Boosted W tagging with Lund jet planes

ATL-PHYS-PUB-2023-017

Jad M. Sardain on behalf of the ATLAS Collaboration

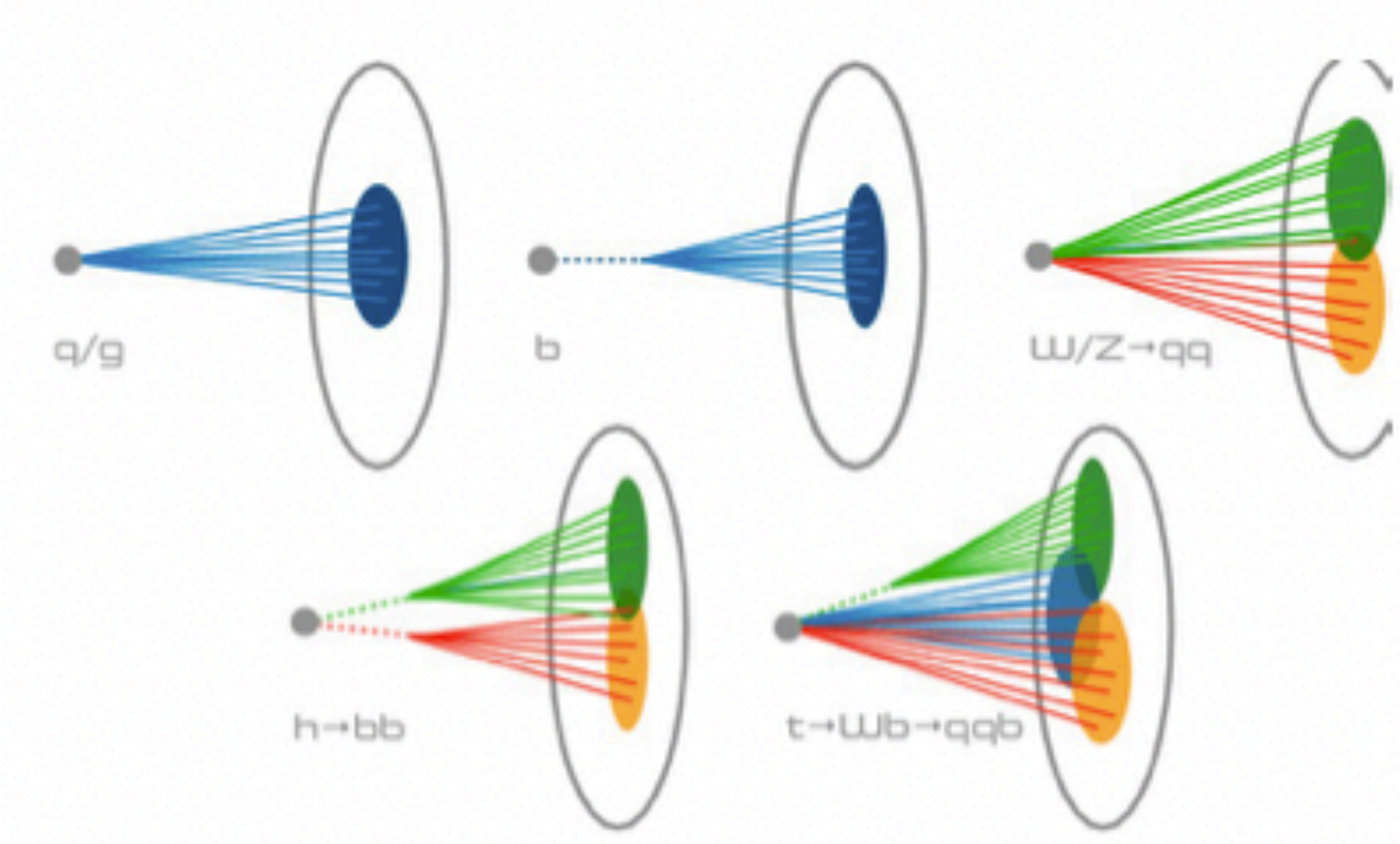
First Lund Jet Plane Institute, July 3-7 2023

Hadronic jet tagging: overview

- Heavy states with large momentum will produce multi-prong topology

- Techniques (ML or not) developed in the last 10 years to distinguish the various jet topologies

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



Jet tagging using Lund plane and ML

- The Lund planes for the W, top and QCD jets already looks quite different
    - Convolutional NN used for identification
    - Results not much better than CNN directly on jet images

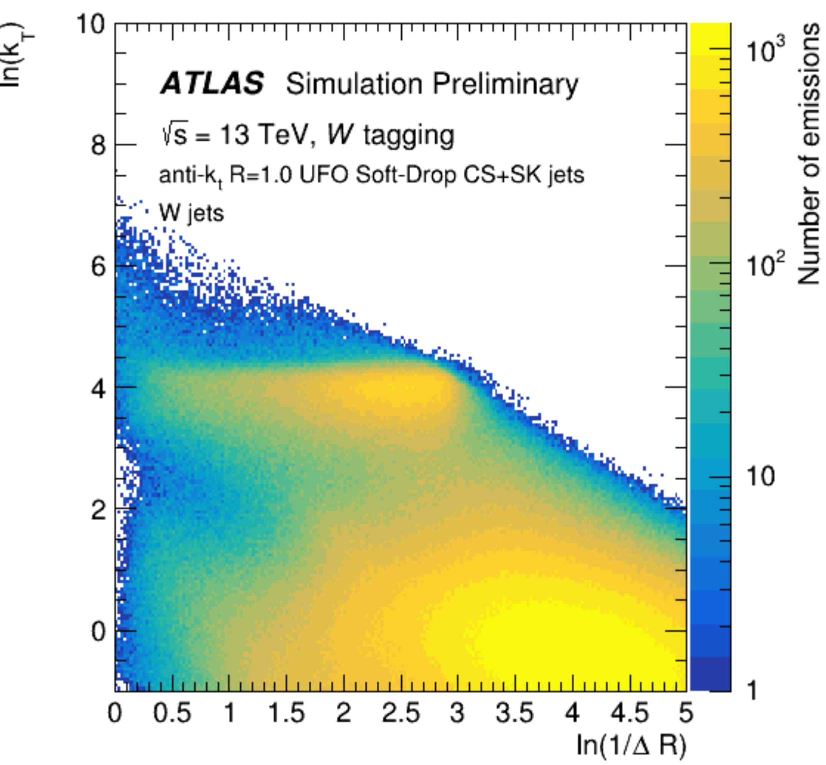
- However, the Lund plane has much more information coming from the sequence that produced it
    - Use GNN where each node has 3 vars:

        z momentum fraction of the branching.

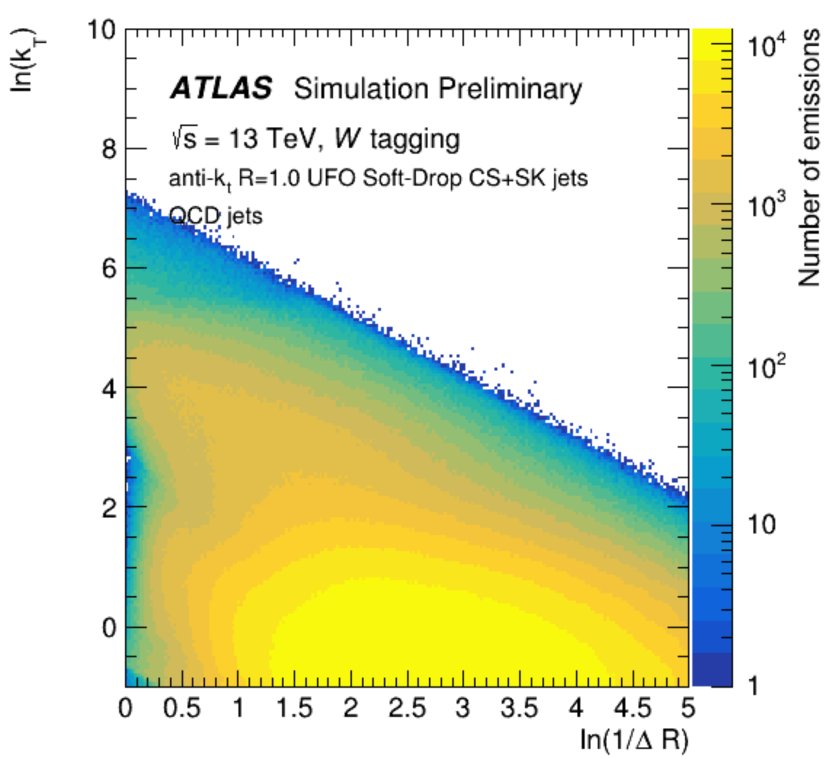
        k_t transverse momentum,

        Δ emission angle
    - Number of tracks per jet as a global feature to help classification

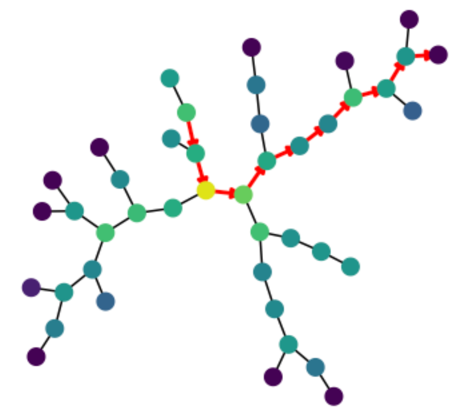
Signal



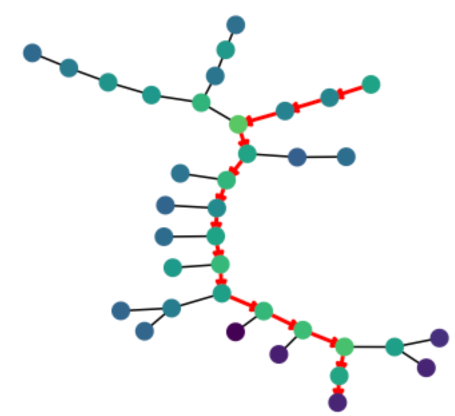
Background



W-jet



QCD



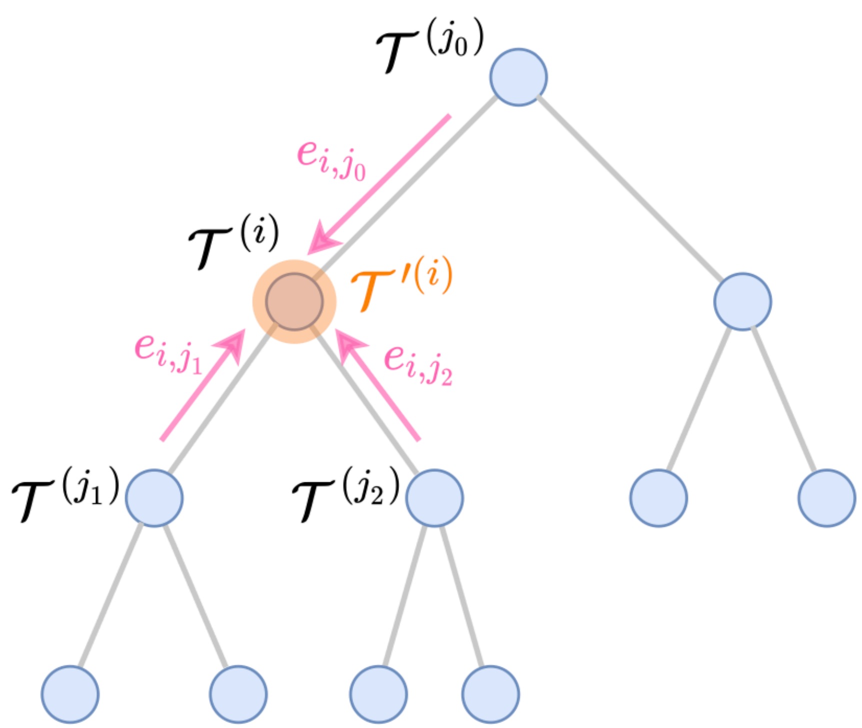
LundNet

- 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]

Analysis

Signal: $W' \to WZ$

- Truth matched to W
- m(truth jet) > 50 GeV
- Number of b-hadrons = 0

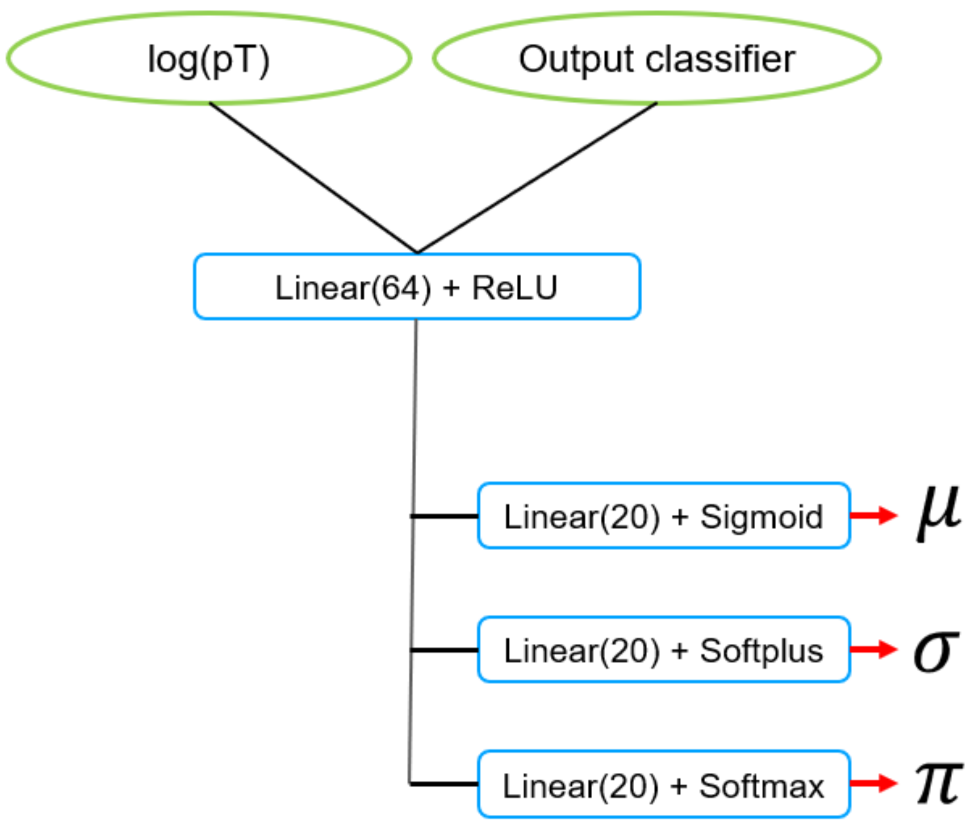
Background: QCD

- Jet is truth-matched
- p_T(truth jet) > 200 GeV
- 200 < p_T(jet) < 3000 GeV, |η(jet)| < 2.0
- 40 < m(jet) < 300 GeV

- 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



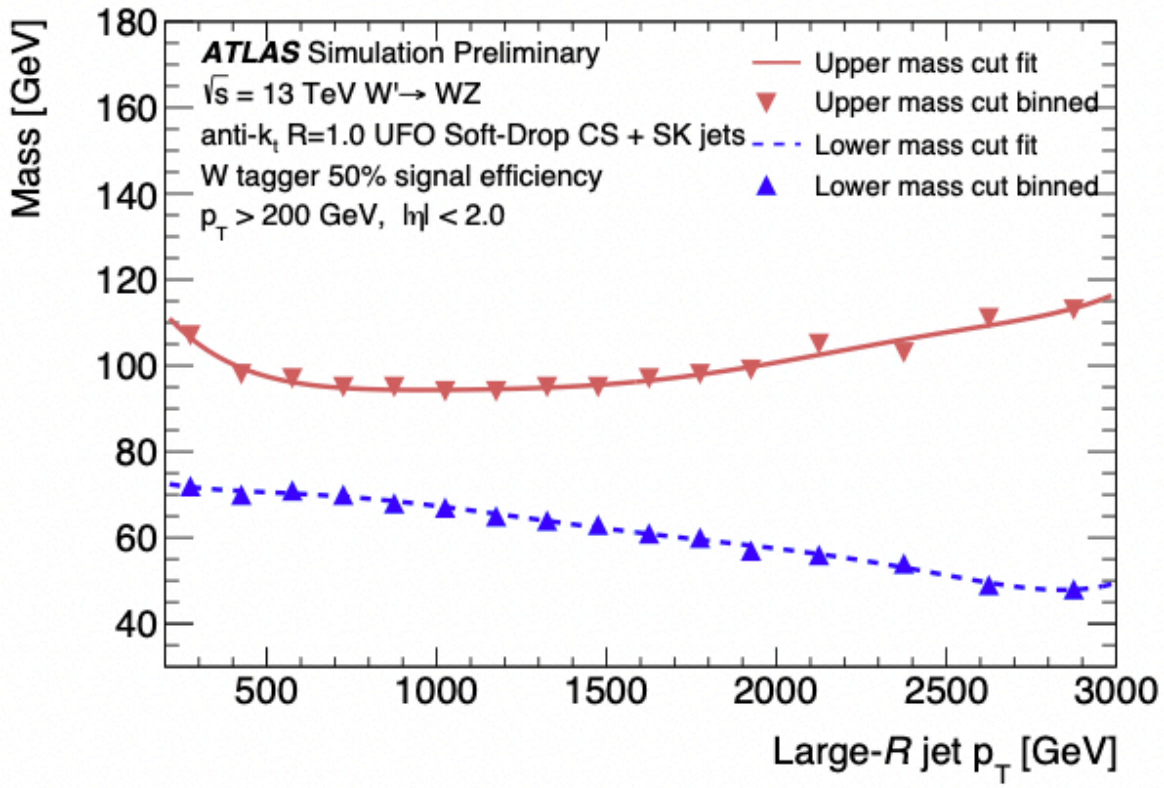
For each Gaussian of 20:
μ : mean, σ : std., π : norm

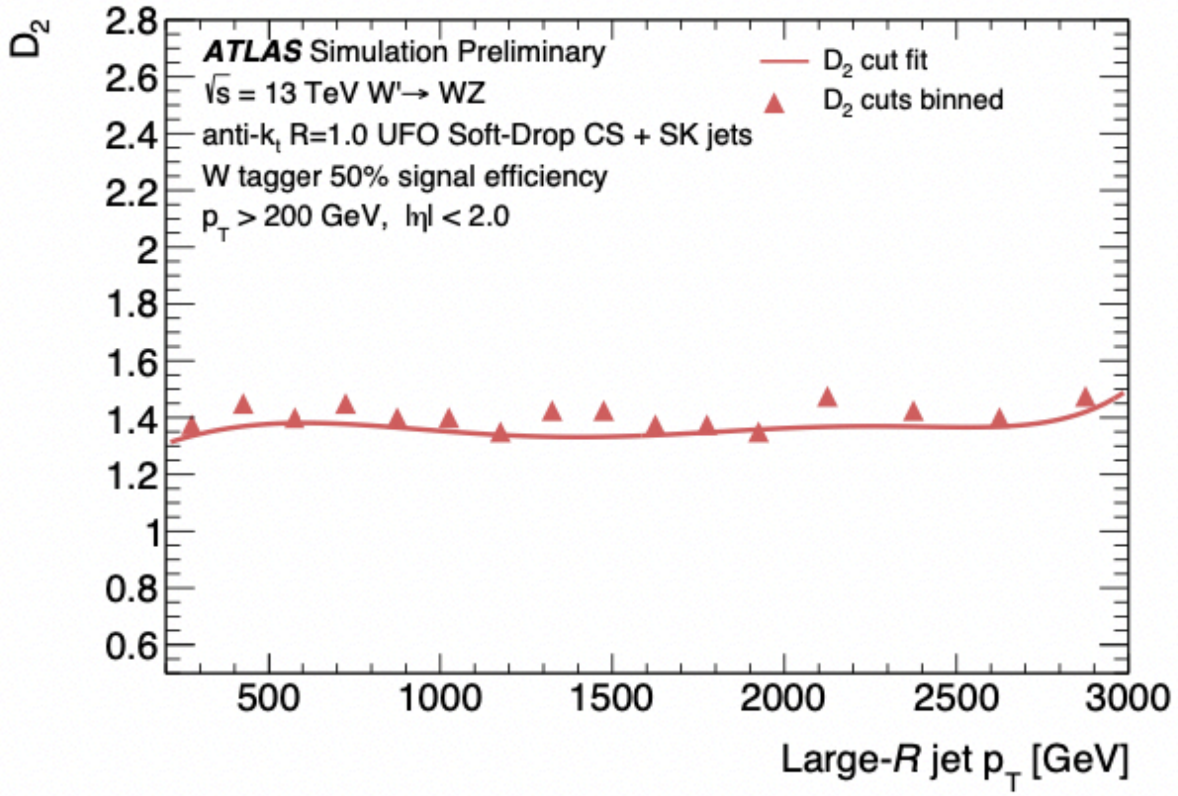
Baseline taggers

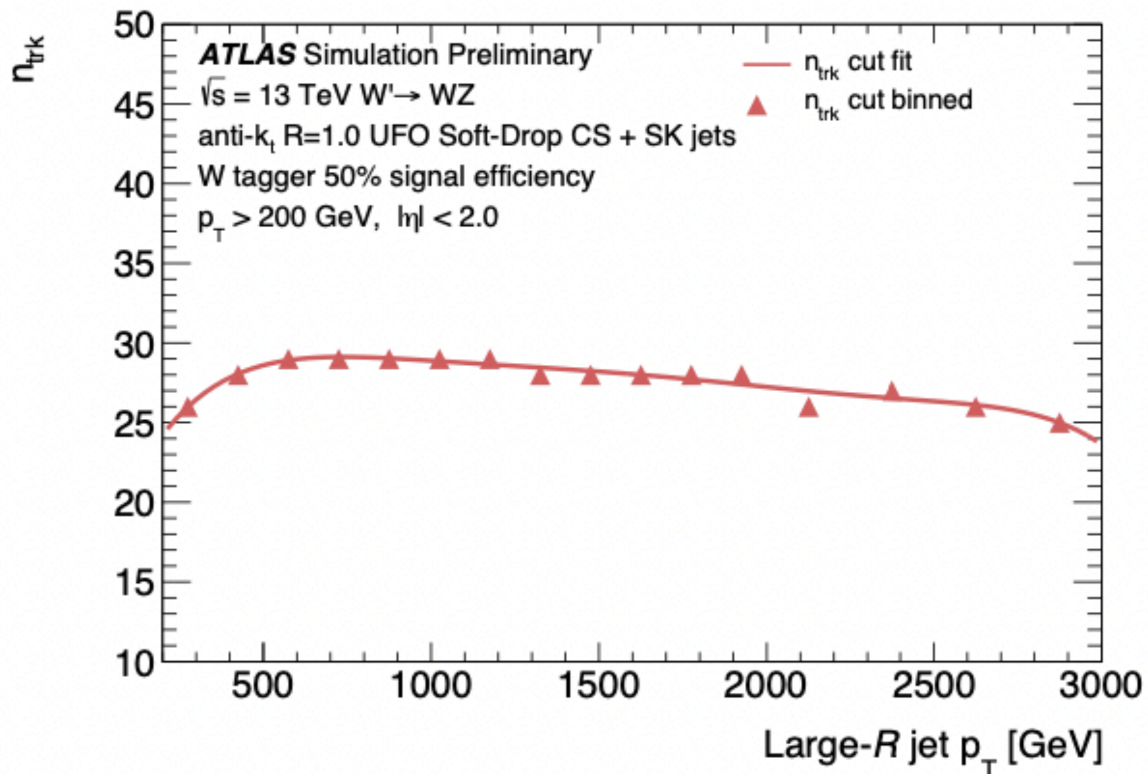
- Two taggers are considered as baseline taggers and are used for comparisons :
  - The “so-called” 3-var tagger, based on number of tracks, mass and D2
  - The DNN/ANN tagger, based on high level observables

Baseline taggers

- Two taggers are considered as baseline taggers and are used for comparisons :
  - The "so-called" 3-var tagger, based on number of tracks, mass and D2
  - The DNN/ANN tagger, based on high level observables







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

- An jet is tagged if:

$$m^{cut}_{low} < m < m^{cut}_{high}$$
$$D_2 < D^{cut}_2$$
$$N_{trk} < N^{cut}_{trk}$$

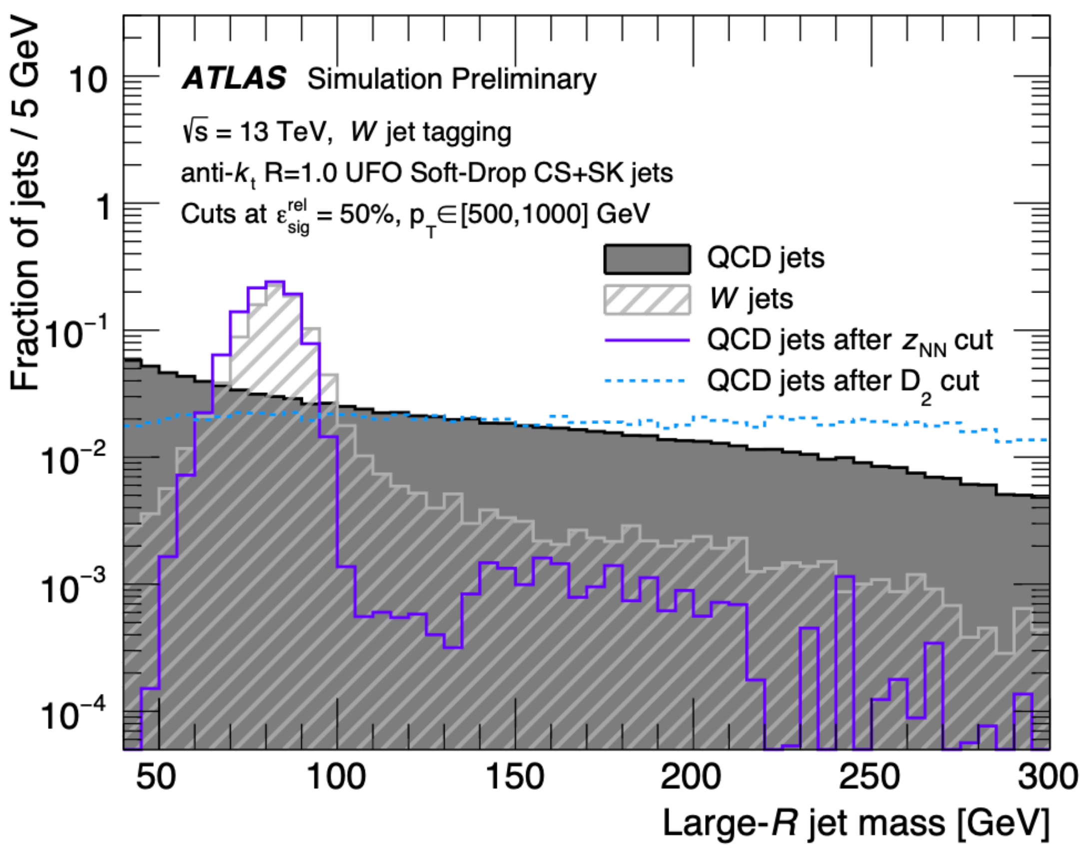
Baseline taggers

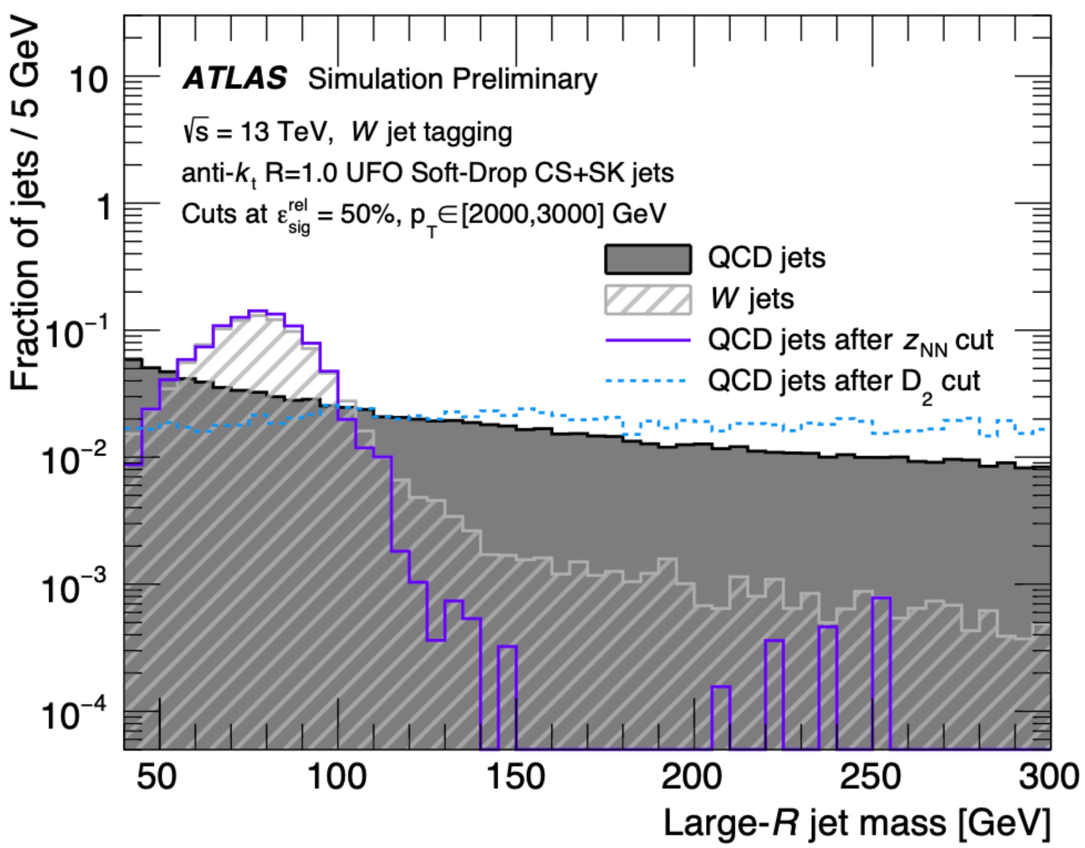
- Two taggers are considered as baseline taggers and are used for comparisons :
  - The “so-called” 3-var tagger, based on number of tracks, mass and D2
  - The DNN/ANN tagger, based on high level observables

- High level observables are used as input features for the classifier

- Same strategy as LundNet, classifier then adversarial

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





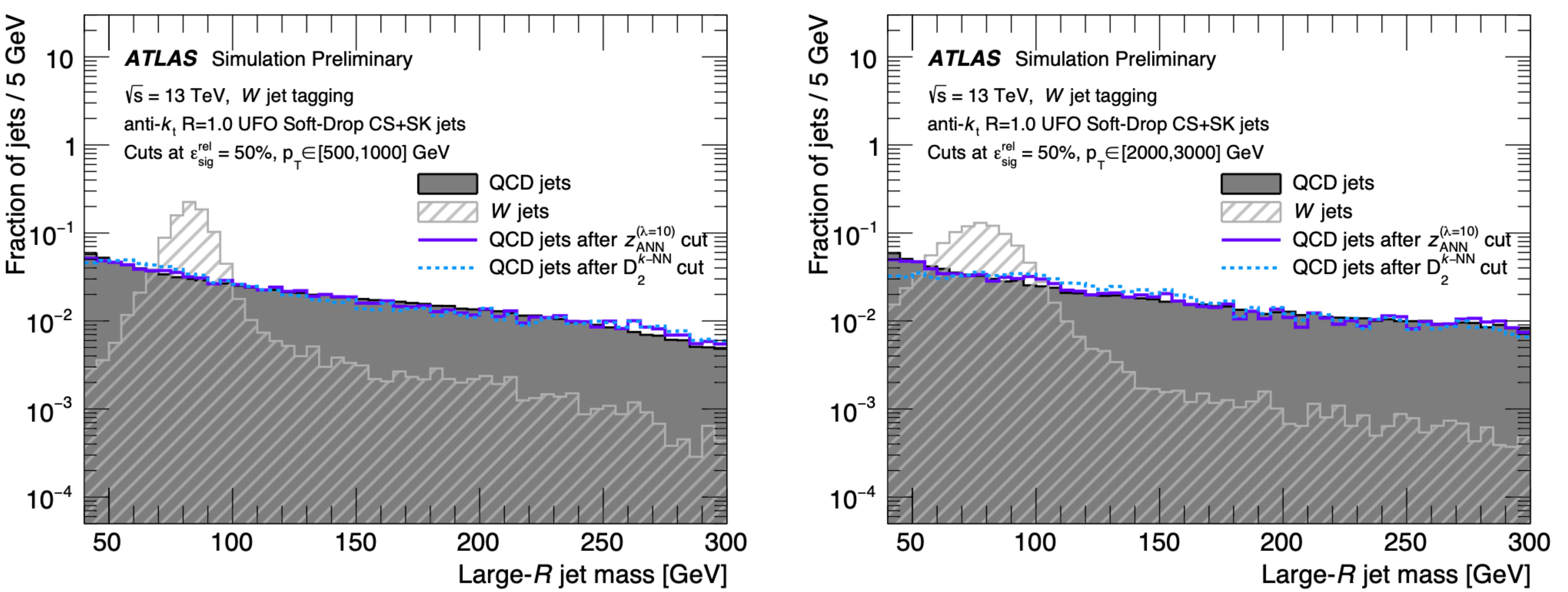
Baseline taggers

- Two taggers are considered as baseline taggers and are used for comparisons :
    - The “so-called” 3-var tagger, based on number of tracks, mass and D2
    - The DNN/ANN tagger, based on high level observables

To decorrelate jet mass and DNN score:
- Apply an additional adversarial neural network (ANN) to the DNN tagger

- ANN trained to infer the jet mass from the DNN score by minimizing L_{ANN}

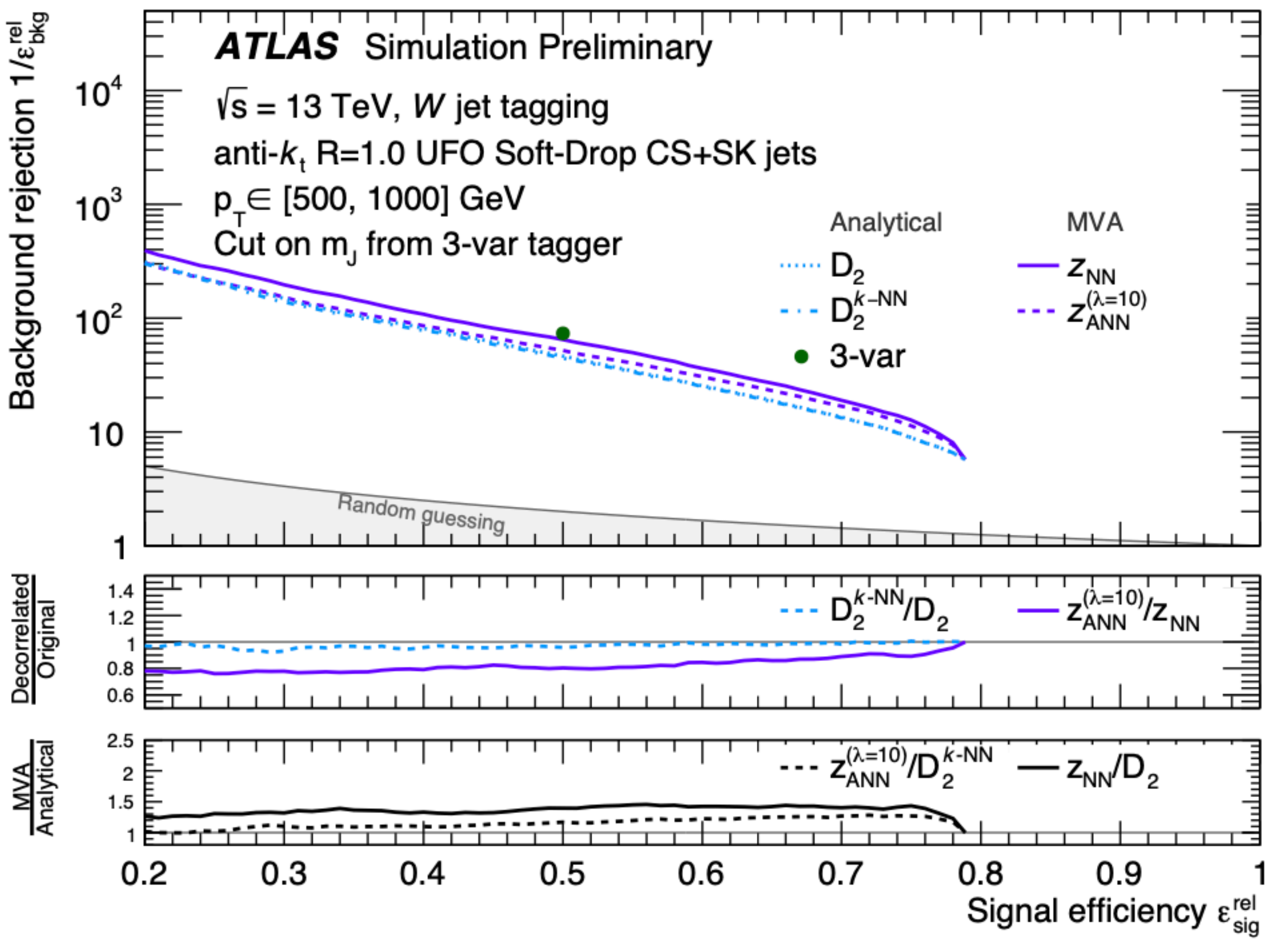
- Loss function of the combined training $L_{total} = L_{DNN} - \lambda L_{ANN}$, with λ being chosen with a compromise between the background rejection and the mass decorrelation

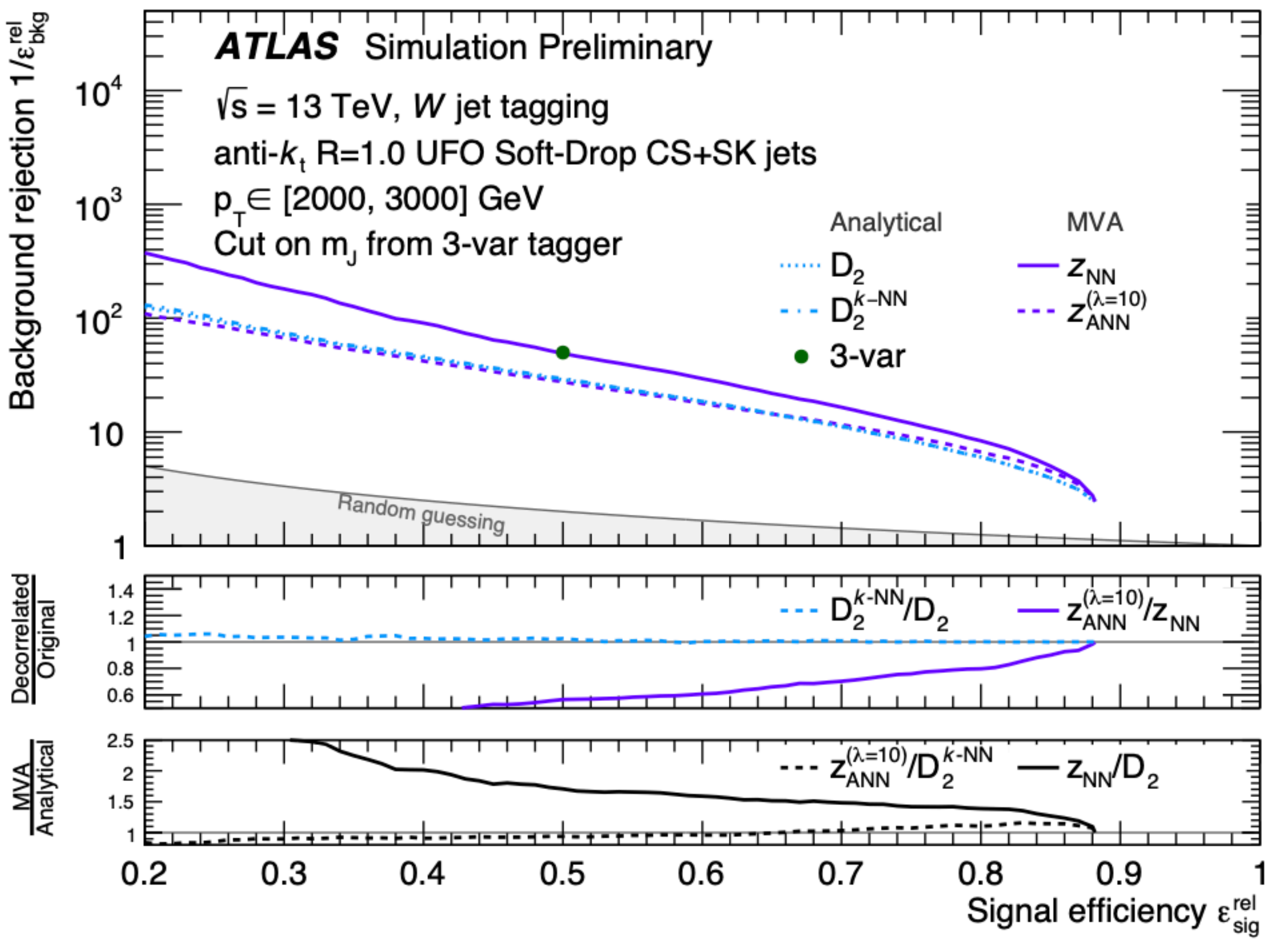


Baseline taggers

- Two taggers are considered as baseline taggers and are used for comparisons :
    - The “so-called” 3-var tagger, based on number of tracks, mass and D2
    - The DNN/ANN tagger, based on high level observables

Background rejection rate comparison of W taggers
- DNN tagger (violet solid) shows the best performance
- Decrease in performance after ANN is expected





Baseline taggers

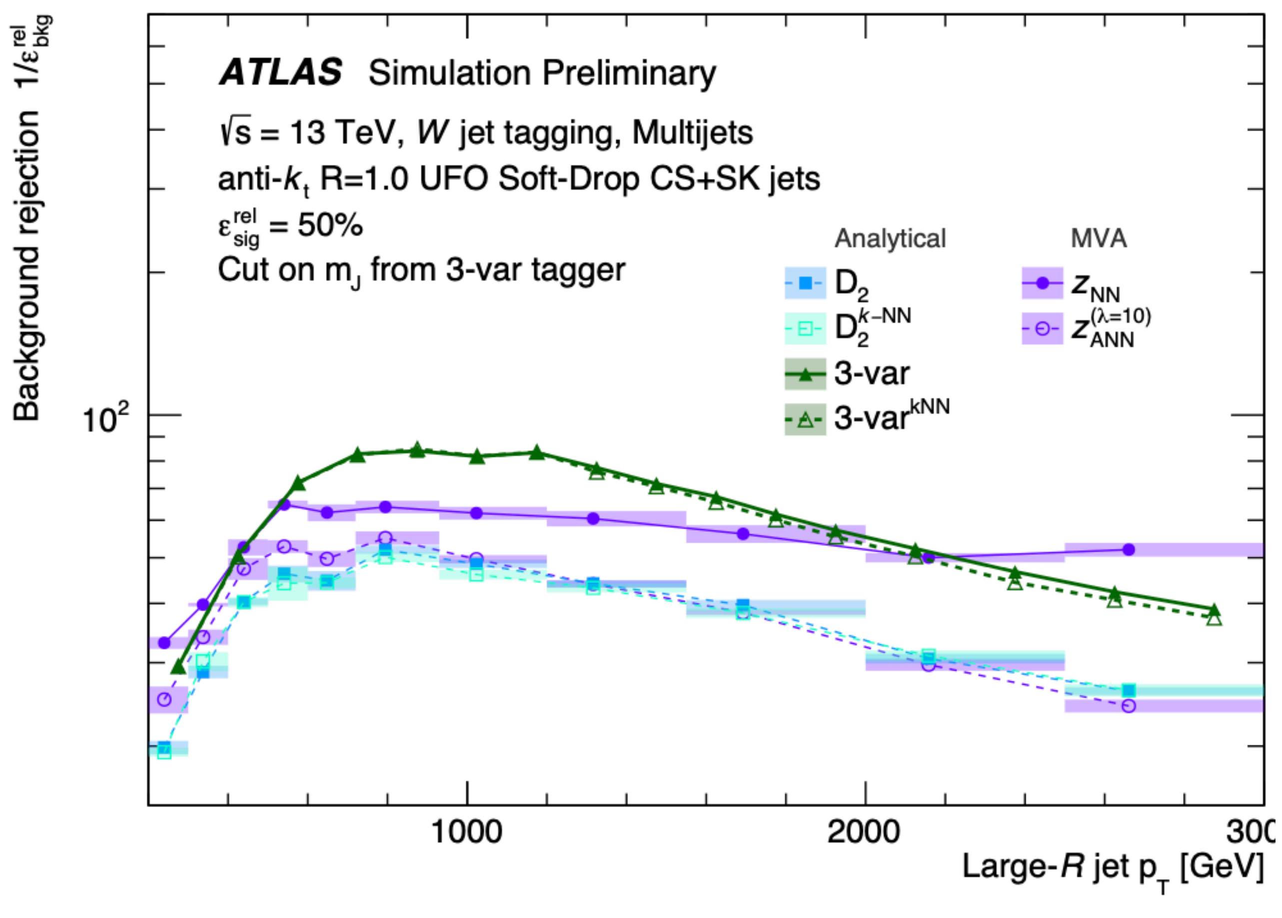
- Two taggers are considered as baseline taggers and are used for comparisons :
  - The "so-called" 3-var tagger, based on number of tracks, mass and D2
  - The DNN/ANN tagger, based on high level observables

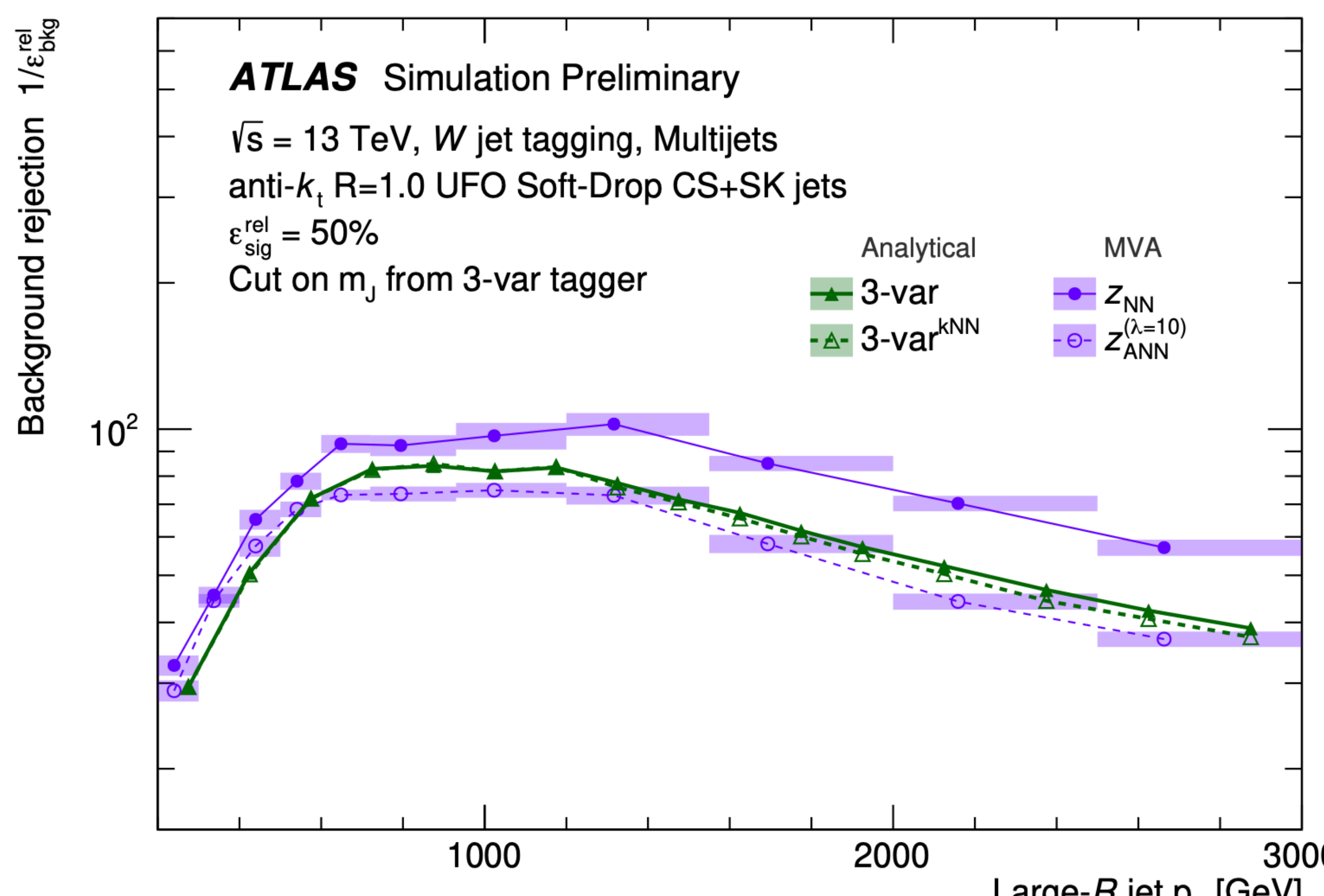
- Adding the information on the number of tracks helped in increasing the background rejection

- Previously, the 3-variable tagger showed better performance than DNN

- Now, DNN performing better than the 3-variable cut based tagger, and the ANN is comparable with the 3-variable tagger performance

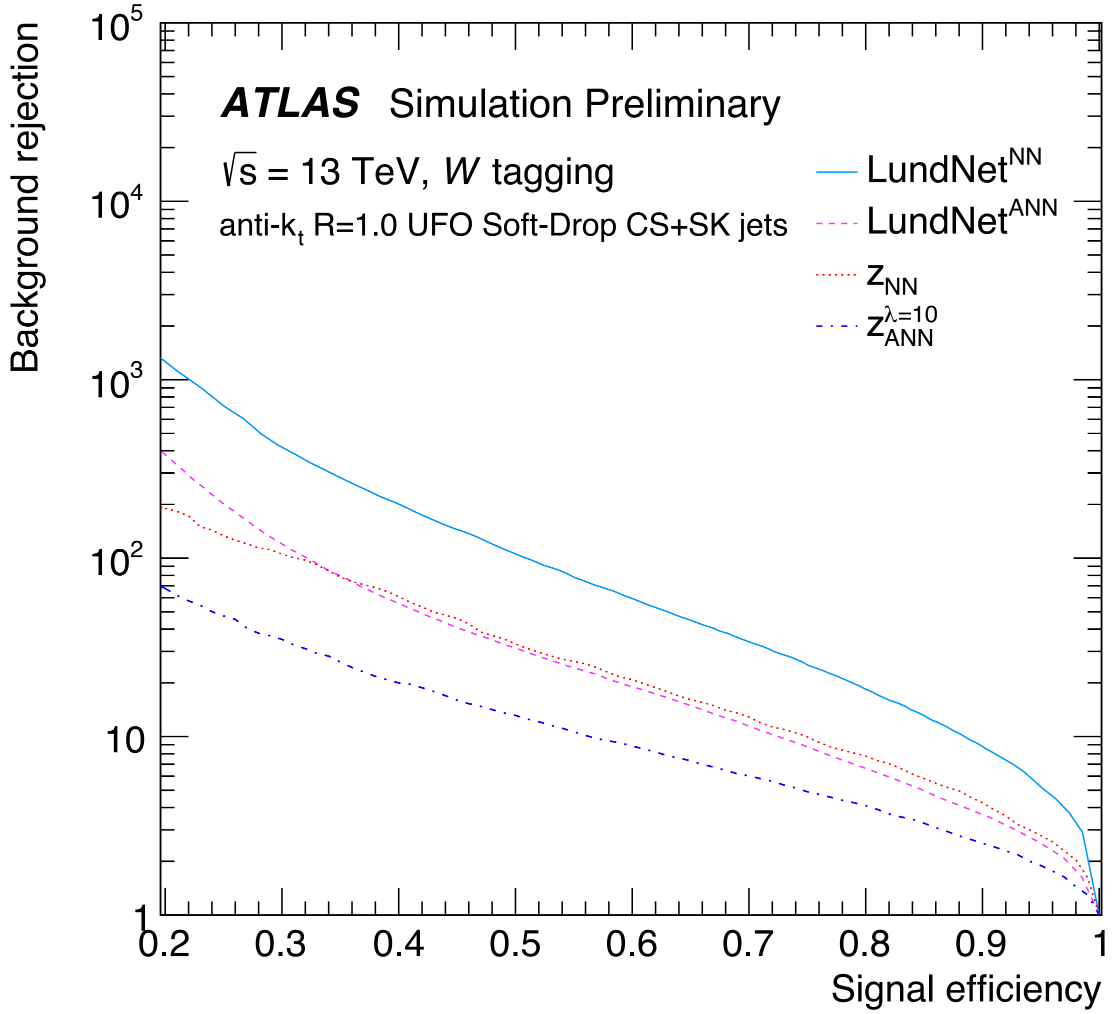
- Reason: Most other feature exploit the 2 prong behavior of the W/Z decay,
  whereas the number of tracks is a good quark/gluon discriminator

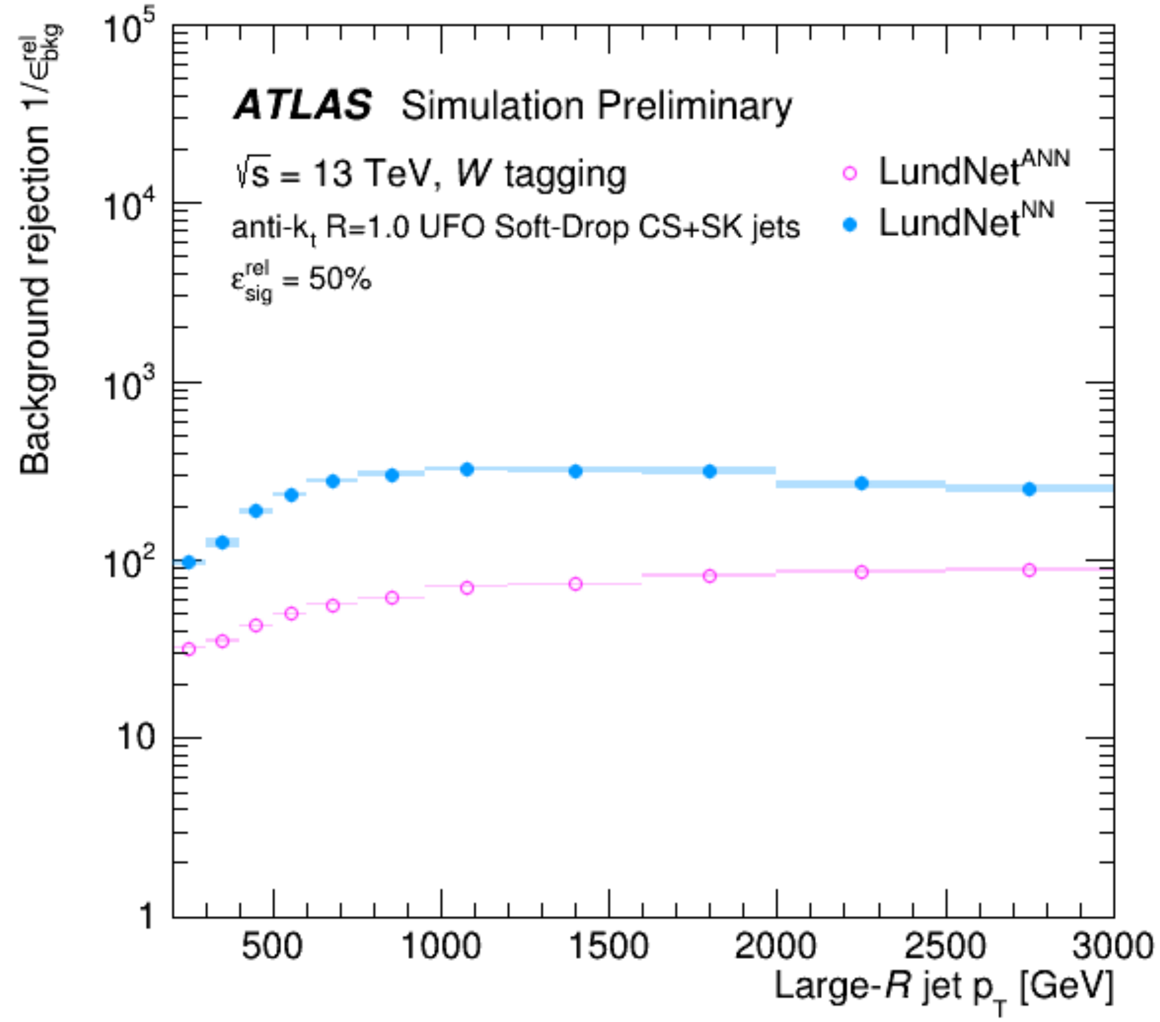




Results LundNet

Let's go back to LundNet





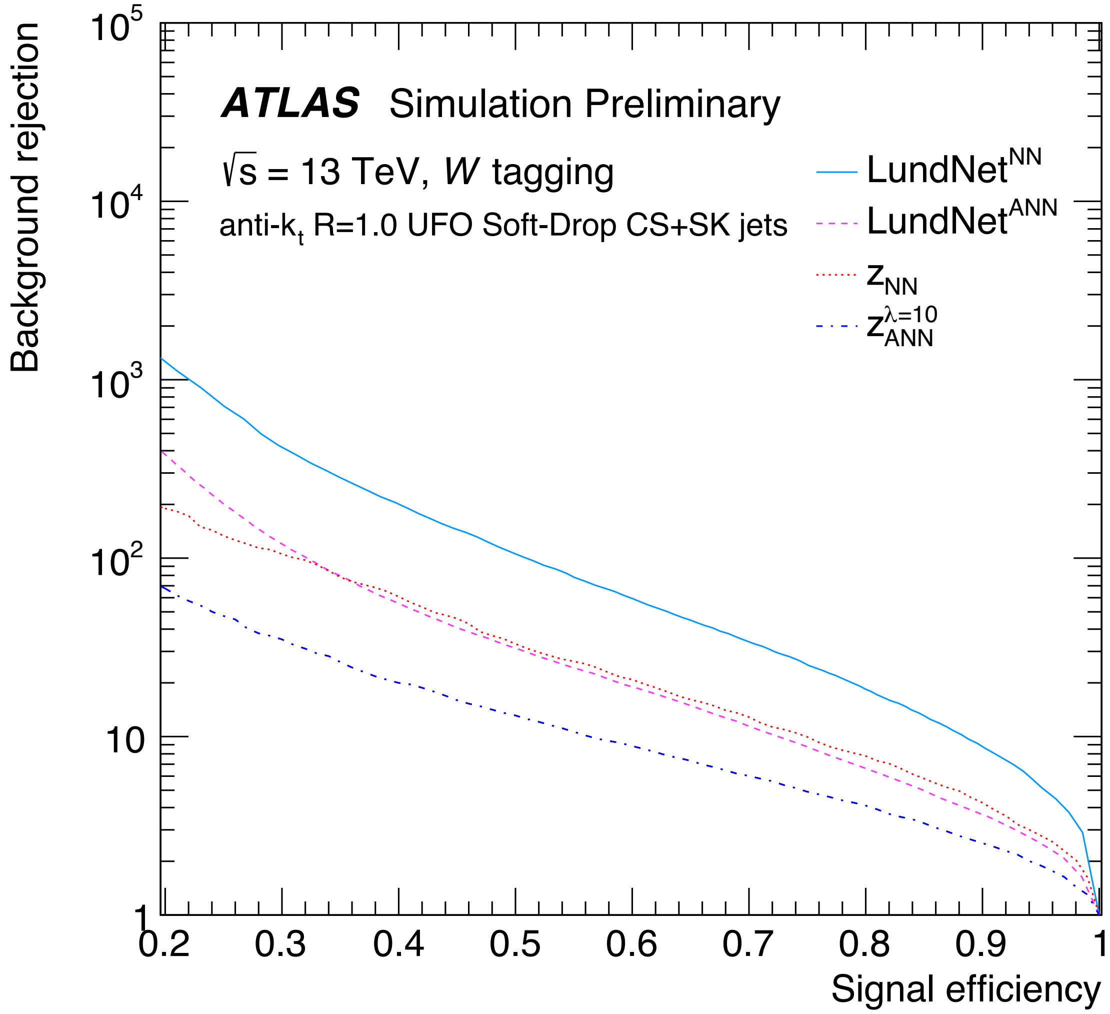
-The LundNet tagger without mass decorrelation achieves the best performance

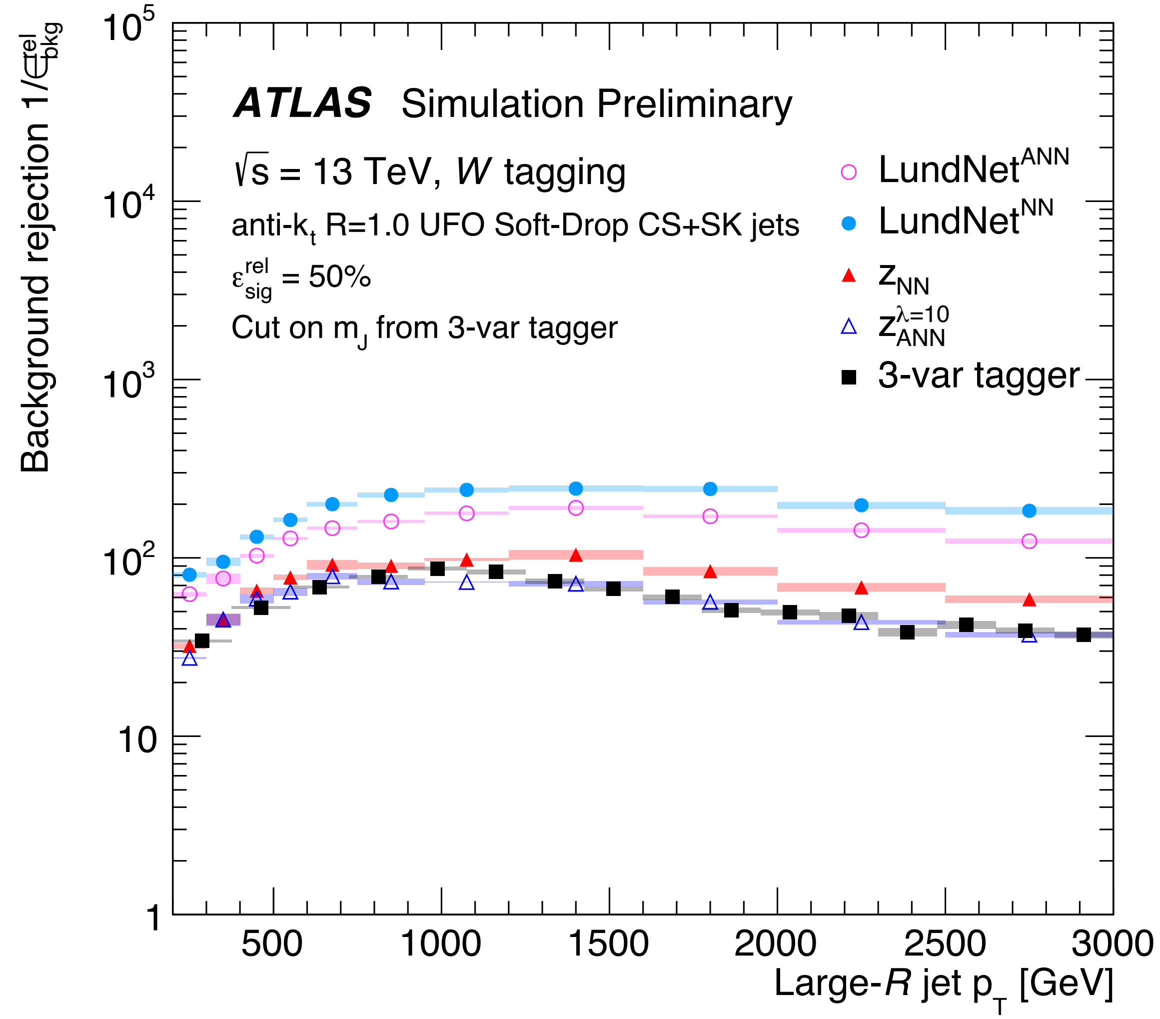
- The adversarial network significantly deteriorates performance (for both LundNet and DNN)

- 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)

Results LundNet

Let’s go back to LundNet



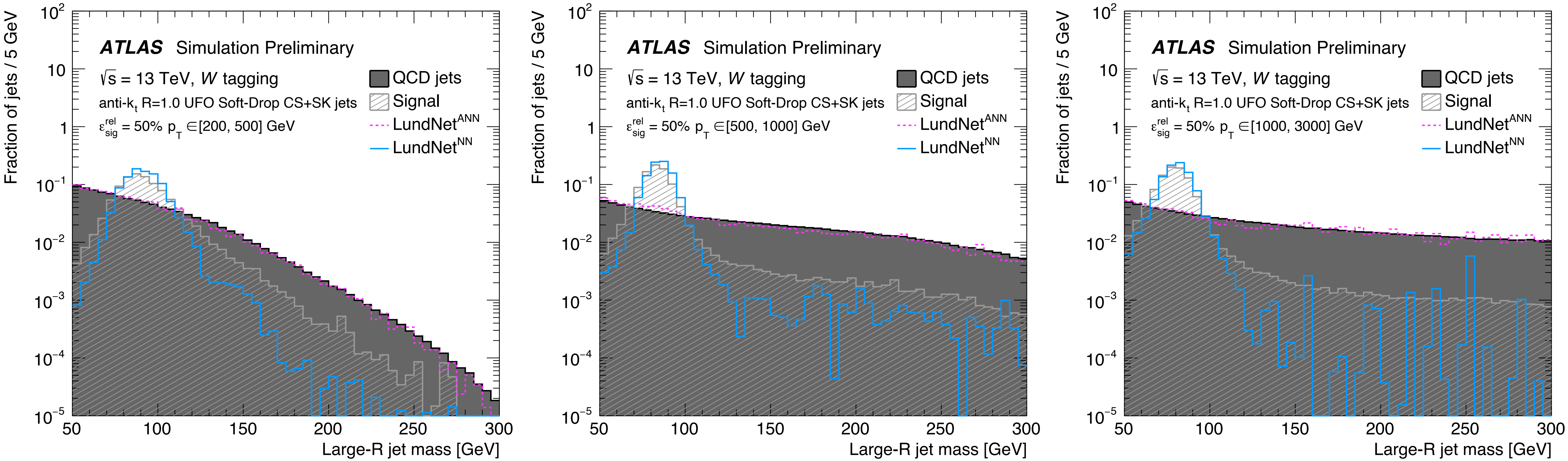


-The LundNet tagger without mass decorrelation achieves the best performance

- The adversarial network 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)

Results LundNet



- Across all p_T ranges:

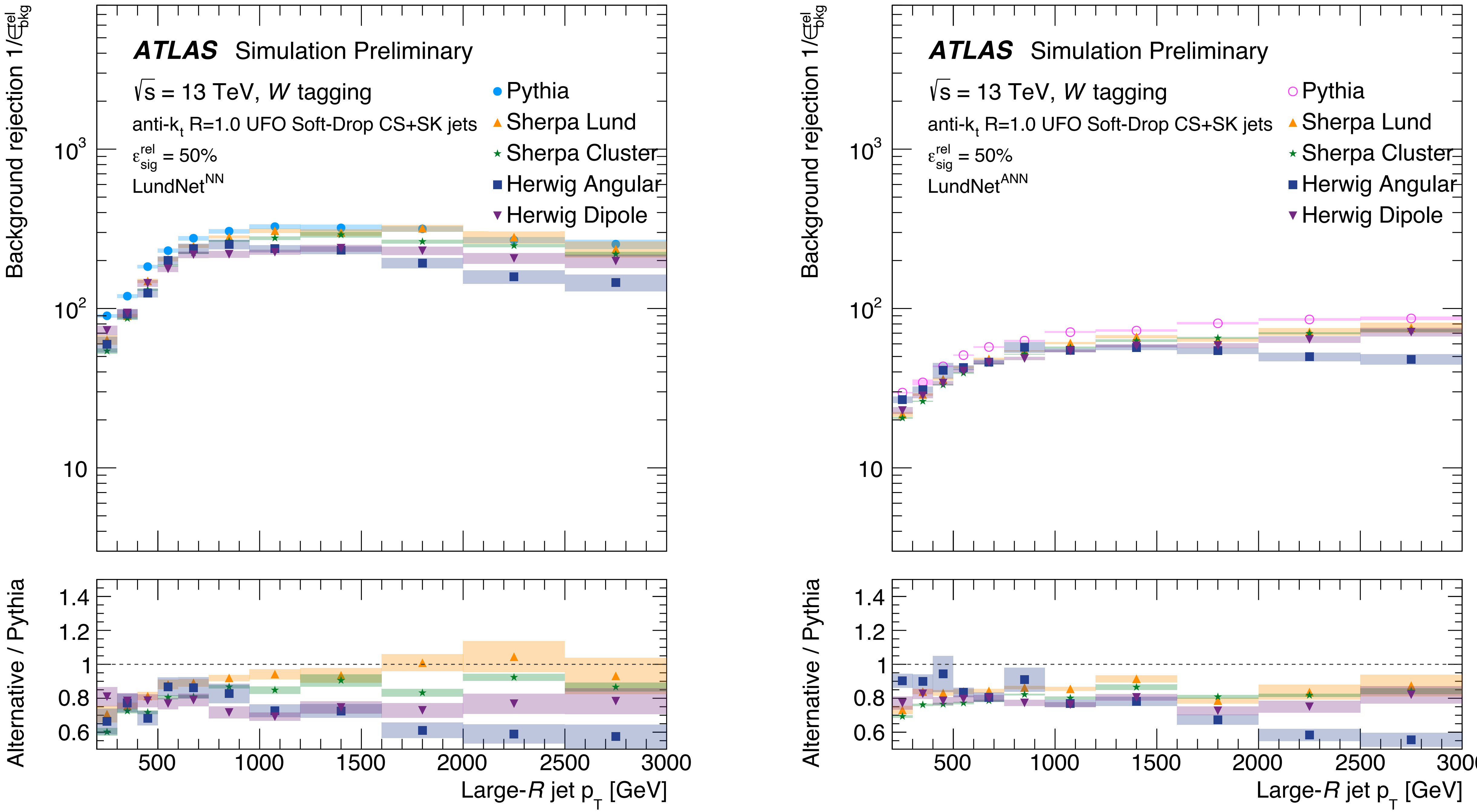
  - LundNet^{NN} is able to retrieve a peak around the W mass

  - LundNet^{ANN} is able to retrieve the shape of the QCD background

- To quantify the agreement, the KL divergence was calculated:
  - Got values < 1% for both comparison:
    LundNet^{NN} with signal
    LundNet^{ANN} with QCD

Backup: results for WP@80%

Results LundNet



- The LundNet tagger shows a decrease in background rejection of 20%-40% for 50% working point

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

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

Conclusions

- Jets are not just an image, they are a process that can be measured by deconstructing the jet clustering algorithm

- This is the ideal field of applications of a GNN

- Results are better than other methods, but mass sculpting shows up in background peaking at m(W)

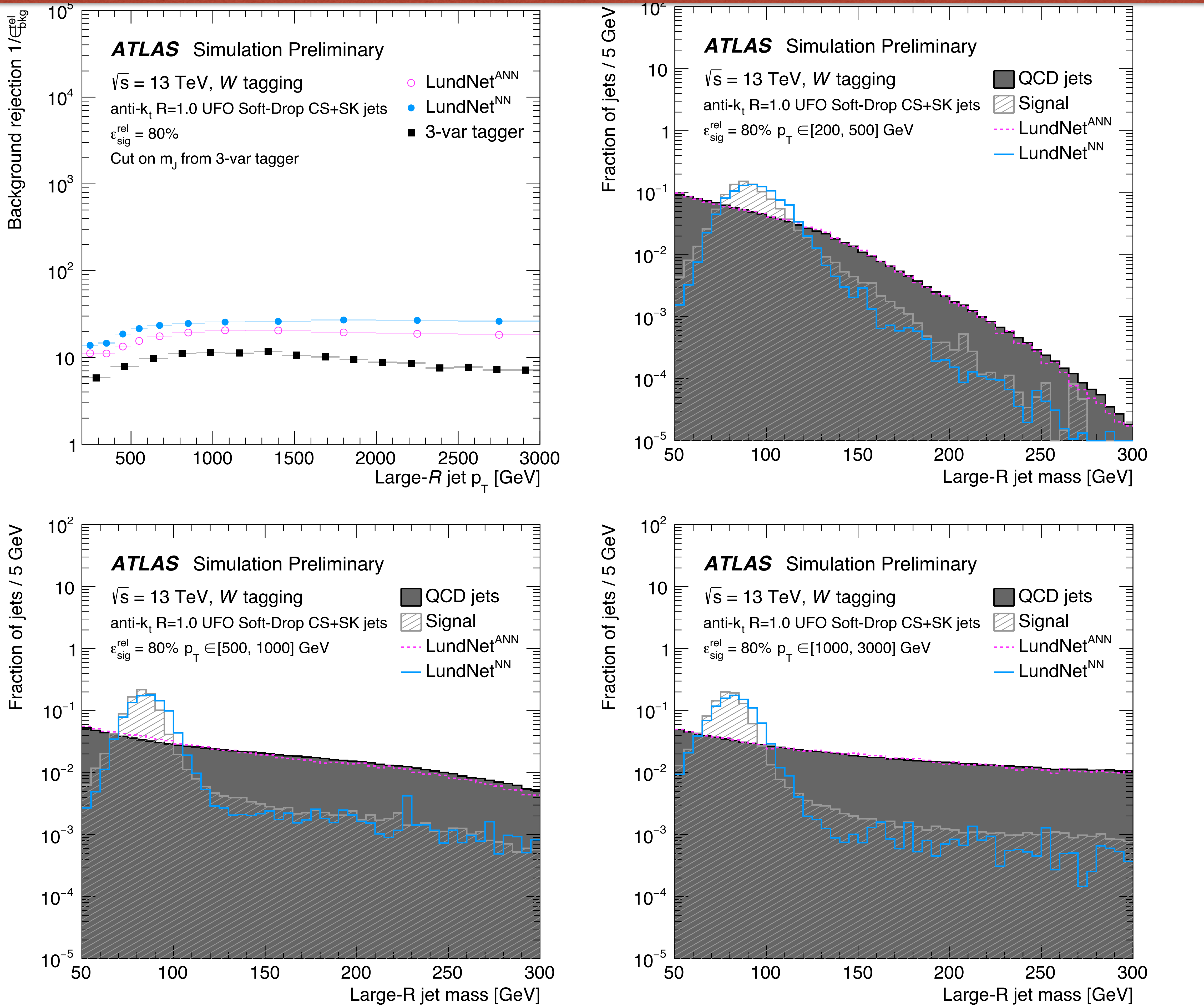
- Use of adversarial network solves the issue but reduces performance

- Good mass decorrelation and background rejection in all p_T intervals

- Mass correlated tagger tests using other MC generators result in good background rejection

Backup

Results for WP@80%



Results for WP@80%