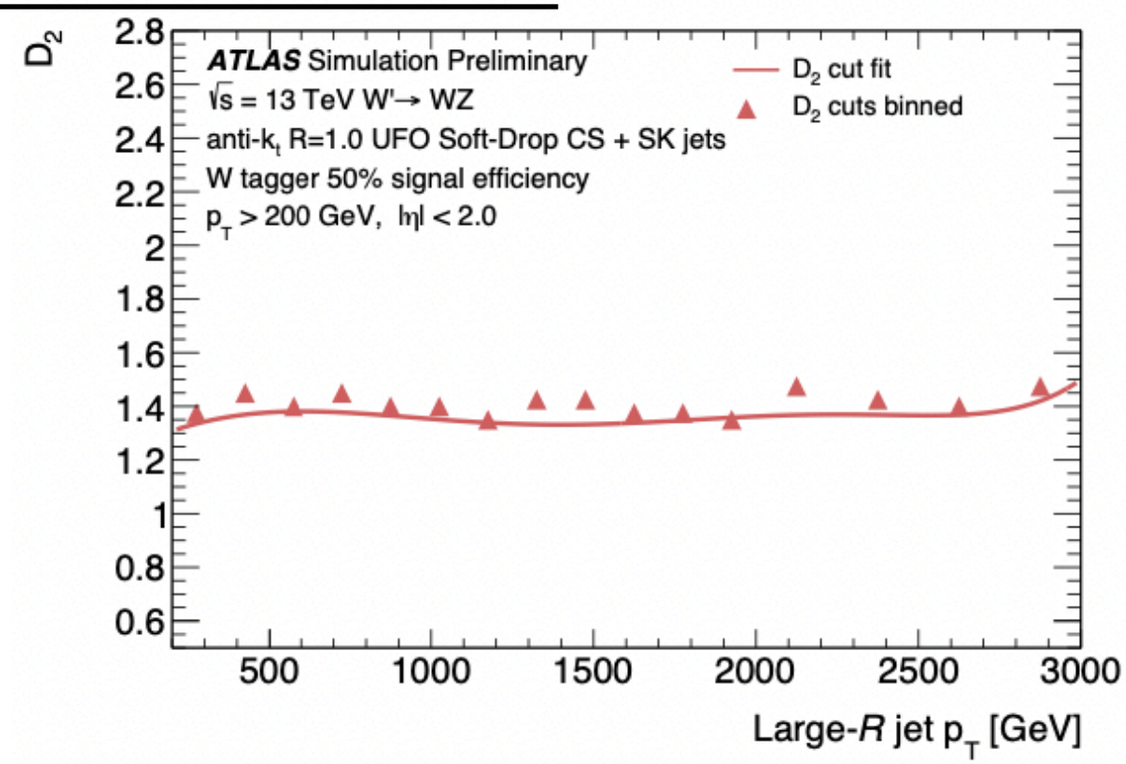
Backup

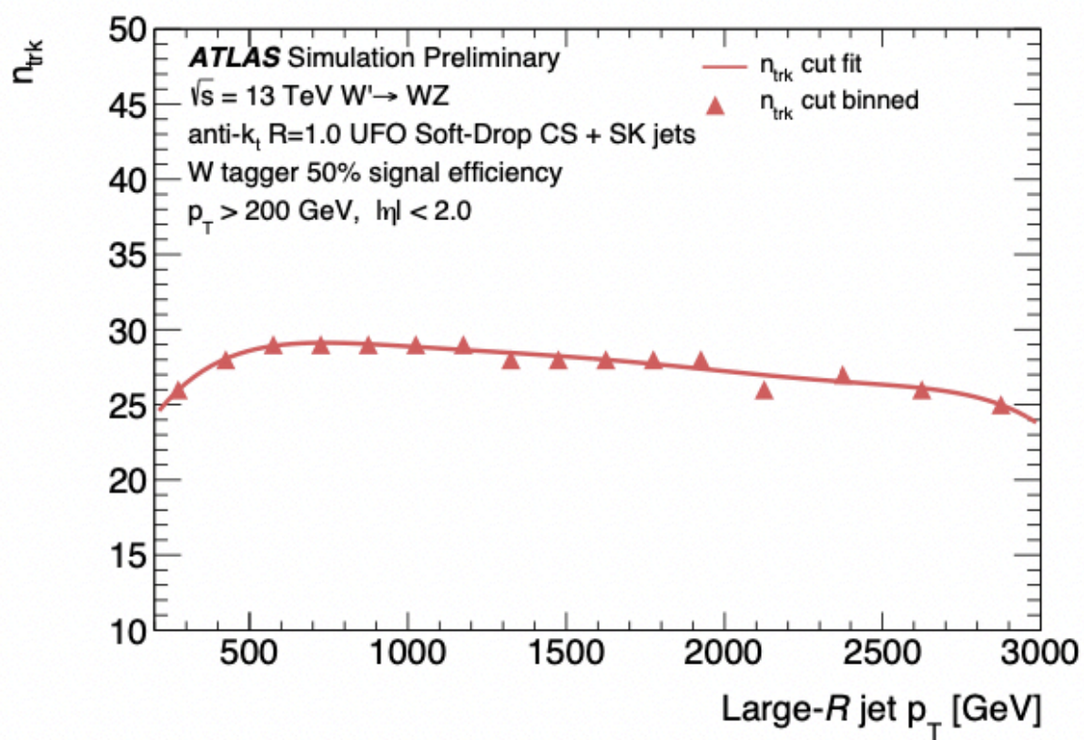
Hadronic jet tagging: overview

ATL-PHYS-PUB-2021-029

- Two taggers are considered as baseline taggers and are used for comparisons :
  - The "so-called" 3-var tagger, using cuts on number of tracks associated to the ungroomed large-R jet, jet mass and the energy correlation ratio (D_2)
  - The DNN/ANN tagger, based on high level observables







- Cuts tuned in order to achieve a certain signal efficiency $\varepsilon^{sig} = N_{sig}^{tagged} / N_{sig}^{total}$ (Example show here for a WP@50%)

- An jet is tagged if:

$$m_{low}^{cut} < m < m_{high}^{cut}$$
$$D_2 < D_2^{cut}$$
$$N_{trk} < N_{trk}^{cut}$$

Baseline taggers

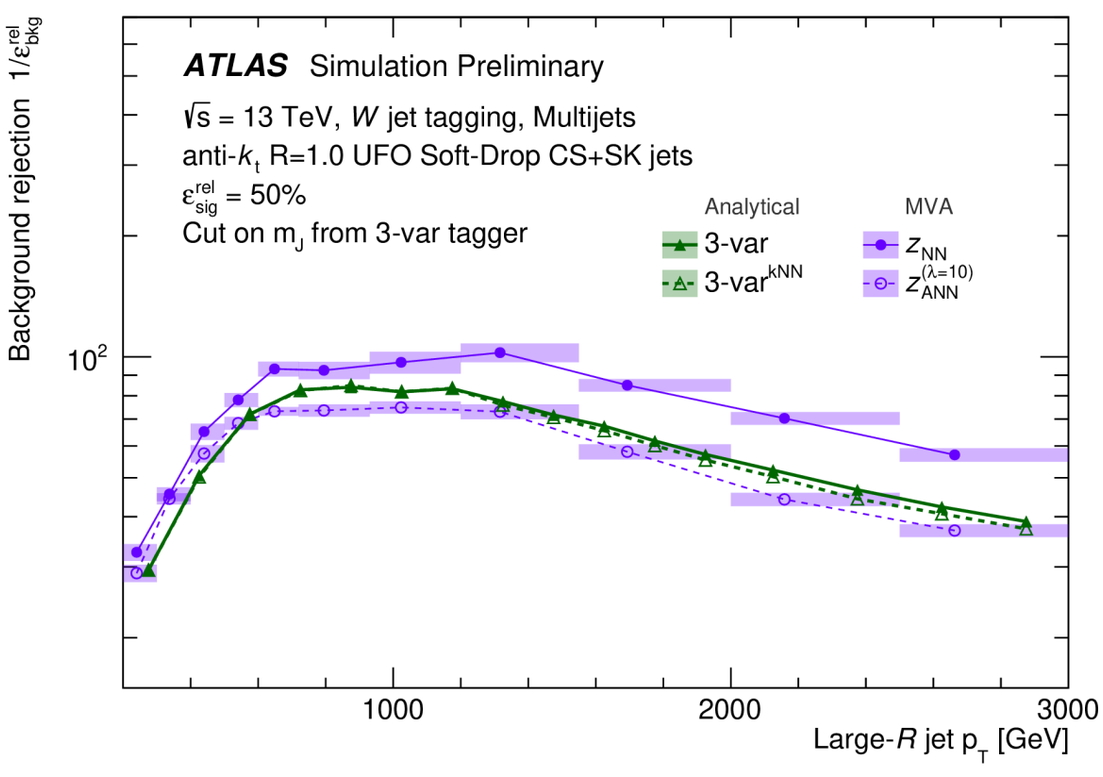
JETM-2022-006

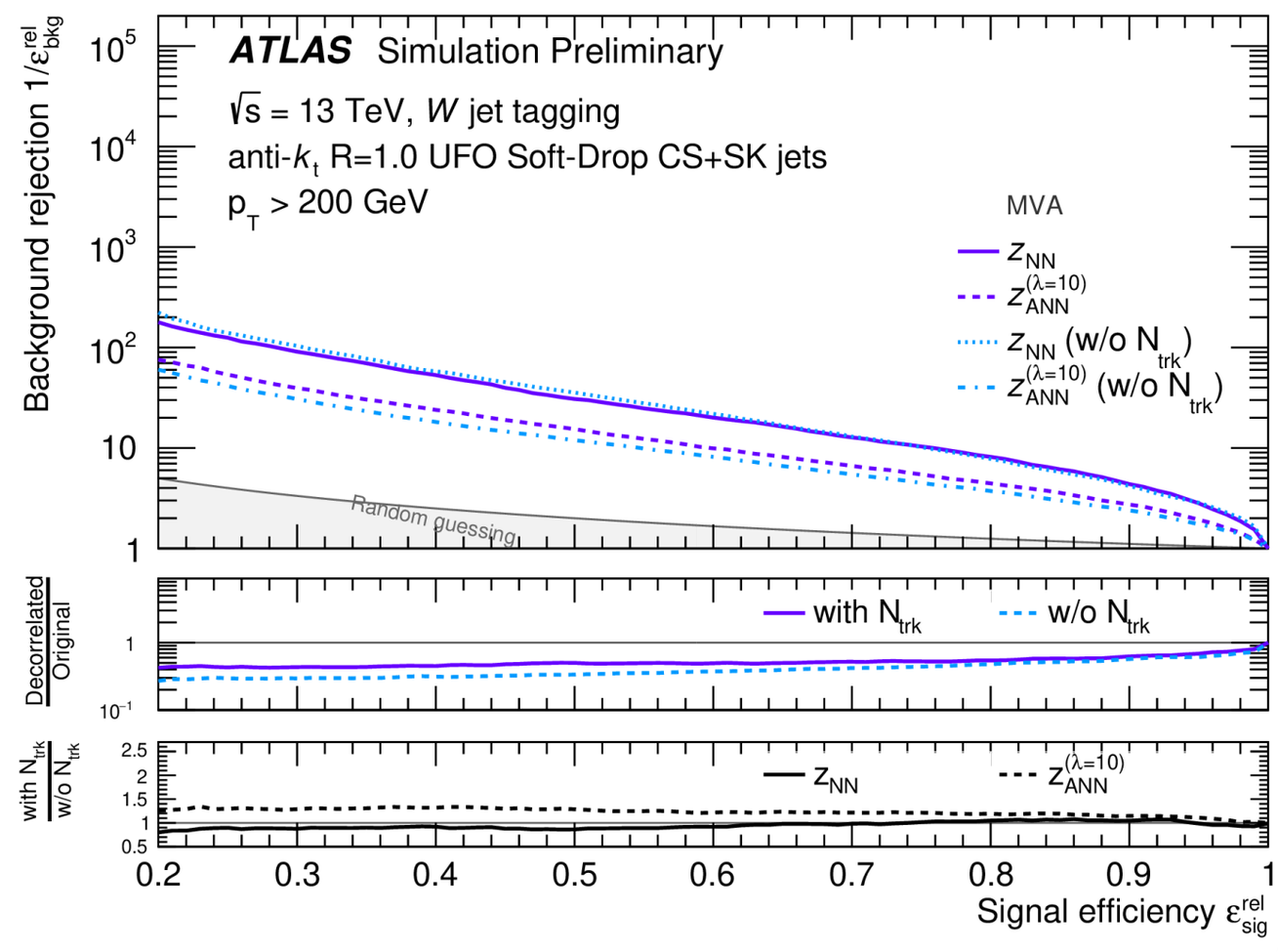
- Two taggers are considered as baseline taggers and are used for comparisons :
  - The "so-called" 3-var tagger, using cuts on number of tracks associated to the ungroomed large-R jet, jet mass and the energy correlation ratio (D_2)
  - The DNN/ANN tagger, based on high level observables

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

- Using high level jet substructure observables as inputs for the Deep Neural Network (variables in table + Ntrk)

- First classifies then applies a mass decorrelation using Adversarial Neural Network





The Lund plane

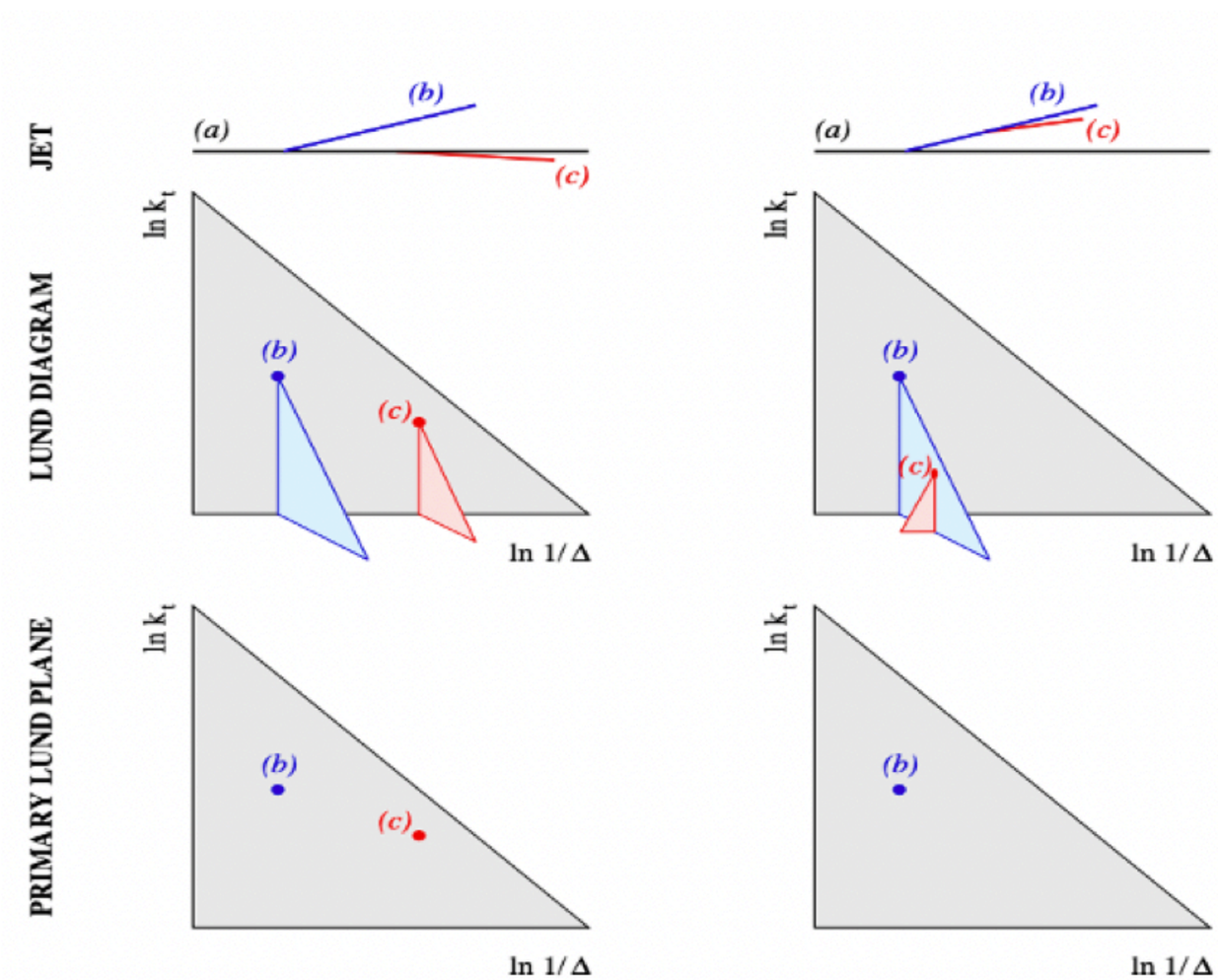
- An abstract representation of the jet formation, initially developed by theorists to better understand it

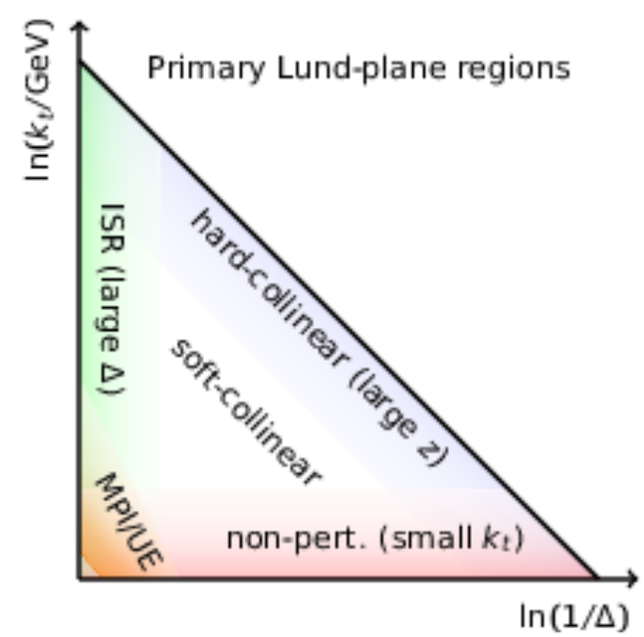
- Each emission represented by a point in the kT-emission angle plane (log scale)

- Hard scattering, collinear and large-angle emissions populate different regions of the plane

- Experimentally, we can have an approximate reconstruction of the Lund Plane
  by running back the CA jet clustering and using the jet merging information

arXiv:2004.03540v2 [hep-ex] (ATLAS)





Constituent based W tagger

Fix link when public ***ATL-COM-PHYS-2023-433***

- Try to maximize the use of the jet constituents' information using state-of-the-art ML/DL algorithms

- A maximum of 200 constituents are considered by all constituent-based taggers

Jet Requirements	W jet requirements
Jet $\lvert\eta\rvert < 2.0$	dR(truth jet, truth W) < 0.75
Jet $p_{\mathrm{T,truth}} > 200$ GeV	Ungroomed truth jet mass > 50 GeV
Number of constituents ≥ 2	Number ghost associated b-hadrons $= 0$
Jet mass > 40 GeV	Truth jet $\sqrt{d_{12}} > 55.25 \times \exp(-2.34 \times 10^{-3} \times$ truth jet $p_{\mathrm{T}})$
dR(jet, truth jet) < 0.75	

- Multiple models are used, each one taking different inputs

- All models are trained to minimize a cross entropy loss with Ranger optimizer

Models	Features
EFN	$\Delta\eta$, $\Delta\phi$, $\ln p_{\mathrm{T}}$
PFN	$\Delta\eta$, $\Delta\phi$, $\ln p_{\mathrm{T}}$, $\ln E$, $\ln \frac{p_{\mathrm{T}}}{\sum_{jet} p_{\mathrm{T}}}$, $\ln \frac{E}{\sum_{jet} E}$, ΔR
ParticleNet	$\Delta\eta$, $\Delta\phi$, $\ln p_{\mathrm{T}}$, $\ln E$, $\ln \frac{p_{\mathrm{T}}}{\sum_{jet} p_{\mathrm{T}}}$, $\ln \frac{E}{\sum_{jet} E}$, ΔR
ParticleTransformer	$\Delta\eta$, $\Delta\phi$, $\ln p_{\mathrm{T}}$, $\ln E$, $\ln \frac{p_{\mathrm{T}}}{\sum_{jet} p_{\mathrm{T}}}$, $\ln \frac{E}{\sum_{jet} E}$, ΔR (E, p_x, p_y, p_z)

- ParticleTransformer: ***arxiv: 2202.03772***
  Transformer designed for particle physics

- ParticleNet: ***Phys.Rev.D 101 (2020) 5, 056019***
  Customized graph neural network architecture for jet tagging with the point cloud approach

- ParticleFlowNetwork (PFN)/Energy Flow Network (EFN): ***JHEP 01(2019) 121***
  Based on Deep Sets Theorem

- ROC curves: ParticleTransformer achieves the best performance, followed by the ParticleNet, PFN, and EFN

- All the constituent-based taggers outperform the high-level-feature-based tagger (noted as z_{NN} in the figure), since benefitting from additional information containing jet constituents

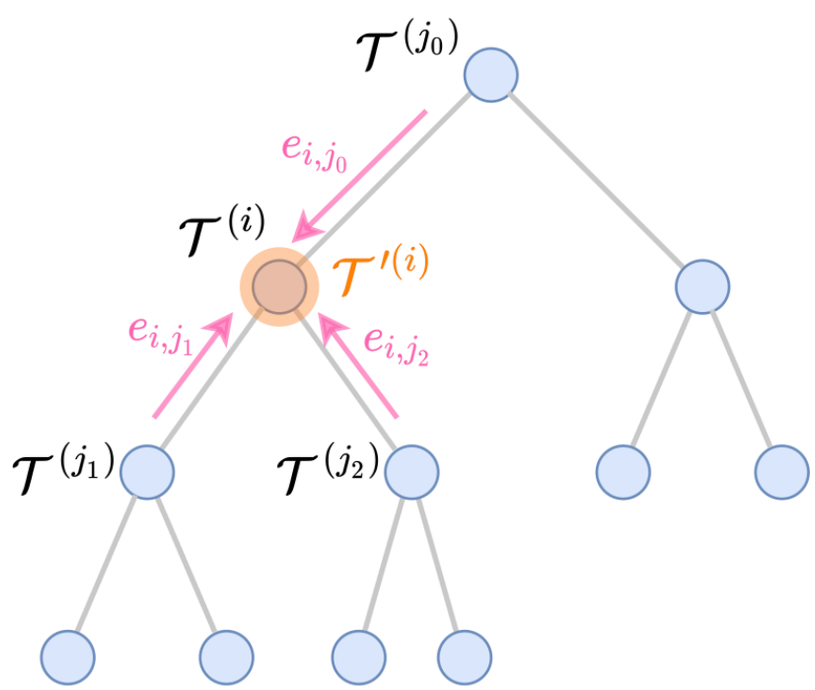
LundNet

- Graph neural network by Frédéric Dreyer used to tag Lund planes when represented as graphs. It is currently the state of the art of Lund tagging and inspired by ParticleNet ***arXiv:2012.08526v2 [hep-ph]***

- It uses the EdgeConv layer. In summary, for a node x_i, we construct a small fully connected neural network. The input is $x_j - x_i = [k_{t,j} - k_{t,i}, \Delta_j - \Delta_i, z_j - z_i]^T$ concatenated with x_i, where x_j is just a node connected to x_i. The output is the edge features, e.

- The EdgeConv block repeats this operation for every node connected to x_i. Then the edge features are aggregated (based on taking the mean) to produce the new node features for x_i.

- LundNet-3 and LundNet-5 are virtually the same model, their difference is the number of Lund variables each node has at the beginning. In our analysis, we only consider LundNet-3.

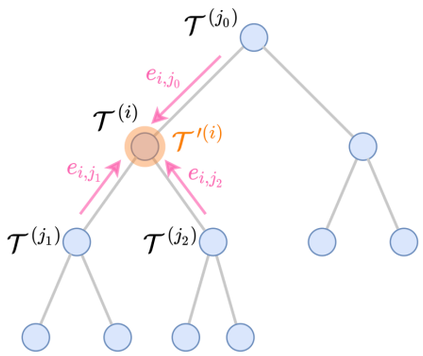


arXiv:2012.08526v2 [hep-ph]

LundNet

- Graph neural network by Frédéric Dreyer et al. used to tag Lund planes when represented as graphs. It is currently the state of the art of Lund tagging and inspired by ParticleNet ***arXiv:2012.08526v2 [hep-ph]***

- LundNet-3 and LundNet-5 are virtually the same model, their difference is the number of Lund variables each node has at the beginning. In our analysis, we only consider LundNet-3.



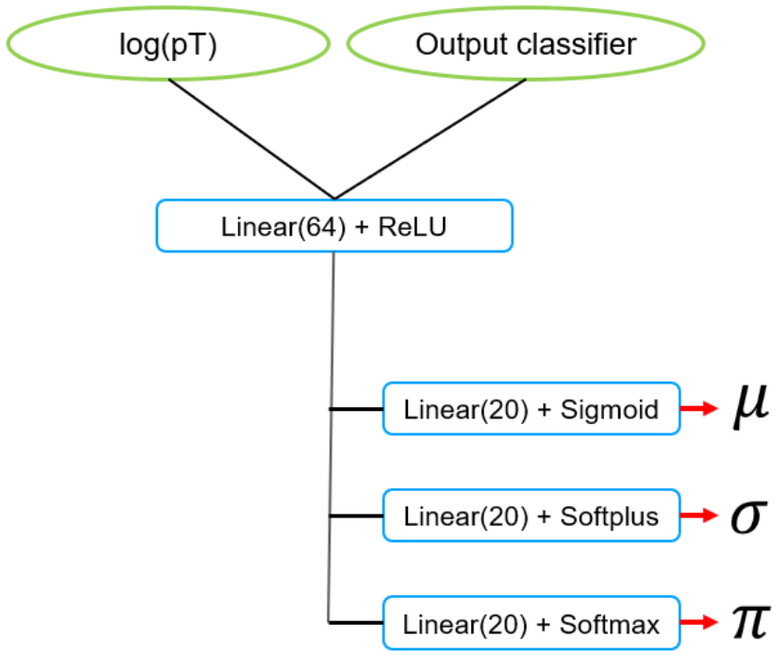
arXiv:2012.08526v2 [hep-ph]

More details

- Outline of the analysis: **1)** Classifier **2)** pre-train adversarial **3)** combined train for mass decorrelation

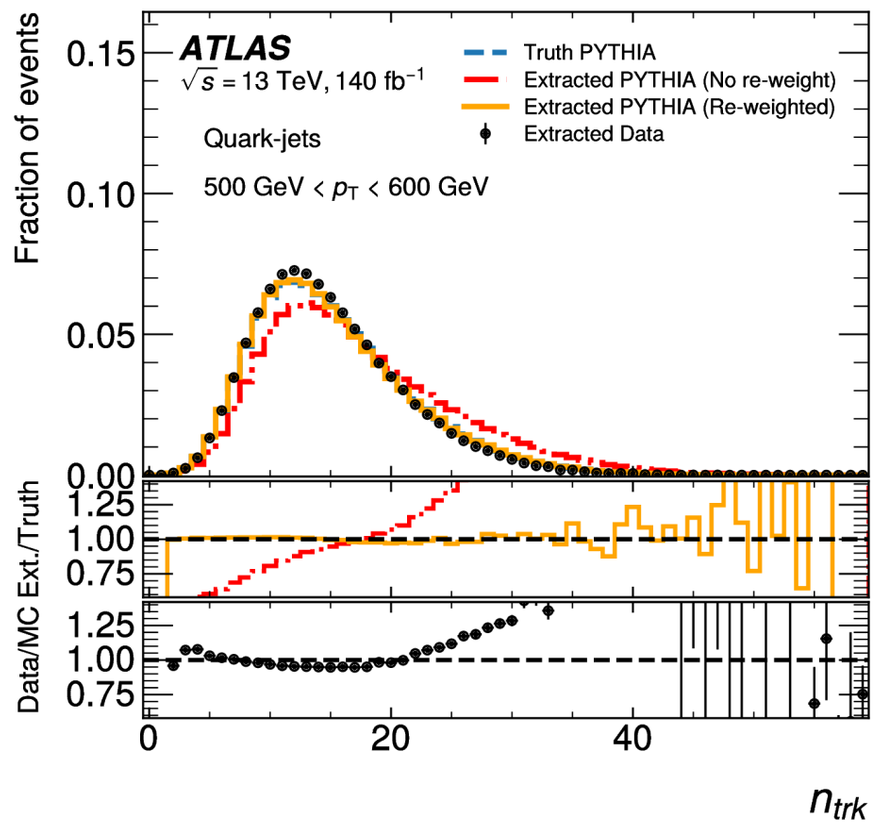
$$\mathscr{L} = w_{clf} \cdot \Sigma_{i\in(s+b)} L_{classifier} \; + w_{adv} \cdot \lambda \cdot \Sigma_{i\in b} \; L_{decor}$$

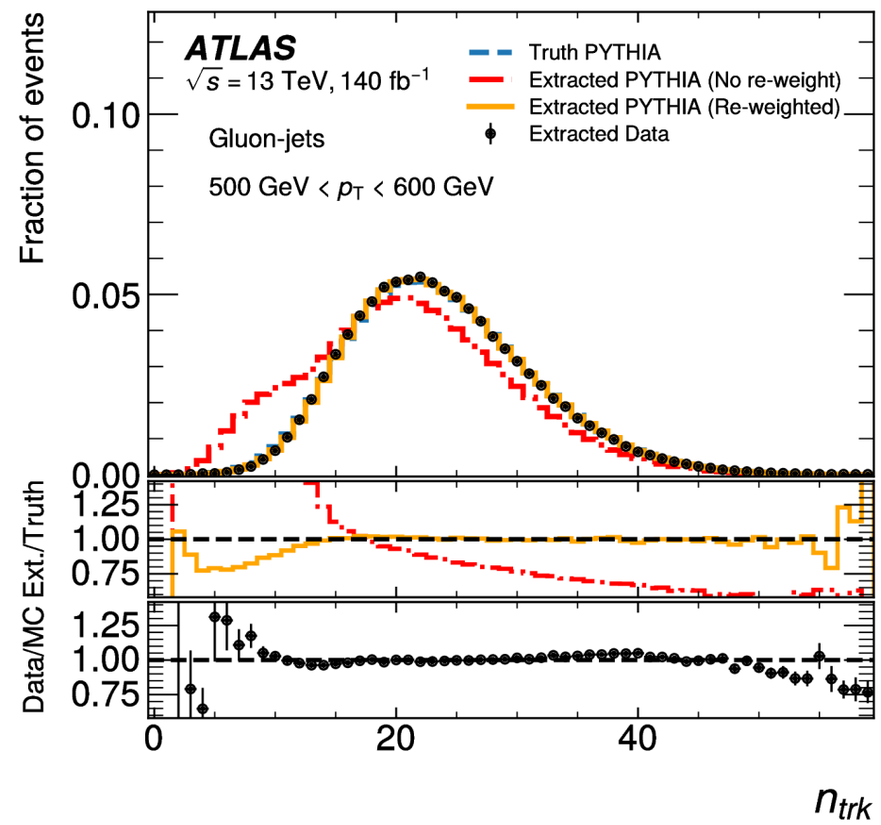
- The adversarial network is a gaussian mixture model that use 20 gaussians to infer the correlation between the output score of the classifier and the mass

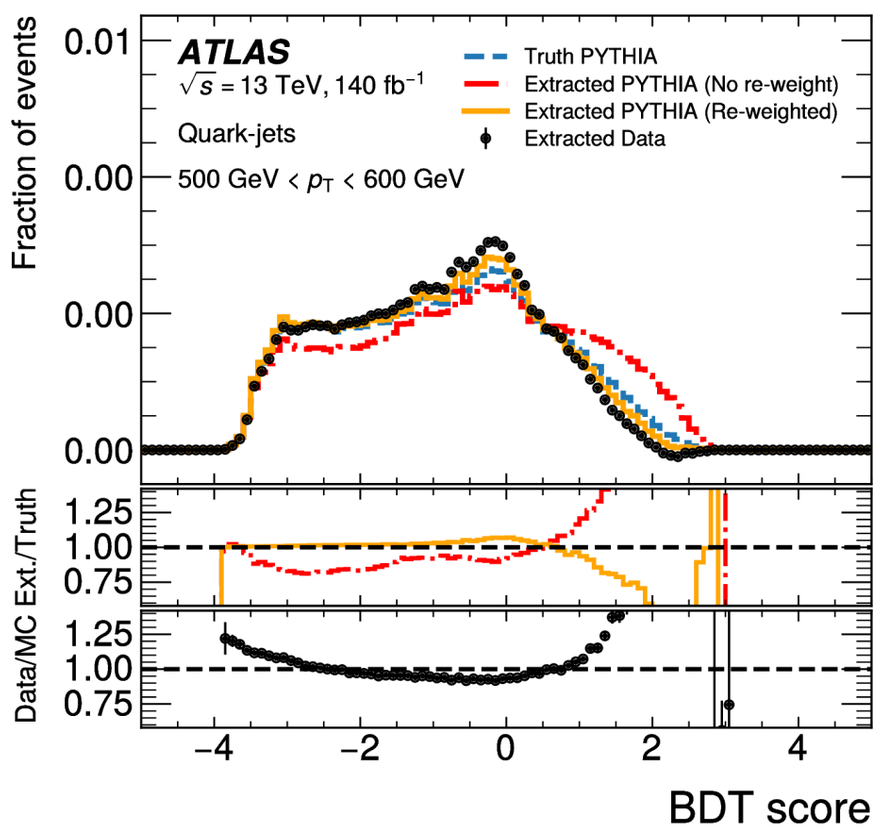


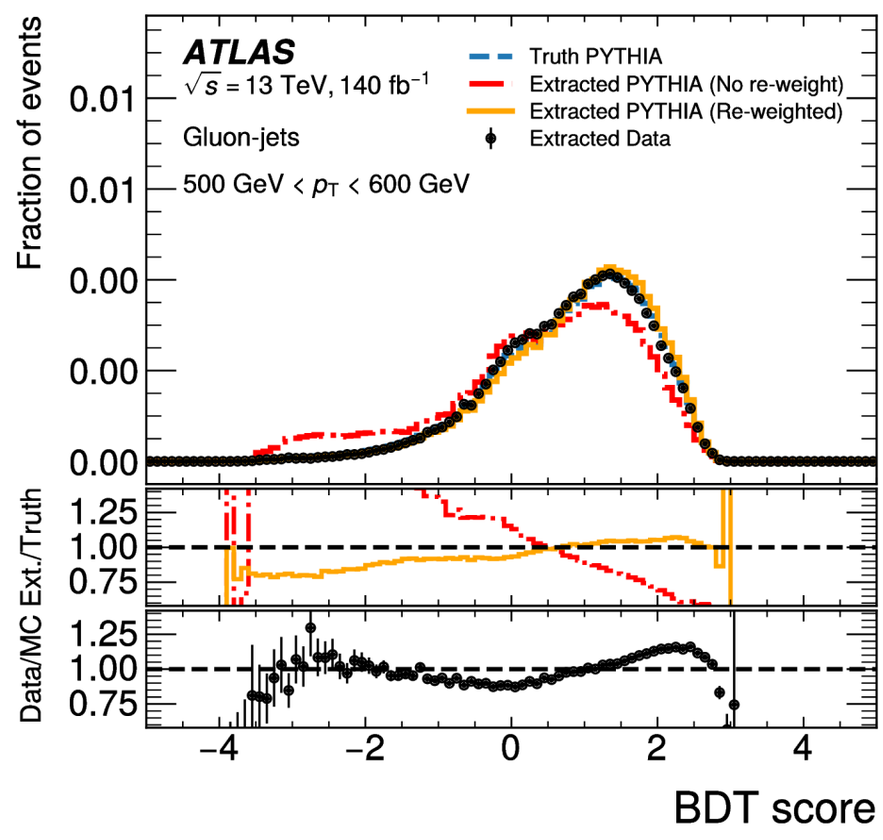
For each Gaussian of 20:
μ : mean, σ : std., π : norm

q/g: performance









Strategy: Hbb/cc

- $H \to b\bar{b}$ can improve the sensitivity of searches for new resonances in Beyond the Standard Model, and searches for $H \to c\bar{c}$ can benefit from the techniques used for $H \to b\bar{b}$

- A new algorithm, GN2X, has been developed to classify large radius jets based on their origin

- Different models have been studies:
  GN2X alone, taking jet kinematics variables + 20 tracking related variables
  GN2X + Subjets: GN2X + subjet kinematics and probabilities for the subjet to be of a certain flavor
  GN2X + Flow: GN2X + flow constituent kinematics

More details

- Baseline taggers to be compared to:
  - VR GN2: tagger using the same inputs as GN2X but training uses R=0.4 jets
  - D_{Xbb}: tagger based on DL1 discriminant

Selection

- R=1 SD+CSSK UFO jets with:
  $200 < p_T < 1500$ GeV
  $|\eta| < 2.0$
  $50 < m_J < 300$ GeV

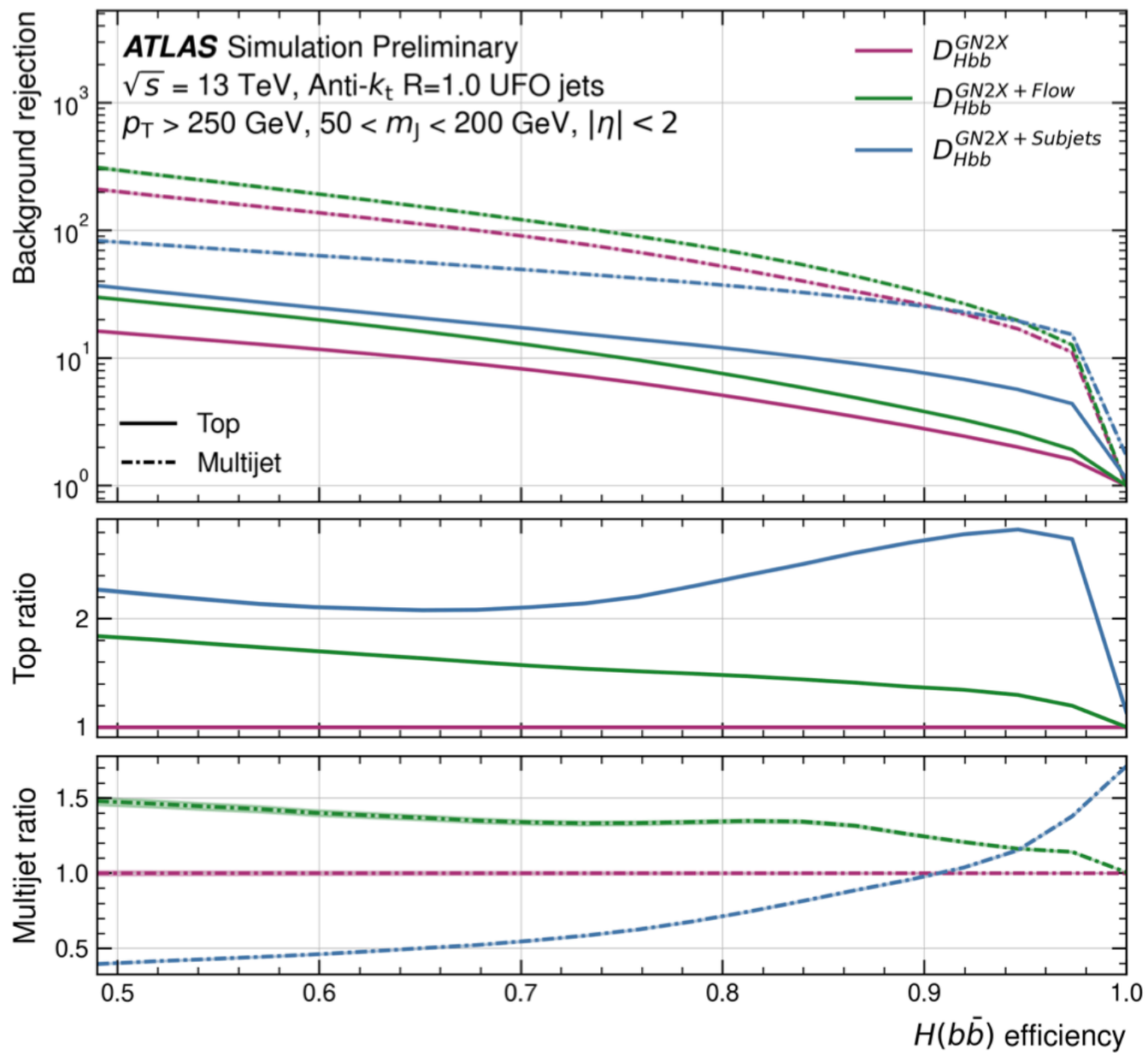
- Tracks are ghost associated and satisfy the following selections

Parameter	Selection
p_T	> 500 MeV
$\|d_0\|$	< 3.5 mm
$\|z_0 \sin\theta\|$	< 5 mm
Silicon hits	≥ 8
Shared silicon hits	< 2
Silicon holes	< 3
Pixel holes	< 2

Hbb/cc inputs

Jet Input	Description
p_T	Large-R jet transverse momentum
η	Signed large-R jet pseudorapidity
mass	Large-R jet mass
Track Input	**Description**
q/p	Track charge divided by momentum (measure of curvature)
$d\eta$	Pseudorapidity of track relative to the large-R jet η
$d\phi$	Azimuthal angle of the track, relative to the large-R jet ϕ
d_0	Closest distance from track to primary vertex (PV) in the transverse plane
$z_0 \sin\theta$	Closest distance from track to PV in the longitudinal plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0 \sin\theta)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
subjetIndex	Integer label of which subjet track is associated to
Subjet Input	**Description** (Used only in GN2X + Subjets)
p_T	Subjet transverse momentum
η	Subjet signed pseudorapidity
mass	Subjet mass
energy	Subjet energy
$d\eta$	Pseudorapidity of subjet relative to the large-R jet η
$d\phi$	Azimuthal angle of subjet relative to the large-R jet ϕ
GN2 p_b	b-jet probability of subjet tagged using GN2
GN2 p_c	c-jet probability of subjet tagged using GN2
GN2 p_u	light-jet probability of subjet tagged using GN2
Flow Input	**Description** (Used only in GN2X + Flow)
p_T	Transverse momentum of flow constituent
energy	Energy of flow constituent
$d\eta$	Pseudorapidity of flow constituent relative to the large-R jet η
$d\phi$	Azimuthal angle of flow constituent relative to the large-R jet ϕ

Hbb/cc performance



- The GN2X + Subjets: **40% increase** in the **top** rejection relative to GN2X by including complementary information on the large-R jet substructure. However, **reduction** in the **multijet** rejection performance.
- The GN2X + Flow model: takes advantage of information on neutral jet components missing from other versions. **60% improvement** in the **multijet** rejection. Further work combining the flow constituents and the subjets is therefore warranted.