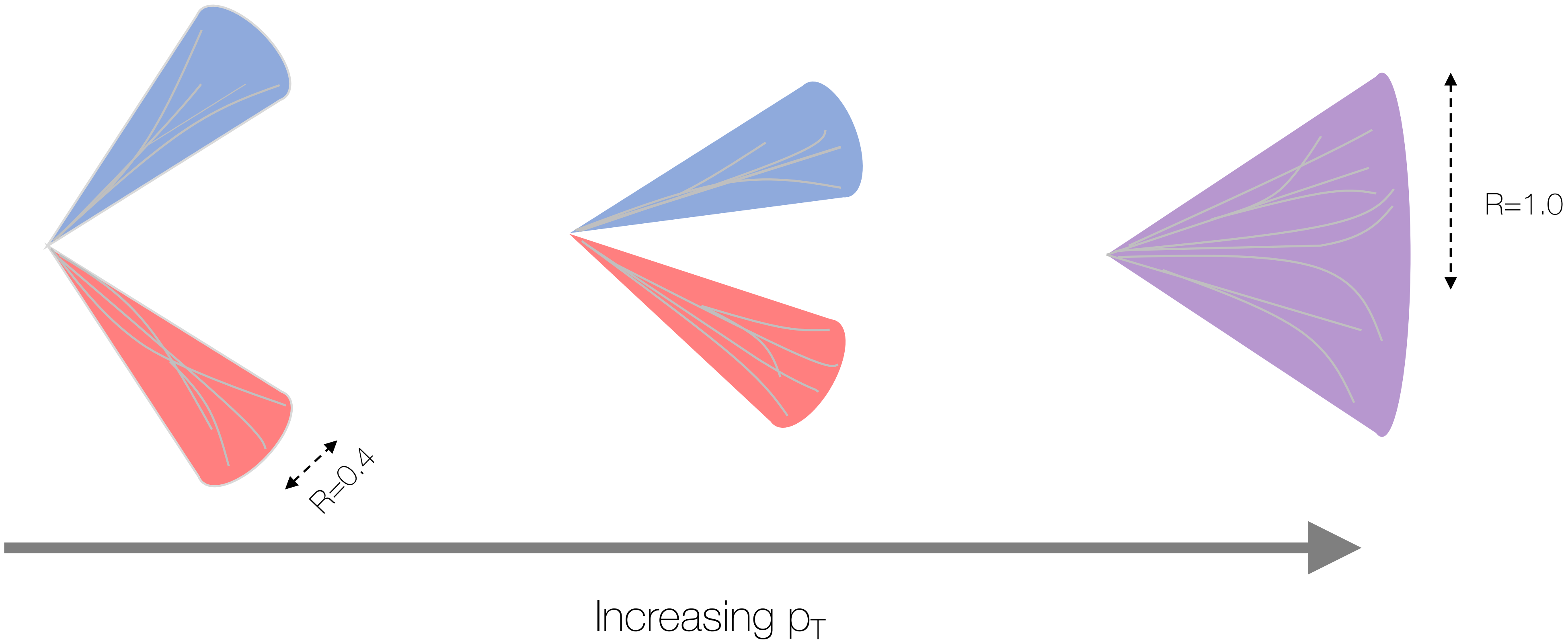
Boosted H→bb/cc Tagging in ATLAS

Jackson Barr
2025-02-20

UCL

DESY.

Boosted Jets

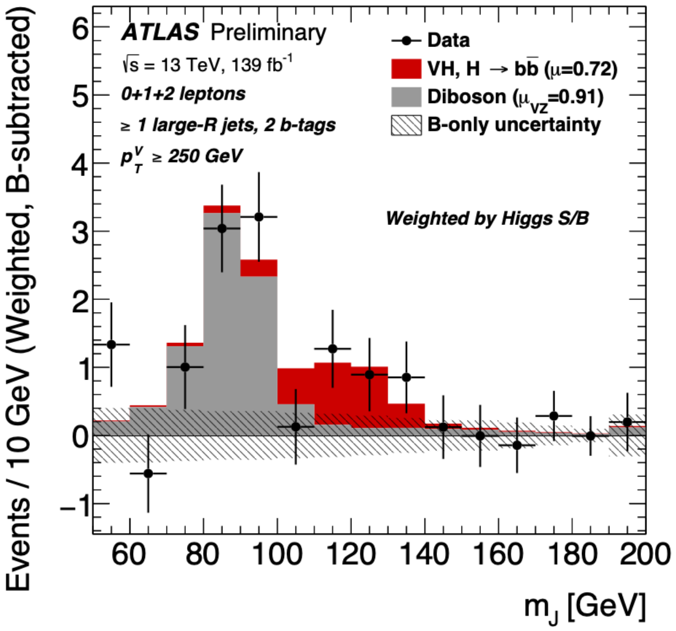


- Standard b-tagging algorithms use R=0.4 anti-k_t jets but as decay products become more collimated at high pT, objects can be reconstructed within a R=1.0 jet
- Rule of thumb for two body decay, $R \approx 2m/p_T$ – Higgs boson decay products are boosted above 250 GeV

Higgs Tagging

- H → bb is the largest branching ratio for the decay of a SM Higgs, boosted bb-tagging important for precision measurements of the Higgs p_T spectra or for searches involving high mass resonances

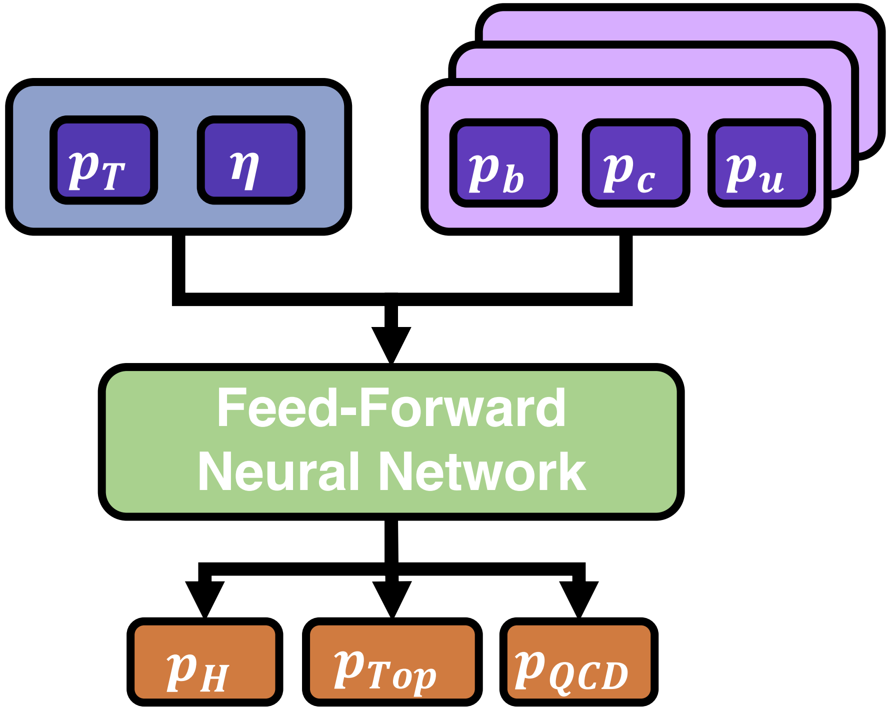
- The boosted regime has also been shown to be very sensitive for H → cc (which still hasn't been observed!)

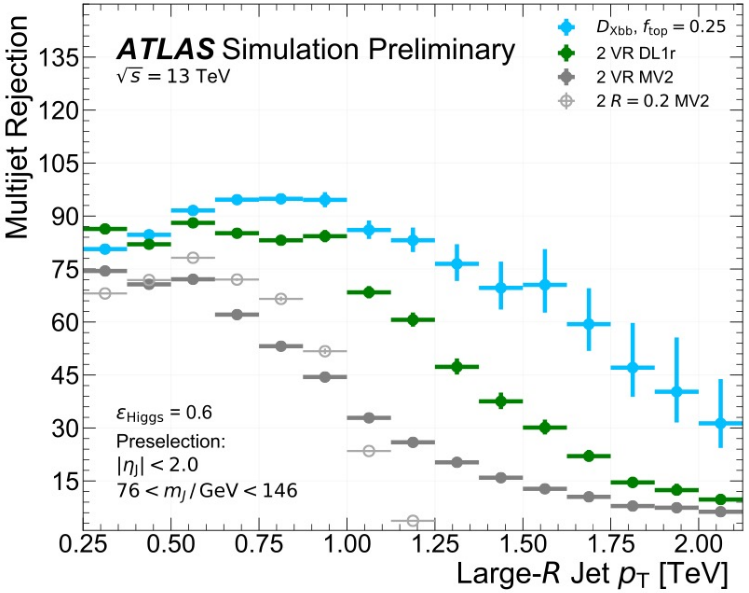


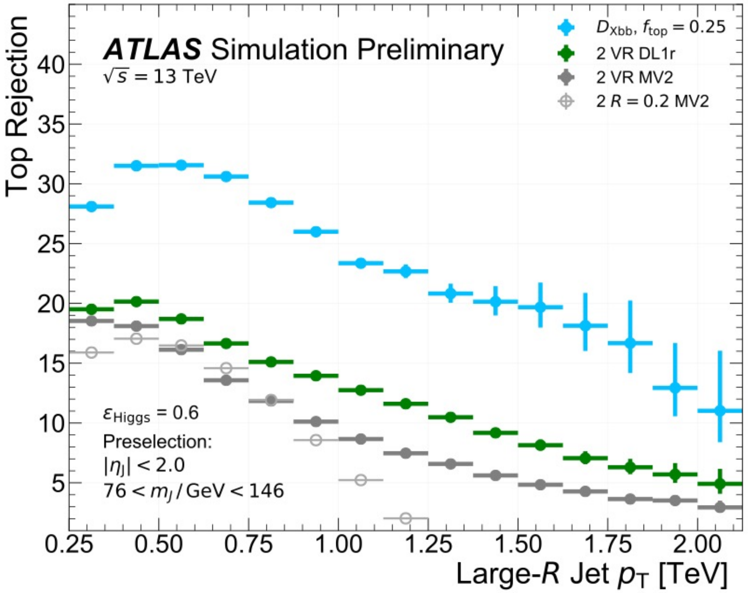
	bb	WW	ττ	ZZ	γγ
bb	34%				
WW	25%	4.6%			
ττ	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
γγ	0.26%	0.10%	0.028%	0.012%	0.0005%

Previous Approach

- The previous standard b-tagging algorithm used in Run 2 was the DL1r tagger – based upon a recurrent neural network
- The previous Xbb tagger consisted of a neural network trained on the DL1r outputs of up to three variable radius sub-jets within the large-R jet along with the jet kinematics





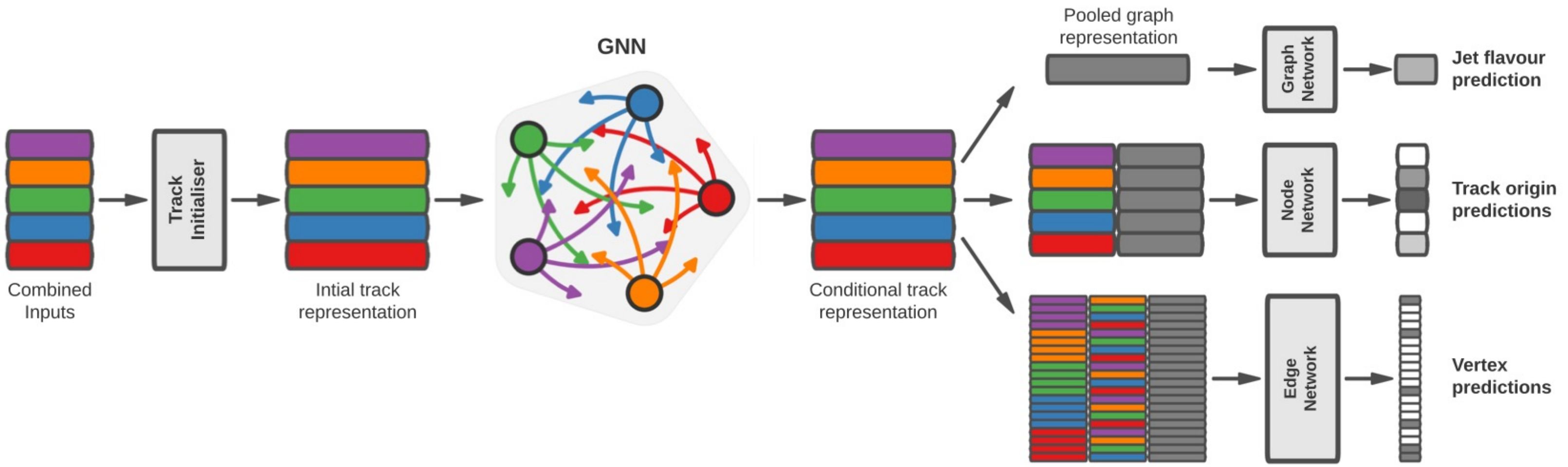


Current Approach – GN2X

- The previous approach b-tags the sub-jets independently and can only learn correlations between their b-tagging scores – GN2X uses all tracks within the jet

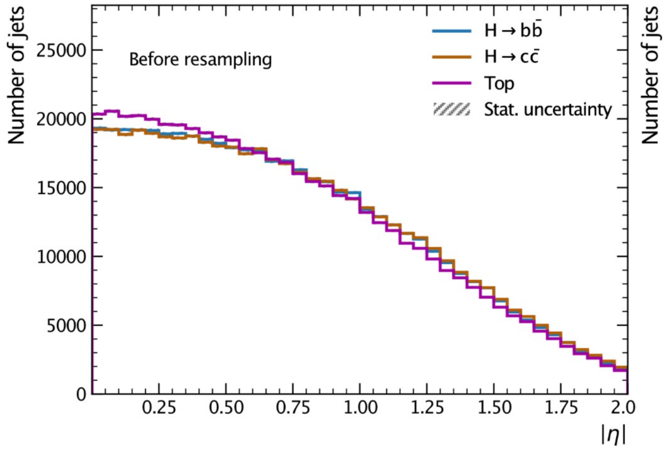
- GN2X is a transformer based model that not only can be used for the identification of boosted Higgs decays but also predicts the origin of the tracks and perform vertex finding

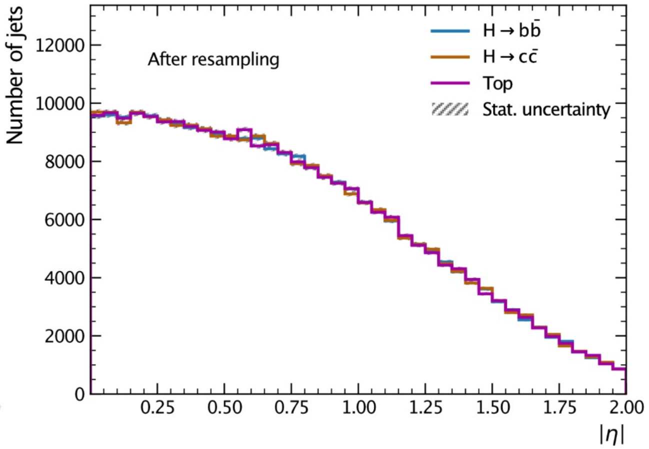
- An important consideration for Higgs tagging is trying to avoid sculpting the background distributions to look like the signal – a Higgs sample with a large decay width is used to give a flatter mass distribution



Inputs and Samples

- Inputs include the jet pT, η and mass along with track level features – heterogeneous versions of model uses information from particle flow objects and/or kinematics of subjets
- A resampling in the jet kinematics is performed to reduce the effects of any kinematic sculpting
- Trained on O(100 m) jets, model with a few million parameters





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 (GN2X + Subjets only)
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 flavour 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

Transformer Models

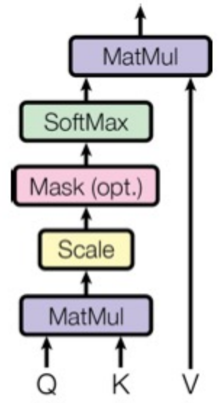
- Transformers are a model architecture that arose from the field of Natural Language Processing (NLP) – architecture used by LLMs like ChatGPT

- The key to the transformer architecture is the Attention Mechanism – calculates the pairwise “relevance” between all input objects

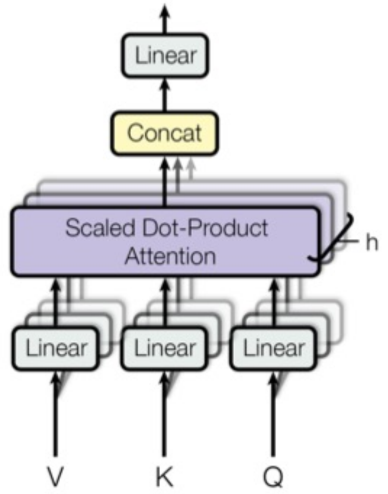
- Each object is then updated by an attention weighted sum of all other objects

- Attention mechanism is a useful inductive bias for this task, tracks originating from a common vertex, or have parent-daughter relationship should have high attention weights, fake tracks and pileup tracks should have lower weights with real PV tracks

Scaled Dot-Product Attention



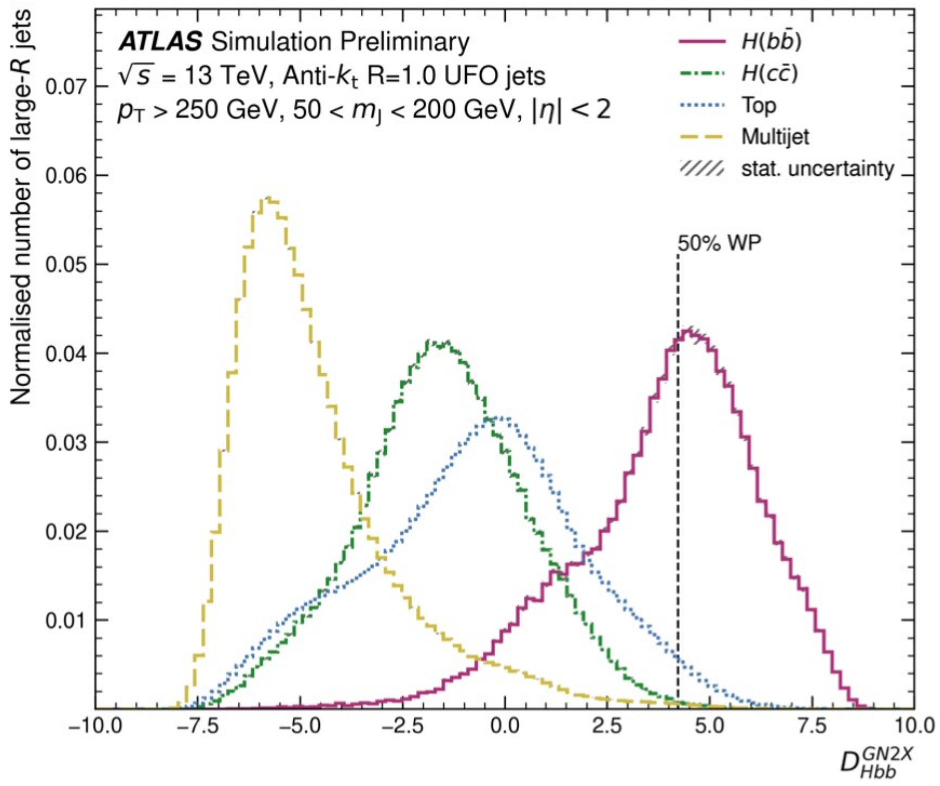
Multi-Head Attention



GN2X Outputs

- GN2X adds a H → cc output class in addition to the H → bb, top and QCD classes from the previous tagger
- A discriminant score is built using a weighted log likelihood ratio
- Auxiliary task outputs are a per track probability score for each type of track and a list of vertices

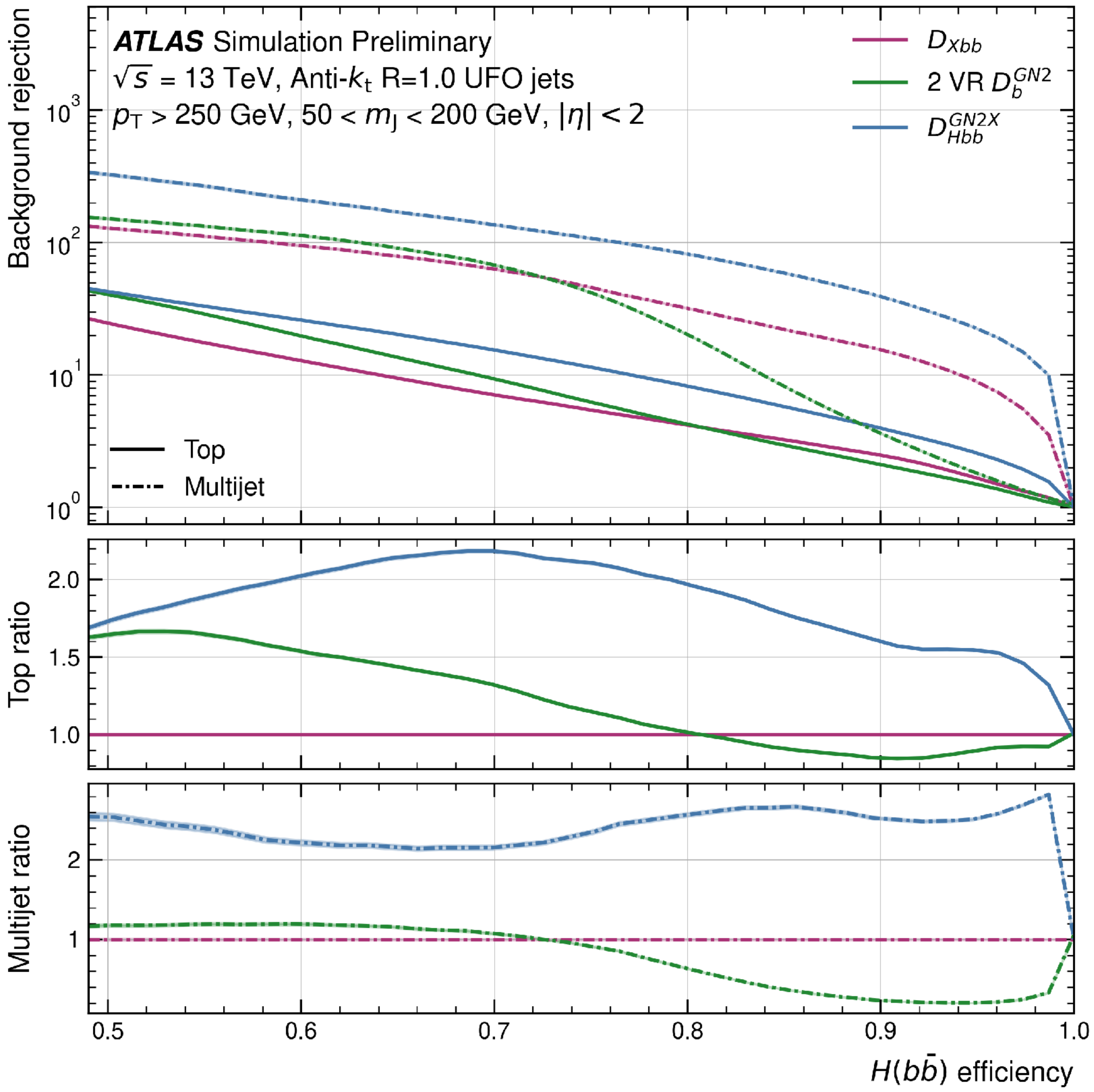
$$D_{\text{Hbb}}^{\text{GN2X}} = \ln\left(\frac{p_{\text{Hbb}}}{f_{\text{Hcc}} \cdot p_{\text{Hcc}} + f_{\text{top}} \cdot p_{\text{top}} + \left(1 - f_{\text{Hcc}} - f_{\text{top}}\right) \cdot p_{\text{QCD}}}\right)$$



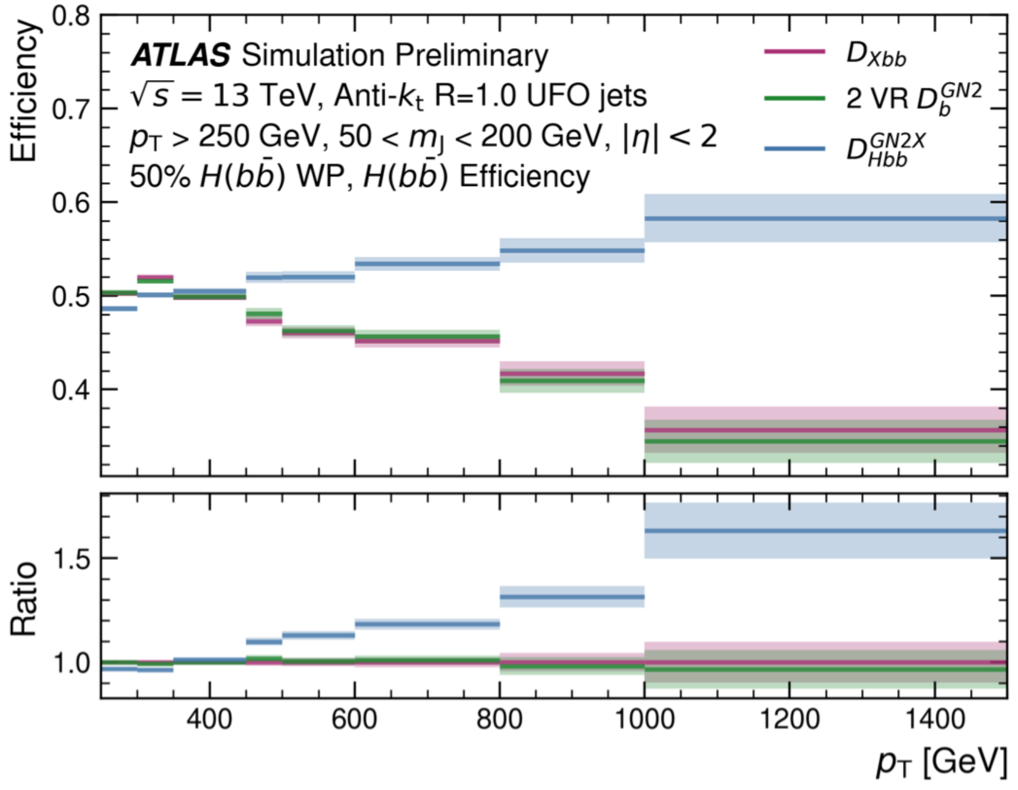
H⟶bb Performance

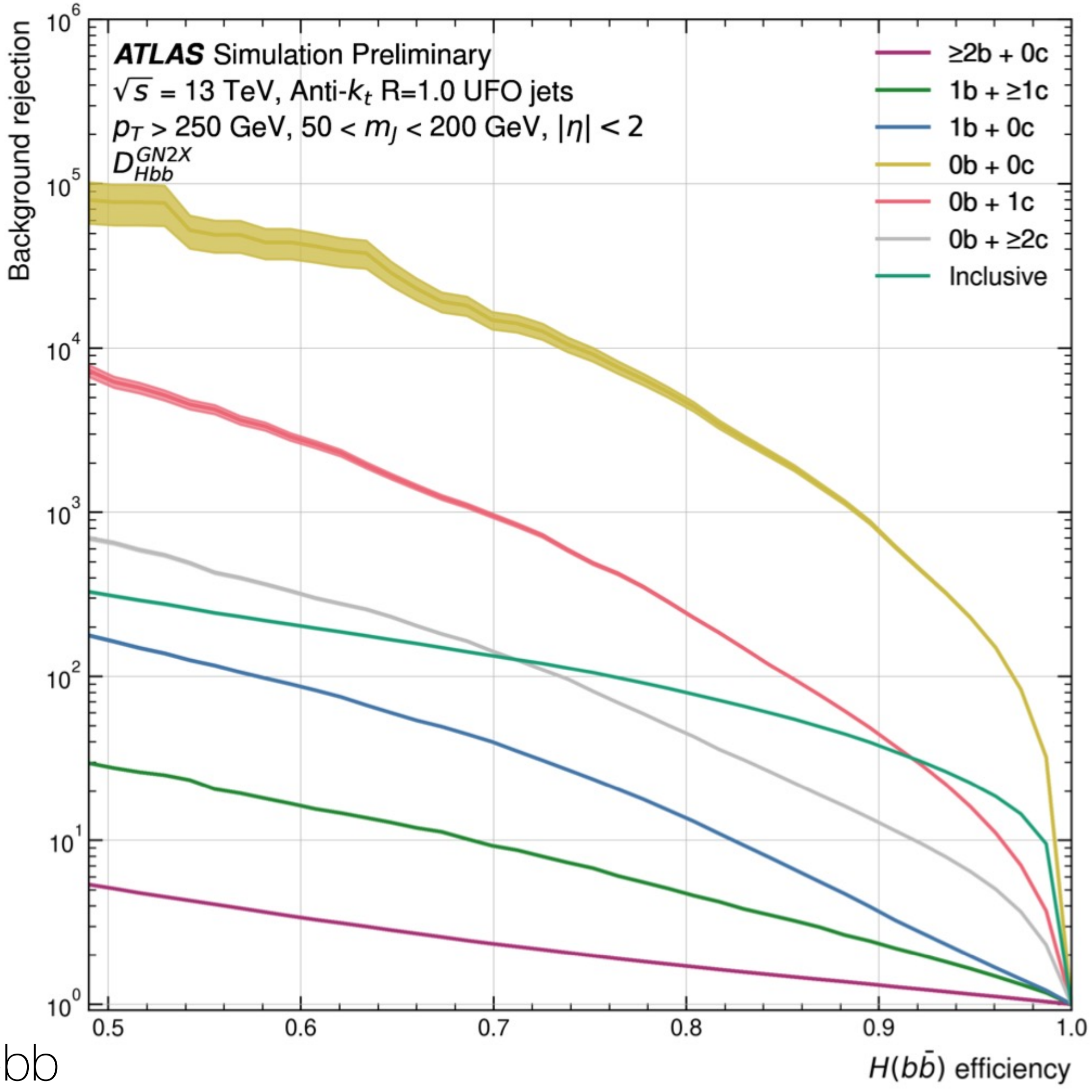
- GN2X significantly improvements in top and QCD rejection over the previous Xbb tagger, more than doubling QCD rejection for all signal efficiencies

- Performance compared to tagging two subjets with GN2 independently to shows how much training on all constituents in the jet inclusively brings over a subjet-based approach



H→bb Performance

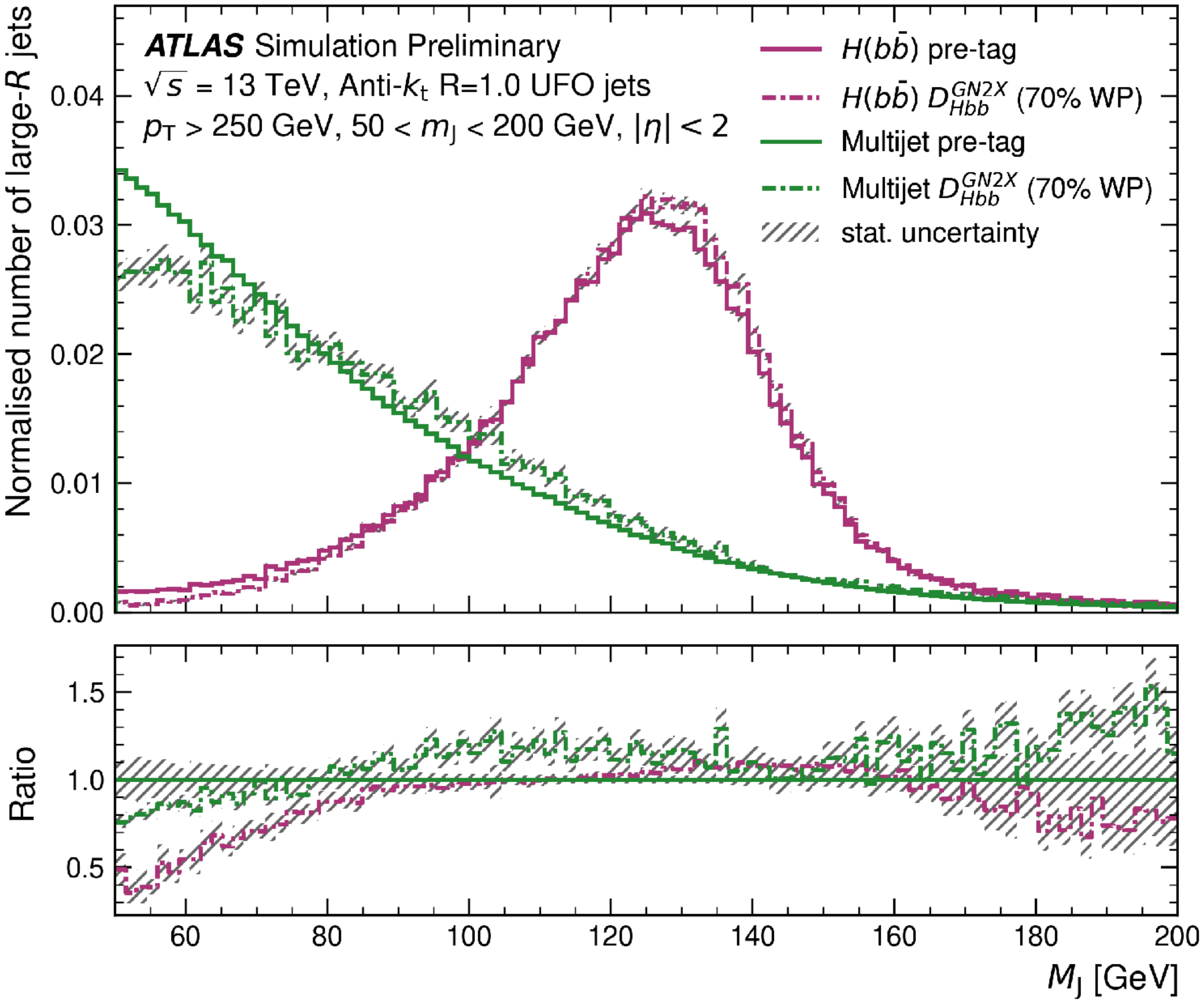




- Post-tag sample composition dominated by real b-hadrons within QCD jets – improving H→bb vs g →bb discrimination vital for further improvements

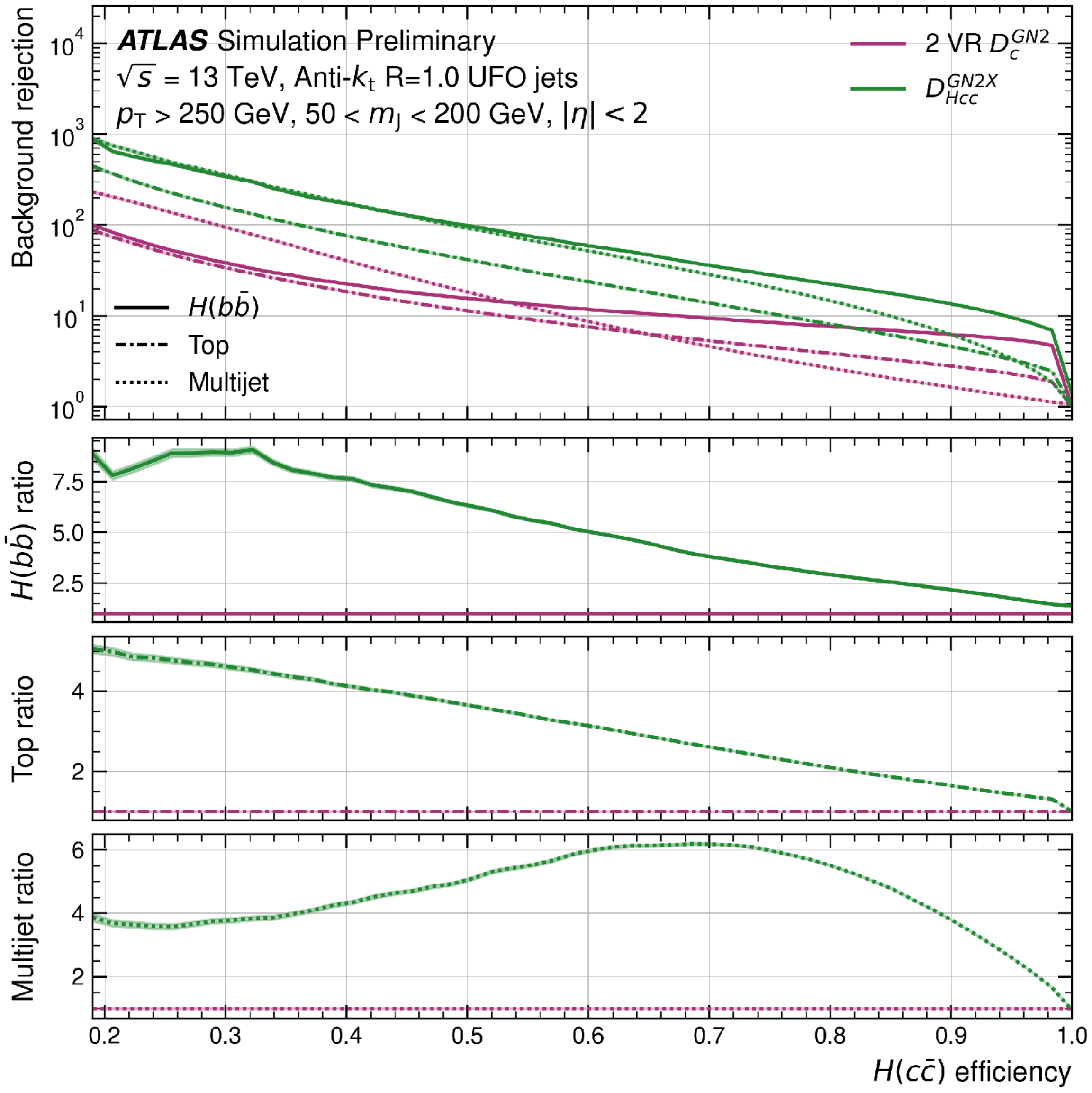
Mass Sculpting

- Jet mass distributions of the QCD background are compared pre- and post-tagging to evaluate the amount of mass sculpting present
- No localised peak at 125 GeV but there is still some residual mass sculpting – mostly a product of changing flavour fractions pre- and post-tag

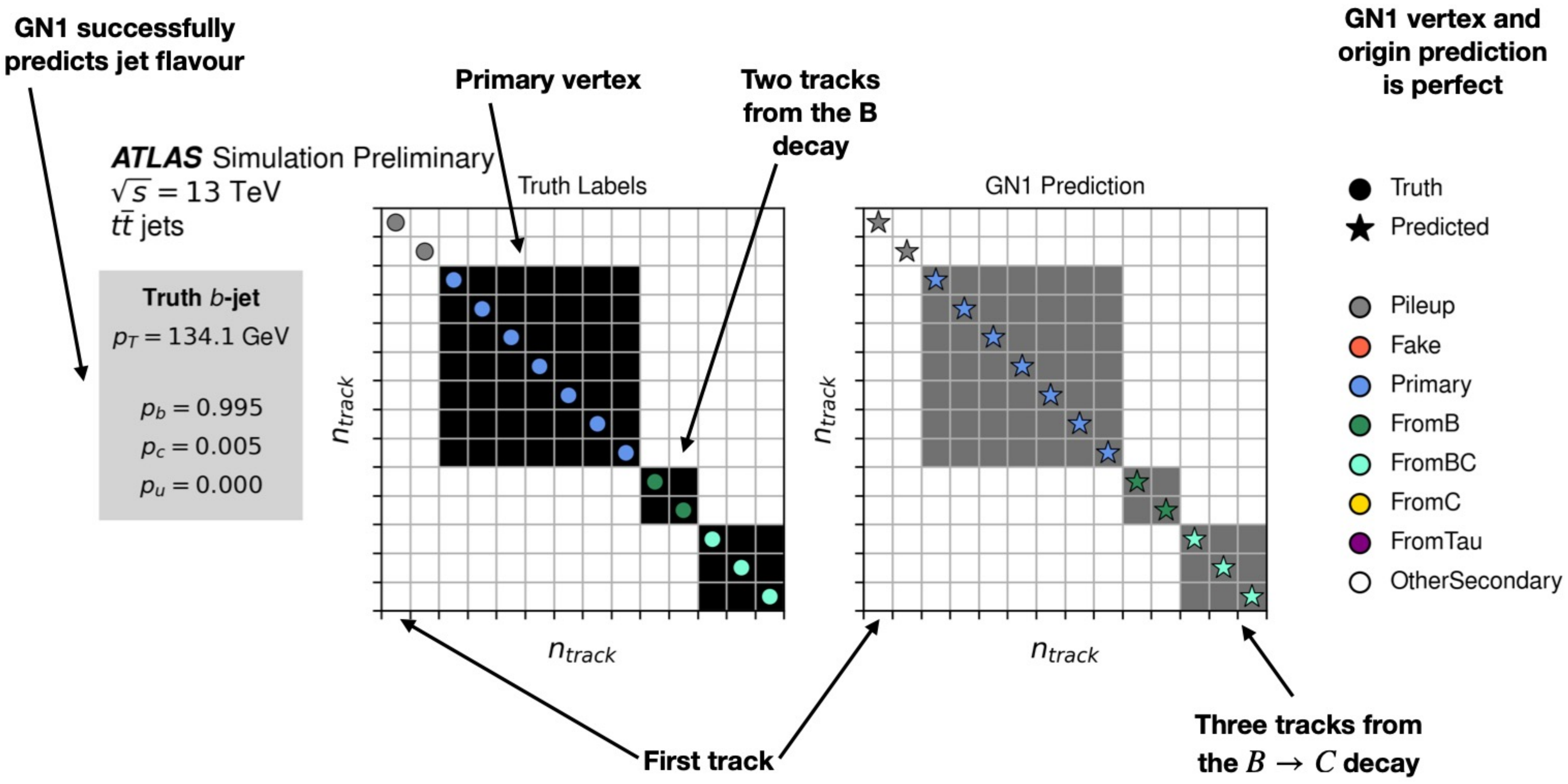


H→cc Performance

- As H → cc is a new addition, the only baseline we can compare to is the c-tagging of the subjets independently
- Larger improvement over baseline compared to the b-tagging case!
- Identification of c-hadrons more challenging than b-hadrons e.g. a 1% QCD mist-tag efficiency corresponds to a ~75% H → bb efficiency, but only a ~50% H → cc efficiency

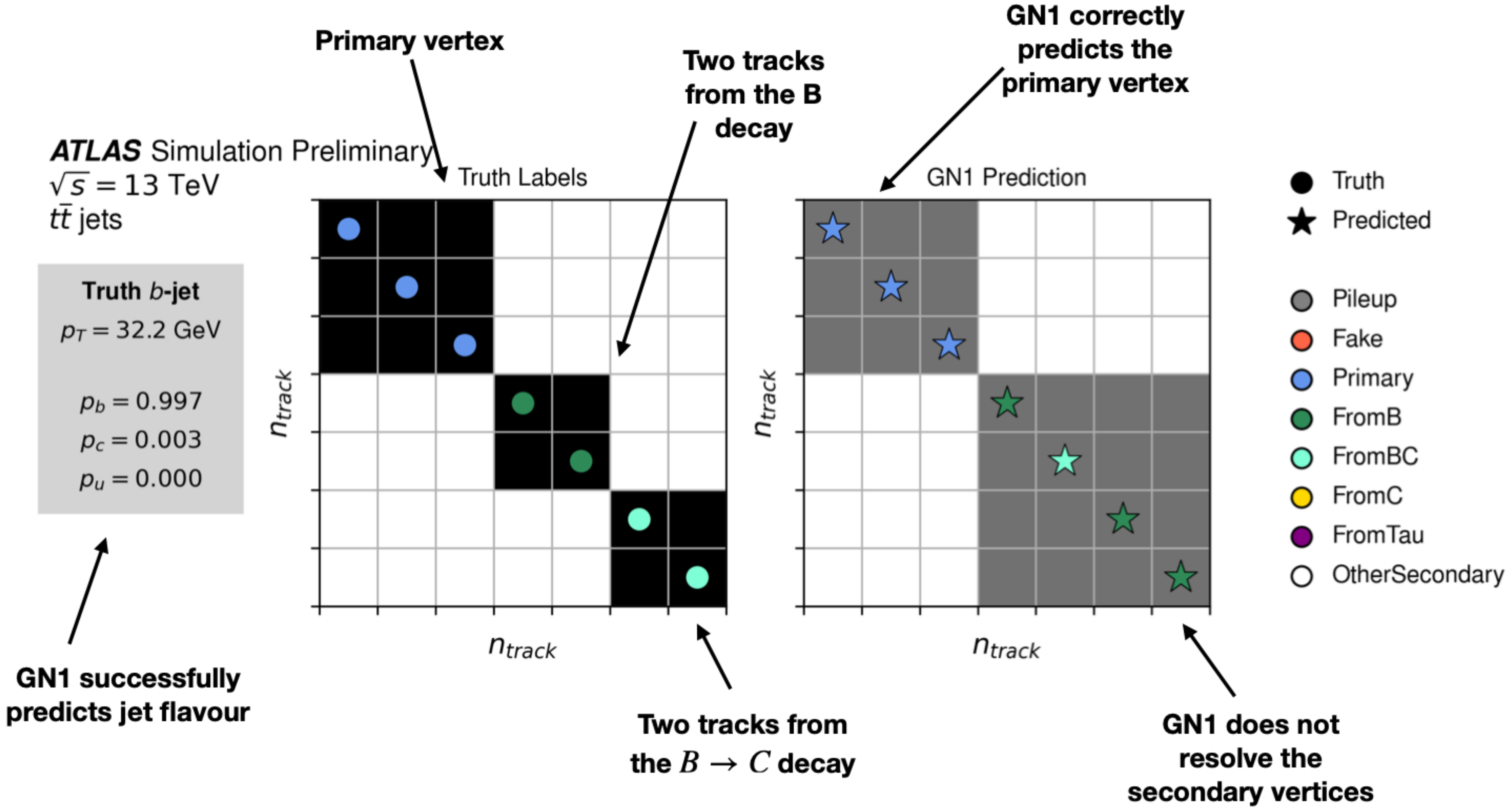


Auxiliary Tasks



Example vertexing performance from previous GN1 model

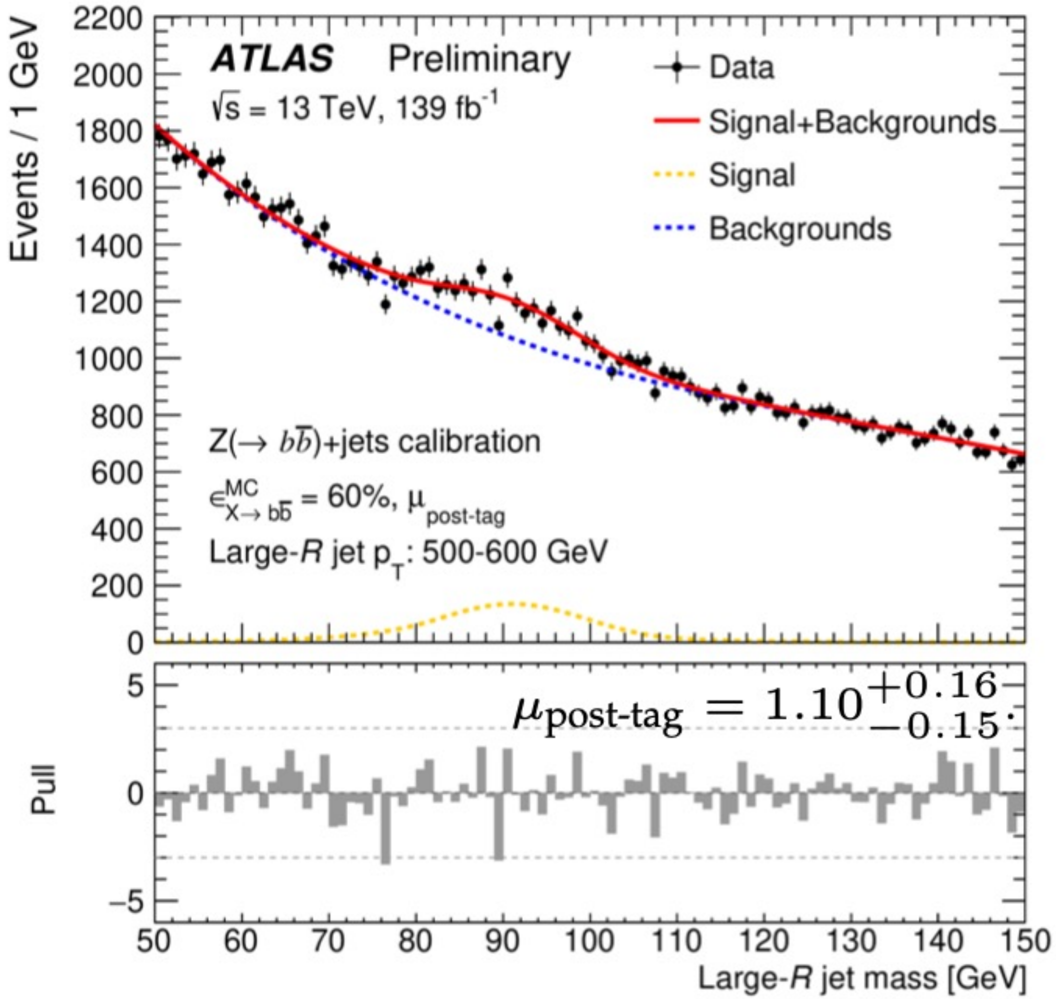
Auxiliary Tasks



Signal Calibration

- MC performance is nice but at the end of the day data is what matters
- Tagger is ~mass agnostic, therefore we assume that we can use a $Z \rightarrow bb$ standard candle as our signal proxy in the calibration
- Alternate calibration approach using $g \rightarrow bb$ has also been explored as well

$$SF = \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{MC}}} = \frac{\frac{N^{\text{data}}_{\text{passed}}}{N^{\text{data}}_{\text{total}}}}{\frac{N^{\text{MC}}_{\text{passed}}}{N^{\text{MC}}_{\text{total}}}} = \frac{\frac{N^{\text{data}}_{\text{passed}}}{N^{\text{MC}}_{\text{passed}}}}{\frac{N^{\text{data}}_{\text{total}}}{N^{\text{MC}}_{\text{total}}}} = \frac{\mu_{\text{post-tag}}}{\mu_{\text{pre-tag}}}$$



Where to Next?*

*As much as I'm allowed to say that is ;)

Heterogenous Inputs

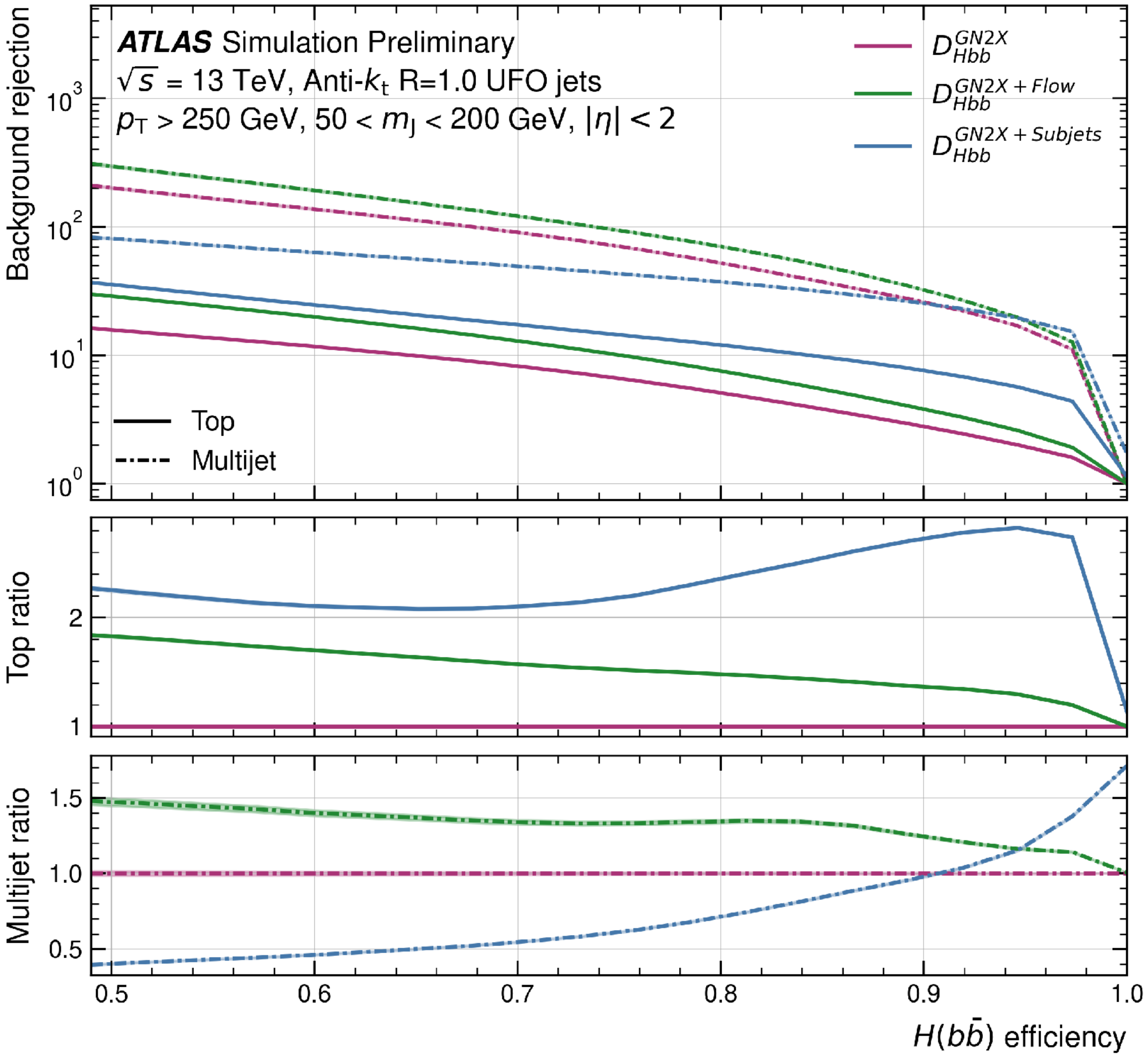
or how ATLAS finally learned pflow is pretty cool for flavour tagging

- Transformers seem here to stay for a while so after doing the easy step of switching to them, how do we further improve as architecture improvements become more minor?

- Two types of approach: put more into our models or try getting more out

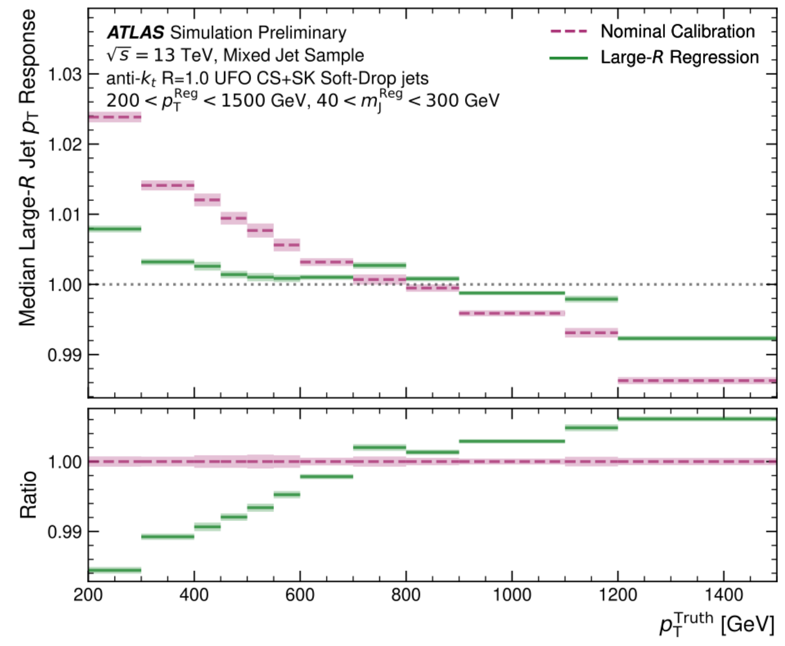
- An example of the first is the inclusion of neutral [PFlow constituents]

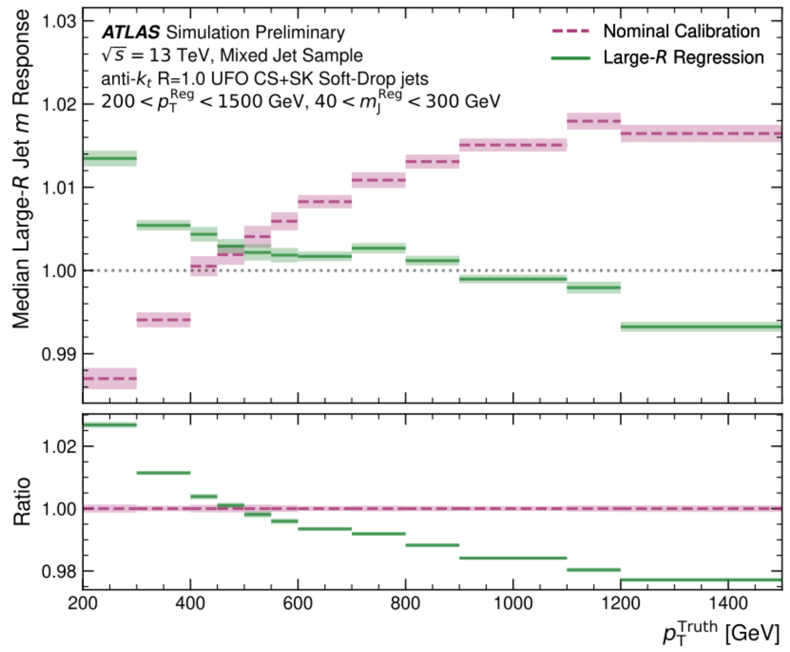
- Neutral constituent information leads to a further ~50% improvement in QCD rejection



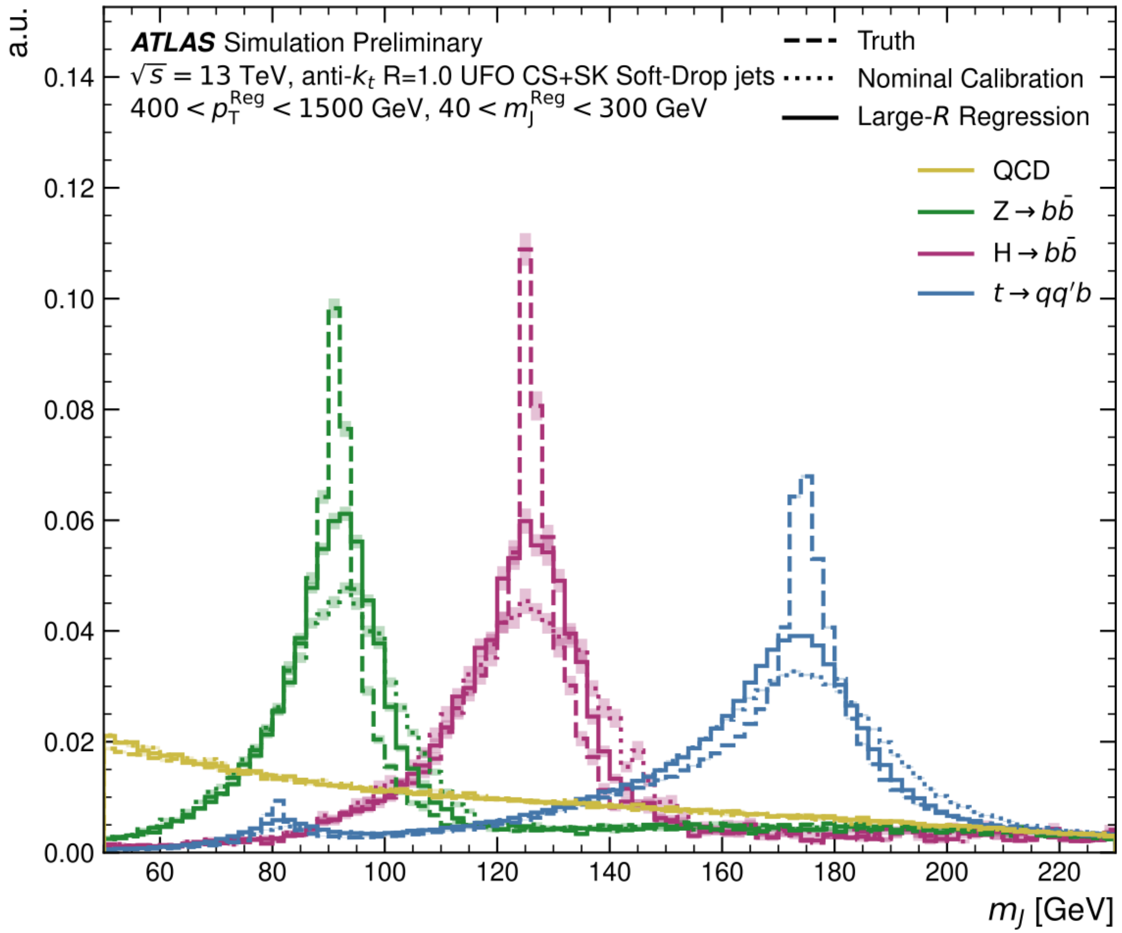
Jet Regression - Response

- We exploit the unique characteristics of b-hadron decays for flavour tagging e.g. displaced tracks, high mass secondary vertices, semi-leptonic decays
- These features also affect the jet mass and p_T and so a dedicated mass and p_T calibration for jets containing b-hadrons is desirable
- A recent result used the GN2 architecture for this task in both small-R and large-R jets
- Improvements in both the p_T and mass responses (difference between truth and reconstructed) are seen across a wide p_T range





Jet Regression - Resolution



- In addition to reducing the bias in the jet response, the dedicated b-jet regression leads to much sharper resolutions

- Performance tested across a range of physics processes with improvements observed everywhere

Summary

- In recent years there's been a lot of effort to harmonise our efforts in single and double b-tagging – we now use the same model architectures and frameworks to train them

- There is a lot of interesting physics that can be done with these models from diHiggs searches to the measurement of the Higgs to second generation quark couplings

- Lots of ongoing further developments being done to further extend the performance and applicability of these taggers to a wider range of signatures

- Expect to see GN2X used in some of the early Run 3 results for those analyses!

Thanks for Listening!