Transforming Flavour Tagging on ATLAS

Greta Brianti on behalf of the ATLAS Collaboration

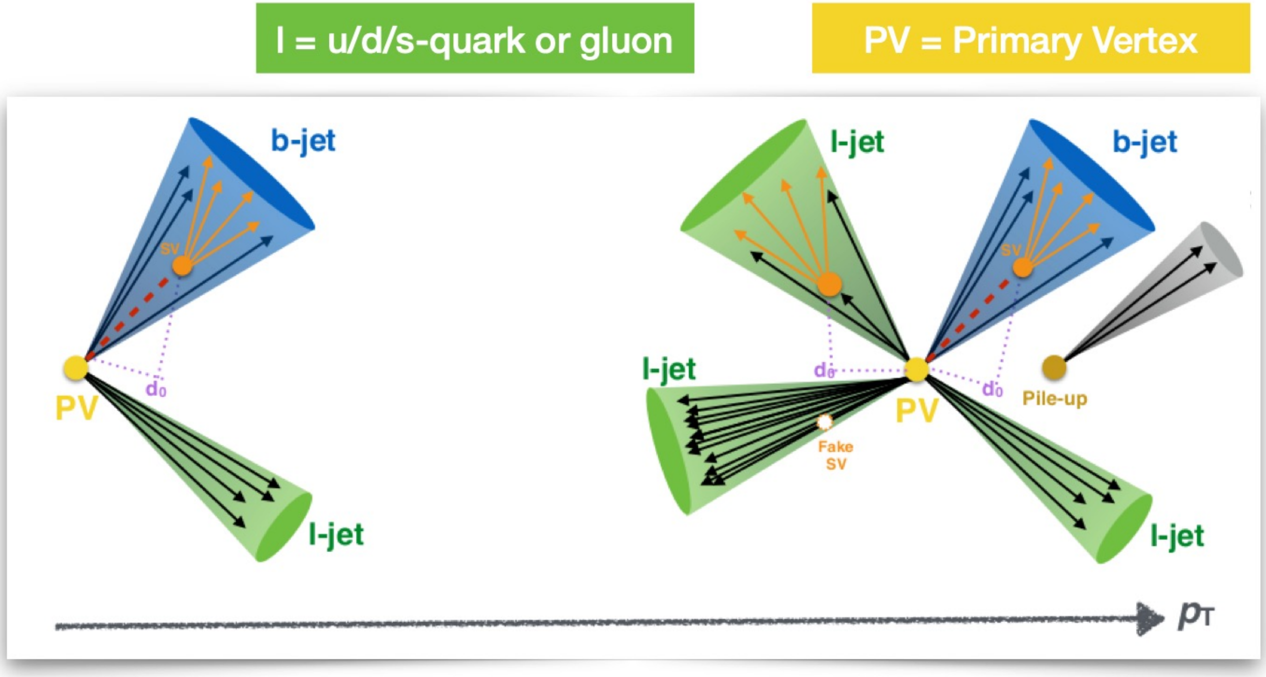
ML4Jets, LPNHE, Paris

05.11.2024

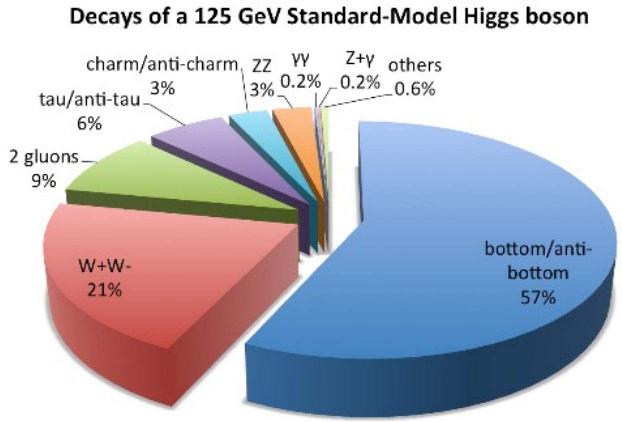
Introduction

FTAG algorithms (aka "tagger") aim to identify the jet flavour using the reconstructed jets and track properties for B-/C-hadron identification.

The relatively long lifetime of B-hadrons (~1.5 ps) can allow for significant **displacement** before decay



[S. Van Stroud]

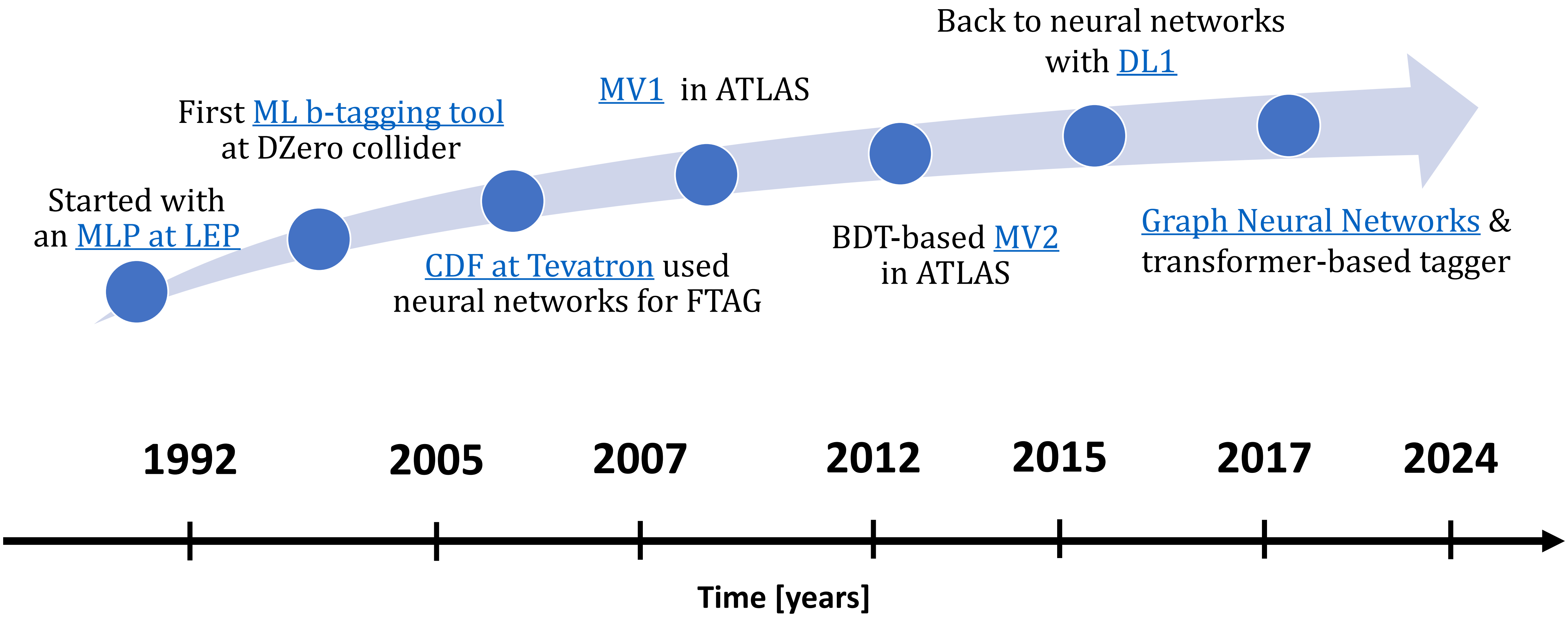


At hadron colliders, they are applied to a wide range of studies such as $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$, and di-Higgs analysis.

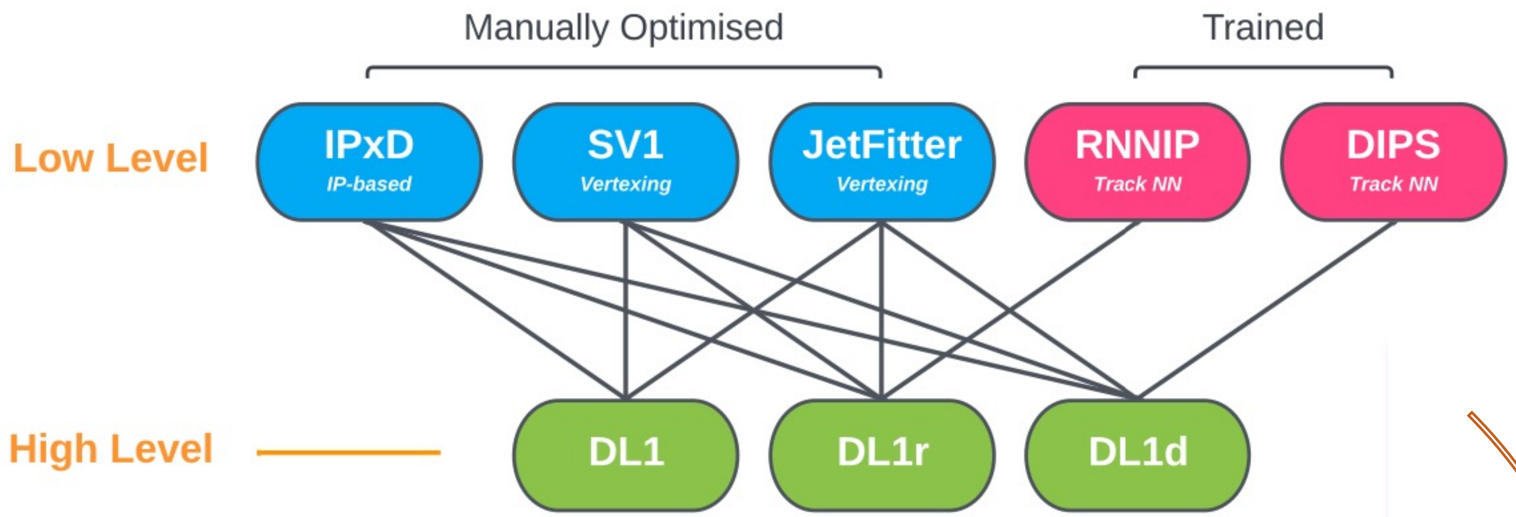
The picture gets complicated at high p_T!

The history of Flavour Tagging

Since MV1, ATLAS FTAG group continually advances, making use of even more sophisticated methods and architectures.



Deep Learning Network approach



Jet and track inputs are fed to low-level taggers :

➢ Use physics knowledge to construct expert variables: IPxD, SV1, JetFitter
➢ Track-based ML models: RNNIP, DIPS

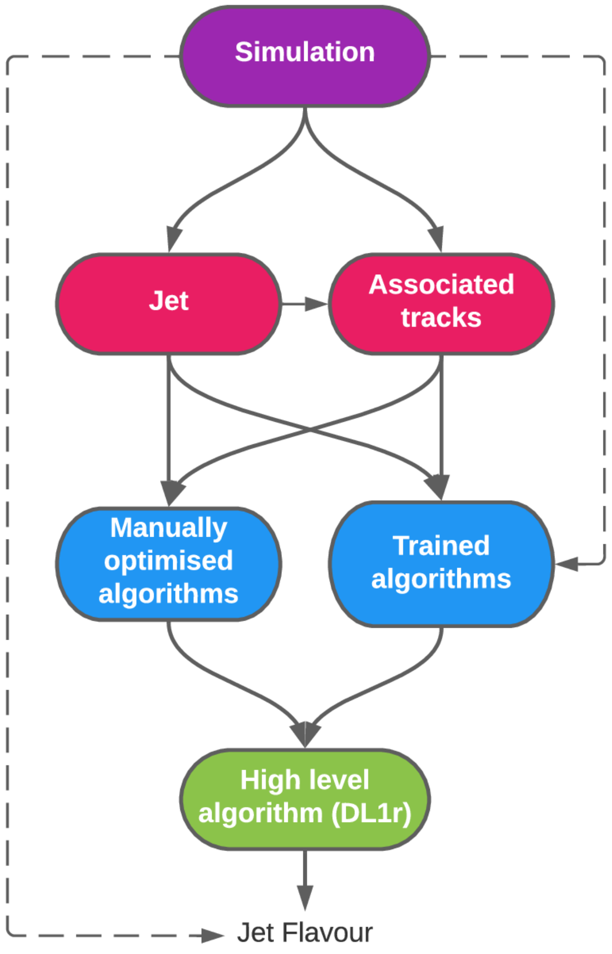
The outputs are fed into high-level taggers, which are BDTs (MV2) or NNs (DL1).

➢ Outputs: probabilities for each flavour class p_b, p_c, p_l

Main challenges

- Complexity in handling reconstructed tracks.
- Dependence on low-level taggers.
- Tuning for different use cases requires many single steps.

A new approach: GN1 and GN2



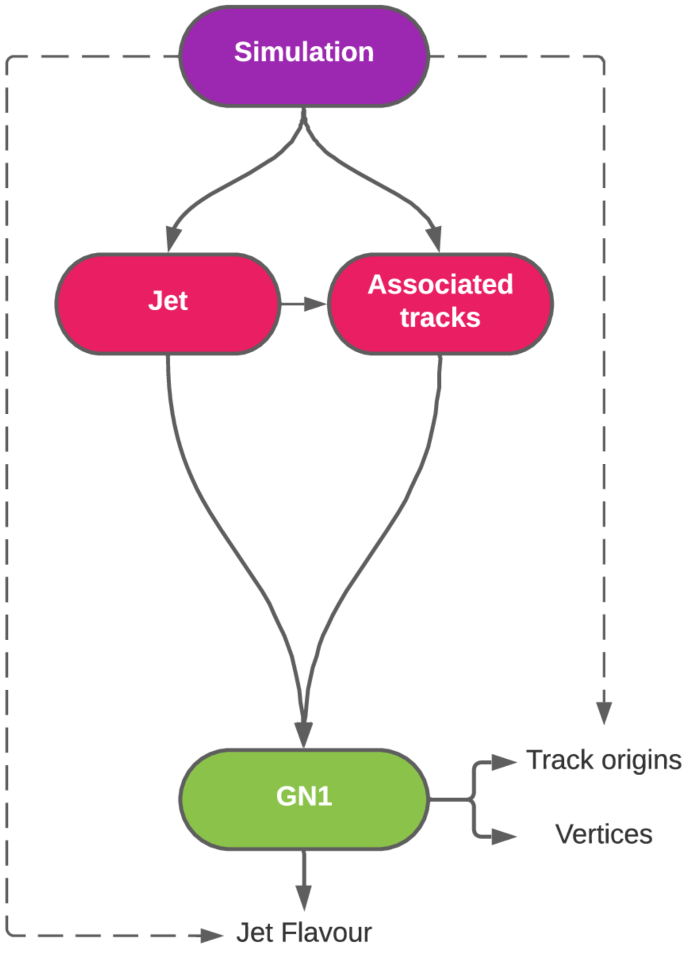
All-in-one GNN-based

GN1 (2022) is an all-in-one GNN-based, inspired by J. Shlomi's Work

GN2 is an upgraded version of GN1: an all-in-one transformer network with significant performance improvement.

GN2 is based on GN1 architecture with

- Optimised training
- Updated architecture
- Increased training statistics



[FTAG-2023-01]

GN2 improvement

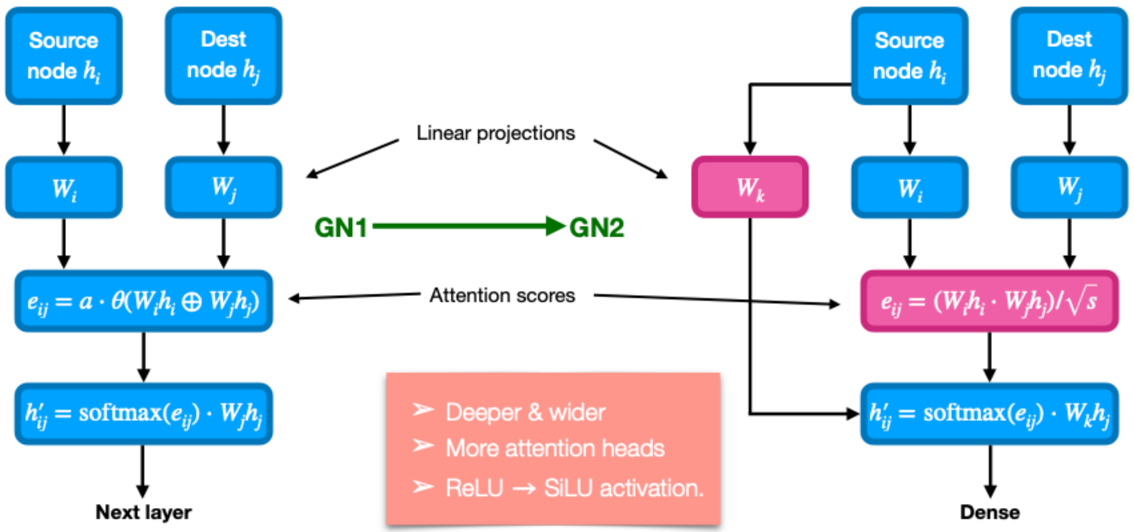
Optimised training

- Learning Rate (LR) Optimisation
  - ➢ Adaptive LR from Adam optimiser not sufficient
  - ➢ Now using a **one-cycle LR scheduler**
- Added LayerNorm and Dropout
  - ➢ Stabilises the training
  - ➢ Allows for expansion to larger model sizes
- New framework: more efficient training
  - ➢ Enables higher statistic training samples
  - ➢ 30M →192M training jets

[S. Van Stroud]

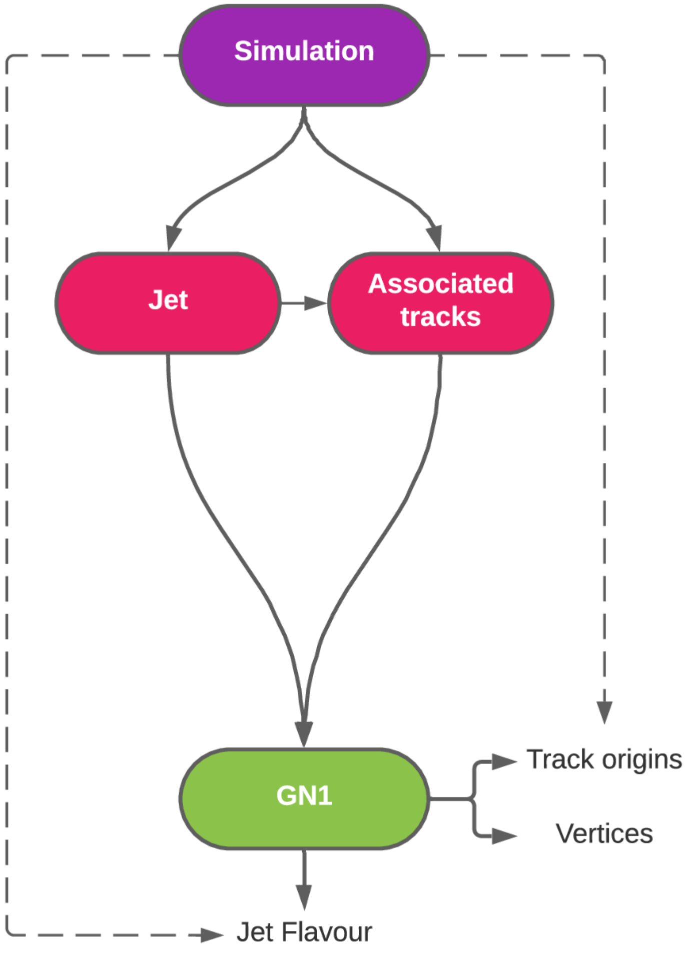
Updated Attention Mechanism

GN2 algorithm is based on the GN1 R&D version of the algorithm **replacing the Graph Attention Network used by GN1 with a Transformer architecture.**



[2105.14491] [1706.03762]

GN2 task



- **Jet flavour prediction**

Classifies jet for flavour, the outcome are p_b, p_c, p_l and p_τ.

From the NN output, it is possible to obtain the b-tag discriminant:

$$D_b = \log \frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_l}$$

Auxiliary tasks:

- **Track origin predicition**

It classifies the track originating from 7 different processes.

- **Vertex prediction**

It predicts if track pair comes from the same vertex.

The auxiliary task enhance the interpretability!

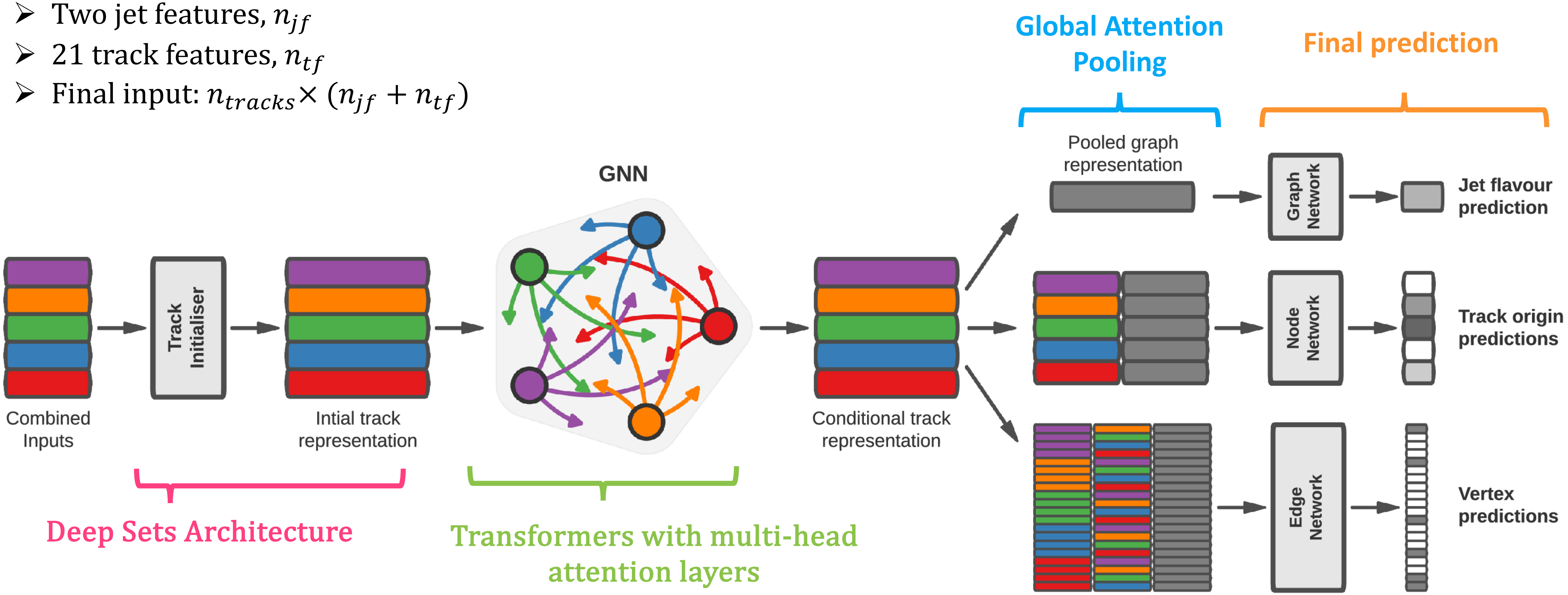
Tasks are trained simultaneously with $\mathcal{L}_{tot} = \mathcal{L}_{jet} + \alpha\mathcal{L}_{trk} + \beta\mathcal{L}_{vtx}$ by optimising the NN weights with gradient descent approach $\nabla\mathcal{L}_{tot}$.

Details in backup

GN2 architecture

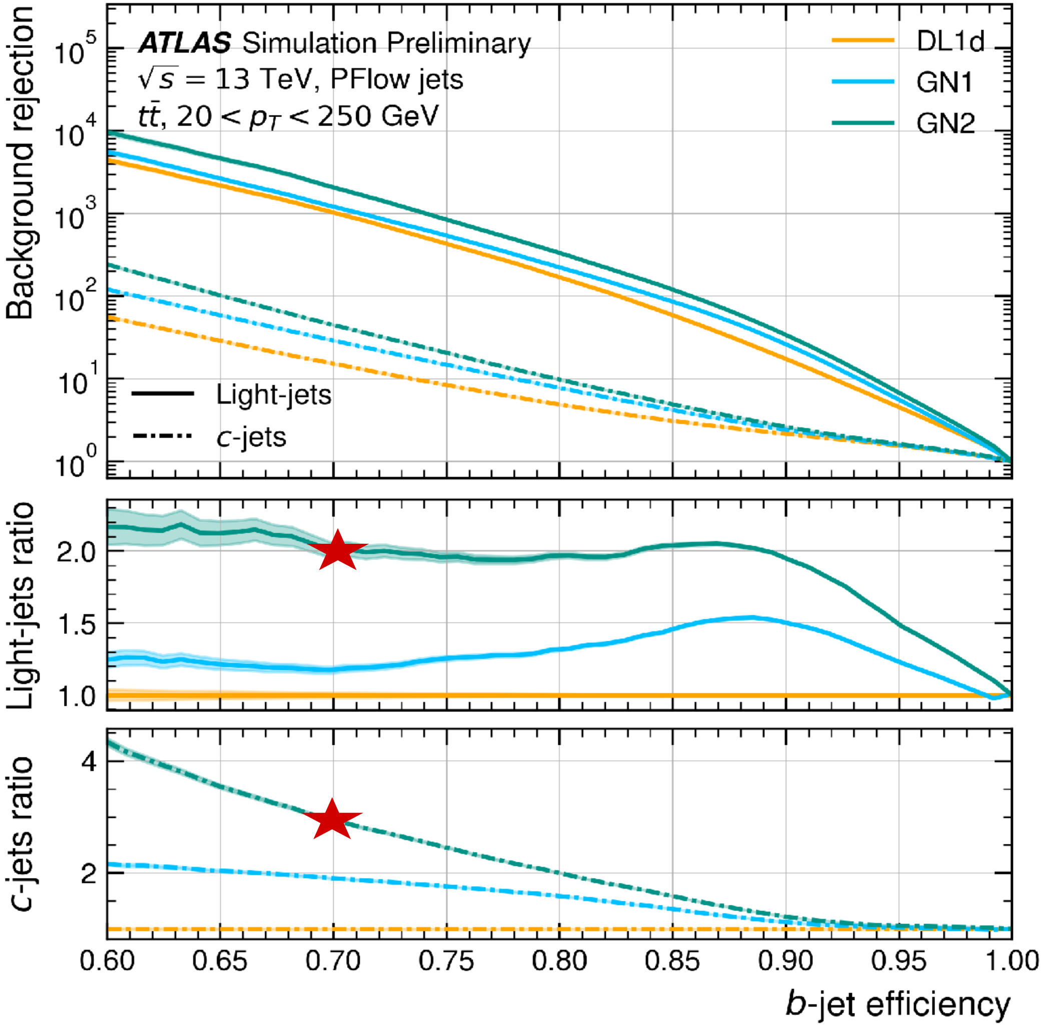
Input:

- Two jet features, n_{jf}
- 21 track features, n_{tf}
- Final input: $n_{tracks} \times (n_{jf} + n_{tf})$



Details in backup

GN2 b-tagging performance at low-p_T



The c-jet and light-jet rejections as a function of the b-jet tagging efficiency for jets with $20 < p_T < 250$ GeV in $t\bar{t}$ simulated sample.

At 70% b-tagging efficiency, we have an improvement with respect to:

- DL1d → ×2 for light rejection and ×2.8 for c-jet rejection
- GN1 → ×2 for light rejection and ×1.5 for c-jet rejection

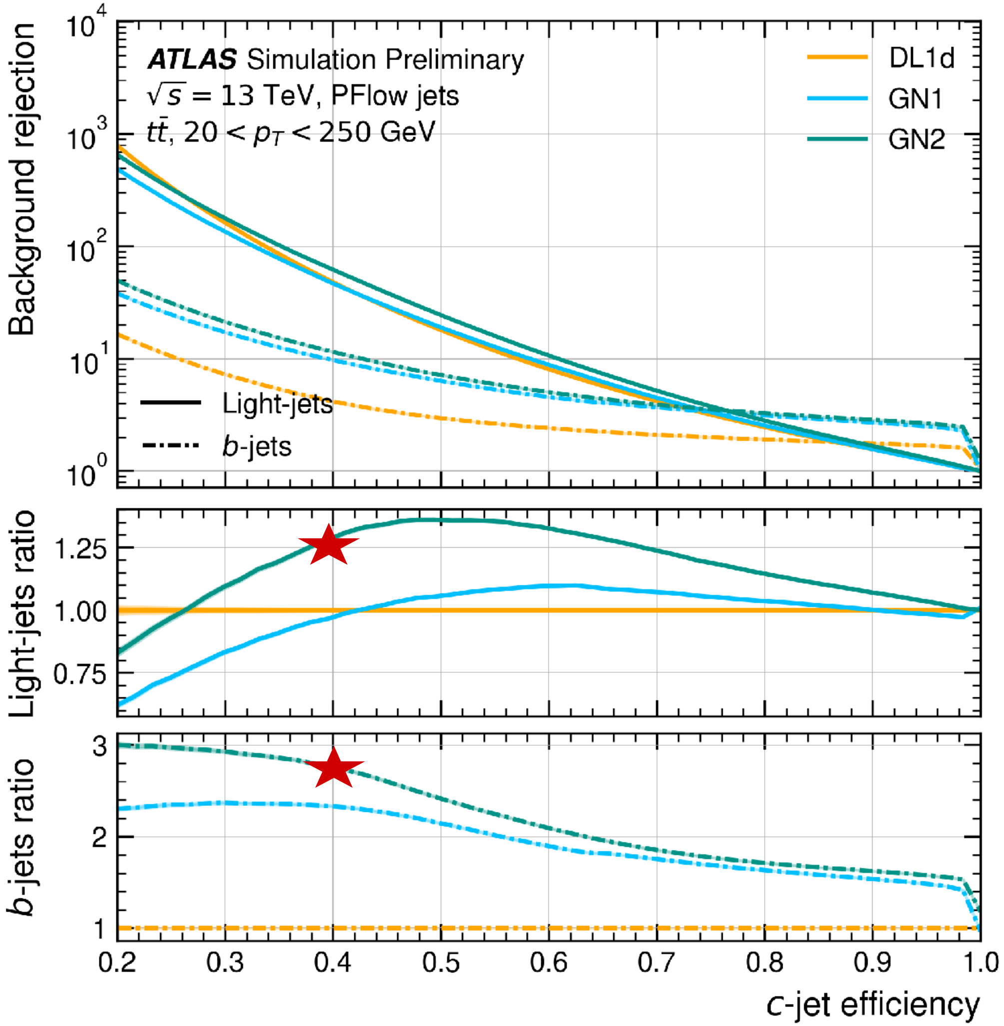
Model	f_c
DL1d	0.018
GN1	0.05
GN2	0.1

B-jet discriminant:

$$D_b = \log \frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_l}$$

[FTAG-2023-01]

GN2 c-tagging performance at low-p_T



[FTAG-2023-01]

The b-jet and light-jet rejections as a function of the c-jet tagging efficiency for jets with $20 < p_T < 250$ GeV in $t\bar{t}$ simulated sample.

At 40% c-tagging efficiency, we have an improvement with respect to:

- DL1d → ×1.3 for light rejection and ×2.7 for b-jet rejection
- GN1 → ×1.35 for light rejection and ×1.3 for b-jet rejection

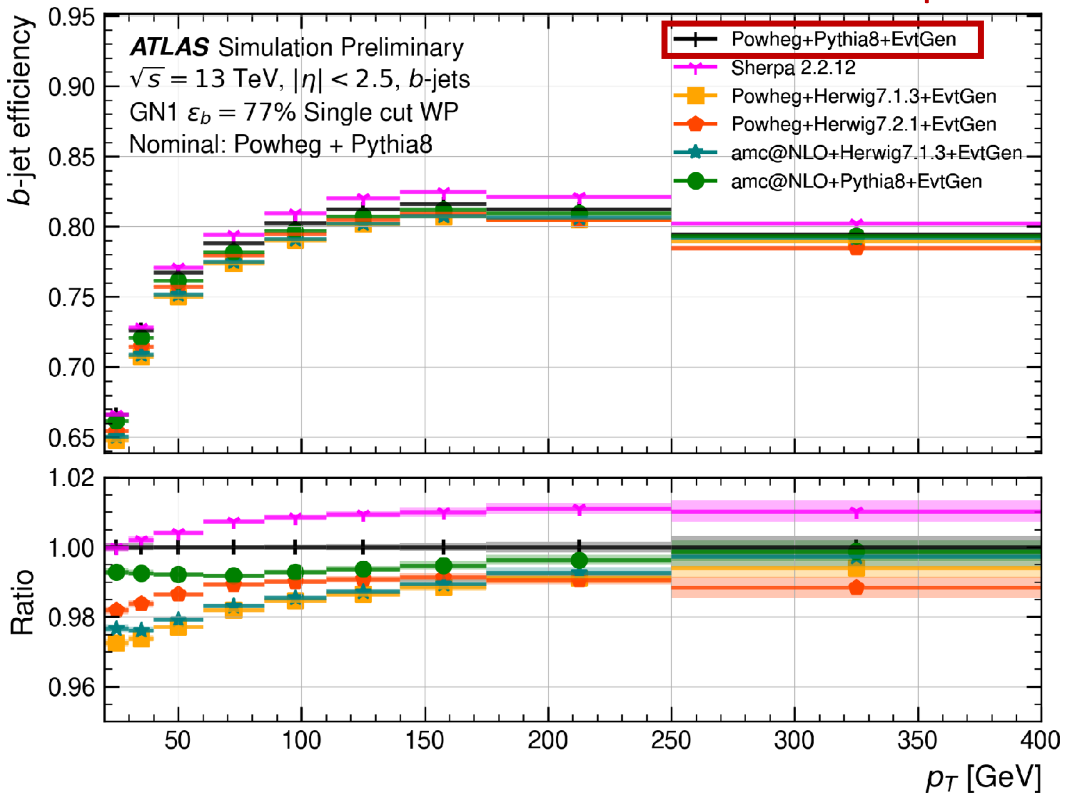
Model	f_b
DL1d	0.04
GN1	0.2
GN2	0.2

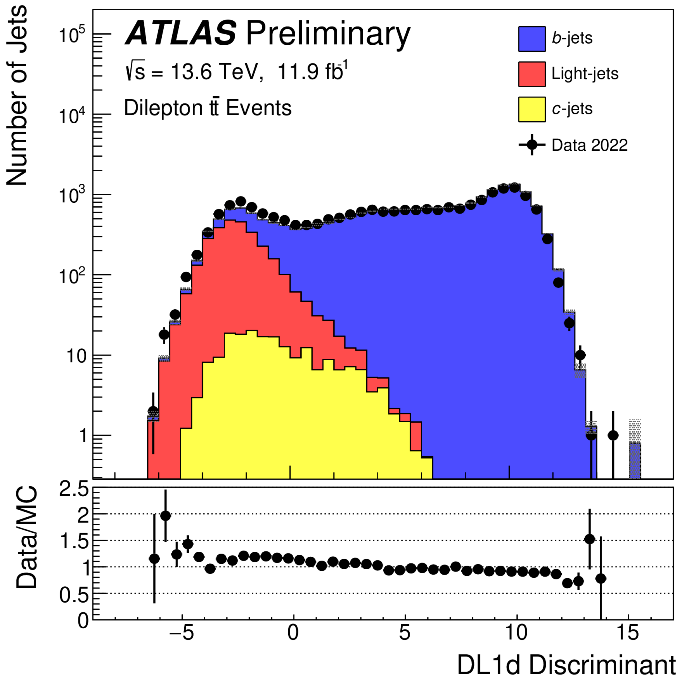
C-jet discriminant:

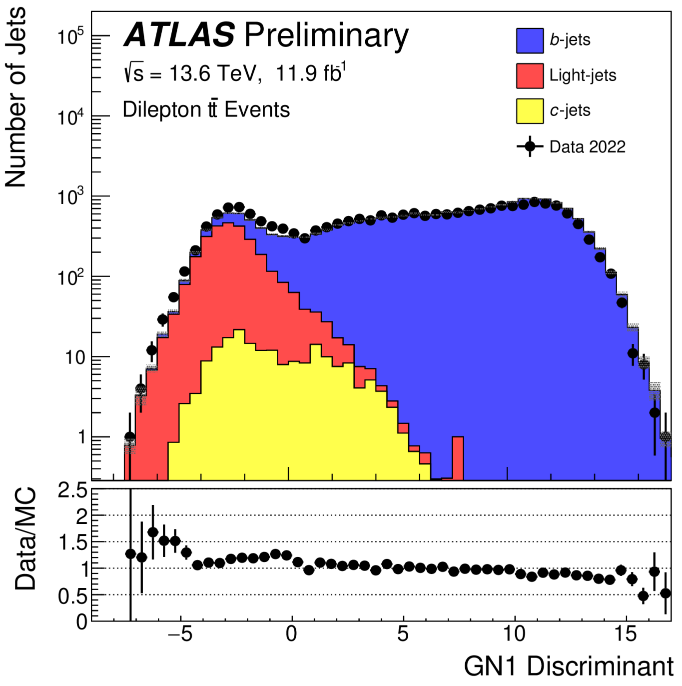
$$D_c = \log \frac{p_c}{f_b p_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_l}$$

Generator dependence and modelling

Nominal sample





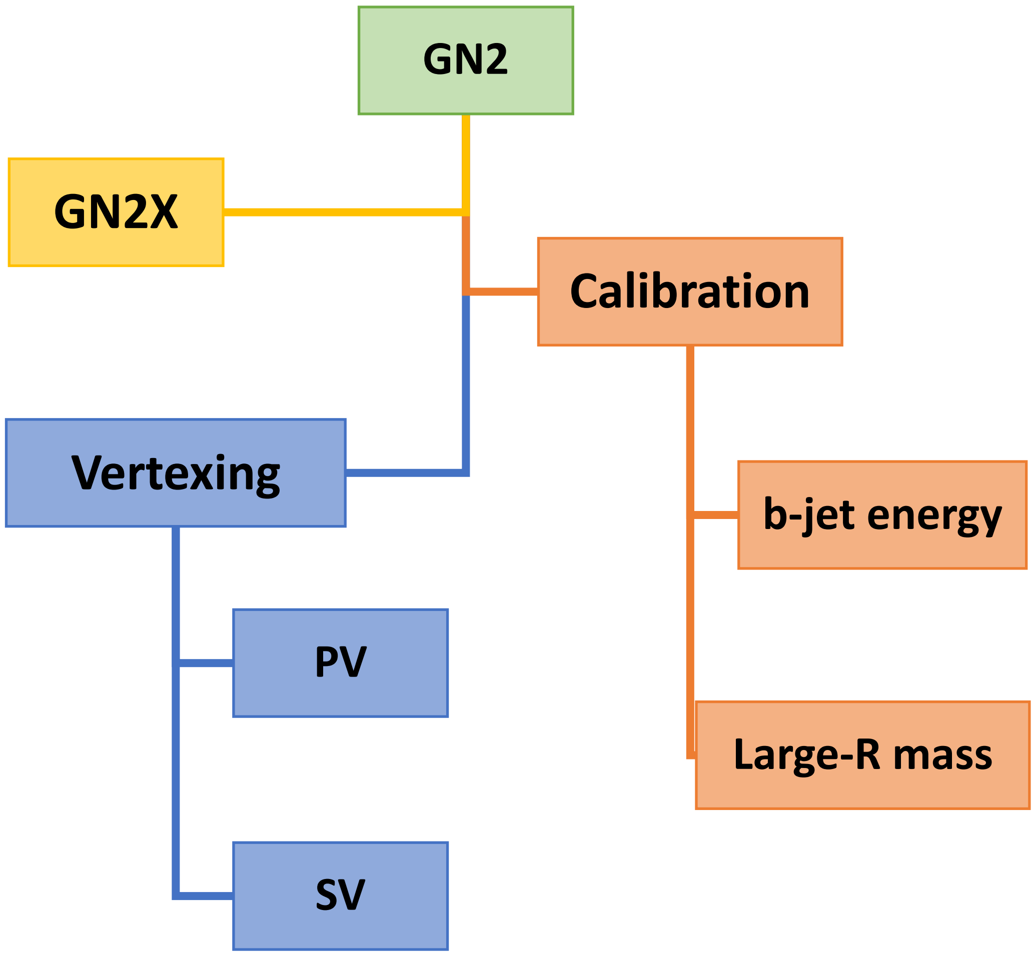


The efficiency of the tagger depends on the MC generator used. The differences between the generators are quantified using **Monte Carlo to Monte Carlo Scale Factors (MC-MC SF)**, defined as the ratio between the efficiency of the nominal and alternative sample.

The disagreement between the simulated samples and data is measured in the **calibration analyses** and corrections applied for use in analyses. The modelling is found to be consistent between the two algorithms.

[FTAG-2023-01]

GN2 application



The ATLAS FTAG developed a **software ecosystem** for GN2, enhancing the versatility and the possible applications of this network:

- Input dataset production, [Training-Dataset-Dumper]
- Network training, [SALT]
- Processing and plotting results, [PUMA-HEP]

Thanks to **clear documentation** and an **easy access** to FTAG tools:

- Successful application in analyses and HL-LHC forecast
- Synergy with other group
- Increased collaboration

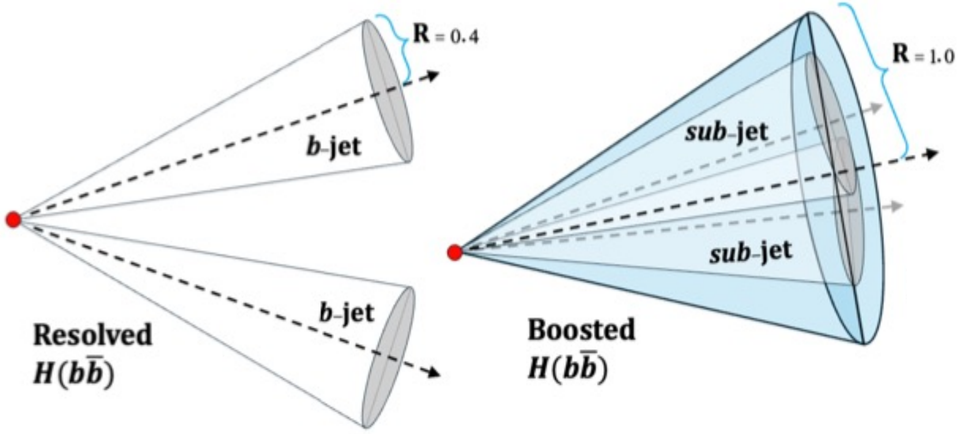
Example of that is Boosted Xbb tagging and its application!

Boosted Higgs Tagging

[ATL-PHYS-PUB-2023-021]

- At higher momenta, **the jets' signature of a boosted decay can no longer be distinguished as separate objects**. For example, the decay products of Higgs bosons with a $p_T \gtrsim$ 250 GeV will be **collimated**.

- A large-R jet clustering reconstructs boosted $H(b\bar{b})$ and $H(c\bar{c})$. Like *b*–jet tagging, the main background is QCD multi–jet production. However, boosted Top quarks are a further background that could "fake" a boosted Higgs.



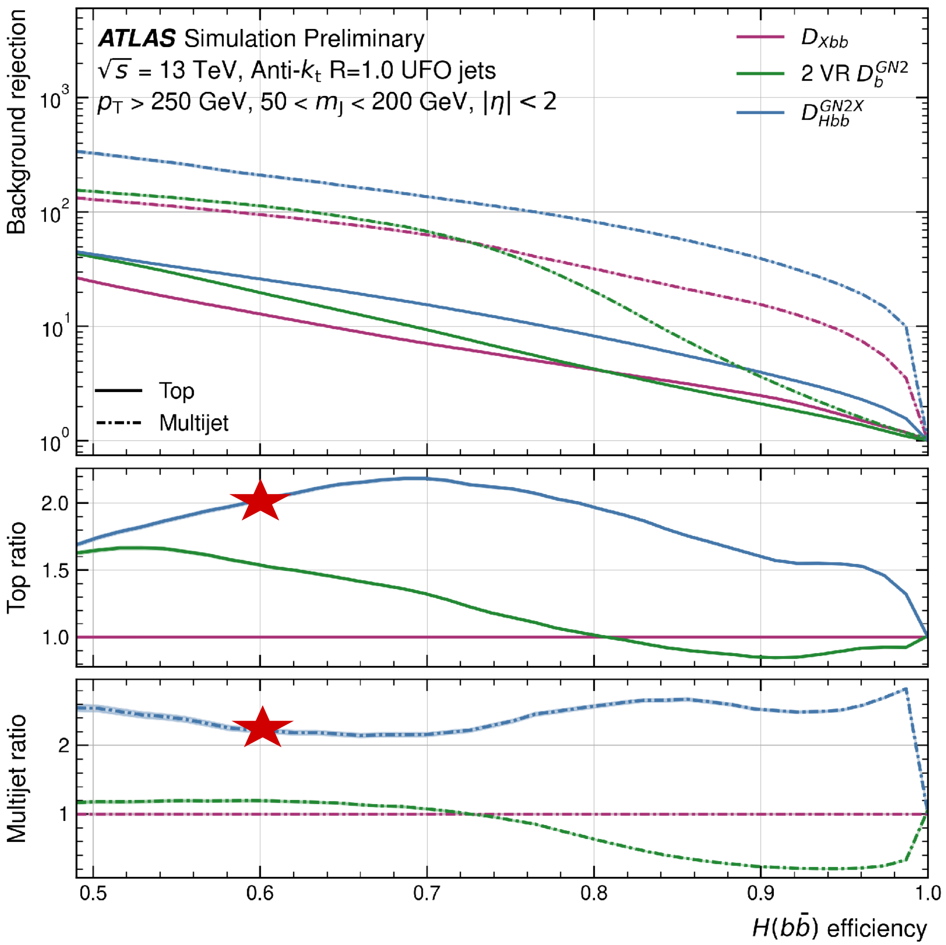
[W.Leinonen]

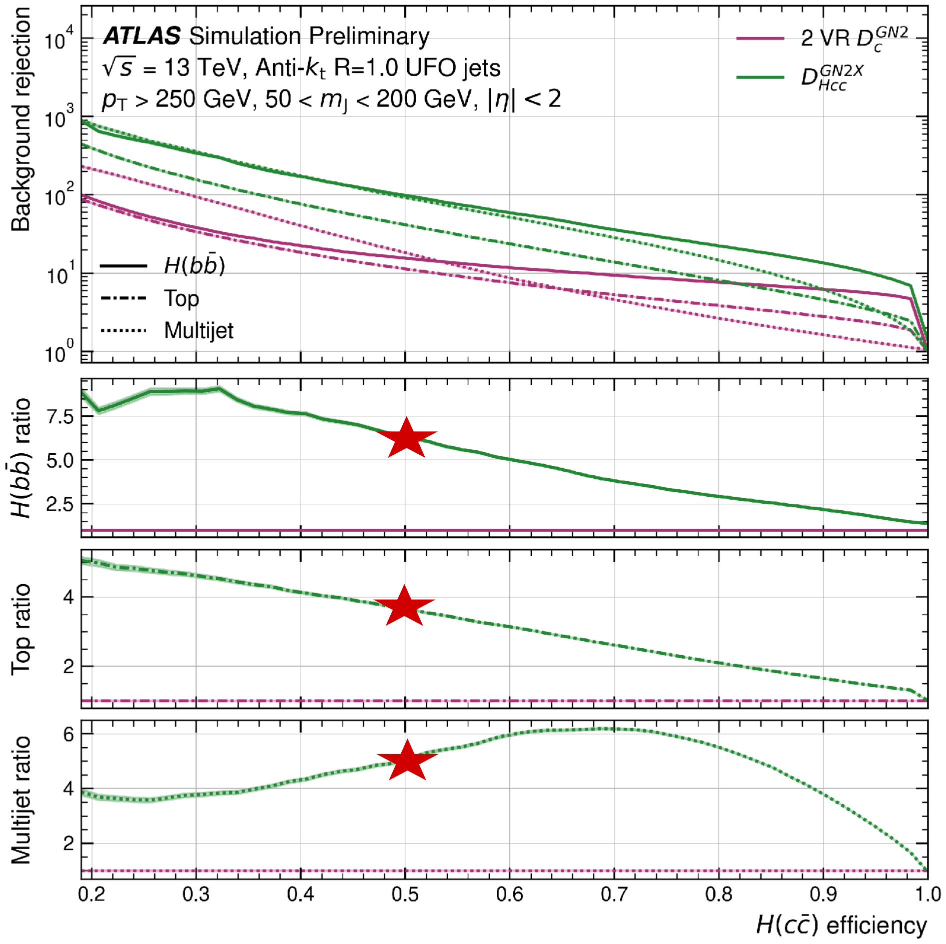
- **A GN2 derivation, GN2X, has been developed to identify these collimated jets in Higgs boson decay.**

- GN2X is a transformer-based Xbb tagger that replaces the previous subjet-based model used within ATLAS. Trained to discriminate between boosted $H \rightarrow b\bar{b}$, $H \rightarrow c\bar{c}$, hadronic top, and QCD jets.

Boosted Higgs Tagging performance

[ATL-PHYS-PUB-2023-021]





At 60% signal efficiency, GN2X to the previous Xbb tagger more than double the top and QCD rejection.

At 40% signal efficiency, GN2X to the previous Xcc tagger ×3 top, ×5 QCD multi-jet and ×6 $H(b\bar{b})$ rejection.

Conclusion

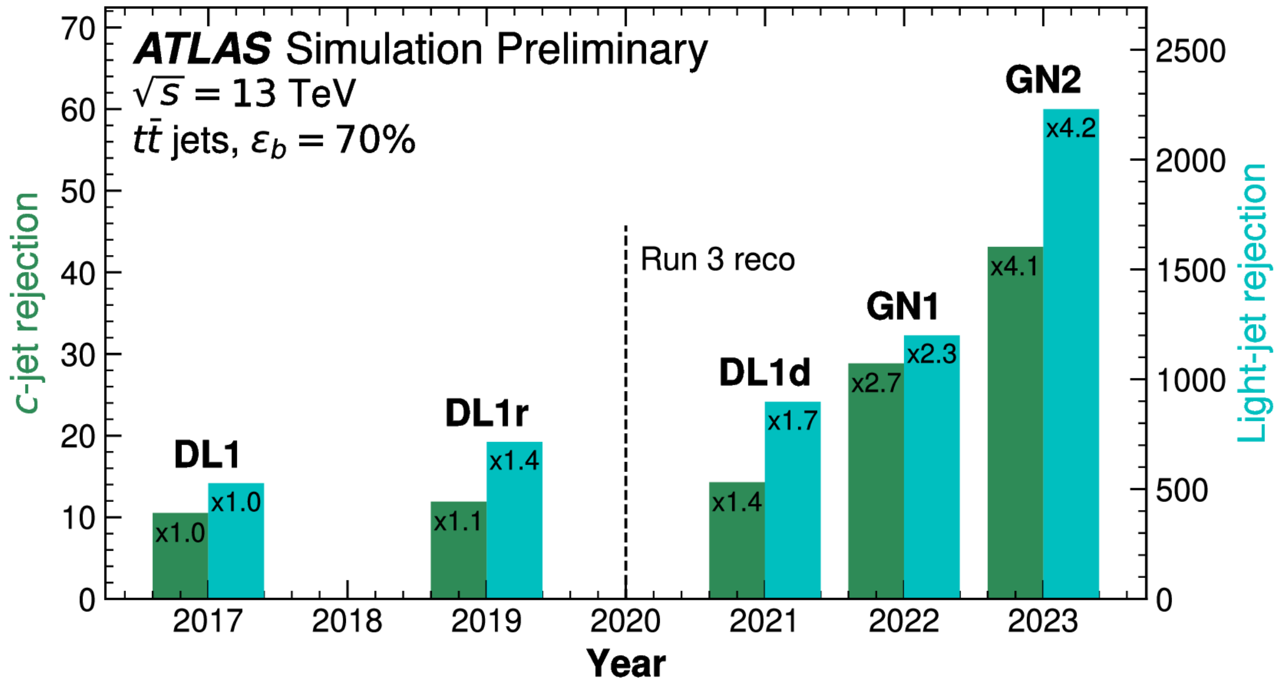
GN2 is the strong successor of GN1:

- Optimised training and updated architecture
- Improvement in rejection over GN1 for b- and c-jets
- Software ecosystem, variability , and easy access to ATLAS FTAG tools
- Strong benefit to the ATLAS physics program
- The new boosted Xbb tagger, GN2X, based on GN2, is being studied.

What's next?

New developments are ongoing in the FTAG group, stay tuned!

Thank you for your attention!



[FTAG-2023-01]

Backup

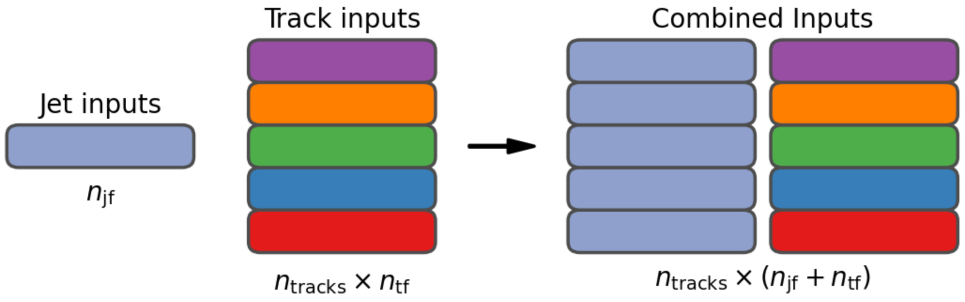
From GN1 to GN2

Type	Name	GN1	GN2
Hyperparameter	Trainable parameters	0.8M	1.5M
Hyperparameter	Learning rate	$1e{-}3$	OneCycle LRS (max LR $4e{-}5$)
Hyperparameter	GNN Layers	3	6
Hyperparameter	Attention Heads	2	8
Hyperparameter	Embed. dim	128	192
Architectural	Attention type	GATv2	ScaledDotProduct
Architectural	Dense update	No	Yes (dim 256)
Architectural	Separate value projection	No	Yes
Architectural	LayerNorm + Dropout	No	Yes
Inputs	Num. training jets	30M	192M

[FTAG-2023-01]

GN2 input

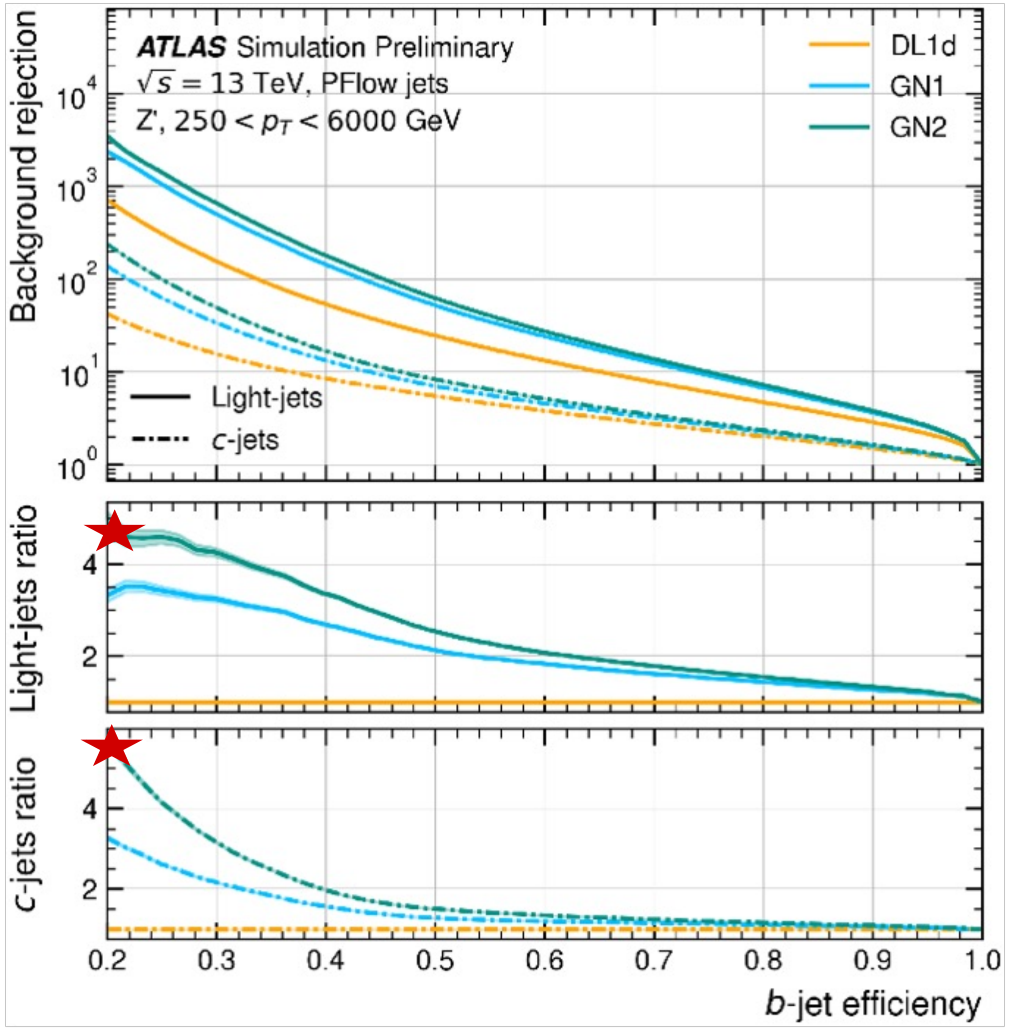
The inputs to GN1 are the two jet features ($n_{jf} = 2$), and an array of ntracks, where each track is described by 21 track features ($n_{tf} = 21$). The jet features are copied for each of the tracks, and the combined jet-track vectors of length 23 form the inputs of GN2.



GN2Lep: possible variant that uses information from semileptonic b-decay leptons via additional track variable indicating if a track has been used in the lepton reconstruction.

Jet Input	**Description**
p_{T}	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	**Description**
q/p	Track charge divided by momentum (measure of curvature)
$\mathrm{d}\eta$	Pseudorapidity of the track, relative to the jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin\theta$	Closest distance from the track to the PV in the transverse 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
$s(z_0)$	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
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used in the reconstruction of an electron or muon (only for GN1 Lep)

GN2 b-tagging performance at high-p_T



[FTAG-2023-01]

The c-jet and light-jet rejections as a function of the b-jet tagging efficiency for jets with $250 < p_T < 6000$ GeV in Z' simulated sample.

At 20% b-tagging efficiency, we have an improvement with respect to:

- DL1d $\rightarrow \times 4.8$ for light rejection and $\times 5.5$ for c-jet rejection
- GN1 $\rightarrow \times 1.2$ for light rejection and $\times 1.75$ for c-jet rejection

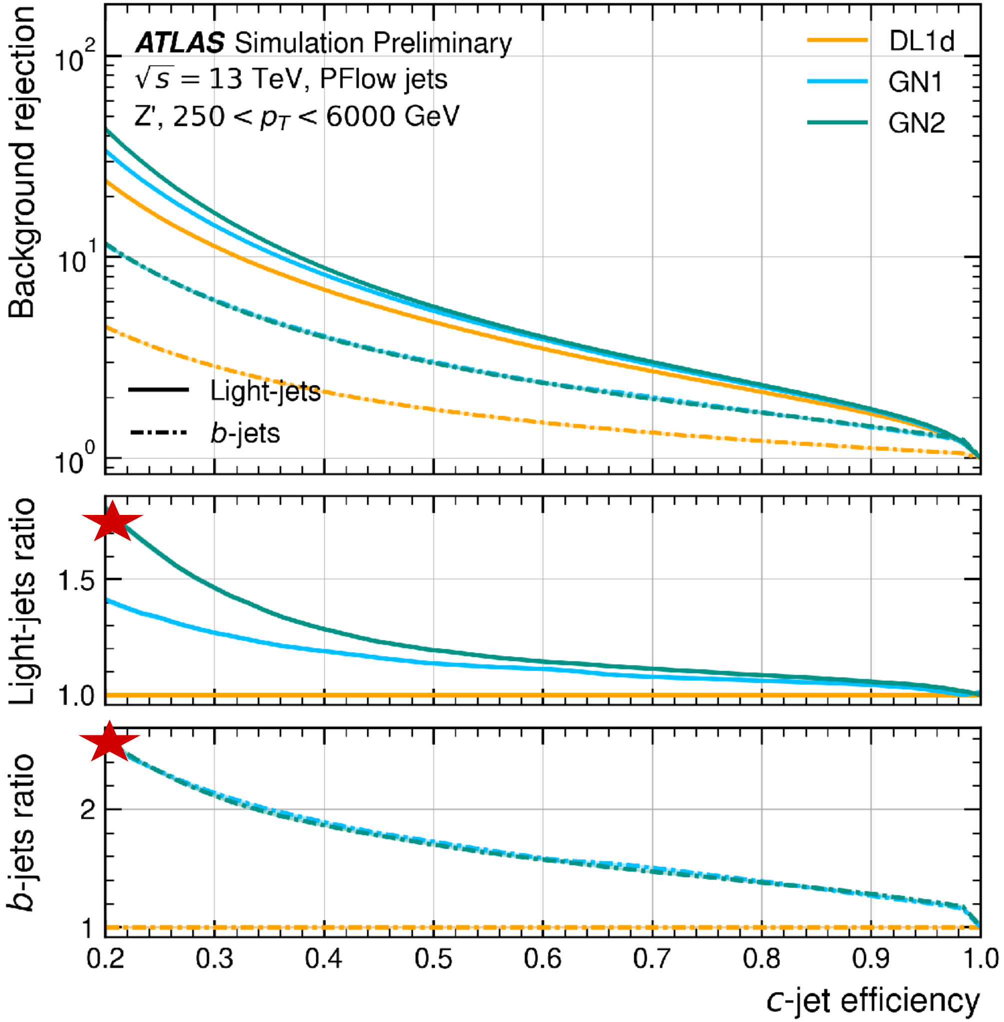
Model	f_c
DL1d	0.018
GN1	0.05
GN2	0.1

B-jet discriminant:

$$D_b = \log \frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_l}$$

GN2 c-tagging performance at high-p_T

ATLAS EXPERIMENT



The b-jet and light-jet rejections as a function of the c-jet tagging efficiency for jets with $250 < p_T < 6000$ GeV in Z' simulated sample.

At 20% c-tagging efficiency, we have an improvement with respect to:

- DL1d $\rightarrow \times 1.8$ for light rejection and $\times 2.6$ for b-jet rejection
- GN1 $\rightarrow \times 1.25$ for light rejection and $\times 1.75$ for b-jet rejection

Model	f_b
DL1d	0.04
GN1	0.2
GN2	0.2

C-jet discriminant:

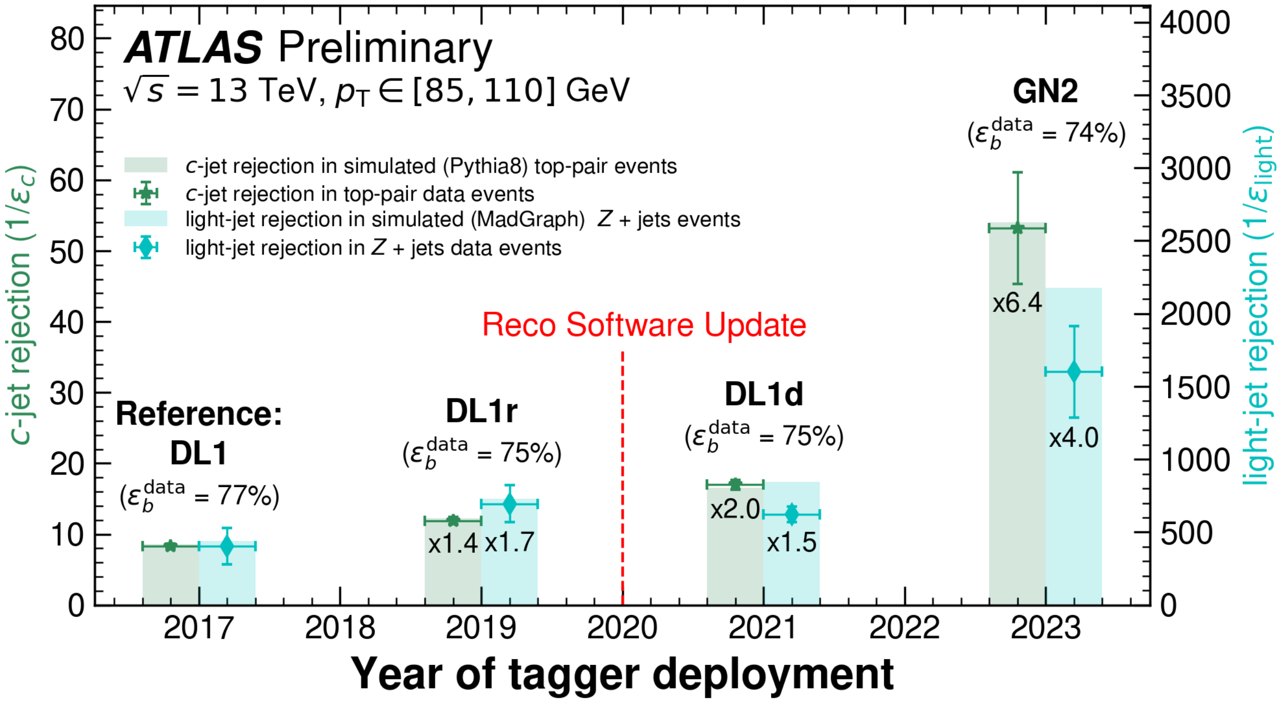
$$D_c = \log \frac{p_c}{f_b p_b + f_\tau p_\tau + (1 - f_b - f_\tau) p_l}$$

[FTAG-2023-01]

Tagging improvement on data

Evolution of the improvement of flavour tagging algorithms. The filled bar represents the expected rejection factors computed on a $t\bar{t}$ MC sample corresponding to a 70%-jet expected efficiency. Data points show the rejections measured by the calibration analyses with their uncertainties. Measured-jet efficiencies are also outlined in the plot.

[FTAG-2023-07]



Latest calibration public results:

- B- and light- jets: [FTAG-2023-04]
- C-jets: [FTAG-2023-05]
- Xbb: [FTAG-2023-06]

Auxiliary task: network size

Network	**Hidden layers**	**Output size**
Node classification network	128, 64, 32	7
Edge classification network	128, 64, 32	1
Graph classification network	128, 64, 32, 16	3

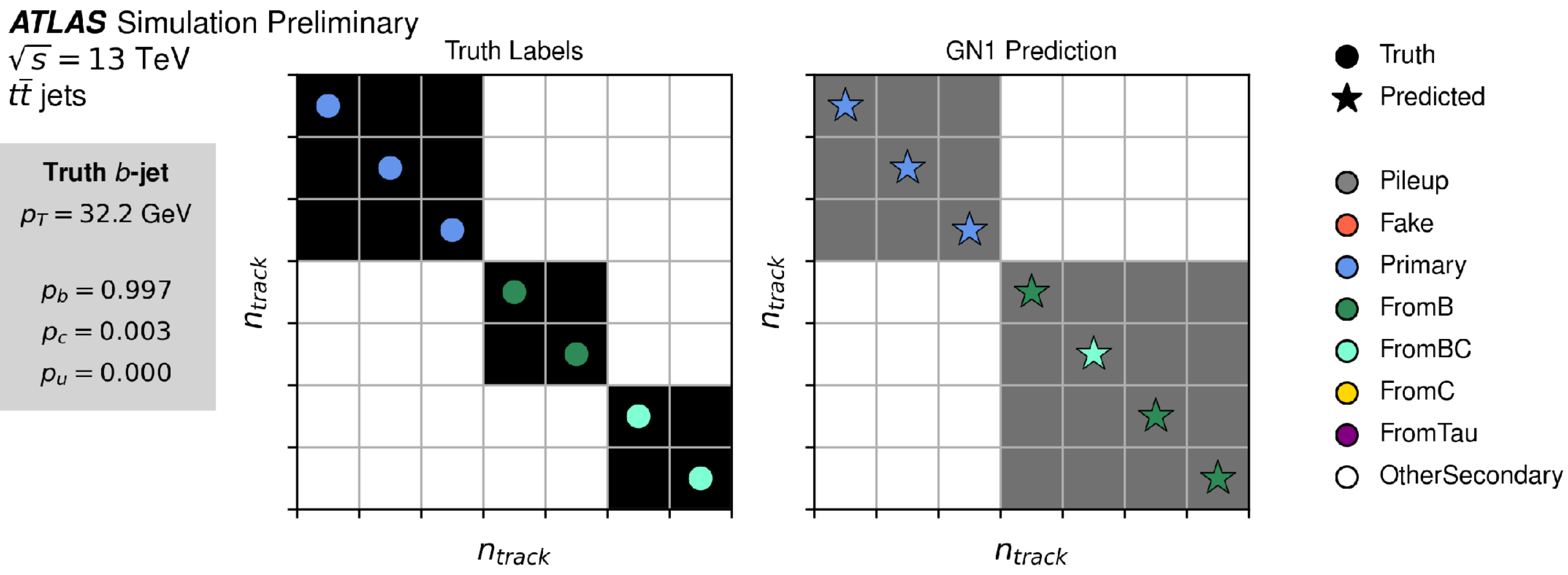
[FTAG-2023-01]

Auxiliary task: Track origin prediction

Truth Origin	**Description**
Pileup	From a pp collision other than the primary interaction
Fake	Created from the hits of multiple particles
Primary	Does not originate from any secondary decay
fromB	From the decay of a b-hadron
fromBC	From a c-hadron decay, which itself is from the decay of a b-hadron
fromC	From the decay of a c-hadron
OtherSecondary	From other secondary interactions and decays

Auxiliary task: interpretability

A schematic view of the true (left) and predicted (right) track origins and vertices in a -jet from the $t\bar{t}$ sample. The filled black (grey) boxes indicate tracks that are grouped into truth (predicted) vertices.



[FTAG-2023-01] , New interpretability studies [FTAG-2024-01]

GN2X input

Input features to the GN2X model.

GN2X + Flow: UFO constituents, including charged and neutral calorimeter information.

GN2X + Subjets: kinematic + b-tagging info VR subjets, where the subjets are tagged using the GN2 tagger.

Features are separated into jet inputs, track inputs, subjet inputs and UFO constituent (flow) inputs.

The subjet and flow inputs are only used in the GN2X + Subjet and GN2X + Flow models respectively.

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

GN2X generator and selection

Signal and background processes for training with corresponding event generator versions, tunes and PDF sets.

Jet type	Process	Event generator and tune	PDF set
$H(b\bar{b})$	$q\bar{q} \to ZH, Z \to \mu^+\mu^-$	PYTHIA 8.306 [17] with A14 [18]	NNPDF2.3LO [19]
$H(c\bar{c})$	$q\bar{q} \to ZH, Z \to \mu^+\mu^-$	PYTHIA 8.306 with A14	NNPDF2.3LO
Top	$Z' \to t\bar{t}$	PYTHIA 8.235 with A14	NNPDF2.3LO
Multijet	Multijet	PYTHIA 8.235 with A14	NNPDF2.3LO

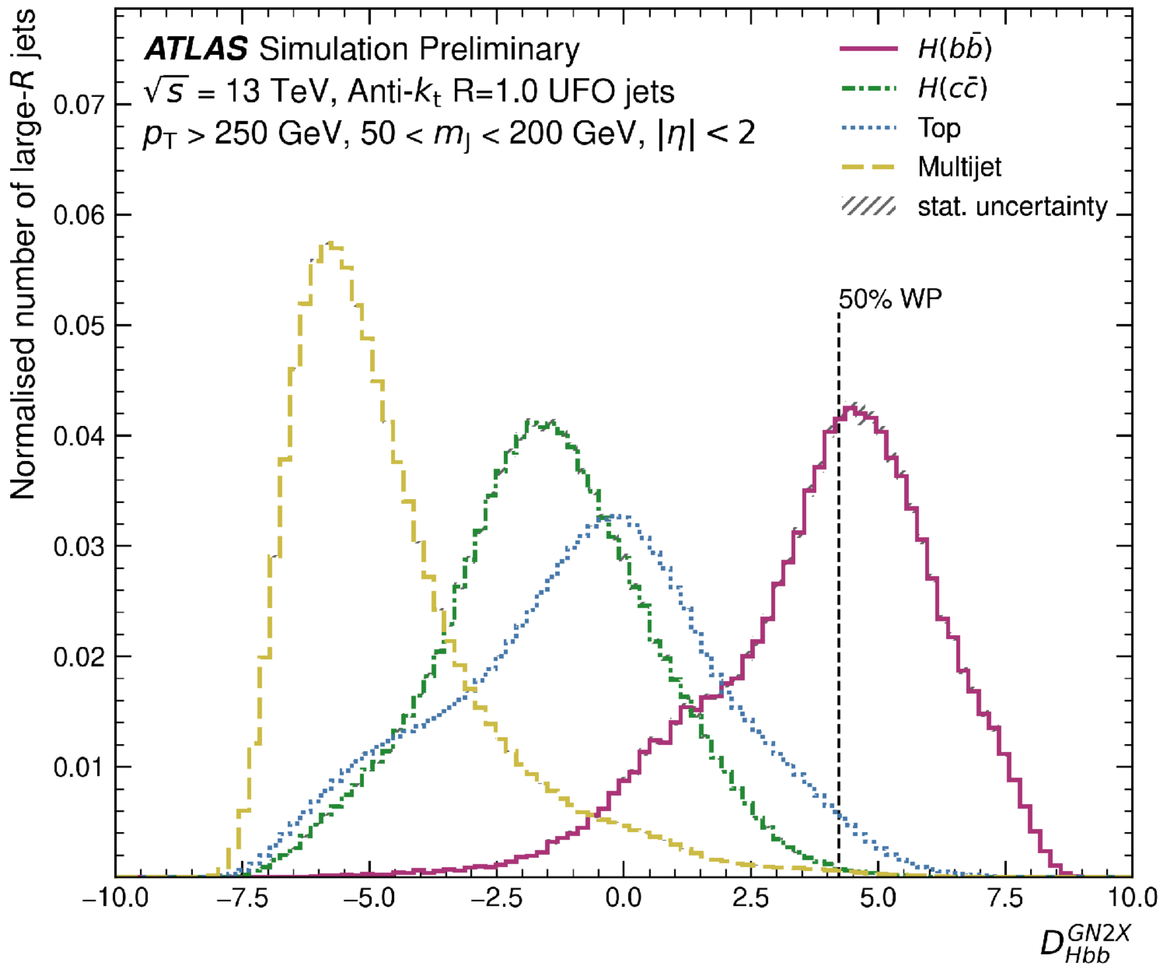
Signal and background processes for evaluation with corresponding event generator versions, tunes and PDF sets. ℓ=e,μ

Jet type	Process	Event generator and tune	PDF set
$H(b\bar{b})$	$q\bar{q}/gg \to ZH, Z \to \ell\ell/\nu\bar{\nu}/q\bar{q}$	POWHEG V2 + PYTHIA 8.212 [20] with AZNLO [21]	NNPDF3.0NLO
$H(c\bar{c})$	$q\bar{q}/gg \to ZH, Z \to \ell\ell/\nu\bar{\nu}/q\bar{q}$	POWHEG V2 + PYTHIA 8.212 with AZNLO	NNPDF3.0NLO
Top	$t\bar{t}$	POWHEG V2 + PYTHIA 8.230 with A14	NNPDF2.3LO
Multijet	Multijet	PYTHIA 8.235 with A14	NNPDF2.3LO

Track selection requirements, where d_0 is the transverse impact parameter (IP) of the track, z_0 is the longitudinal IP to the primary vertex, and θ is the track polar angle. Shared hits are hits used in the reconstruction of multiple tracks that have not been classified as split by the cluster-splitting neural networks.

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

GN2X WP selection on discriminant



Discriminant Hbb of GN2X

$$D_{Hbb}^{GN2X} = \log \frac{p_{Hbb}}{f_{Hcc}p_{Hcc} + f_{top}p_{top} + (1 - f_{Hcc} - f_{top})p_{QCD}}$$

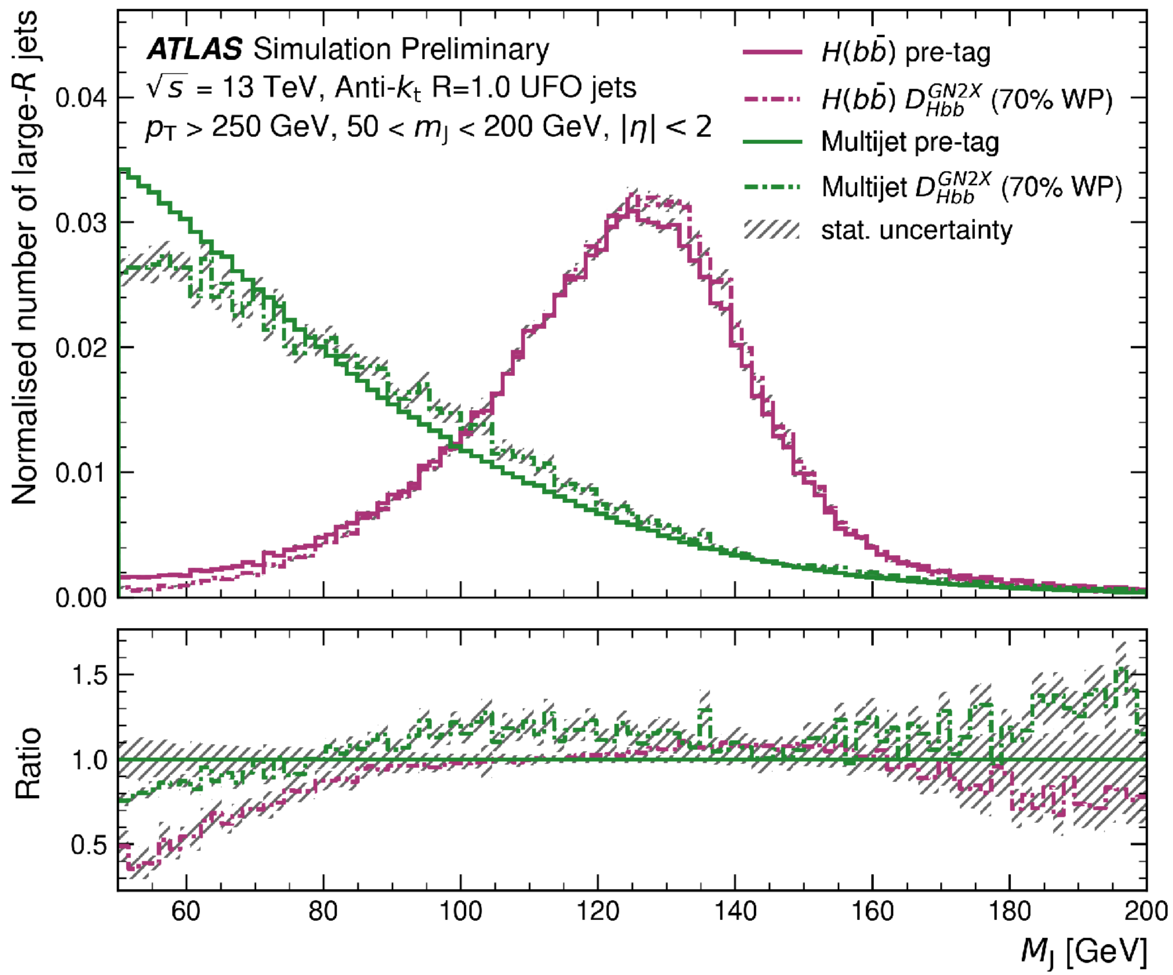
Where $f_{Hcc} = 0.02, f_{top} = 0.25$

Discriminant Hcc of GN2X

$$D_{Hcc}^{GN2X} = \log \frac{p_{Hcc}}{f_{Hbb}p_{Hbb} + f_{top}p_{top} + (1 - f_{Hcc} - f_{top})p_{QCD}}$$

Where $f_{Hbb} = 0.3, f_{top} = 0.25$

GN2X mass sculpting



Large-R jet mass distributions for $H(b\bar{b})$ and multijet samples, before and after applying a 70% $H(b\bar{b})$ efficiency D_{Hbb}^{GN2X} cut. The distribution is shown for the SM evaluation samples.

One of the major challenges that GN2X faces is the distribution-level mass sculpting effect on backgrounds such as QCD jets. GN2X is trained on mass decorrelated Higgs sample, in which the Higgs boson decay width is artificially enlarged (nominally, the Higgs width $\Gamma_{Higgs} \sim 4$ MeV to minimize correlations between jet mass and other features from being exploited by the network, and a kinematic resampling alters relative MC statistics in regions of phase-space to ensure similar kinematic distributions between all classes of jet $H(b\bar{b})$, $H(c\bar{c})$,Top, QCD).