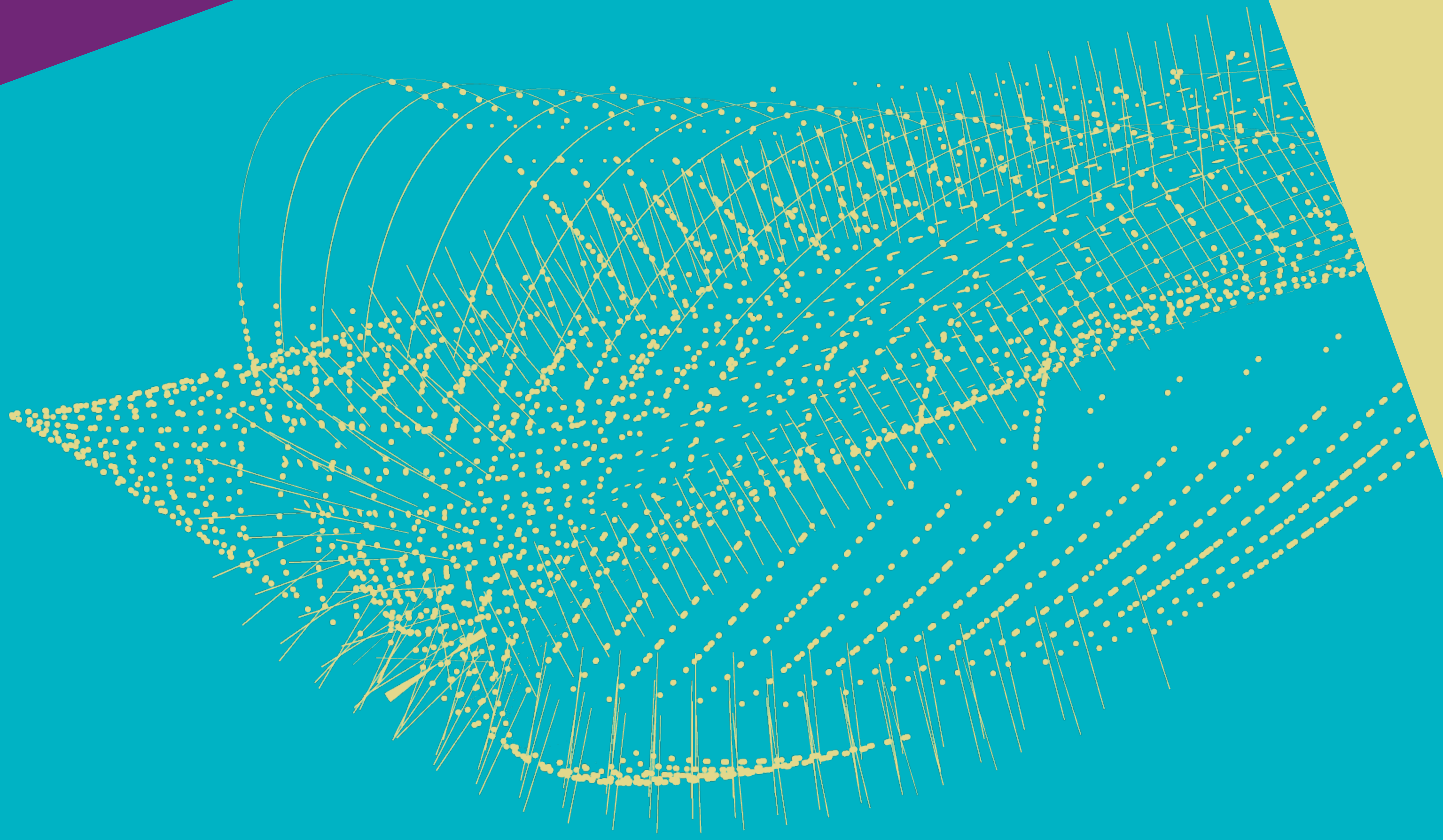


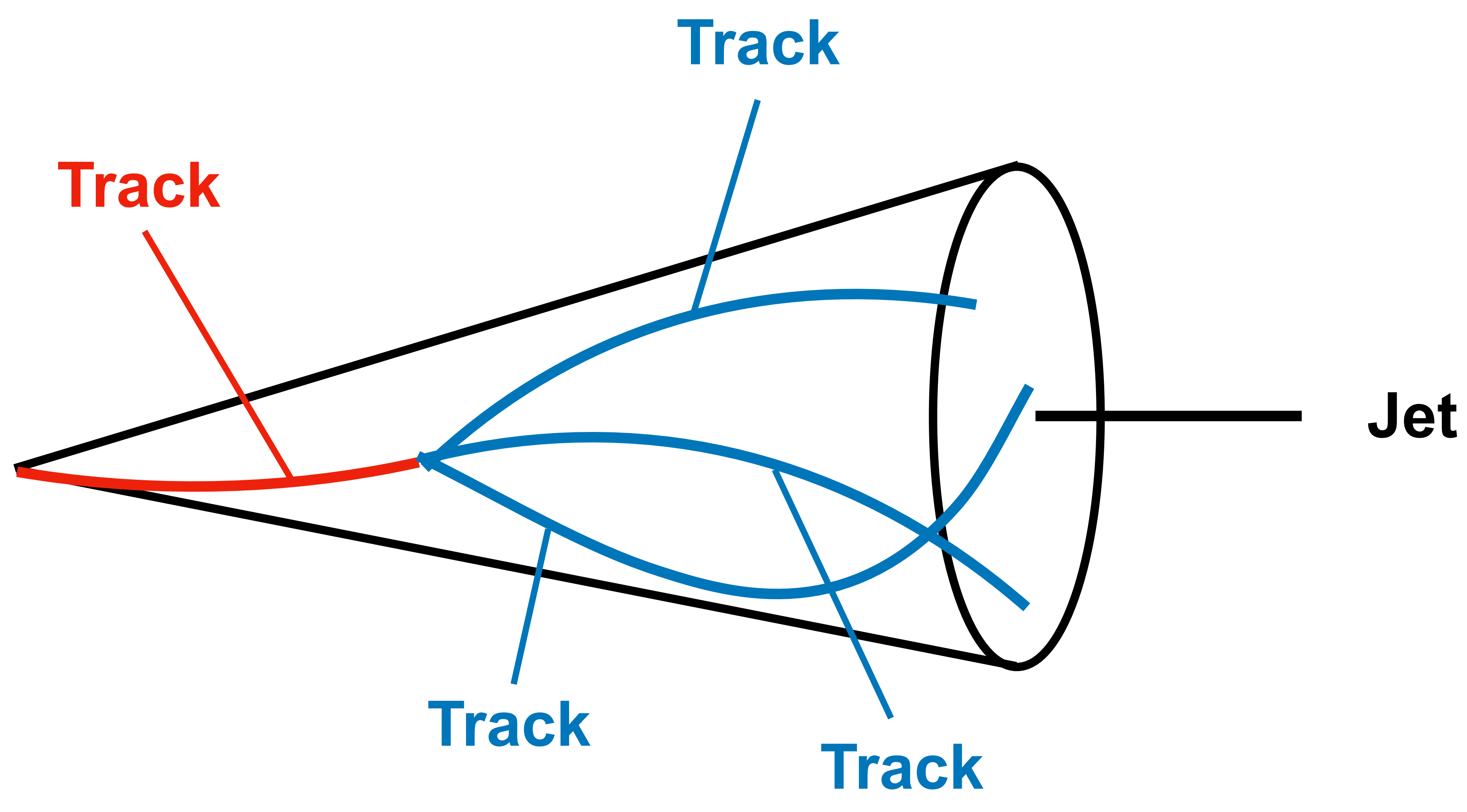
Boosted flavour-tagging within the ATLAS experiment

Jurjan Bootsma on behalf of the ATLAS collaboration

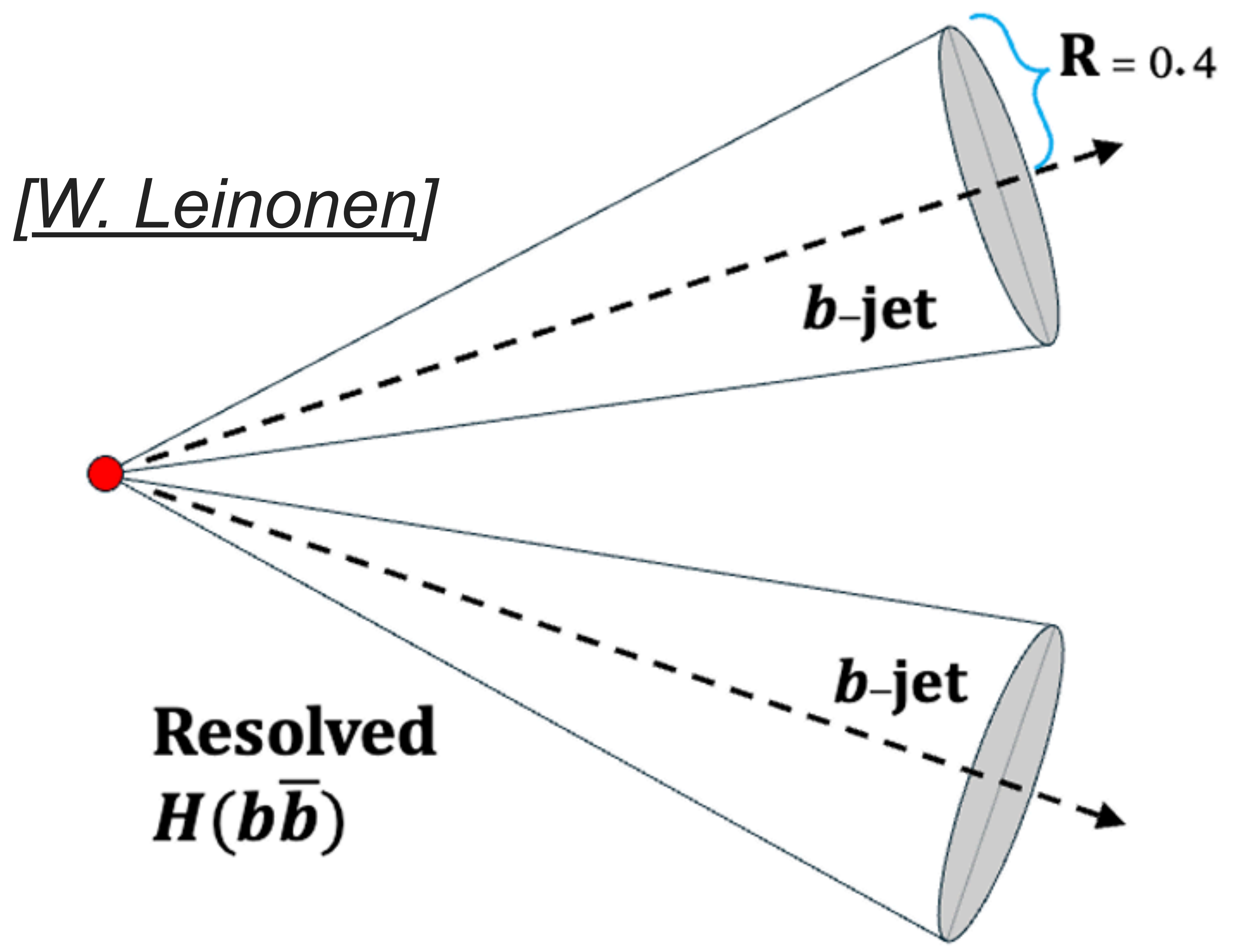
HHH Workshop 2025
Dubrovnik
30-09-2025

Flavour tagging

- Flavour tagging of hadronic jets
  - Identification of original quark flavour
- Important for Higgs decays
  - Crucial additional background suppression
  - Decays like $H \to b\bar{b}$ and $H \to c\bar{c}$
- Use different track features within jet
  - Use machine learning for optimal identification
- Rapid advancements
  - Run-2 flavour-tagger: DL1r
  - Run-3 flavour-tagger: GN2

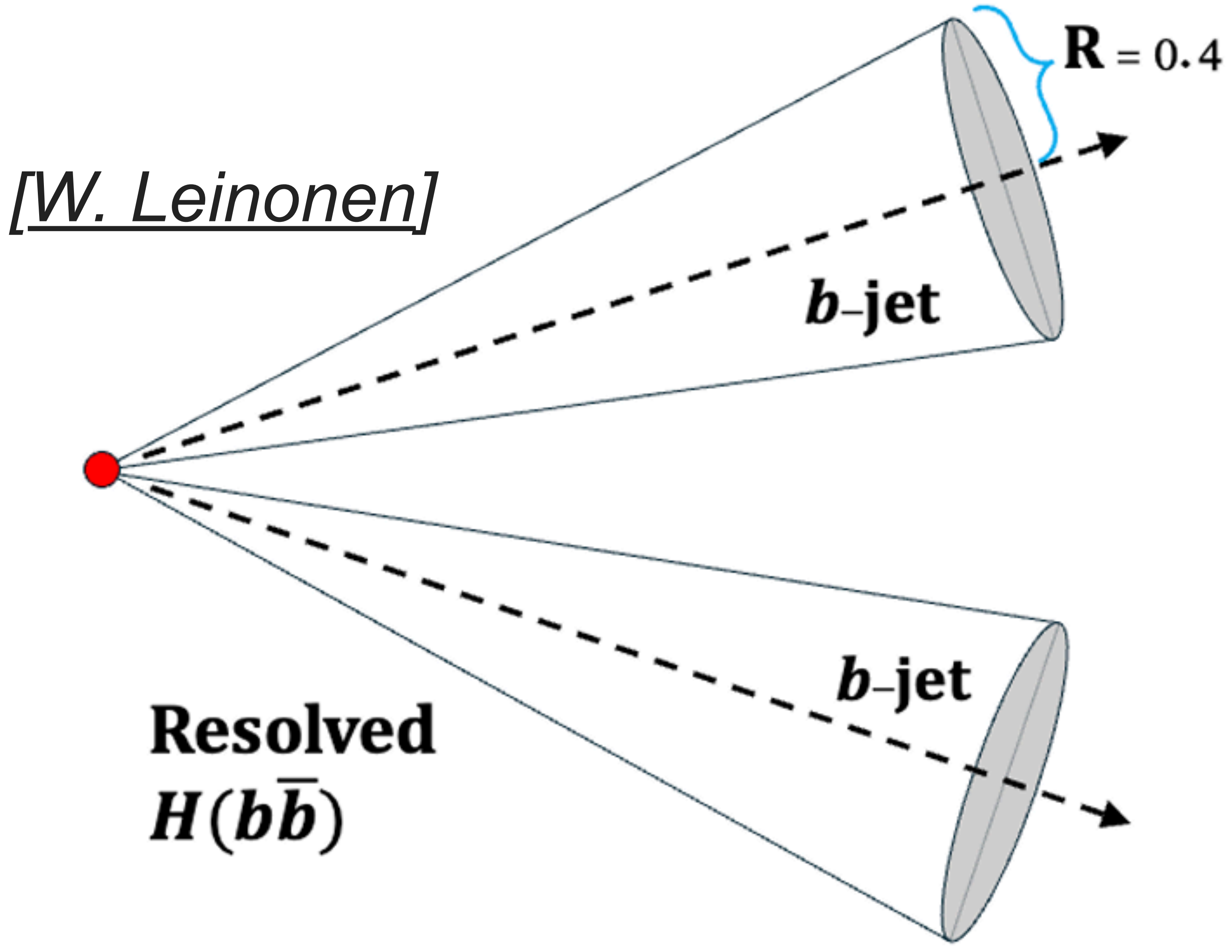


Lorentz-boosted regime



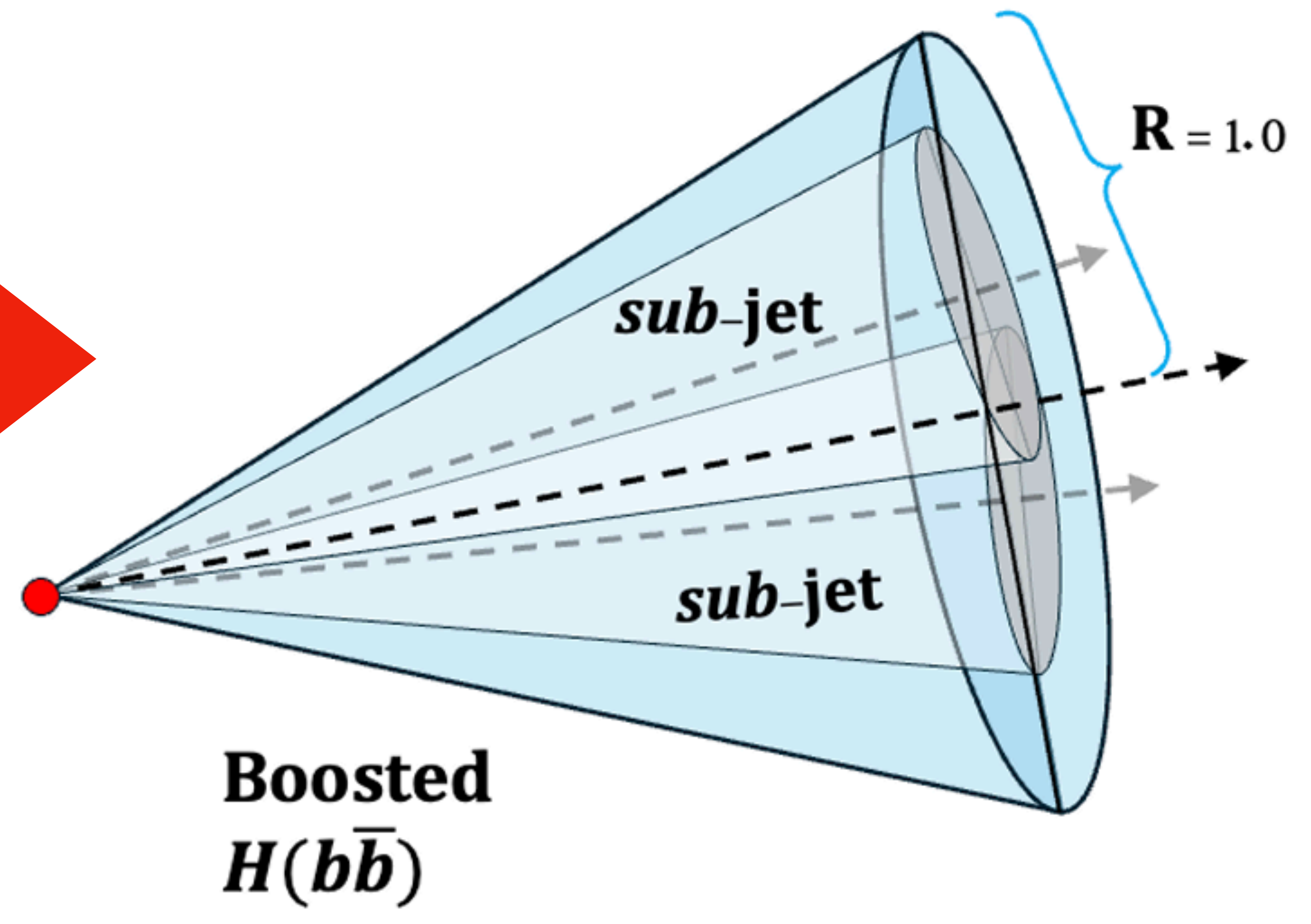
- Resolved flavour-tagging
  - Higgs boson with low p_T
  - Reconstruct as resolved small-R jets

Lorentz-boosted regime



Increasing p_T

$$\Delta R(jet1, jet2) \approx \frac{2m_H}{p_T^H}$$



- Resolved flavour-tagging
  - Higgs boson with low p_T
  - Reconstruct as resolved small-R jets

- Boosted flavour-tagging
  - Higgs boson with high p_T
  - Jets become merged (large-R jet)
  - Requires different reconstruction

Opportunities to triple Higgs

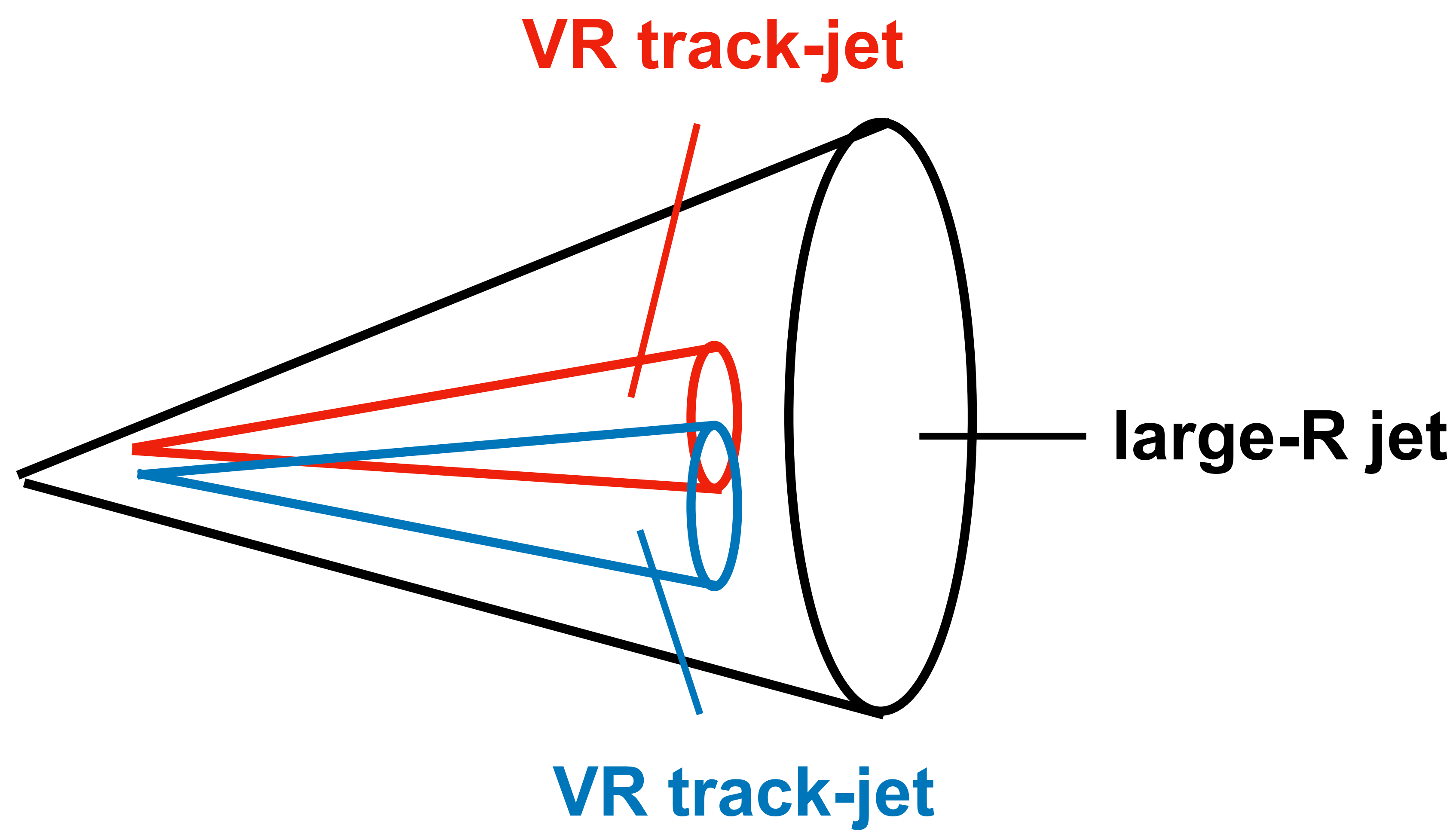
- Flavour-tagging important for triple Higgs
  - Sensitive channels rely on *b*-jet identification
- Could improve identification high-p_T Higgs candidates

Run-2 boosted flavour-tagging approaches

1. DL1r VR track-jet *b*-tagging
2. Xbb tagging

Resolved flavour-tagger: DL1r

- Both Run-2 approaches depend on resolved small-R jet flavour tagging
- DL1r: small-R jet flavour-tagger
  - Architecture: Deep Neural Network (DNN)
  - Input: variables describing the jets
    - Jet kinematic variables
    - Lower-level tagger variables
- Boosted approaches
  - Use DL1r to tag Variable-Radius (VR) subjets (track-jets)
  - Output: p_b, p_c, p_u

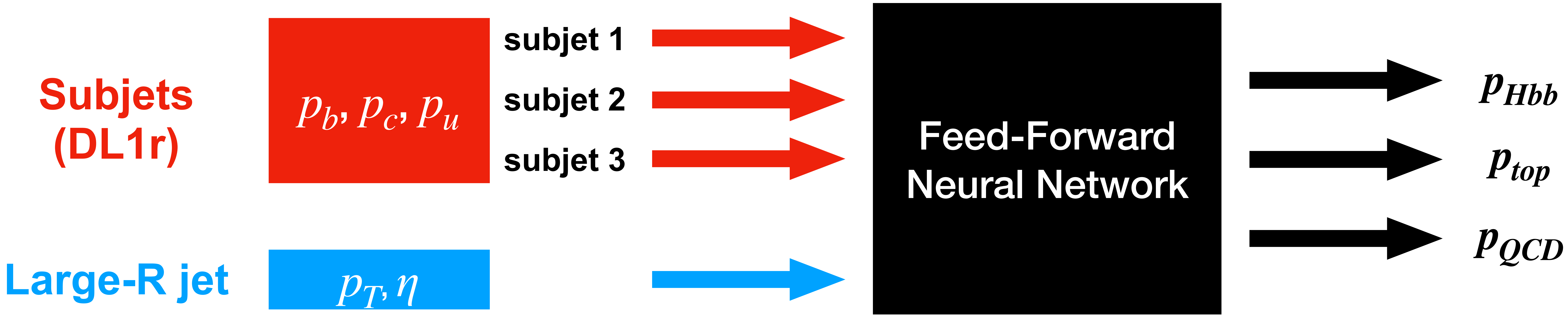


DL1r VR track-jet *b*-tagging

- Create 1D discriminant per subjet
  - With DL1r outputs: p_b, p_c, p_u
- Event selection
  - Based on constraints on discriminants of the subjets

$$D^{b}_{\mathrm{DL1r}} = log\left(\frac{p(b)}{f_c p(c) + (1 - f_c) p(u)}\right)$$

Previous Xbb flavour tagger



- Previous Xbb tagger
  - NN-based architecture
  - DL1r outputs as input

New Run-3 boosted flavour-tagging approach

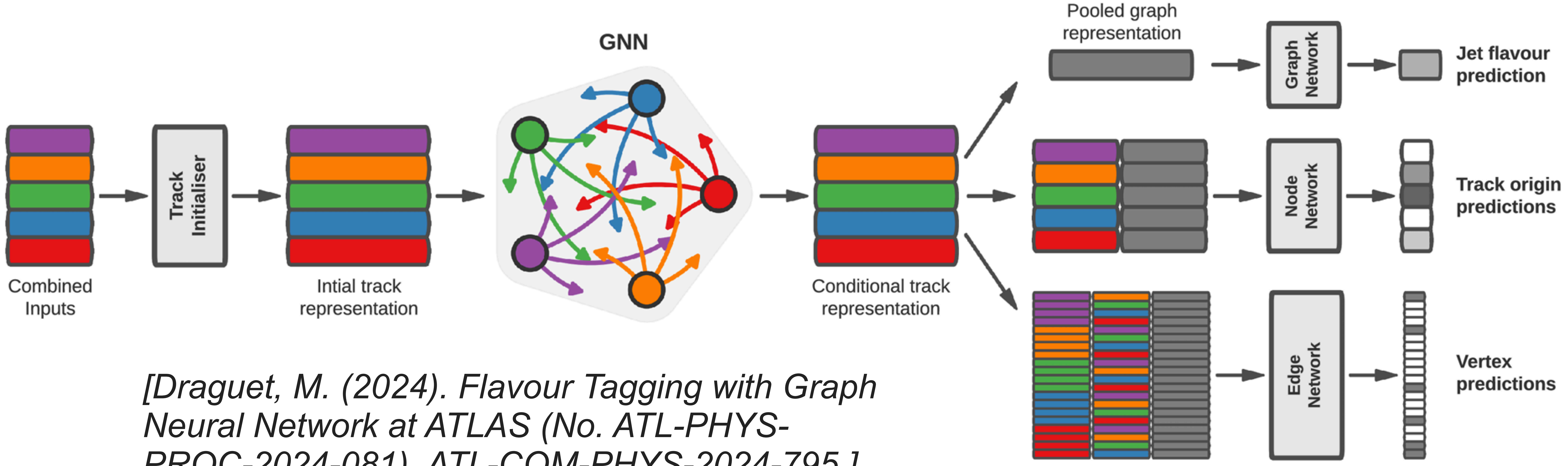
GN2X

[ATLAS collaboration. (2023). *Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into bb and cc in ATLAS* (No. PUBDB-2023-08047). LHC/ATLAS Experiment.]

[ATLAS Collaboration. (2025). Transforming jet flavour tagging at ATLAS. arXiv preprint arXiv:2505.19689.]

Current boosted flavour-tagger: GN2X

GN2X model



[Draguet, M. (2024). Flavour Tagging with Graph Neural Network at ATLAS (No. ATL-PHYS-PROC-2024-081). ATL-COM-PHYS-2024-795.]

- GN2X
  - New boosted flavour-tagger
  - Transformer-based architecture
  - Track-level input

GN2X input

- Large-R jet input
  - p_T
  - η
  - m_J
- Track-level input
- GN2X uses more information and correlations between the tracks

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)
dη	Pseudorapidity of track relative to the large-R jet η
dϕ	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)

[*] = *[ATLAS collaboration. (2023). Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into bb and cc in ATLAS (No. PUBDB-2023-08047). LHC/ATLAS Experiment.]*

GN2X training: auxiliary tasks

- Advantage GN2X: auxiliary tasks
  - Improve jet classification
  - Improve interpretability of the model
- Auxiliary task 1
  - Predict track origins track-by-track
- Auxiliary task 2
  - Determine which tracks originate from same vertex
- Model without auxiliary tasks shows decreased performance

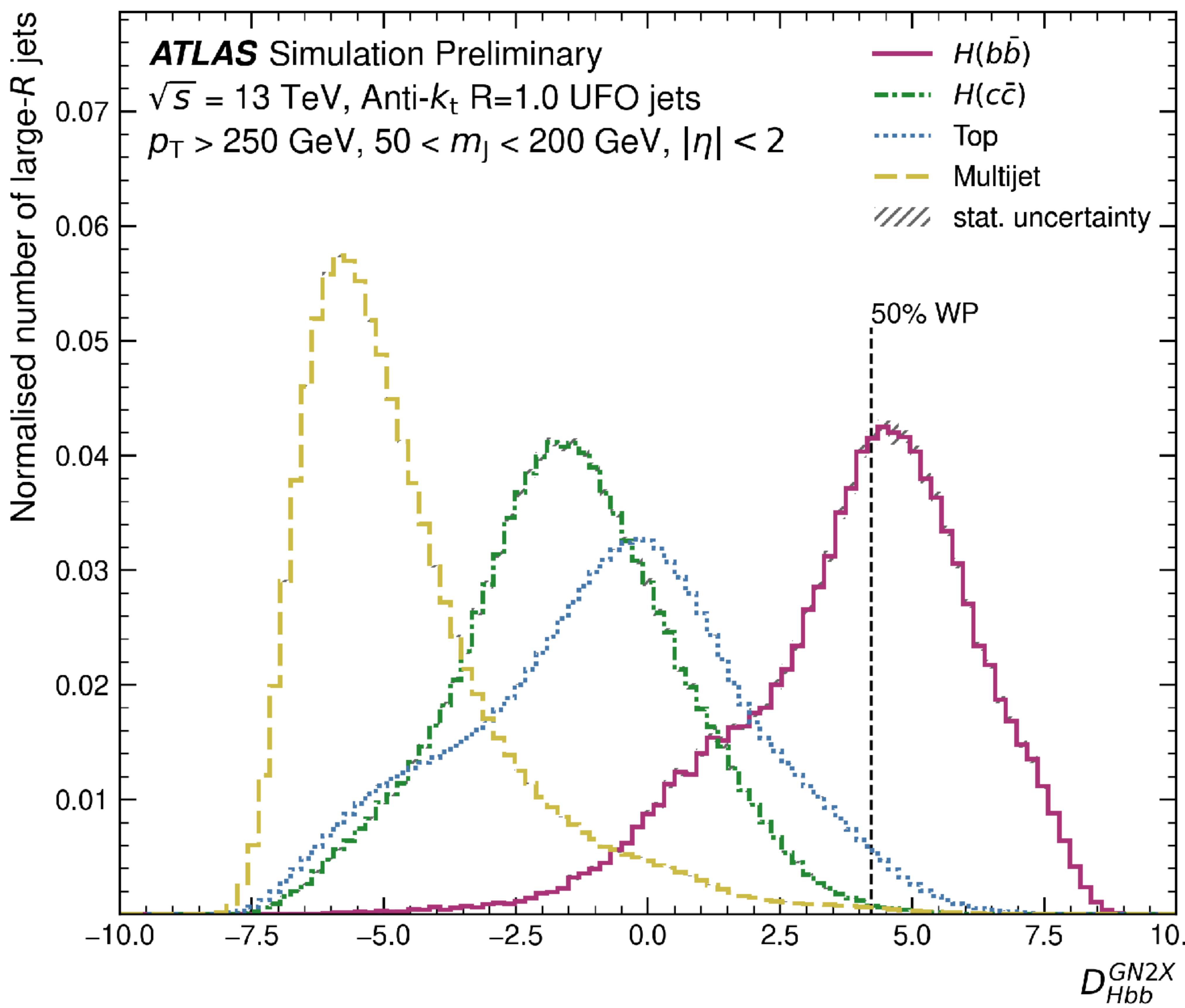
$$L = L_{\text{jet}} + \alpha L_{\text{vertexing}} + \beta L_{\text{tracking}}$$

GN2X discriminant

GN2X model

- GN2X provides probability scores
  - $p_{Hbb}, p_{Hcc}, p_{top}, p_{QCD}$
- Construct 1D discriminant
  - $f_{Hcc} = 0.02,\ f_{top} = 0.25$
- Cut on samples using discriminant
  - Working point (WP) corresponding to ϵ_{Hbb}
  - Tight WP: high background rejection
  - Loose WP: high signal efficiency

[*]

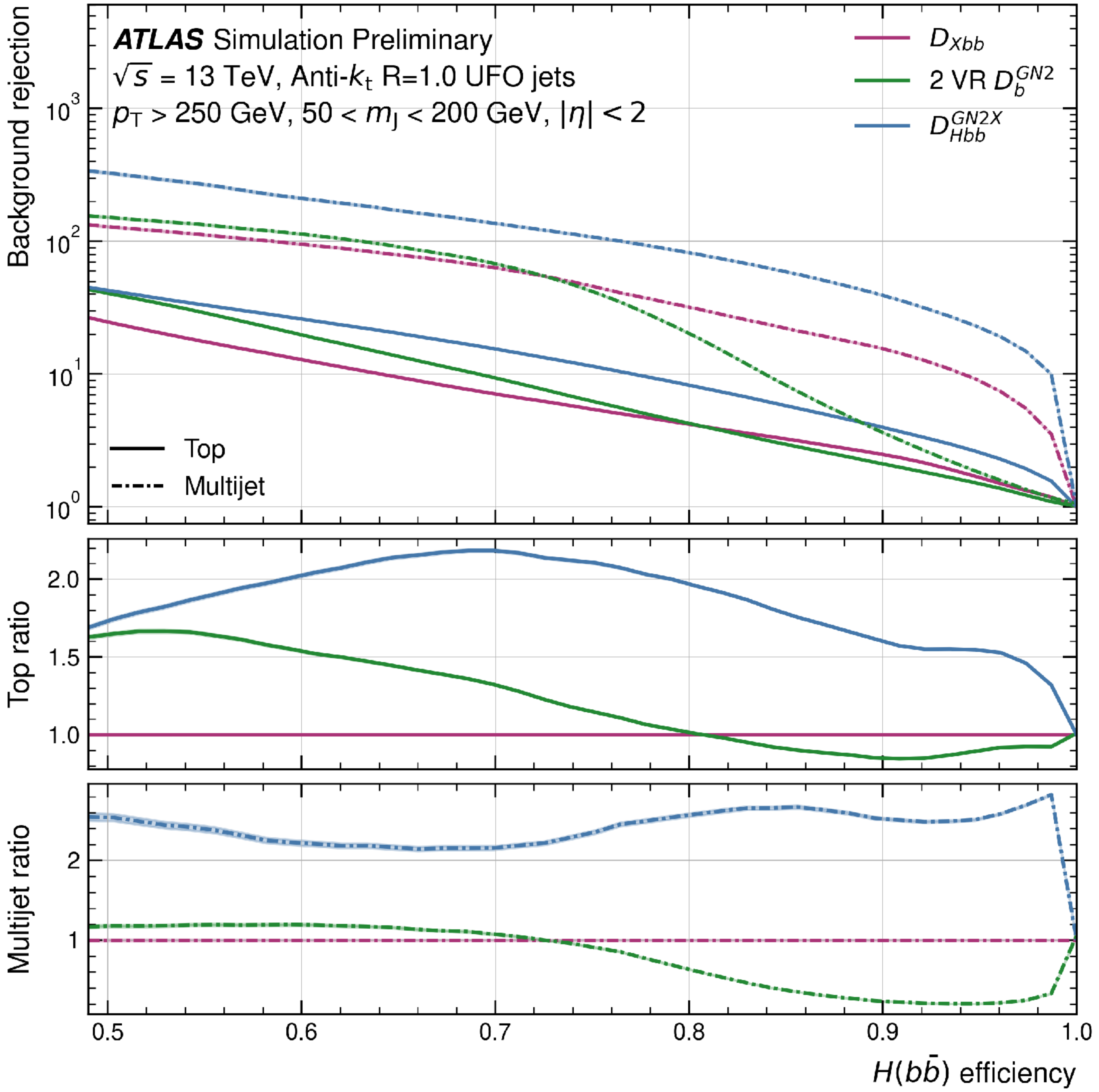


$$D_{GN2X}^{Hbb} = log(\frac{p_{Hbb}}{f_{Hcc}p_{Hcc} + f_{top}p_{top} + (1 - f_{Hcc} - f_{top})p_{QCD}})$$

GN2X performance

- GN2X Hbb-tagging performance plot
  - Background rejection (top and multijet)
  - As function of Hbb efficiency
- GN2X outperforms previous approaches across efficiency range
- Rejection improvements at 50%
  - Top: 1.6x
  - Multijet: 2.5x

[*]

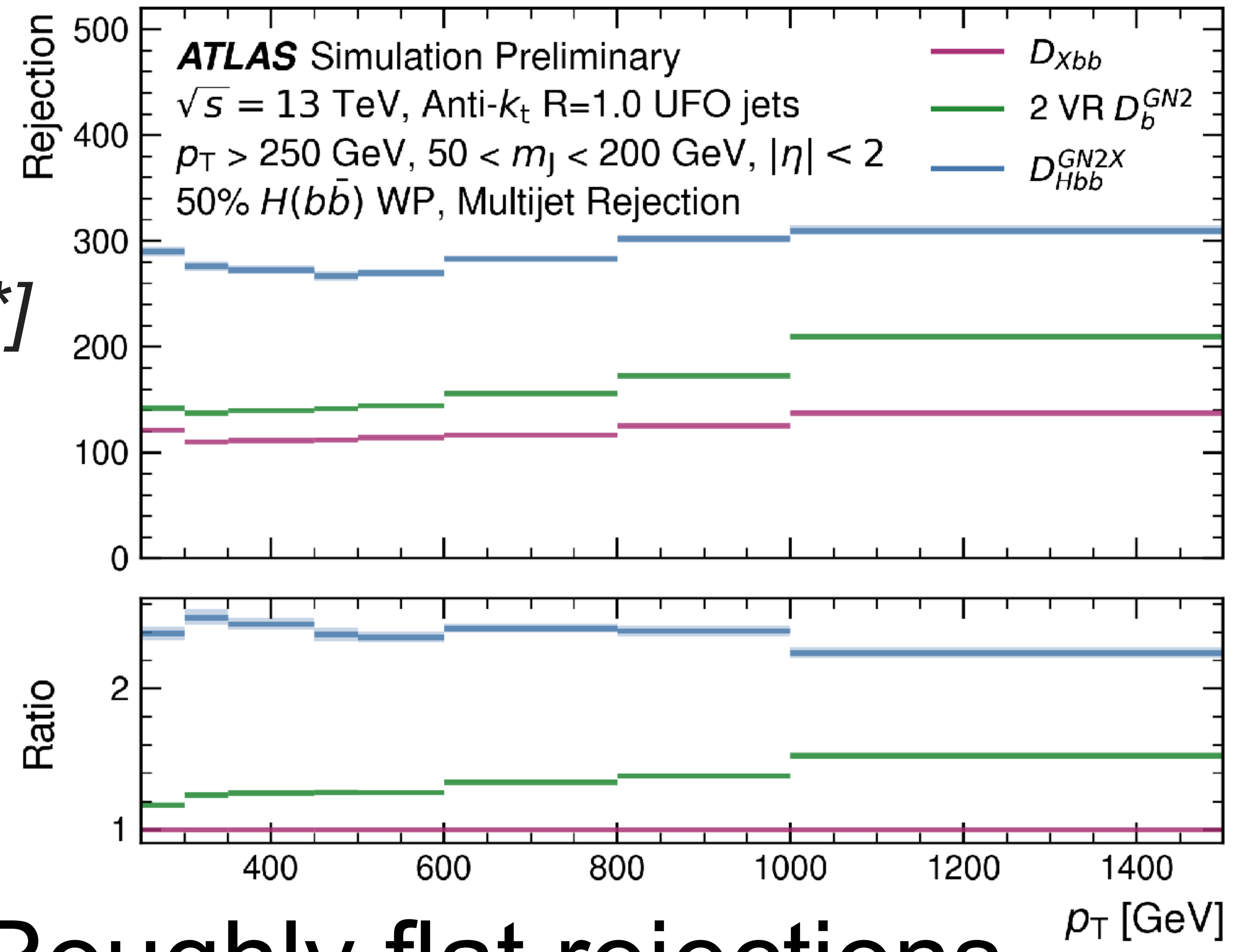


[ATLAS collaboration. (2023). Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into bb and cc in ATLAS (No. PUBDB-2023-08047). LHC/ATLAS Experiment.]

Background rejection as function of p_T

Multijet rejection at 50% Hbb WP

[*]



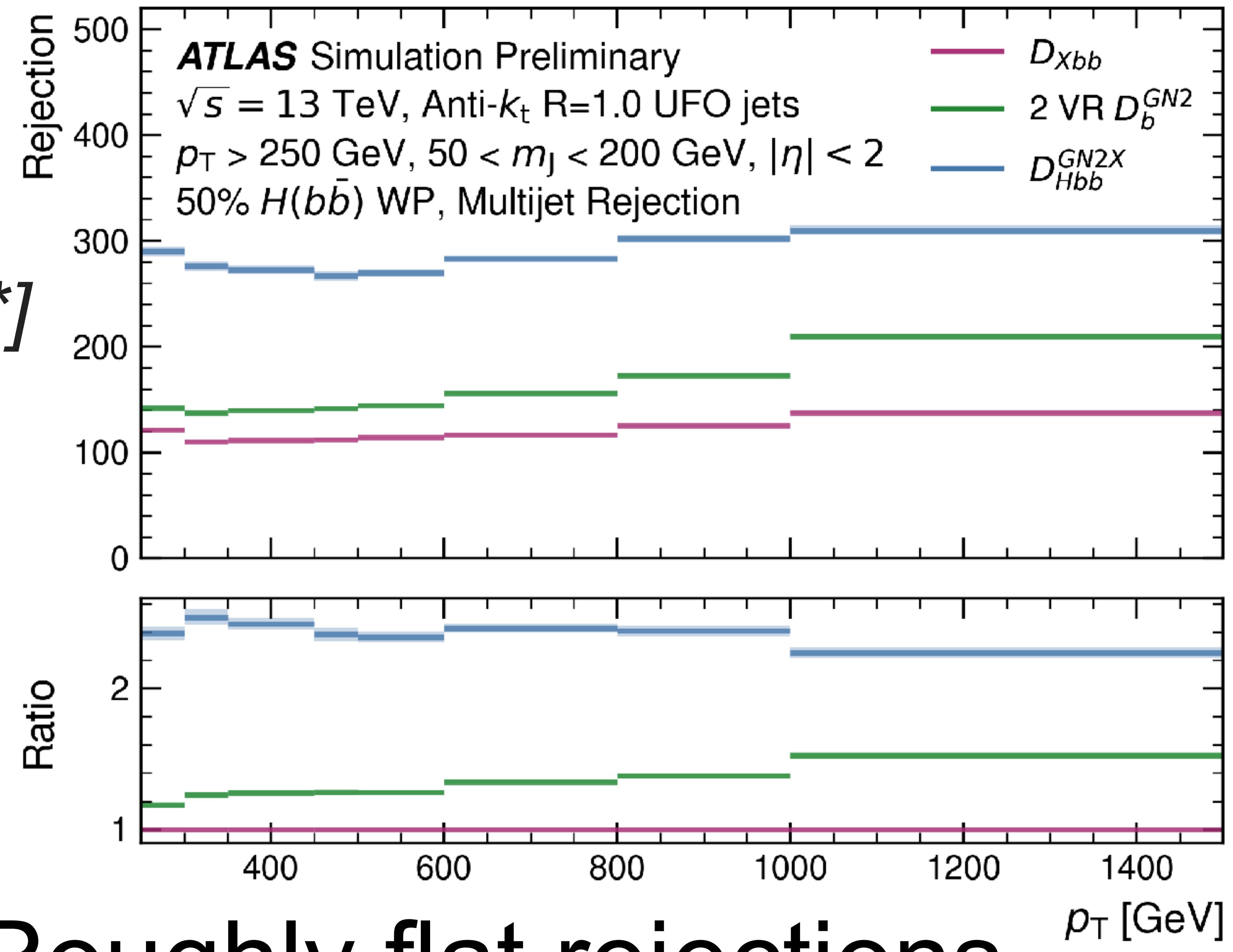
- Roughly flat rejections
  - Reflects minimal multijet sculpting as function of p_T

Background rejection as function of p_T

GN2X performance

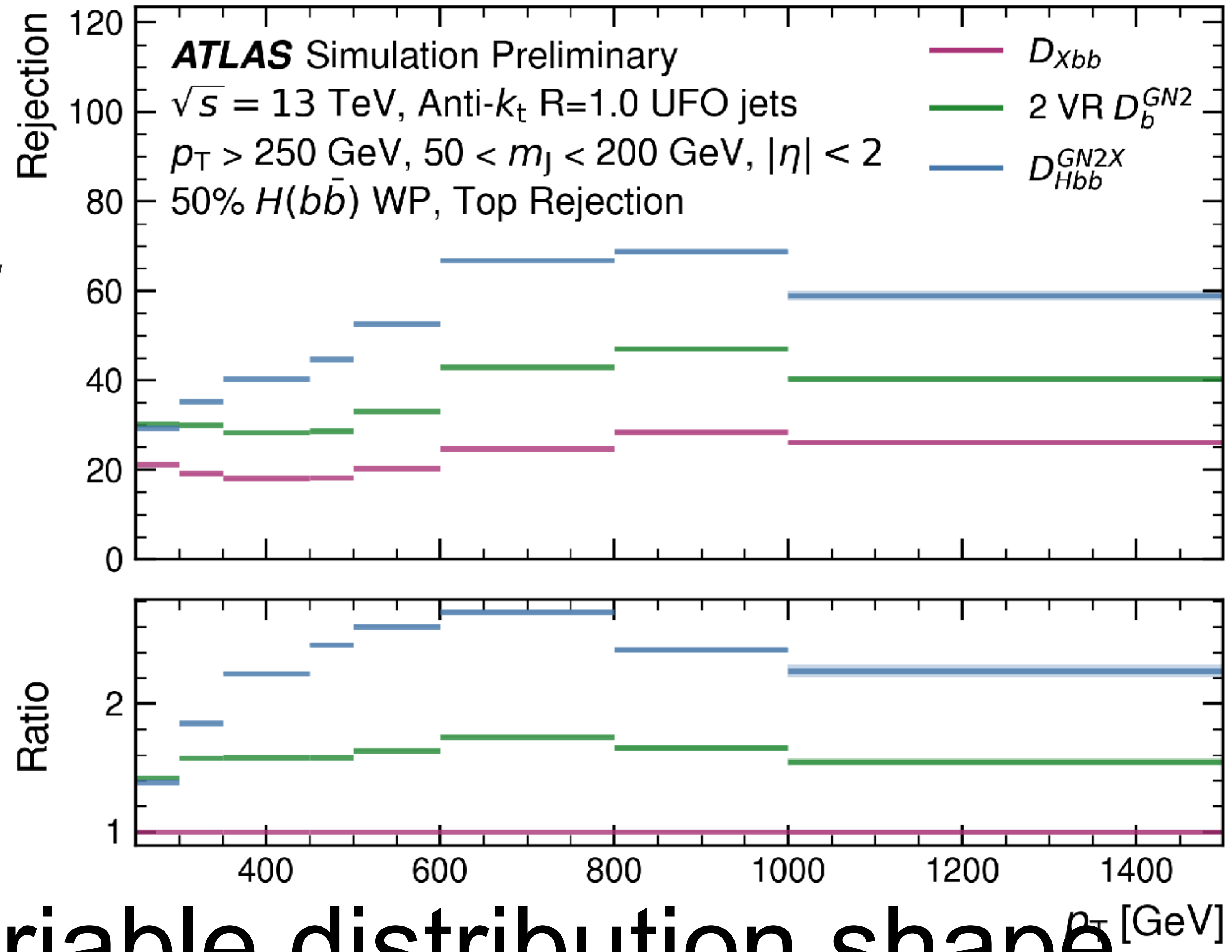
Multijet rejection at 50% Hbb WP

[*]



Top rejection at 50% Hbb WP

[*]



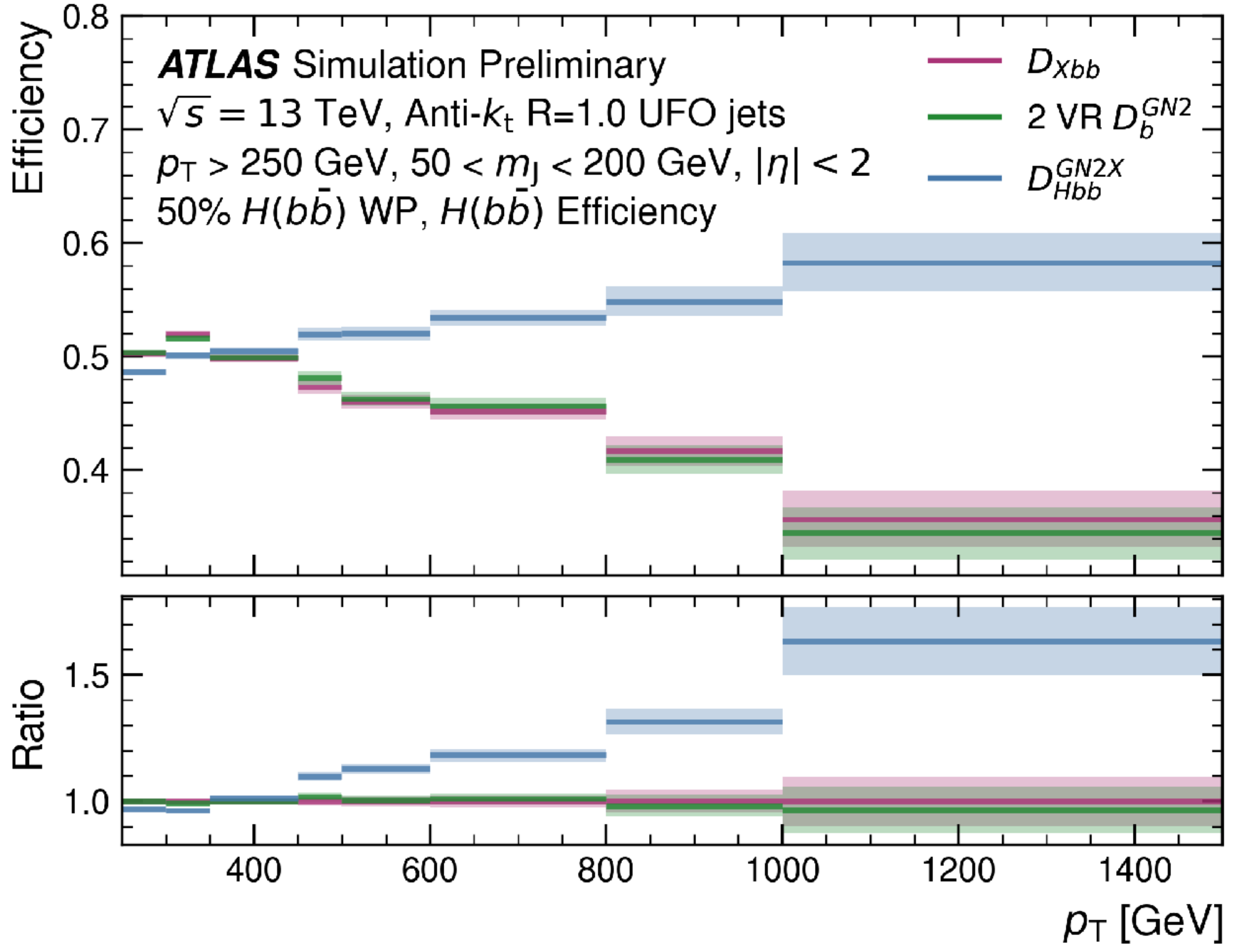
- Roughly flat rejections
  - Reflects minimal multijet sculpting as function of p_T
- Variable distribution shape
  - Reflects change in top compositions
    - W(qq) → fully top-contained jets

Nikhef

Signal efficiency as function of p_T

- $H(\rightarrow b\bar{b})$ efficiency compared to p_T
- Previous approaches
  - Efficiency decrease as function of p_T
    - Subjet resolution effects
- GN2X
  - Efficiency increase up to p_T = 1.5 TeV

[*]

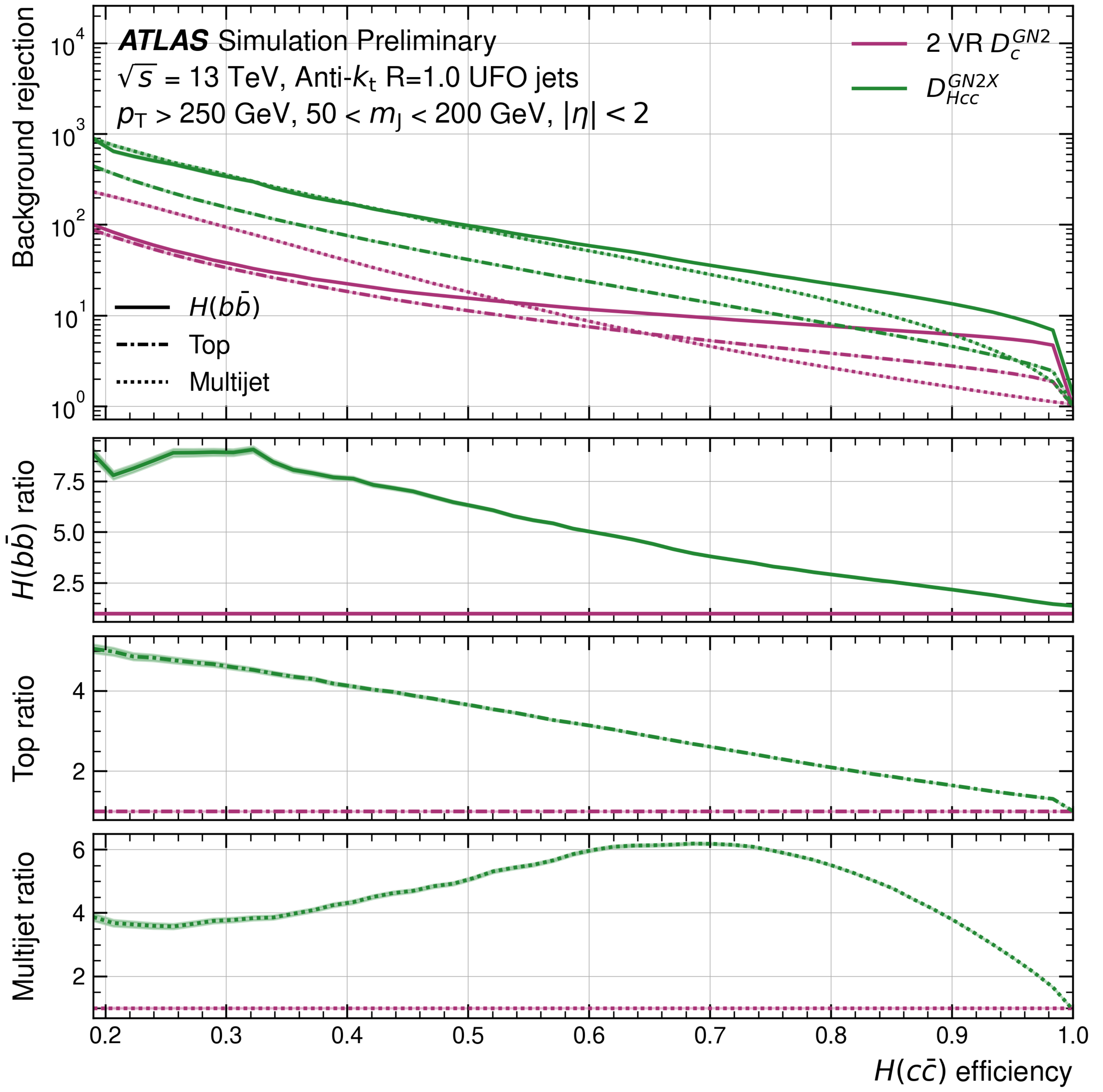


GN2X Hcc-tagging

- GN2X also provides Hcc-tagging
- Event selection based on Hcc-discriminant
- GN2X outperforms previous approaches across efficiency range
- Rejection improvements at 50%
  - Top: 3x
  - Multijet: 5x
  - $H(\rightarrow b\bar{b})$: 6x

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

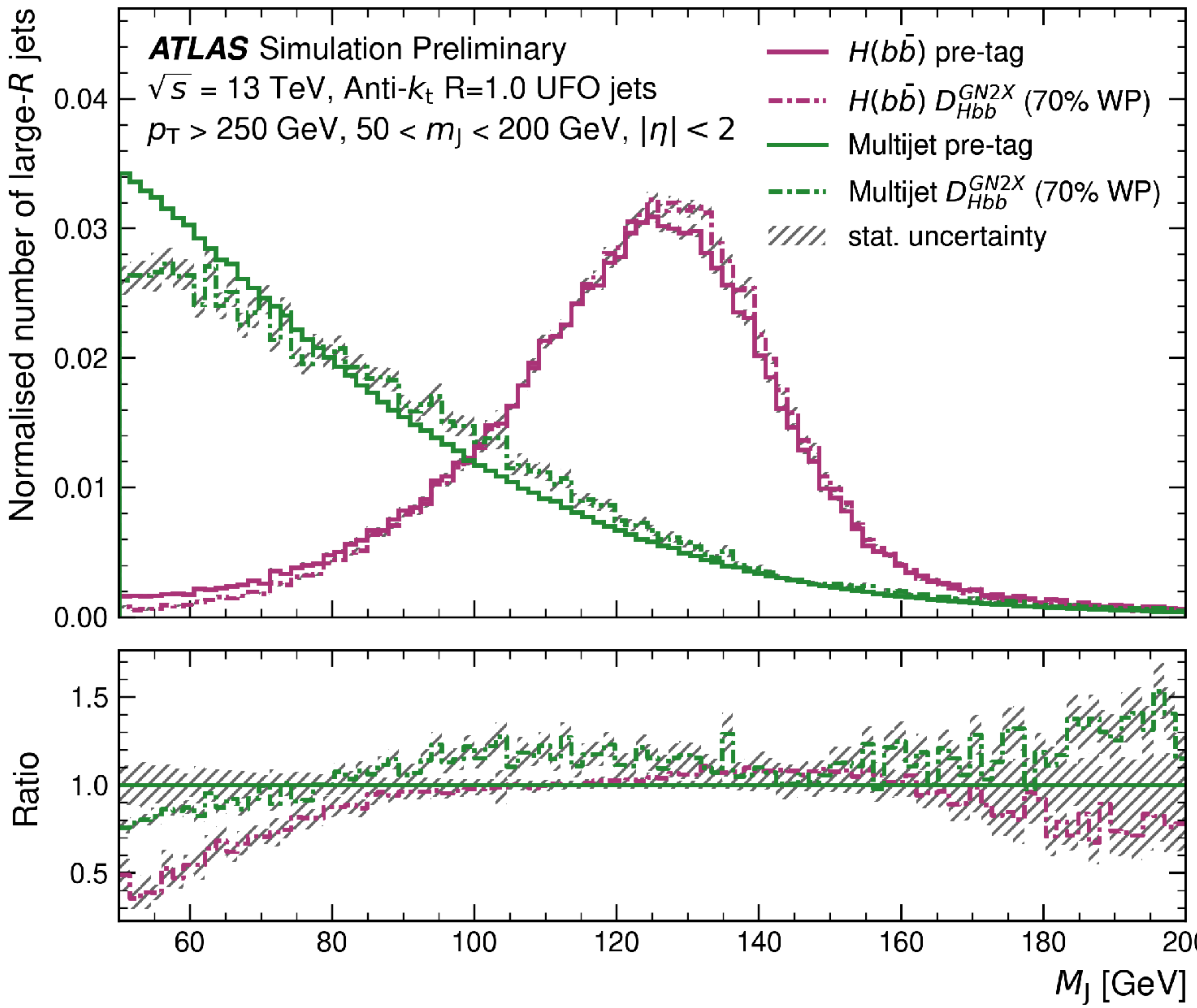
[*]



Mass sculpting

- Mass sculpting problem
  - Tagging efficiency dependent on the large-R jet mass
    - Due to changing flavour compositions
  - Results in different pre-tag and post-tag distributions
- Important for calibration

[*]



Signal calibration

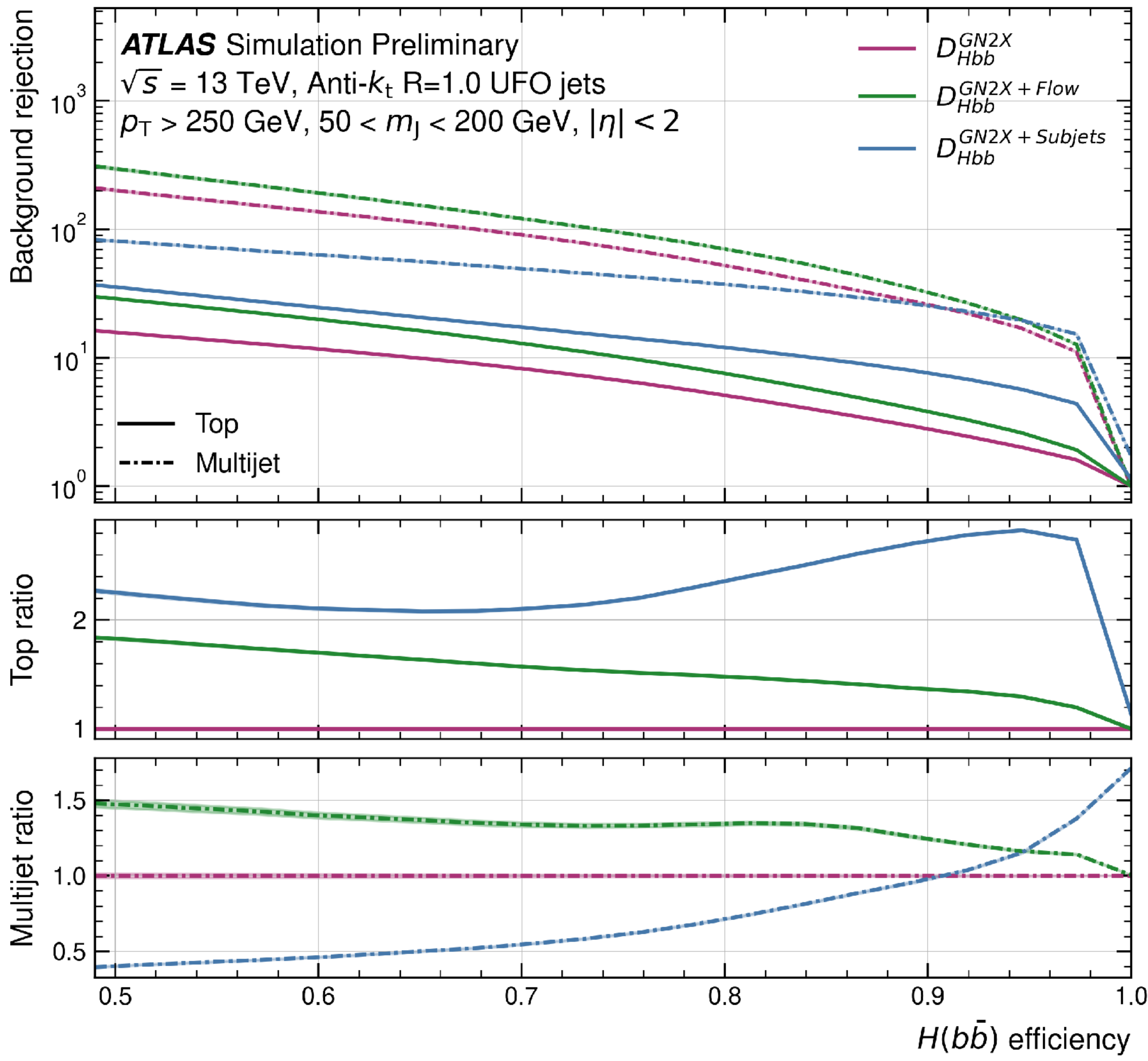
- Previous plots show MC-based performance
  - But we need application to data
- Use $Z(\rightarrow b\bar{b})$ as standard for calibration
  - Should be similar resonance as Higgs due to mass independence
- Scale factors to correct for MC-data difference

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

Outlook: heterogeneous inputs

- Improving GN2X
  - Increase model input
- Method 1: Add subjet tagging information
  - Use resolved tagger GN2 to tag subjets
  - Use kinematic + GN2 scores as input
- Method 2: Add Flow objects
  - Use additional calorimeter information

[*]



Summary

- Boosted flavour-tagging has **big potential** for ATLAS analyses
  - Also triple Higgs: high-p_T Hbb tagging
- State-of-the-art ATLAS boosted flavour-tagger: **GN2X**
- Model improvements
  - Transformer architecture
  - Direct track-level input
  - Auxiliary tasks
- Model MC performance (compared to previous Xbb approach)
  - Background rejection at 50% WP
    - Top: **1.6x**
    - Multijet: **2.5x**

Back-up

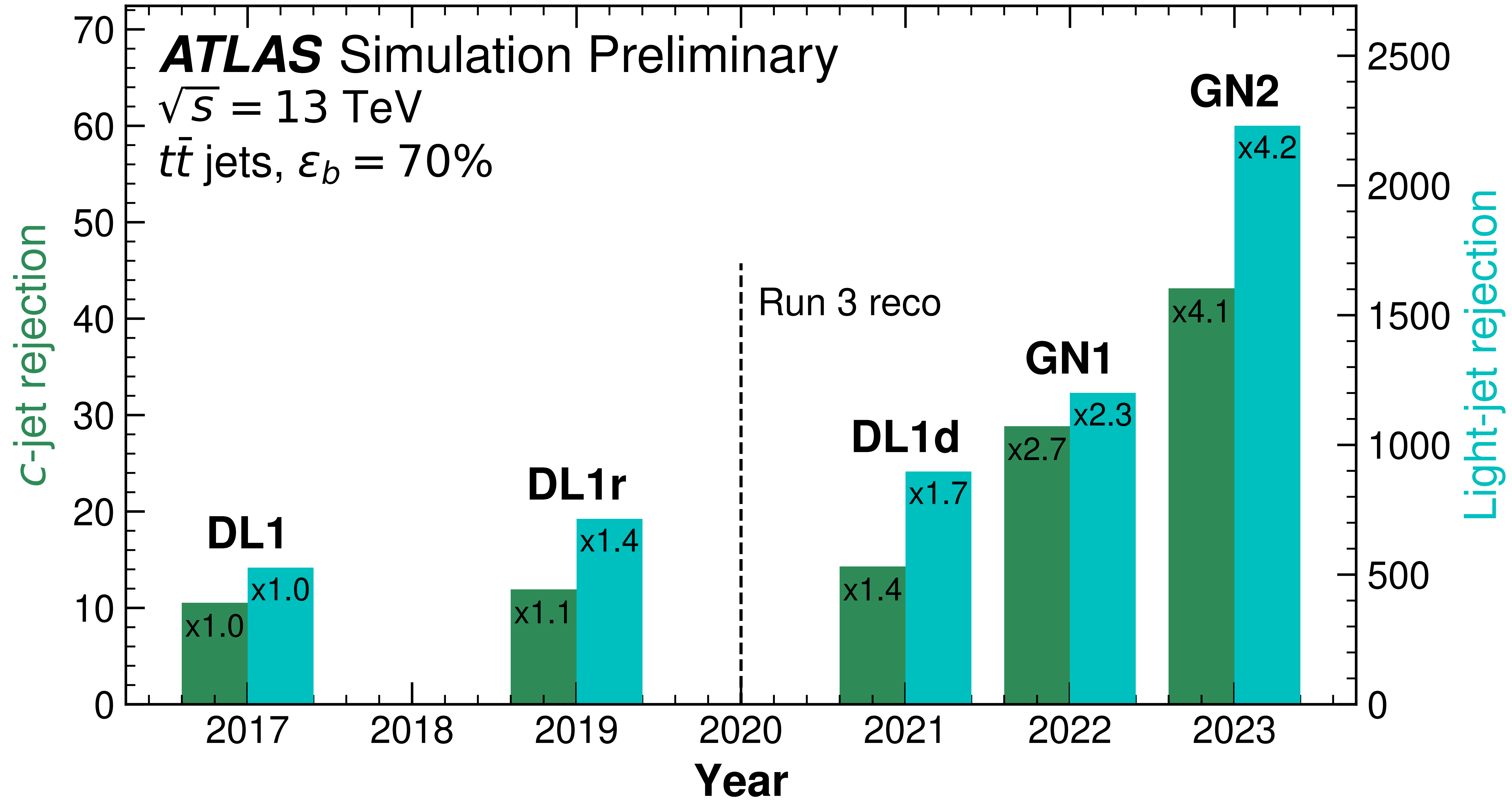
DL1r input

Input	Variable	Description	SVKine	JFKine	DL1	DL1r
Kinematics	p_{T}	Jet p_{T}	✓	✓	✓	✓
	η	Jet $\|\eta\|$	✓	✓	✓	✓
IP2D, IP3D	$\log(P_{\mathrm{b}}/P_{\mathrm{light}})$	Likelihood ratio of the b-jet to light-flavour jet hypotheses			✓	✓
	$\log(P_{\mathrm{b}}/P_{\mathrm{c}})$	Likelihood ratio of the b-jet to c-jet hypotheses			✓	✓
	$\log(P_{\mathrm{c}}/P_{\mathrm{light}})$	Likelihood ratio of the c-jet to light-flavour jet hypotheses			✓	✓
RNNIP	P_{b}	b-jet probability				✓
	P_{c}	c-jet probability				✓
	P_{light}	light-flavour jet probability				✓
SV1	$m(\mathrm{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass	✓		✓	✓
	$f_E(\mathrm{SV})$	Jet energy fraction of the tracks associated with the secondary vertex	✓		✓	✓
	$N_{\mathrm{TrkAtVtx}}(\mathrm{SV})$	Number of tracks used in the secondary vertex	✓		✓	✓
	$N_{\mathrm{2TrkVtx}}(\mathrm{SV})$	Number of two-track vertex candidates	✓		✓	✓
	$L_{xy}(\mathrm{SV})$	Transverse distance between the primary and secondary vertices	✓		✓	✓
	$L_{xyz}(\mathrm{SV})$	Distance between the primary and secondary vertices	✓		✓	✓
	$S_{xyz}(\mathrm{SV})$	Distance between the primary and secondary vertices divided by its uncertainty	✓		✓	✓
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})(\mathrm{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.	✓		✓	✓
JetFitter	$m(\mathrm{JF})$	Invariant mass of tracks from displaced vertices		✓	✓	✓
	$f_E(\mathrm{JF})$	Jet energy fraction of the tracks associated with the displaced vertices		✓	✓	✓
	$\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})(\mathrm{JF})$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices		✓	✓	✓
	$S_{xyz}(\mathrm{JF})$	Significance of the average distance between PV and displaced vertices		✓	✓	✓
	$N_{\mathrm{TrkAtVtx}}(\mathrm{JF})$	Number of tracks from multi-prong displaced vertices		✓	✓	✓
	$N_{\mathrm{2TrkVtx}}(\mathrm{JF})$	Number of two-track vertex candidates (prior to decay chain fit)		✓	✓	✓
	$N_{\text{1-trk vertices}}(\mathrm{JF})$	Number of single-prong displaced vertices		✓	✓	✓
	$N_{\geq\text{2-trk vertices}}(\mathrm{JF})$	Number of multi-prong displaced vertices		✓	✓	✓
	$L_{xyz}(2^{\mathrm{nd}})(\mathrm{JF})$	Distance of 2^{nd} vertex from PV		✓	✓	✓
	$L_{xy}(2^{\mathrm{nd}})(\mathrm{JF})$	Transverse displacement of the 2^{nd} vertex		✓	✓	✓
	$m_{\mathrm{Trk}}(2^{\mathrm{nd}})(\mathrm{JF})$	Invariant mass of tracks associated with the 2^{nd} vertex		✓	✓	✓
	$E(2^{\mathrm{nd}})(\mathrm{JF})$	Energy of the tracks associated with the 2^{nd} vertex		✓	✓	✓
	$f_E(2^{\mathrm{nd}})(\mathrm{JF})$	Jet energy fraction of the tracks associated with the 2^{nd} vertex		✓	✓	✓
	$N_{\mathrm{TrkAtVtx}}(2^{\mathrm{nd}})(\mathrm{JF})$	Number of tracks associated with the 2^{nd} vertex		✓	✓	✓
	$\eta_{\mathrm{trk}}^{\mathrm{min,max,avg}}(2^{\mathrm{nd}})(\mathrm{JF})$	Min., max. and avg. pseudorapidity of tracks at the 2^{nd} vertex		✓	✓	✓

[Aad, G., Abbott, B., Abbott, D. C., Abeling, K., Abidi, S. H., Aboulhorma, A., ... & Ballabene, E. (2023). ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset. The European Physical Journal C, 83(7), 681.]

Resolved flavour-tagging improvements

Backup



Samples

Backup

Training samples

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

Evaluation samples

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

Track selection requirements

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

Model 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

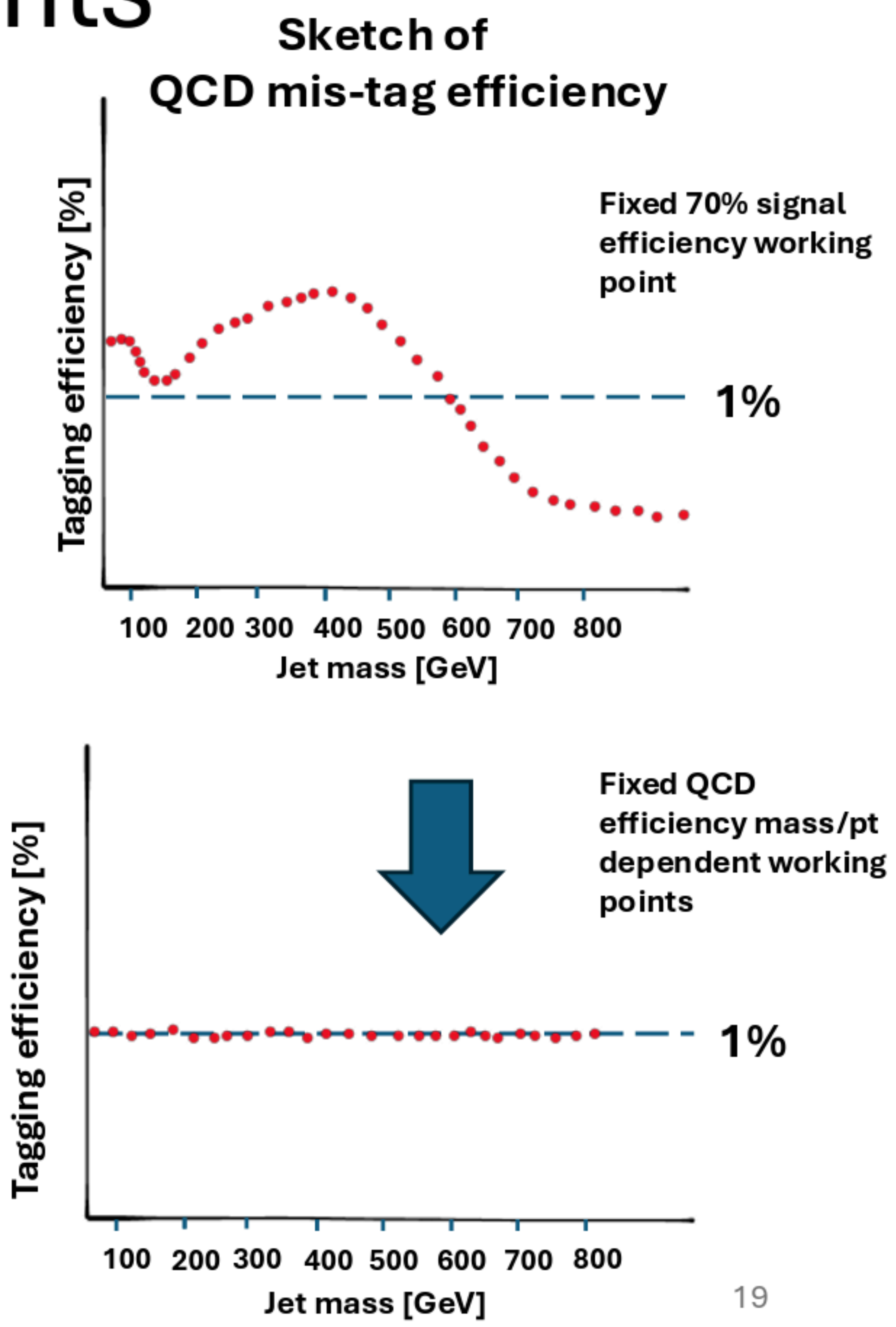
Flat-mass working points

Backup

Flat QCD Working Points

By defining mass/pt dependent fixed-efficiency QCD working points, we can remove this residual sculpting **manually**.

For the GN2X tagger, H(bb) mis-tag rates of 0.3%, 0.45%, 0.75%, and 1.25% are used to produce **equivalent H(bb) signal-efficiency selections** of 50%, 60%, 70%, and 80%.



3/25/2025 19

[Talk by Waltteri Leinonen https://cds.cern.ch/record/2929790]