Flavour Tagging with Graph Neural Network with the ATLAS Detector

Neelam Kumari

On behalf of the ATLAS experiment

BOOST 2024 Genova, Italy

29th July 2024

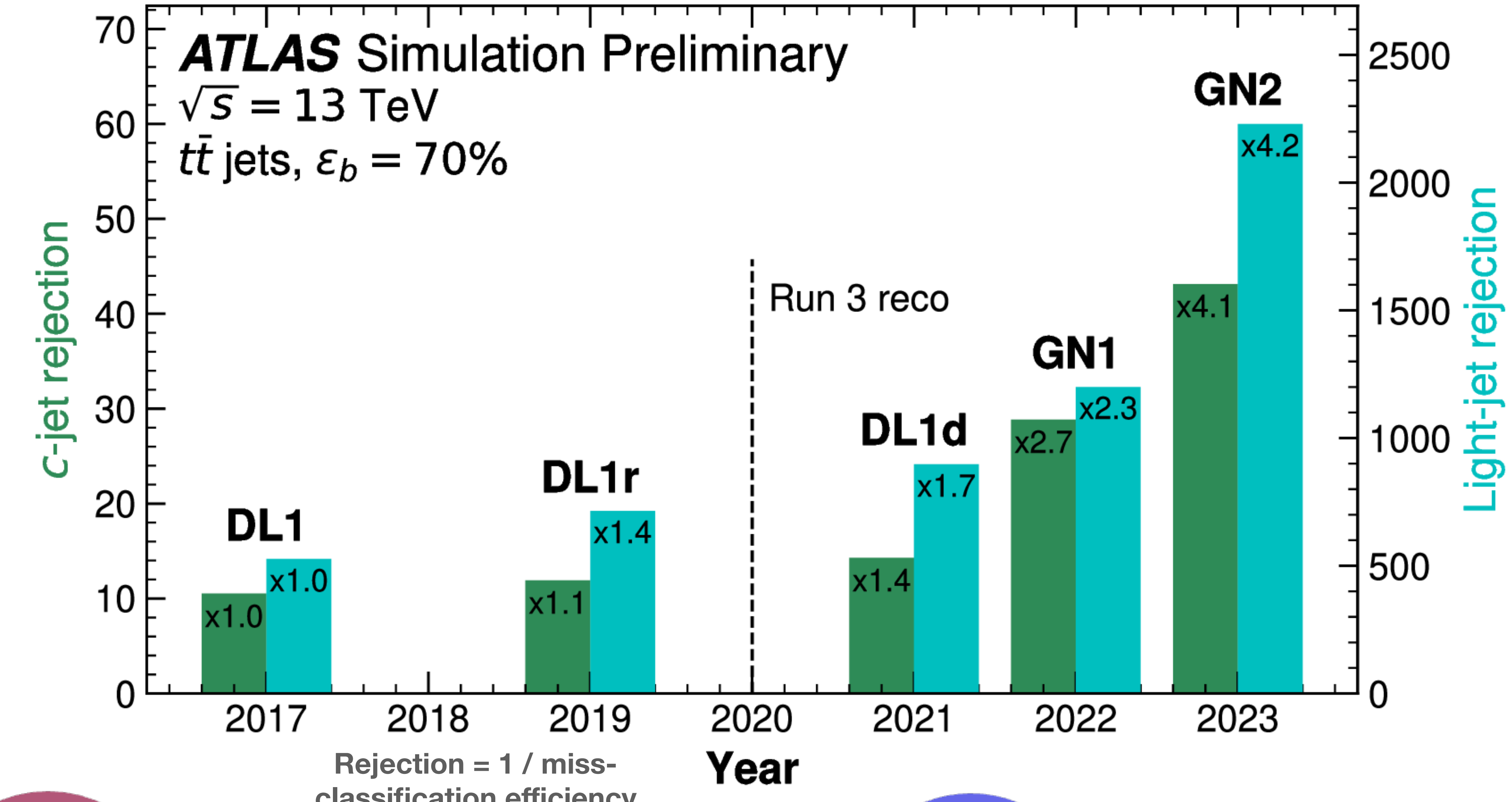
Odyssey of ATLAS Flavour Tagging

- FTAG algorithms (aka "tagger") hypothesise jet flavour ⟹ Use reconstructed jets & track properties for B/C hadron identification

- Used in many analyses: H→ bb, H→ cc & di-Higgs ….

ATLAS FTAG group continually advances, making use of ever more sophisticated methods & architectures like deep-learning

FTAG-2023-01



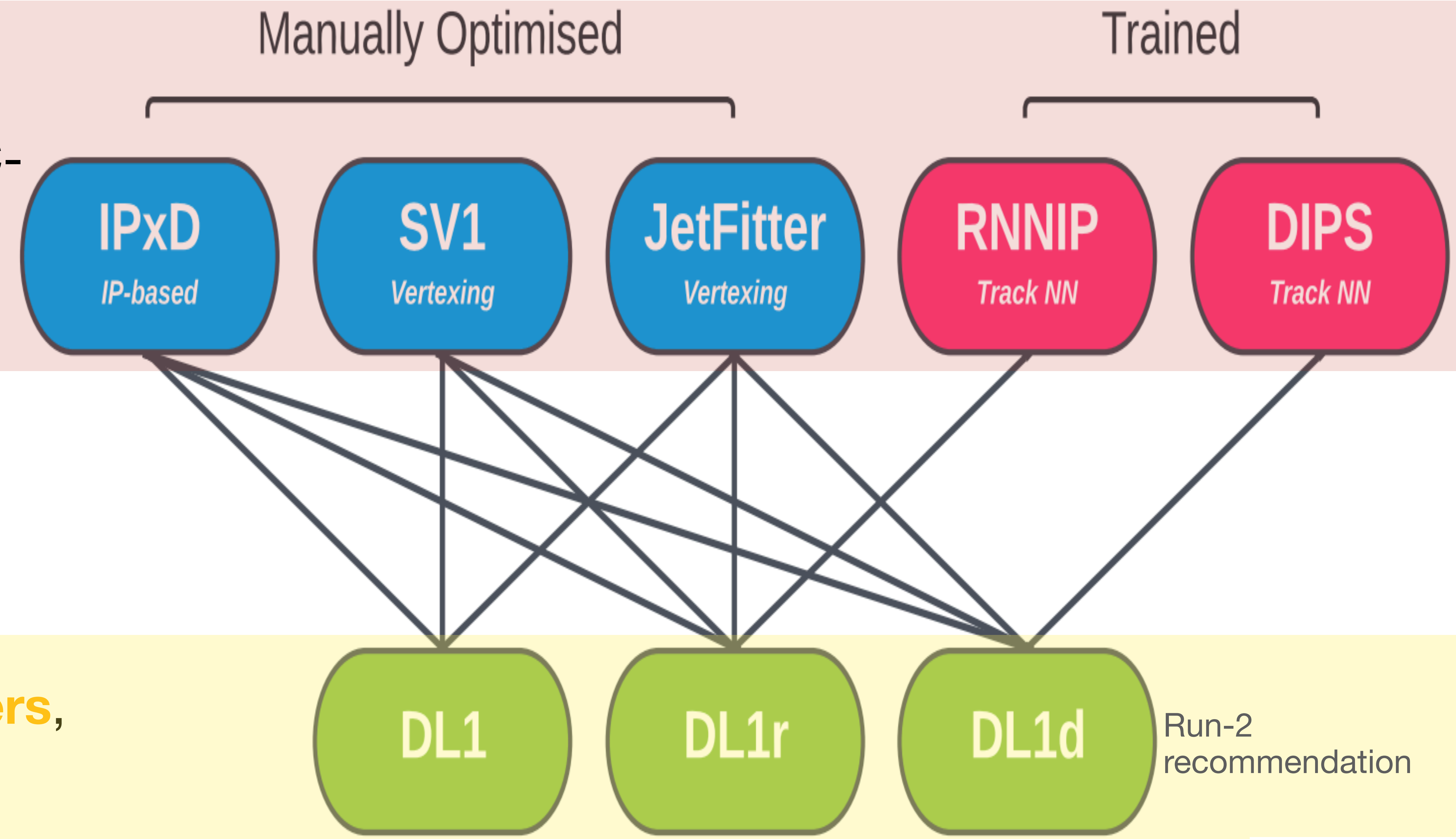
Rejection = 1 / miss-classification efficiency

Today in FTAG we are all about transformers

Transformers we're training:
GN2: Small-R jets
GN2X: Large-R jets (X→bb tagging…)

Previous approach…

- Jet and track inputs are fed to **low level taggers**
- Manually optimised taggers, exploit different B/C-hadron decay properties: ***IPxD, SV1, JetFitter***
- Track-based ML models: ***RNNIP, DIPS***



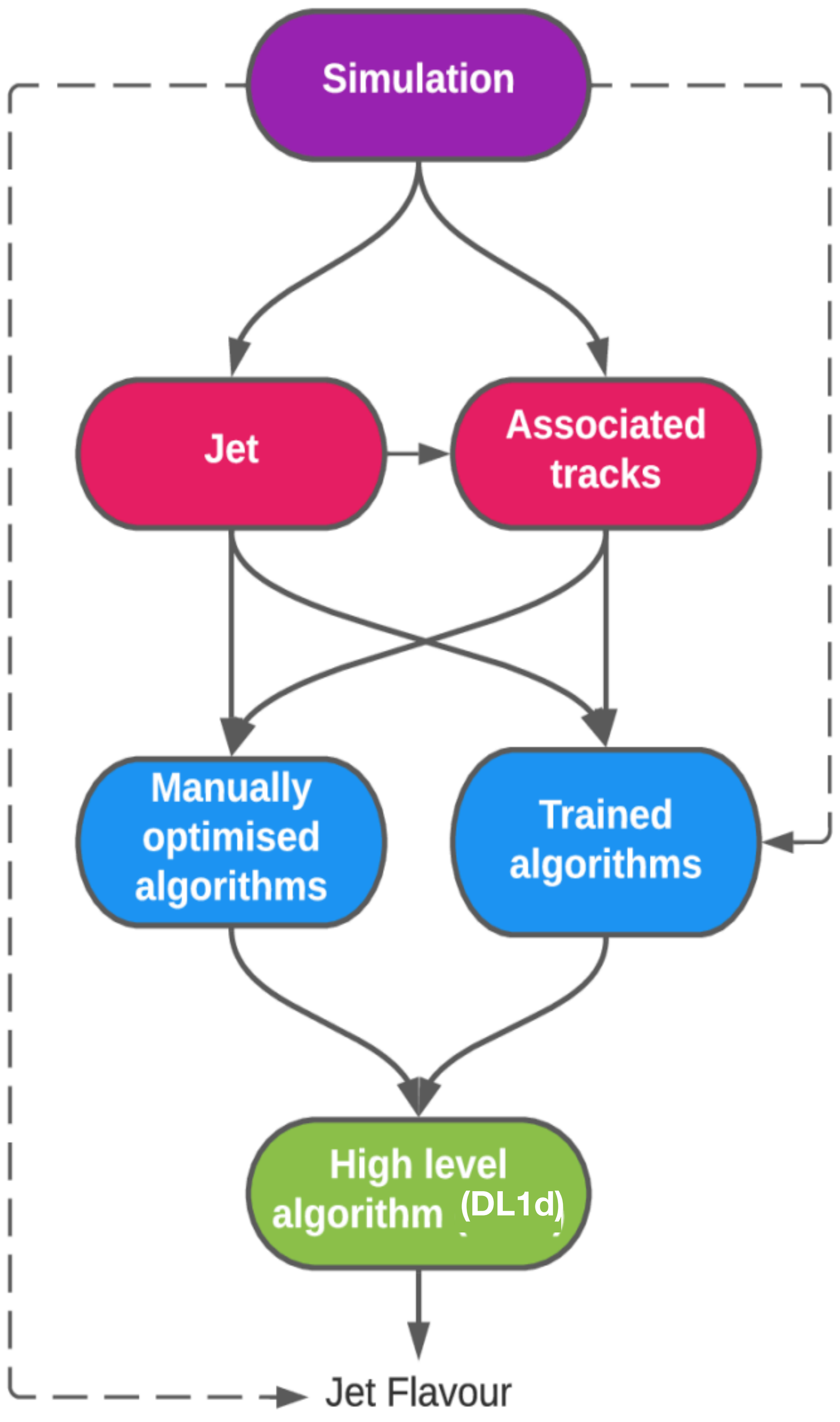
Low-level outputs are fed into **high level taggers**, ***MV2*** (BDTs) or ***DL1*** (NNs)

Output probability p_b, p_c, p_l

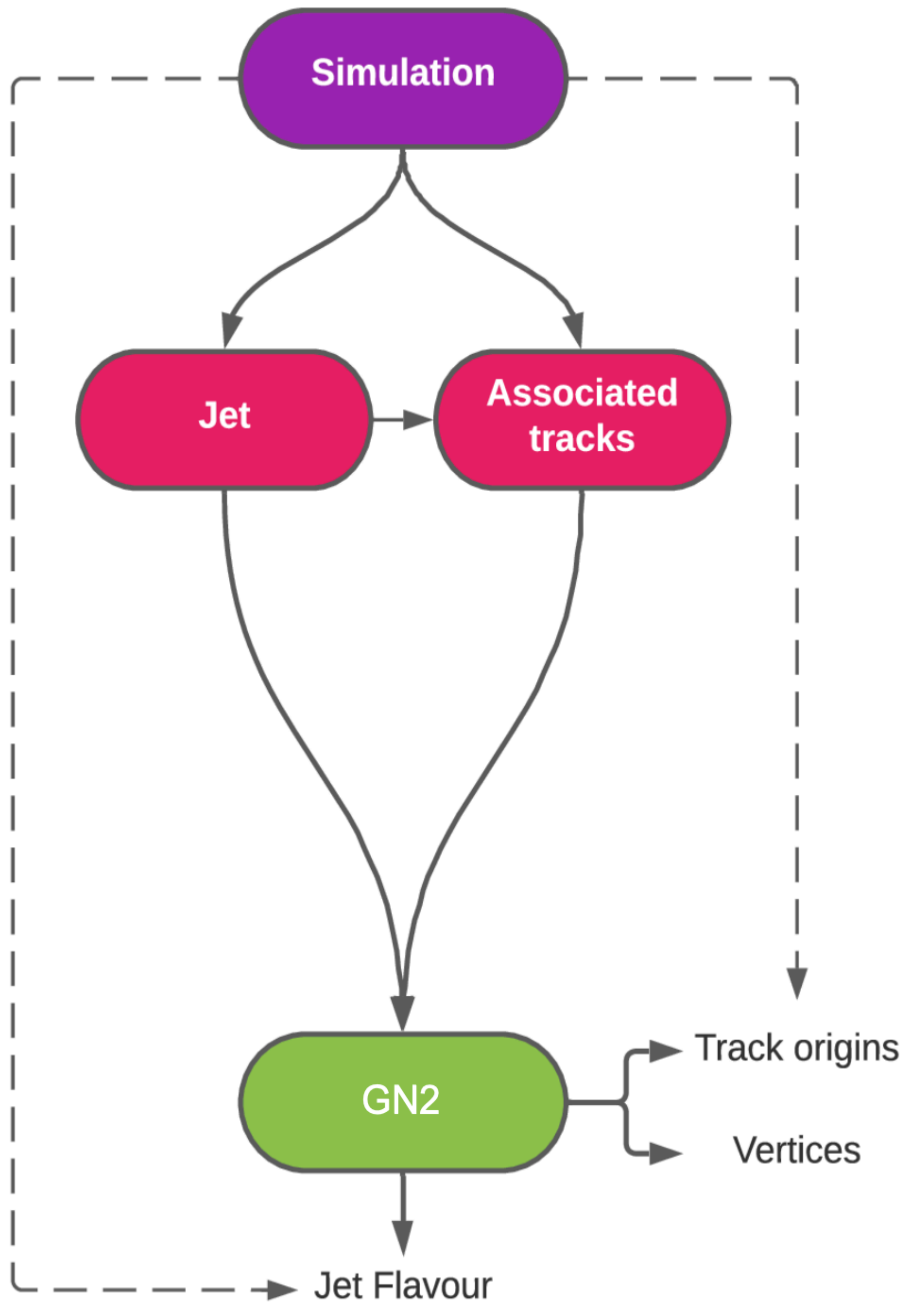
Challenges….

- Complexity in handling reconstructed tracks
- Dependence on Low-Level Tagger
- Tuning for different use-cases requires lots of single steps before final tagger

GN2 Flavour Tagging



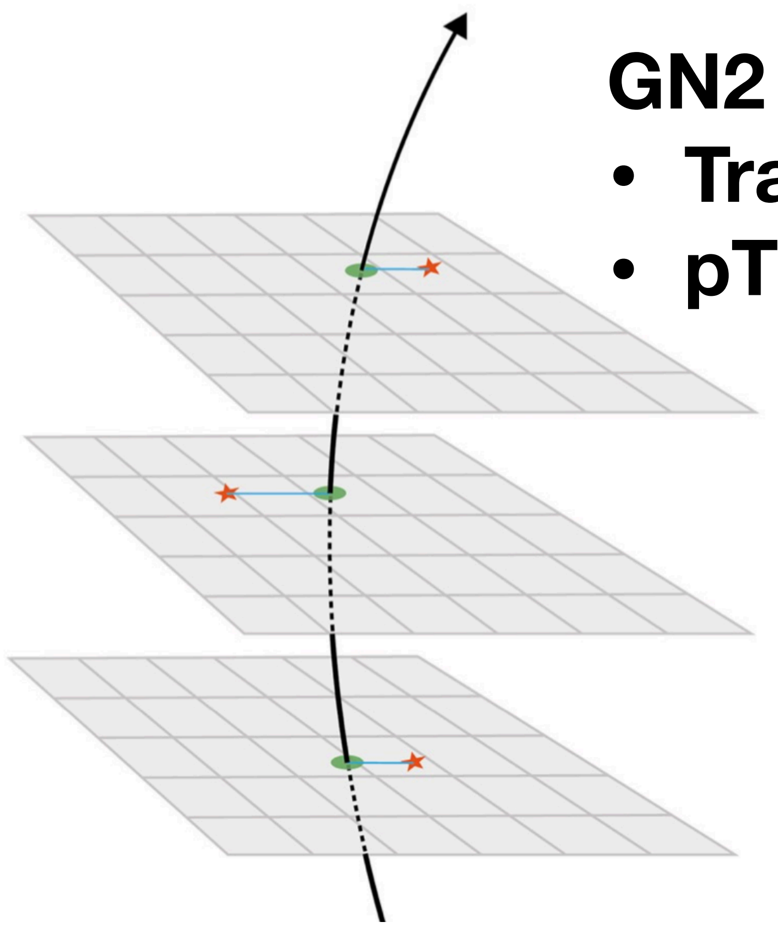
- GN1 [ATL-PHYS-PUB-2022-027](): All-in-one GNN-based (Inspired by J. Shlomi's work -[arXiv:2008.02831]())
- Use track information and jet kinematics directly ⟹ naturally adapts to variable **#unordered input tracks**
- Tasks include **jet flavour, vertexing, and track origin prediction**, trained simultaneously
  - **Auxiliary targets enhance interpretability**
- Easily optimised for diverse use cases and track/jet improvements.



GN2 is an upgraded version of GN1: all-in-one transformer network with significant state-of-Art performance enhancements

GN2 is based on **GN1** architecture with **Optimised training , Updated architecture, Increased training statistics**

GN2 architecture

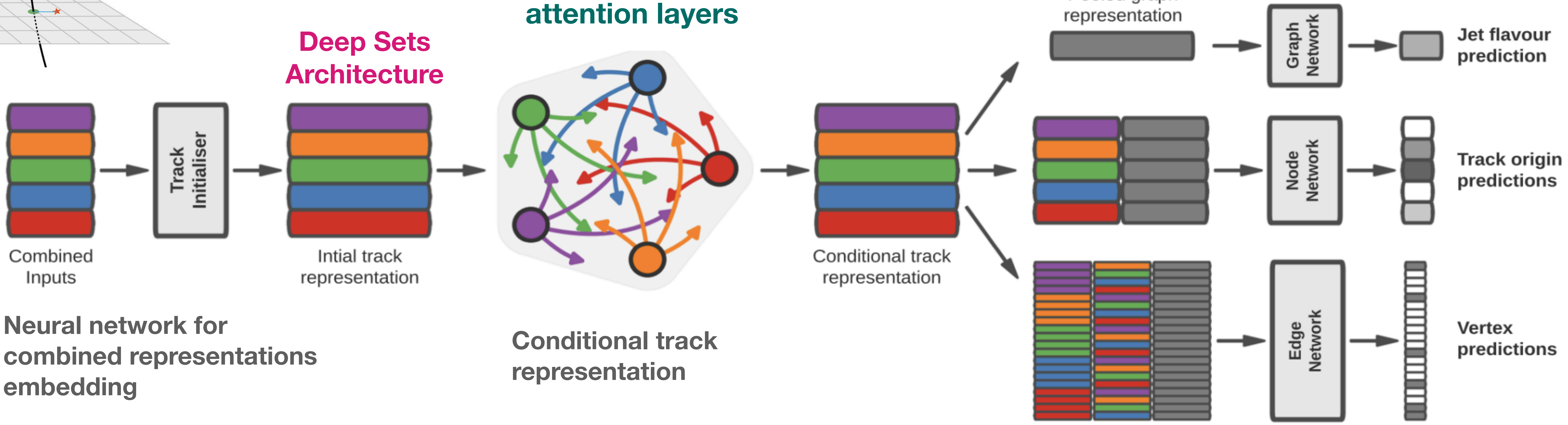


GN2 Inputs:
- **Tracks + jets variables**
- **pT & η resampled for each flavour**

Global Attention Pooling

Transformers, with multi-head attention layers

Deep Sets Architecture



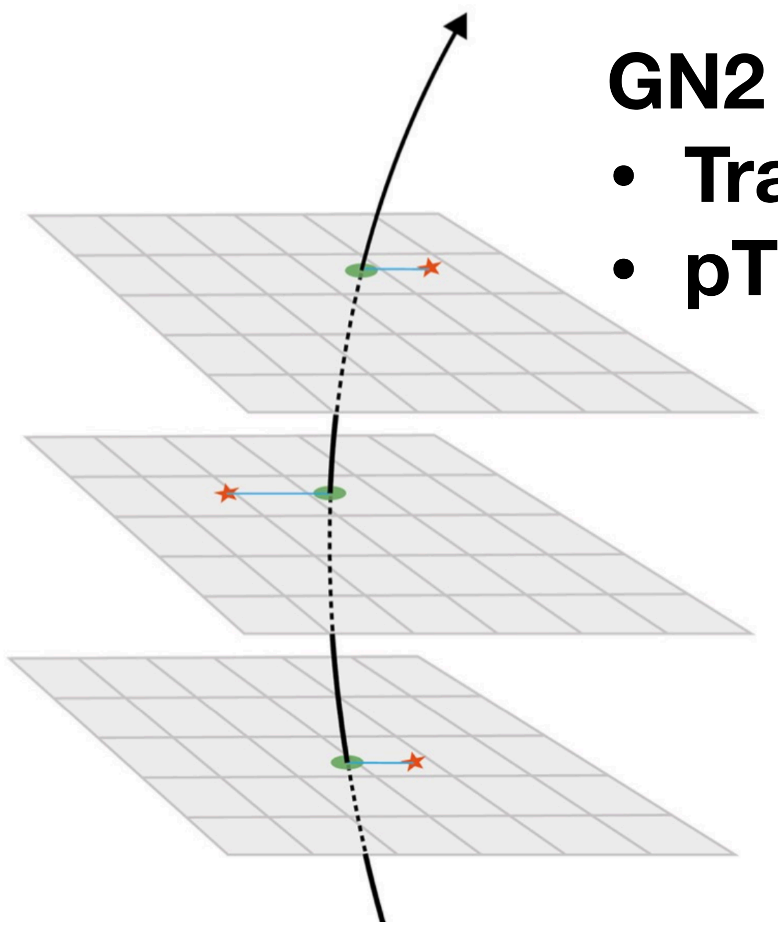
Neural network for combined representations embedding

Conditional track representation

Auxiliary Task Heads

ATL-PHYS-PUB-2022-027

GN2 architecture



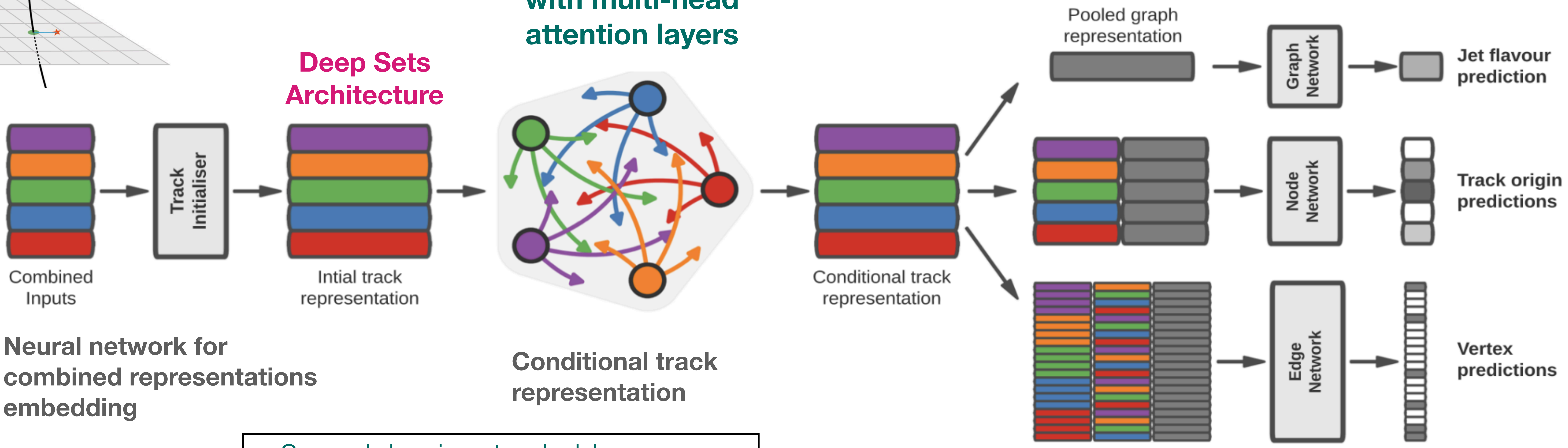
GN2 Inputs:
- **Tracks + jets variables**
- **pT & η resampled for each flavour**

Large Multimodal Multitask Transformer Model with over 2600k parameters (GN1: 800K, DL1d: 130k)

Global Attention Pooling

Transformers, with multi-head attention layers

Deep Sets Architecture



Neural network for combined representations embedding

Conditional track representation

- One-cycle learning rate scheduler
- LayerNorm + DropOut to stabilise training
- Large training dataset: 192M training jets (GN2)

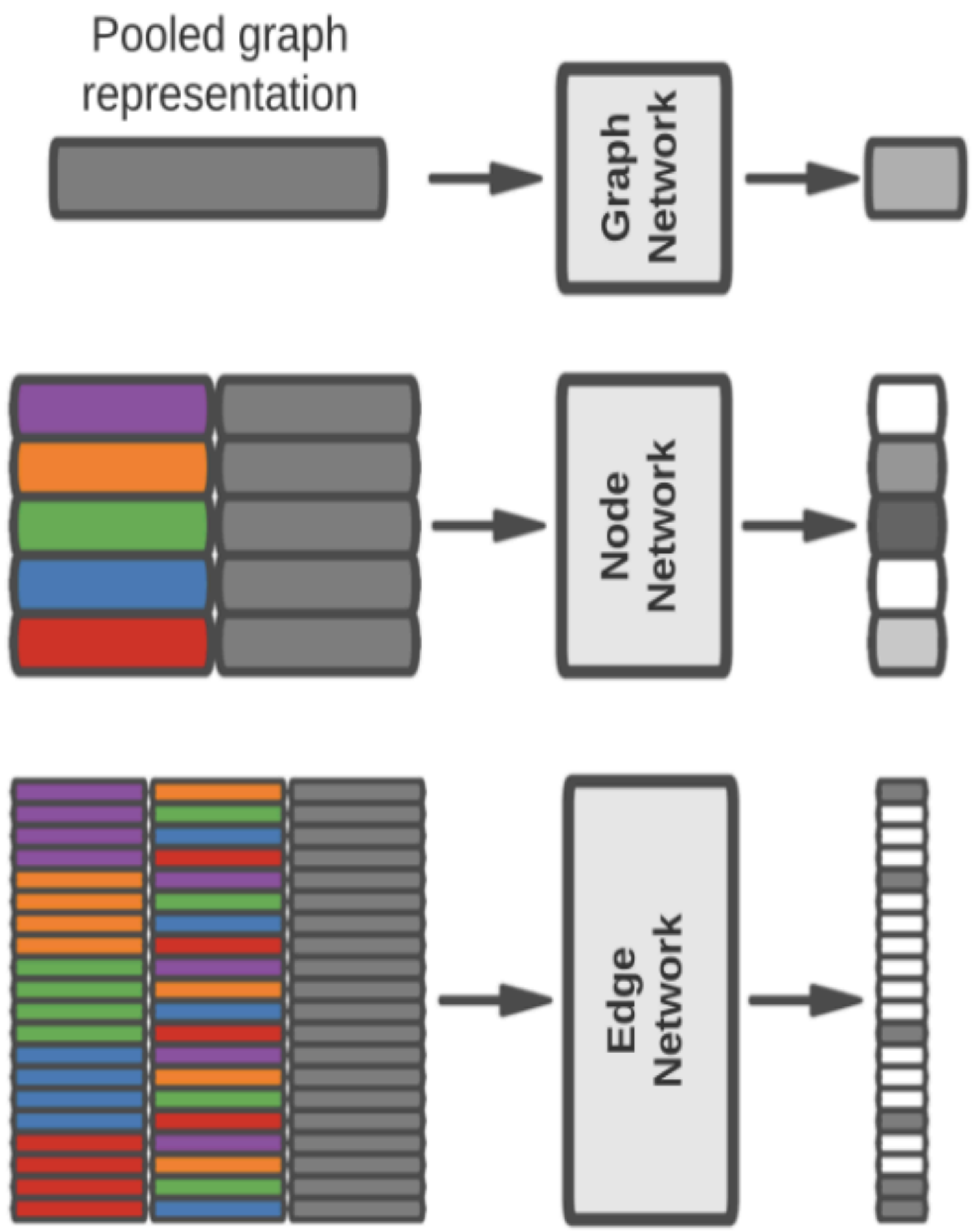
Auxiliary Task Heads

FTAG-2023-01

ATL-PHYS-PUB-2022-027

GN2 training performed using the Salt framework

GN2 tasks



Jet flavour prediction $\rightarrow \begin{pmatrix} p_b \\ p_c \\ p_u \\ p_\tau \end{pmatrix}$

NN output creates a b-tag discriminant

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

Track origin prediction

Classifies track originating from pileup, primary, B-/C-hadron decay etc.

Vertex prediction

2-tracks vertex origin: predict if track pair comes from same vertex

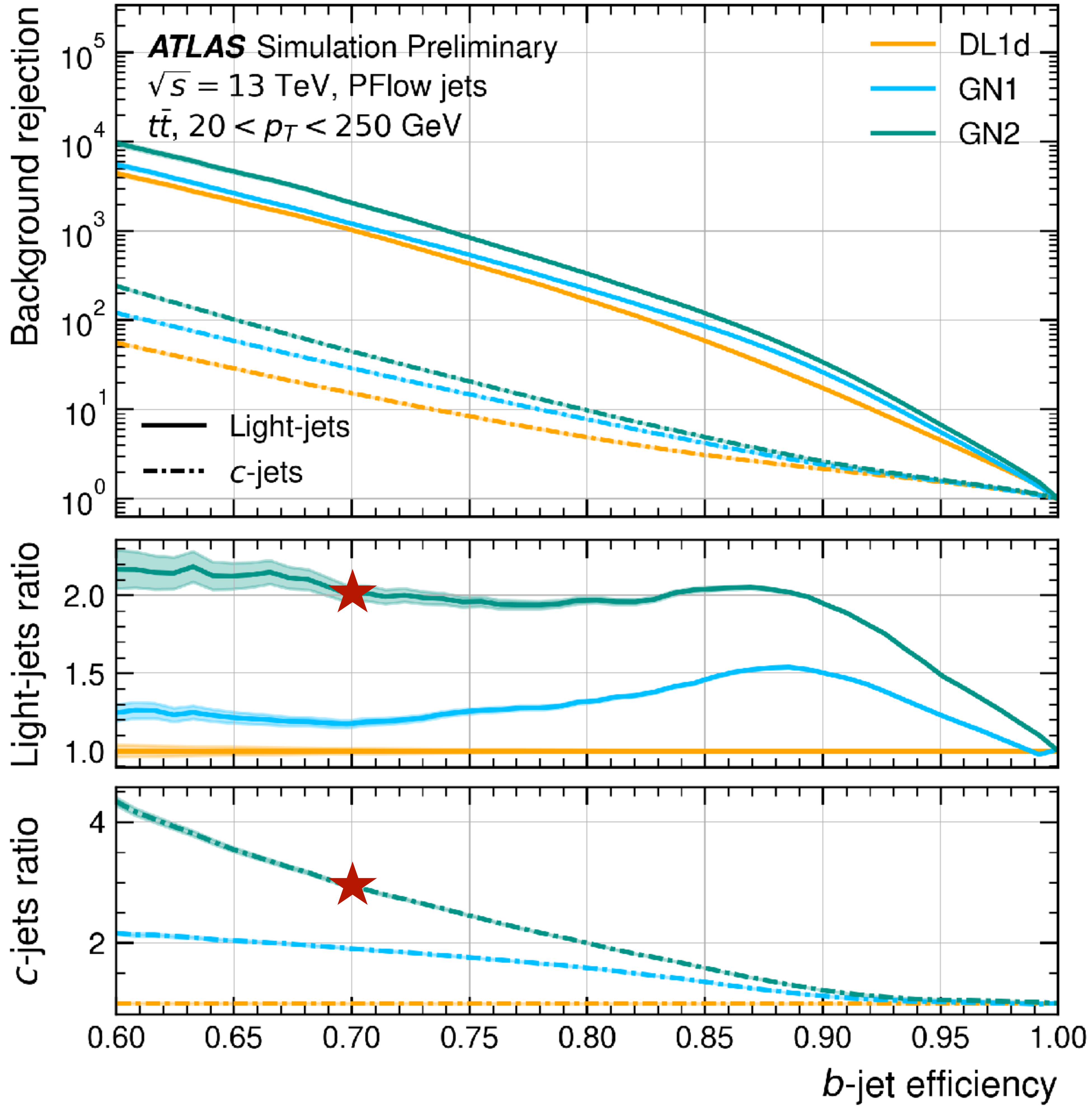
$$\mathcal{L}_{tot} = \mathcal{L}_{jet} + \alpha \mathcal{L}_{trk} + \beta \mathcal{L}_{vtx}$$

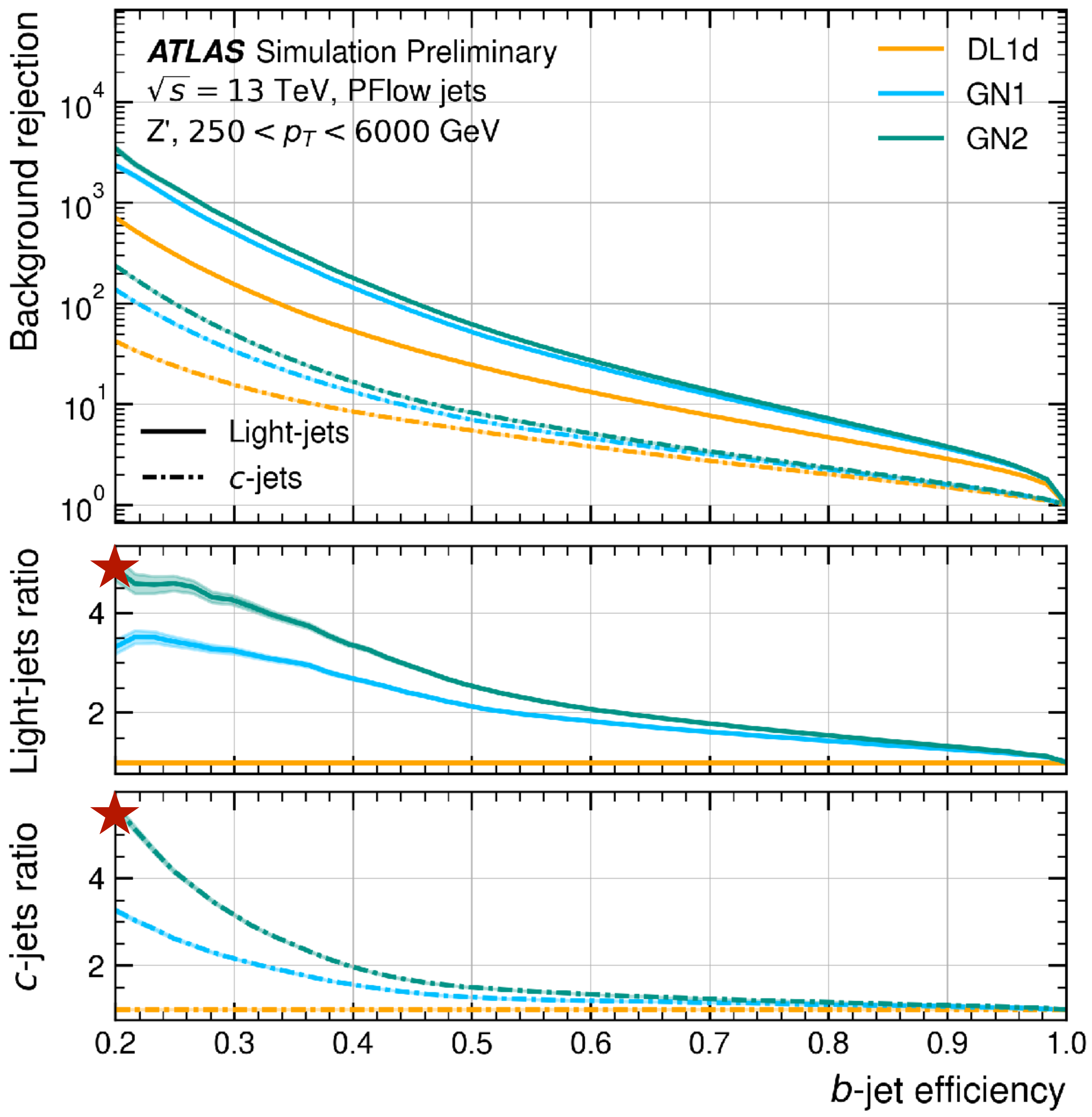
$\alpha = 0.5$ $\beta = 1.5$

NN weights optimised by gradient descent: $\nabla \mathcal{L}$

FTAG-2023-01

GN2 b-tagging performance





Model	fc
DL1d	0.018
GN1	0.05
GN2	0.1

FTAG-2023-01

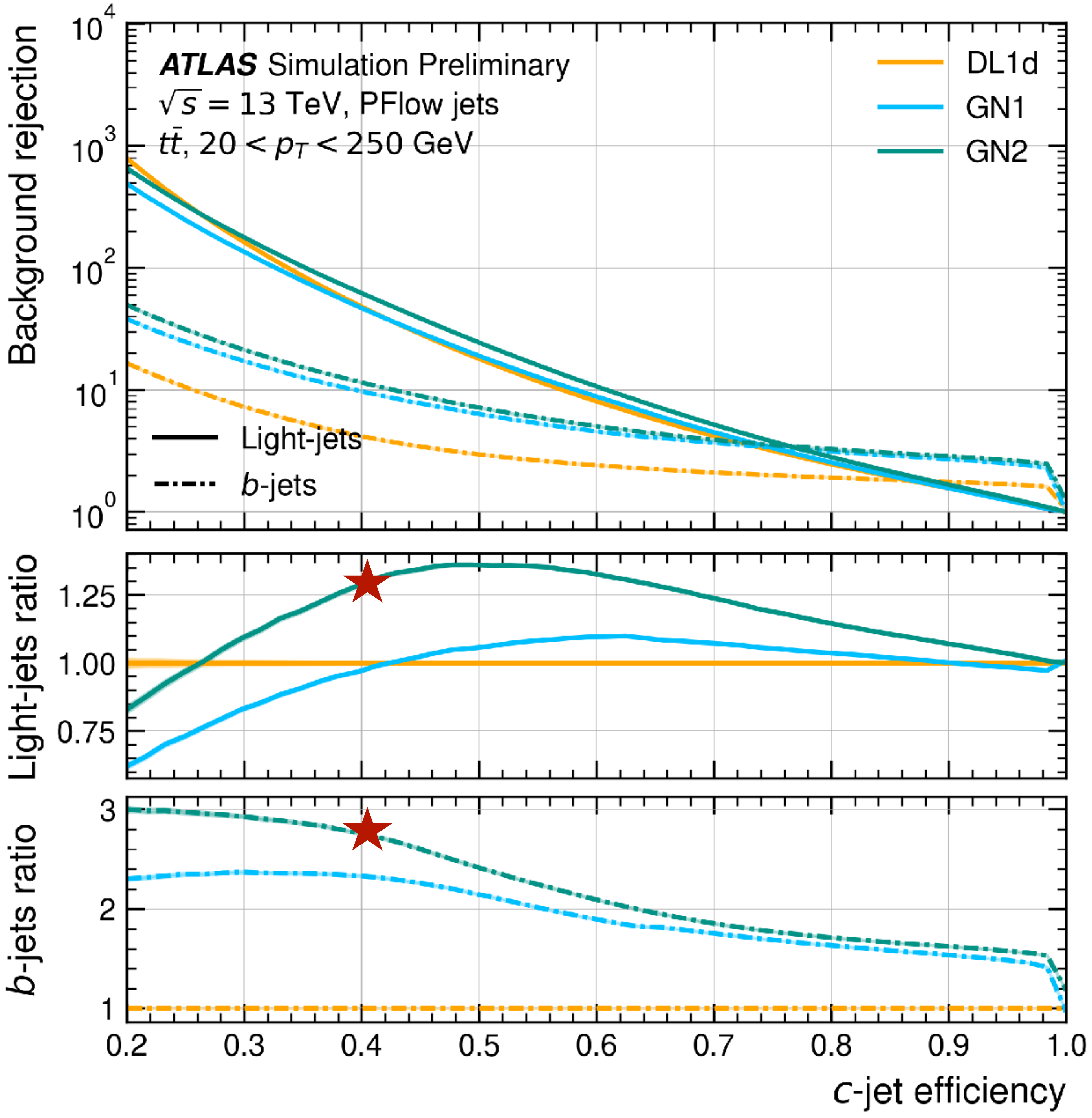
Perf at Low pT @70% b-jet eff

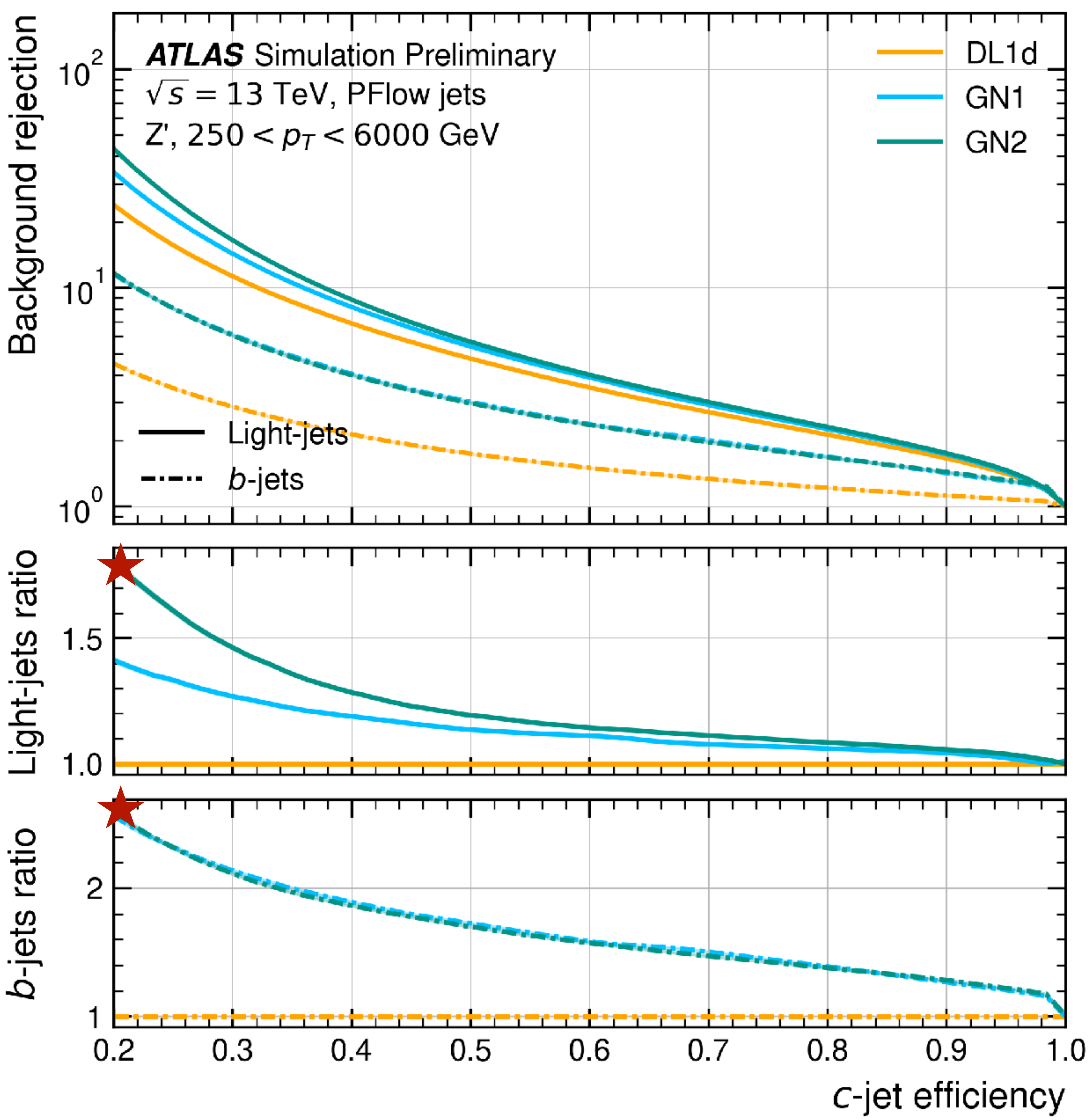
x2 light-jet rej, x2.8 c-jet rej

Perf at hight pT @20% b-jet eff

x4.8 light-jet rej, x5.5 c-jet rej

GN2 c-tagging performance





Model	fc
DL1d	0.018
GN1	0.05
GN2	0.1

FTAG-2023-01

Perf at low pT @40% c-jet eff

x1.3 light-jet rej, x2.7 c-jet rej

Perf at hight pT @20% b-jet eff

x1.8 Light-jet rej, x2.6 c-jet rej

Building GN2 ecosystem….

S. Van Stroud

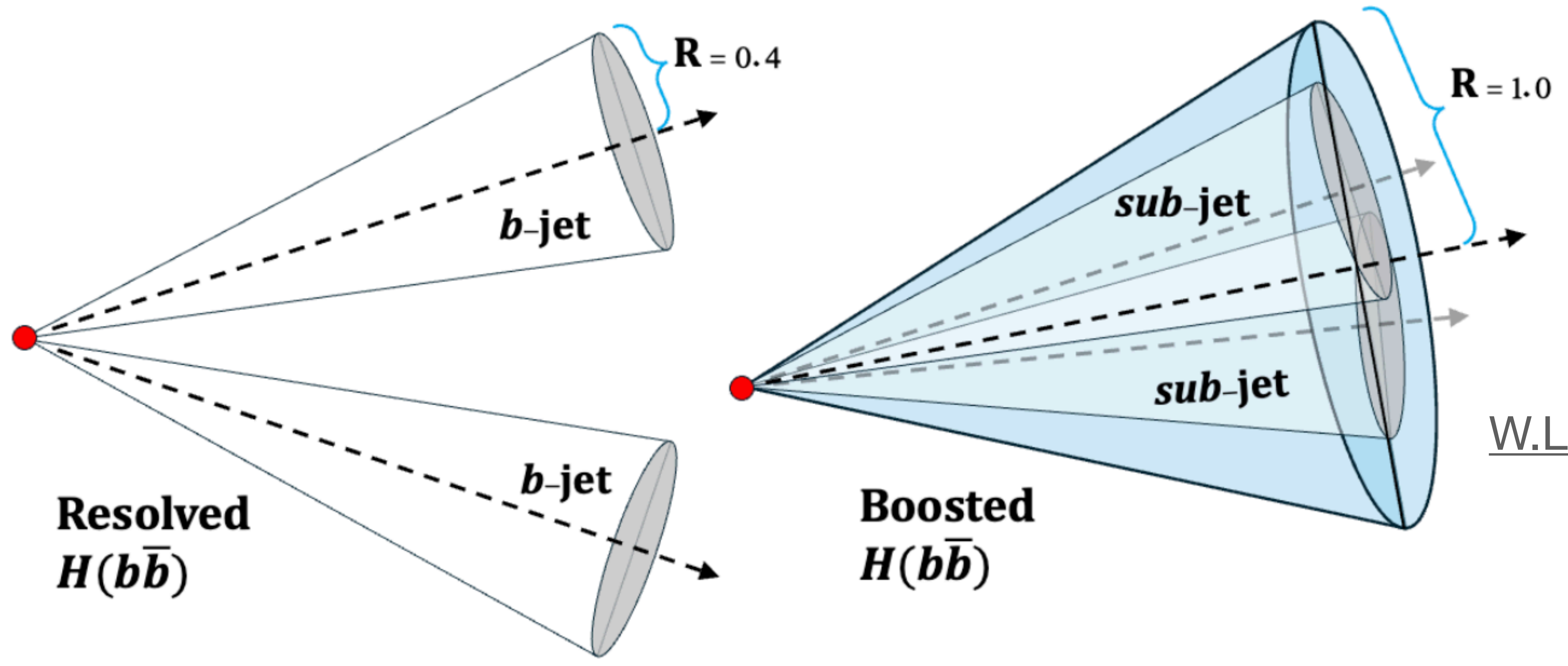
Taking an ecosystem-wide view brings many benefits:

- Key software frameworks used for GN2: Training Data Dumper (**TDD**), **Umami Pre-processing (UPP), Puma-HEP, SALT**
- Successful application in various CP, analyses and HL-LHC forecast
- Synergy with other CP groups: Access to FTAG tools becomes easier.
- Increased collaboration
- Reduced barriers: Clear **documentation**

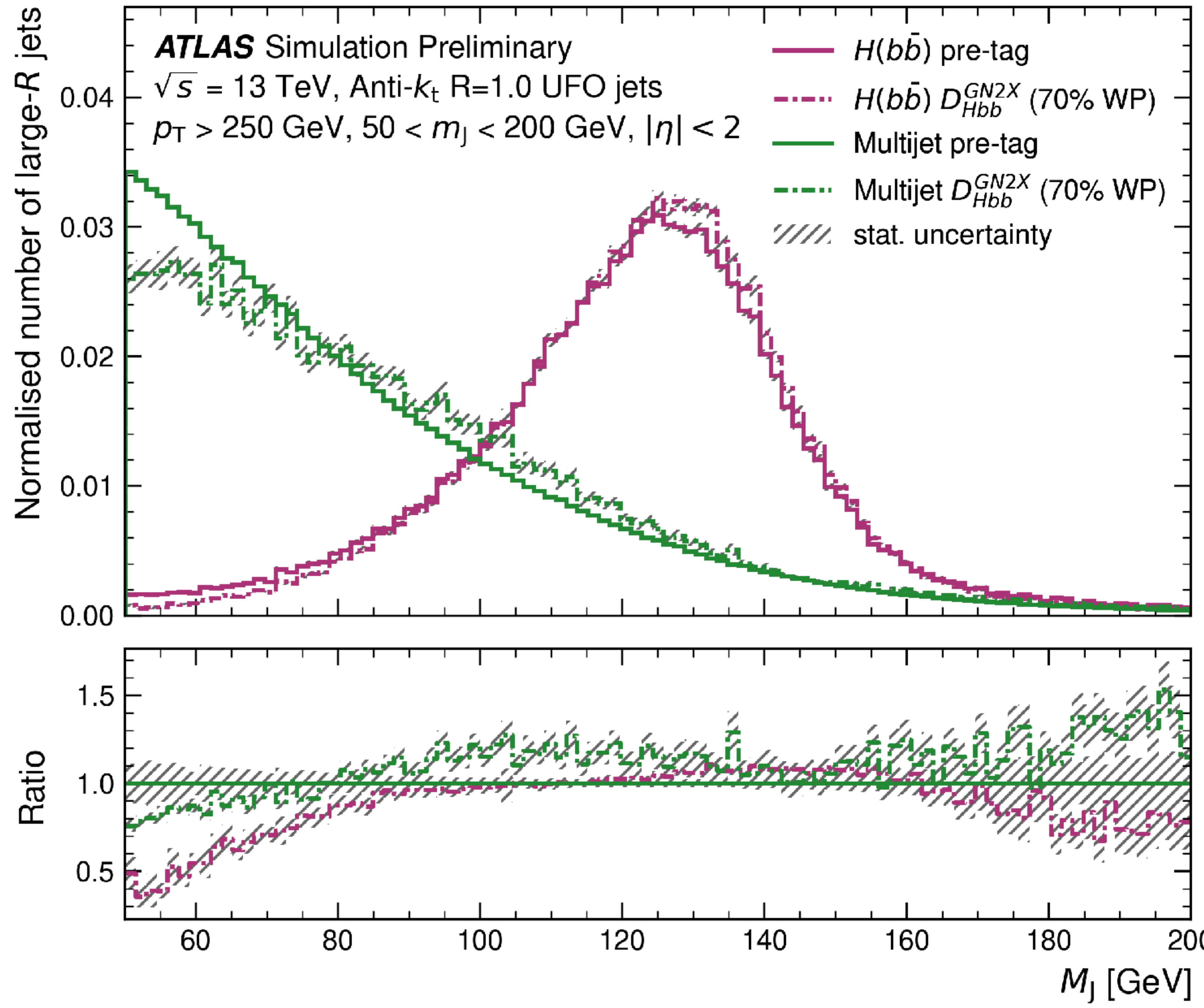
- Boosted Xbb tagging: [ATL-PHYS-PUB-2023-021]: GN2X Similar architecture as GN2 with four output classes (Hbb, Hcc, top, or multijet)

Boosted Higgs Tagging

- GN2X is a transformer based Xbb tagger that replaces the previous subjet based model used within ATLAS
- Trained to discriminate between boosted H → bb, H → cc, hadronic top and QCD jets
- GN2X + Subjets: kinematic + b-tagging info VR subjets, where the subjets are tagged using the GN2 tagger
- GN2X + Flow uses UFO constituents which includes the use of charged and neutral calorimeter information.



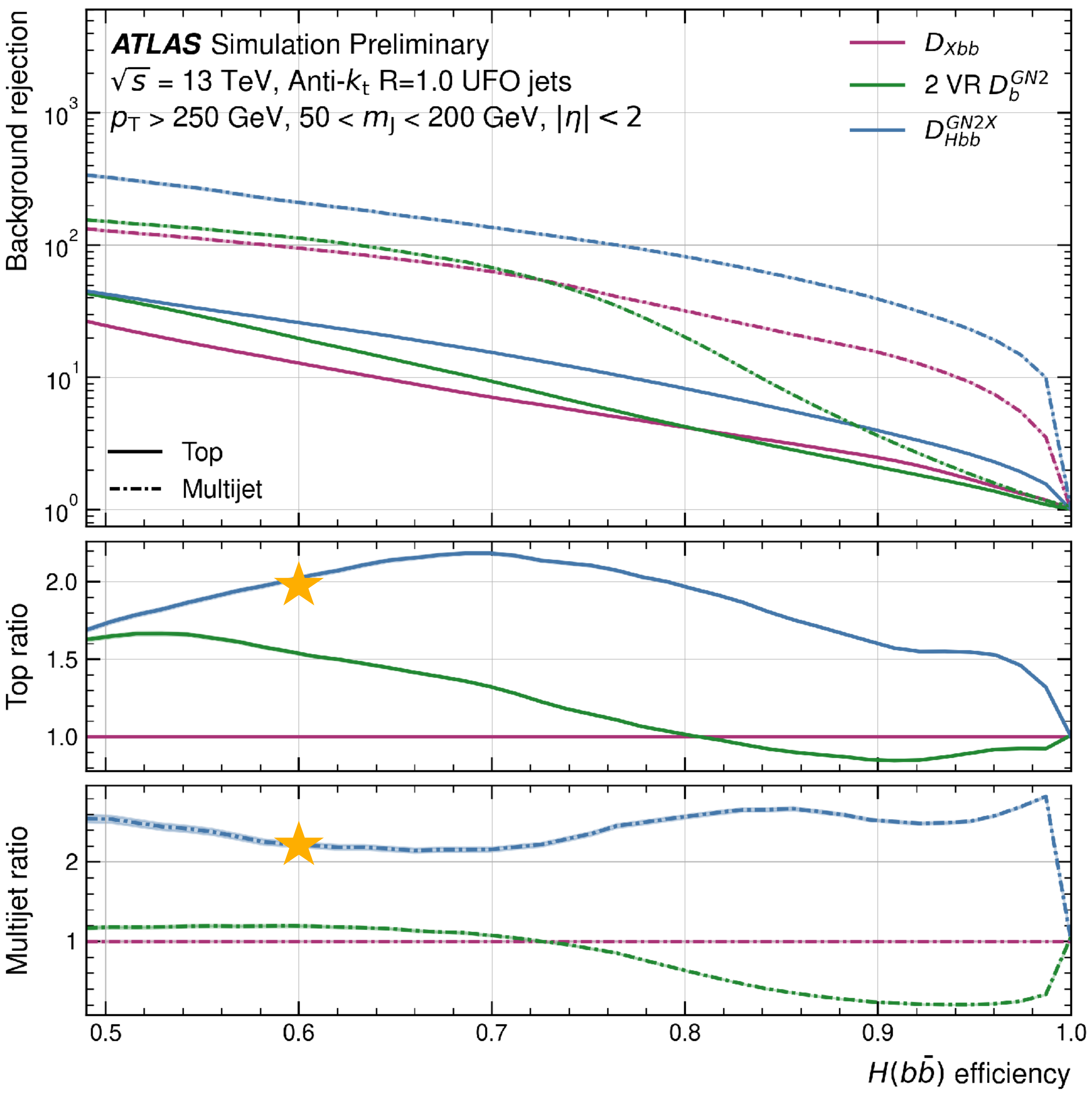
W.Leinonen

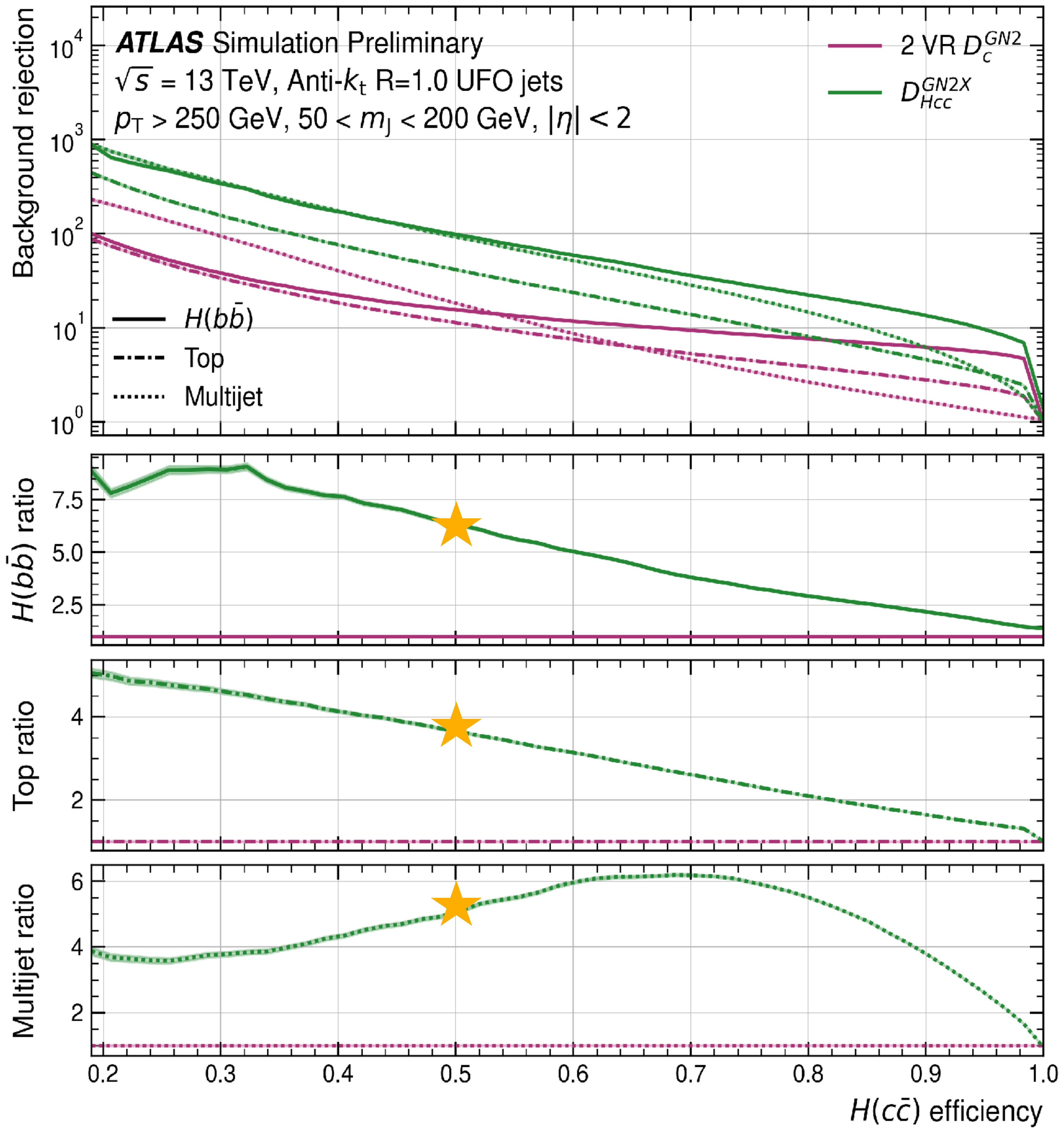


- GN2X is trained on mass decorrelated Higgs sample
- Modified Higgs sample with increased decay width to reduce background mass sculpting $\implies$ keeps sculpting within 20% in bulk of distribution

ATL-PHYS-PUB-2023-021

Boosted Higgs Tagging





ATL-PHYS-PUB-2023-021

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

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

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

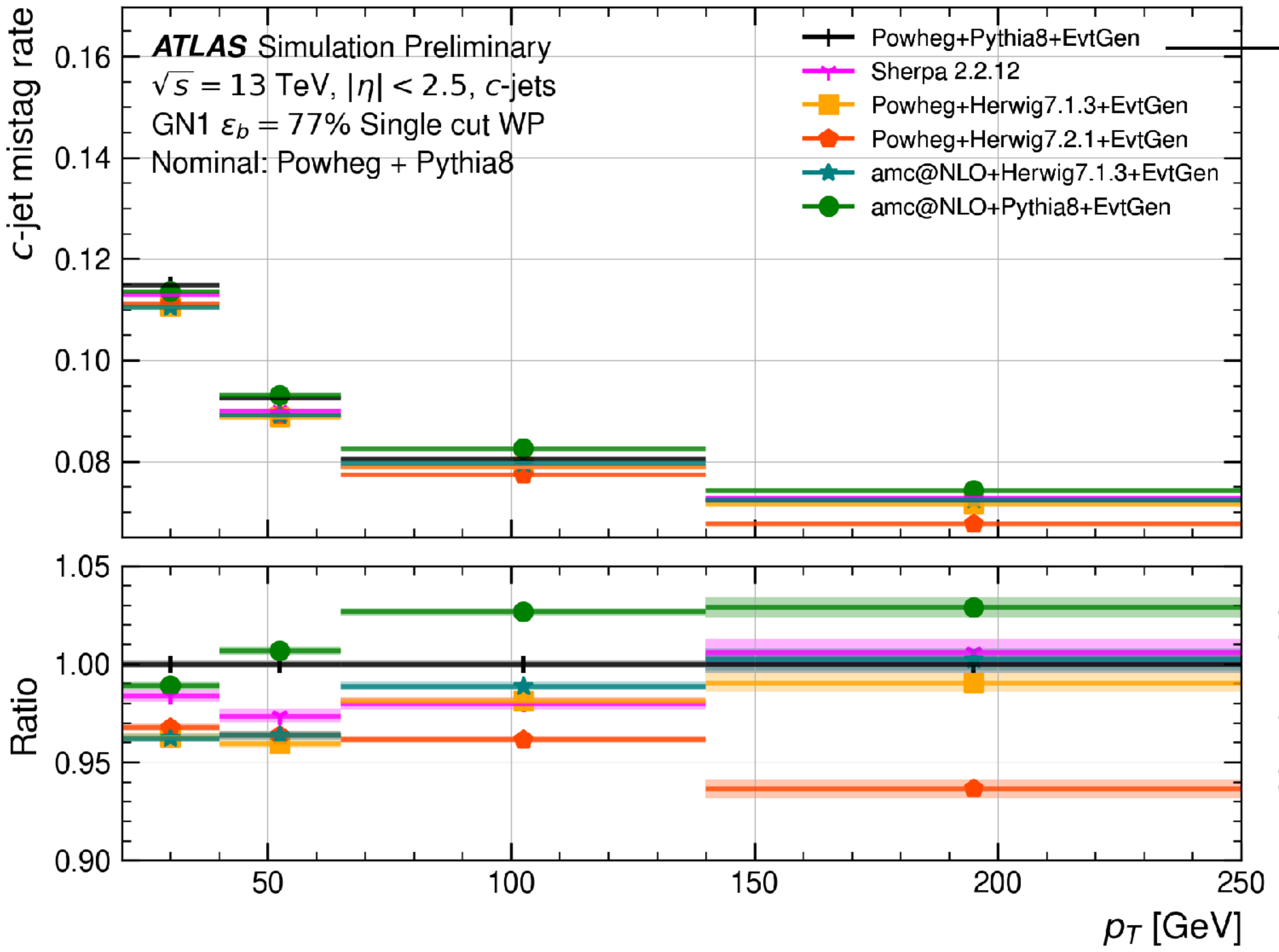
$$f_{top} = 0.25, f_{Hbb} = 0.3$$

@60% signal efficiency, GN2X achieves more than double the top and QCD rej

@50% signal eff, 3x top jet rej, 5x multij-et rej and a 6x improvement in the $H(b\bar{b})$ rej

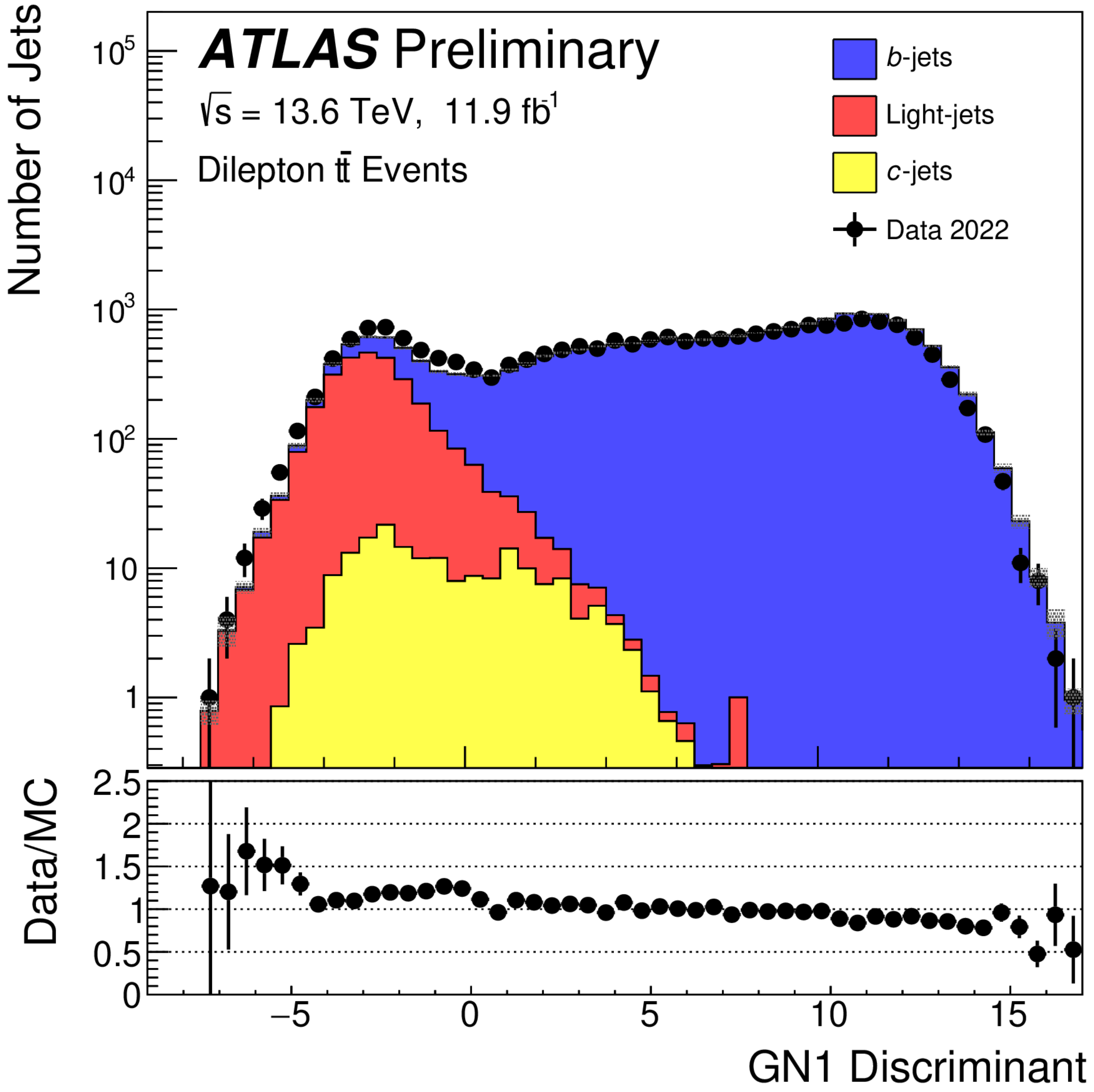
Calibration efforts underway, future boosted Higgs searches with hadronic decays greatly benefit from GN2-based model

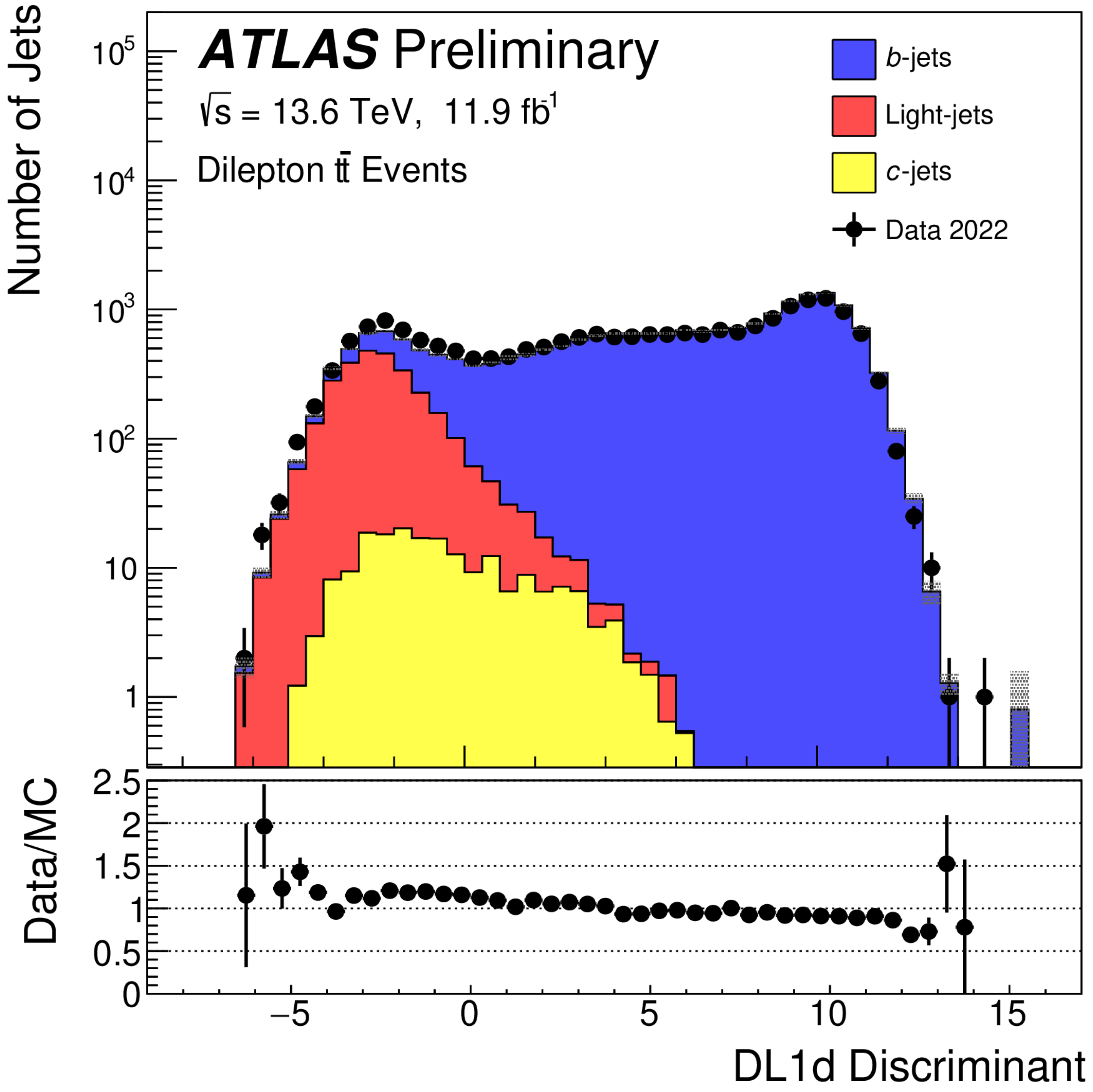
MC/Data comparison



Simulated sample trained on

- Overall generator dependence ~ O(3-6%)





- Ongoing calibration for GN2

- Overall good agreement similar to DL1d

FTAG-2023-01

Conclusion

- New GN2 algorithms with transformer architecture provides significant boost in rejection power over DL1d: @70% **GN2 >2x rejection**
  - GN2 soon be the recommended ATLAS flavour tagger
  - The GN2 architecture can be re-used in many places: Trigger, Upgrade, Xbb, etc
- Developments for GN3 ongoing: targeted for **ATLAS / CMS FTAG workshop 2024**
- Latest FTAG CMS results ICHEP24

New tagger journey…

2021 Early proof of concept studies [1] [2]

2022 Use in the trigger [1]

2022 GN1 Pub Note [ATL-PHYS-PUB-2022-027]

2023 Xbb GN2x Pub Note [ATL-PHYS-PUB-2023-021]

2023 GN2 CERN seminar

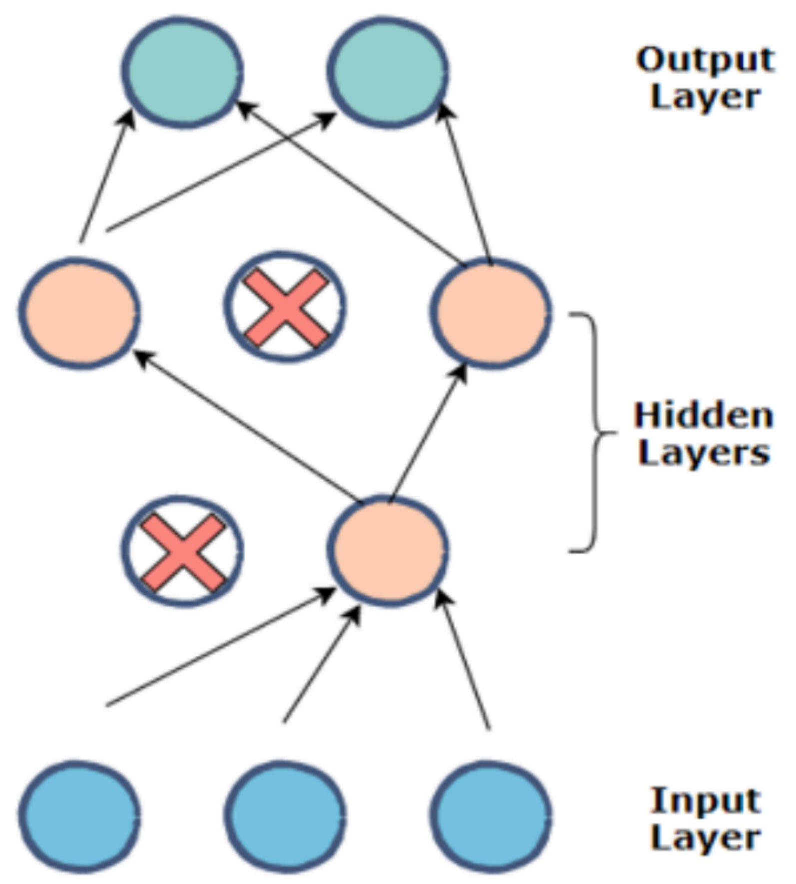
Soon!

GN2: Recommended ATLAS flavour tagger

Thank you!

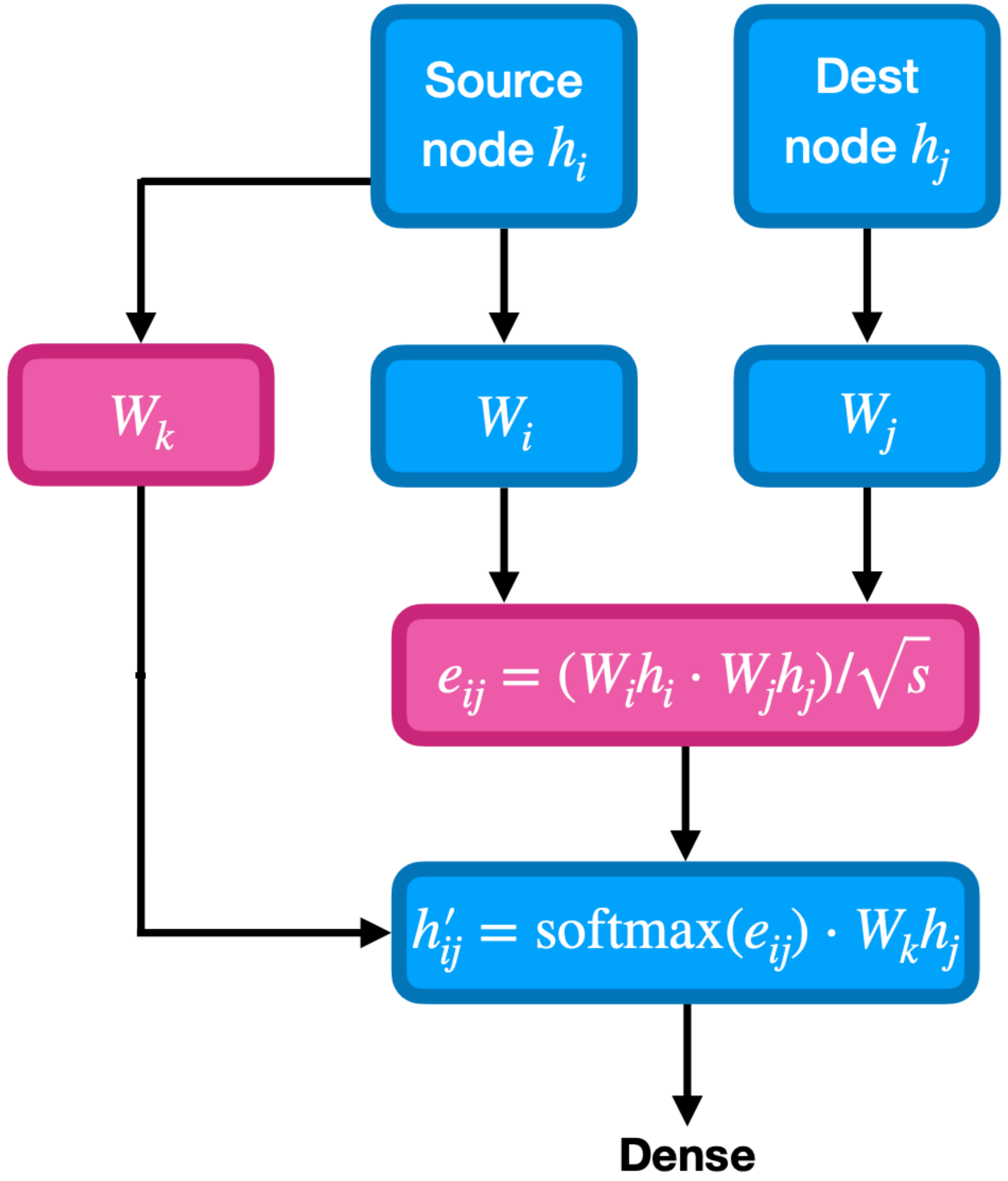
Backup

GN2 architecture



drop out networks

GN2 follows more closely the **transformer** architecture [1706.03762]



Data Science Seminar(Talk by Samuel)

GN2 inputs

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 tasks

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

Model	fc
DL1d	0.0018
GN1	0.05
GN2	0.1

GN2X

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)
$\mathrm{d}\eta$	Pseudorapidity of track relative to the large-R jet η
$\mathrm{d}\phi$	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
$\mathrm{d}\eta$	Pseudorapidity of subjet relative to the large-R jet η
$\mathrm{d}\phi$	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

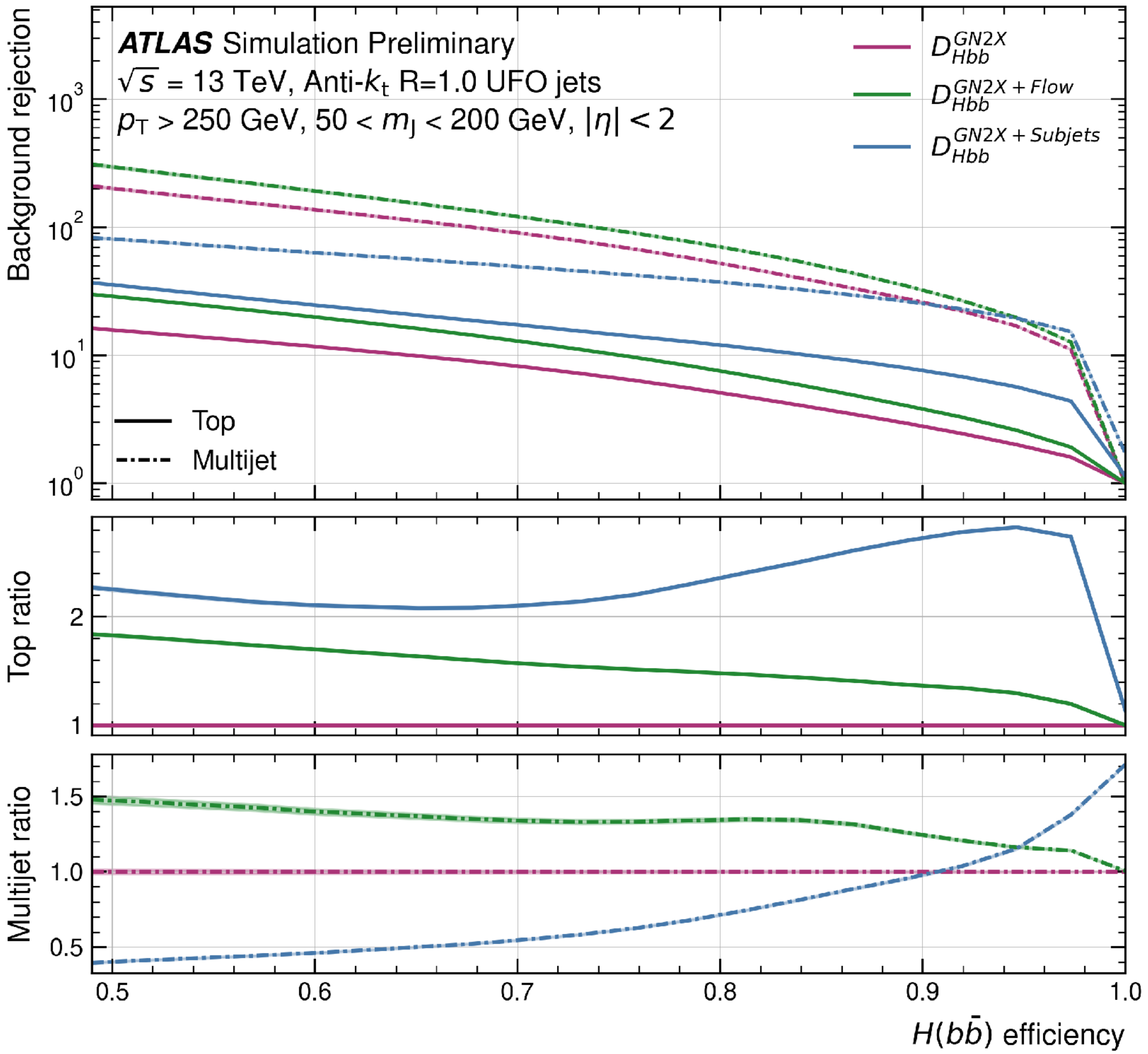
Flow Input	**Description** (Used only in GN2X + Flow)
p_{T}	Transverse momentum of flow constituent
energy	Energy of flow constituent
$\mathrm{d}\eta$	Pseudorapidity of flow constituent relative to the large-R jet η
$\mathrm{d}\phi$	Azimuthal angle of flow constituent relative to the large-R jet ϕ

GN2X

Table 3: Track selection requirements, where d_0 is the transverse impact parameter (IP) of the track, z_0 is the longitudinal IP with respect to the primary vertex and θ is the track polar angle. Shared hits are hits used in the reconstruction of multiple tracks which have not been classified as split by the cluster-splitting neural networks [36]. A hole is a missing hit, where one is expected, on a layer between two other hits on a track.

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

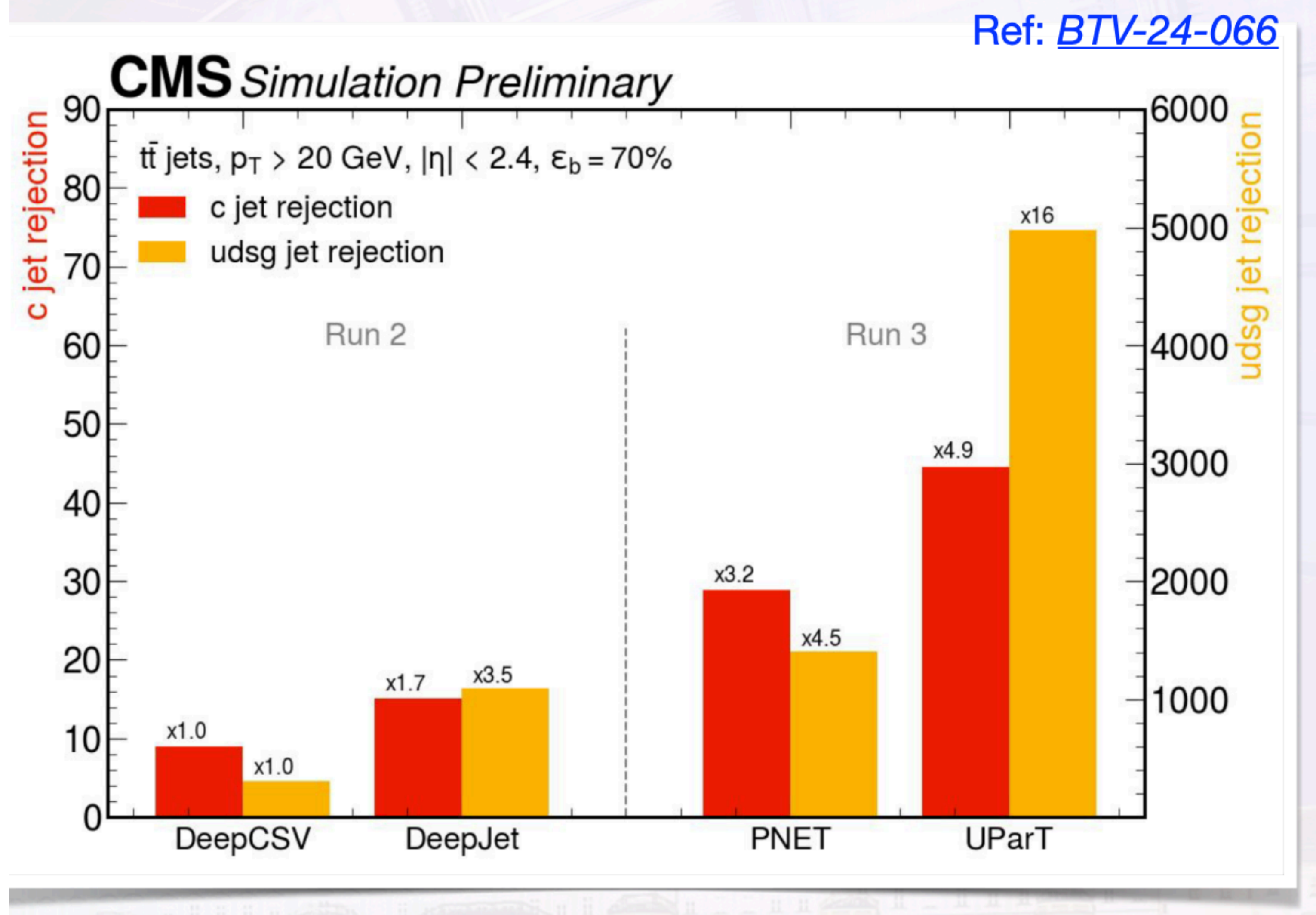
GN2X



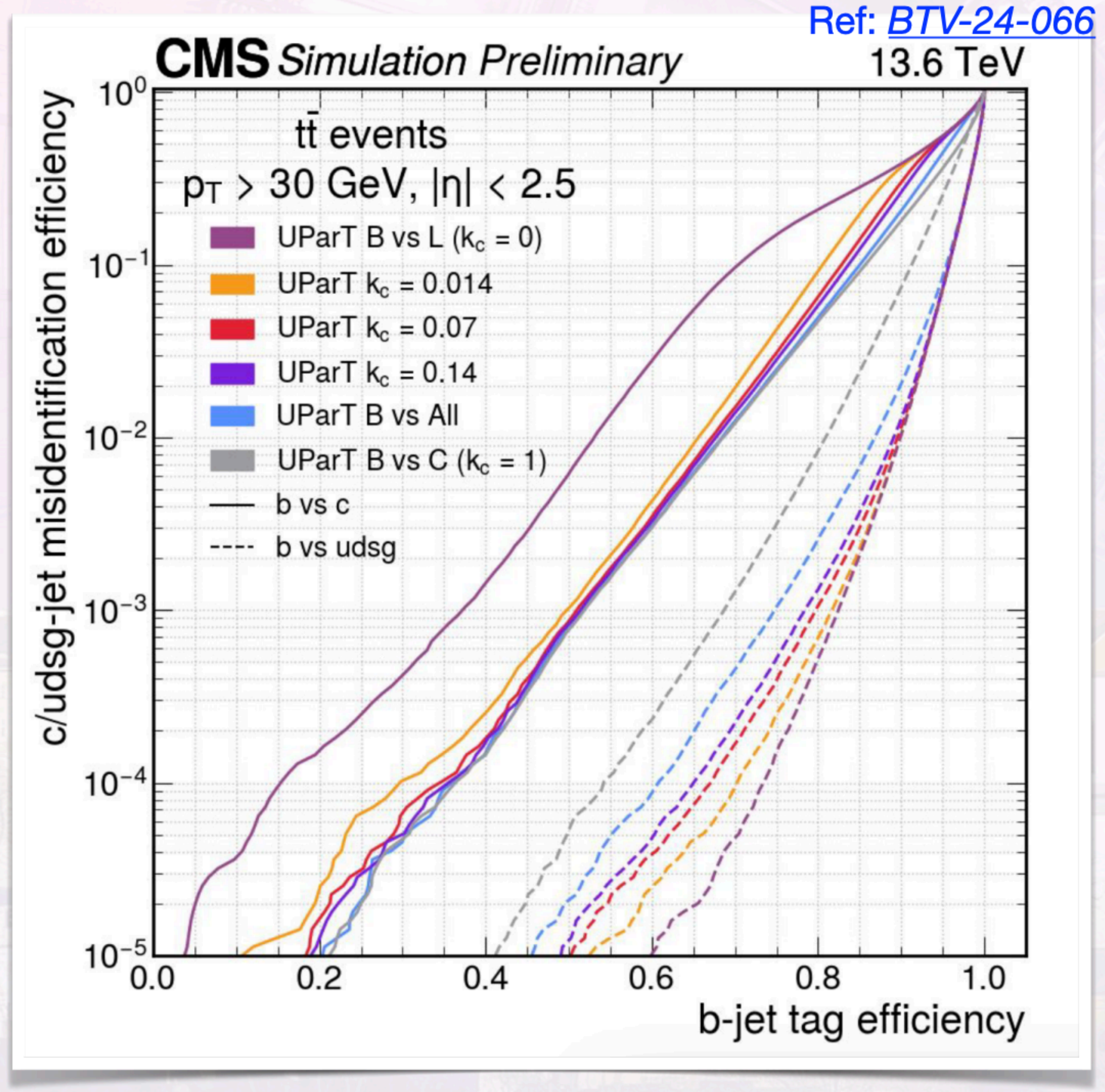
ATLAS Simulation Preliminary
$\sqrt{s}$ = 13 TeV, Anti-k_t R=1.0 UFO jets
p_T > 250 GeV, 50 < m_J < 200 GeV, |η| < 2
D_{Hbb}^{GN2X}
$D_{Hbb}^{GN2X+Flow}$
$D_{Hbb}^{GN2X+Subjets}$
Top
Multijet
Background rejection
Top ratio
Multijet ratio
$H(b\bar{b})$ efficiency

CMS

ICHEP24

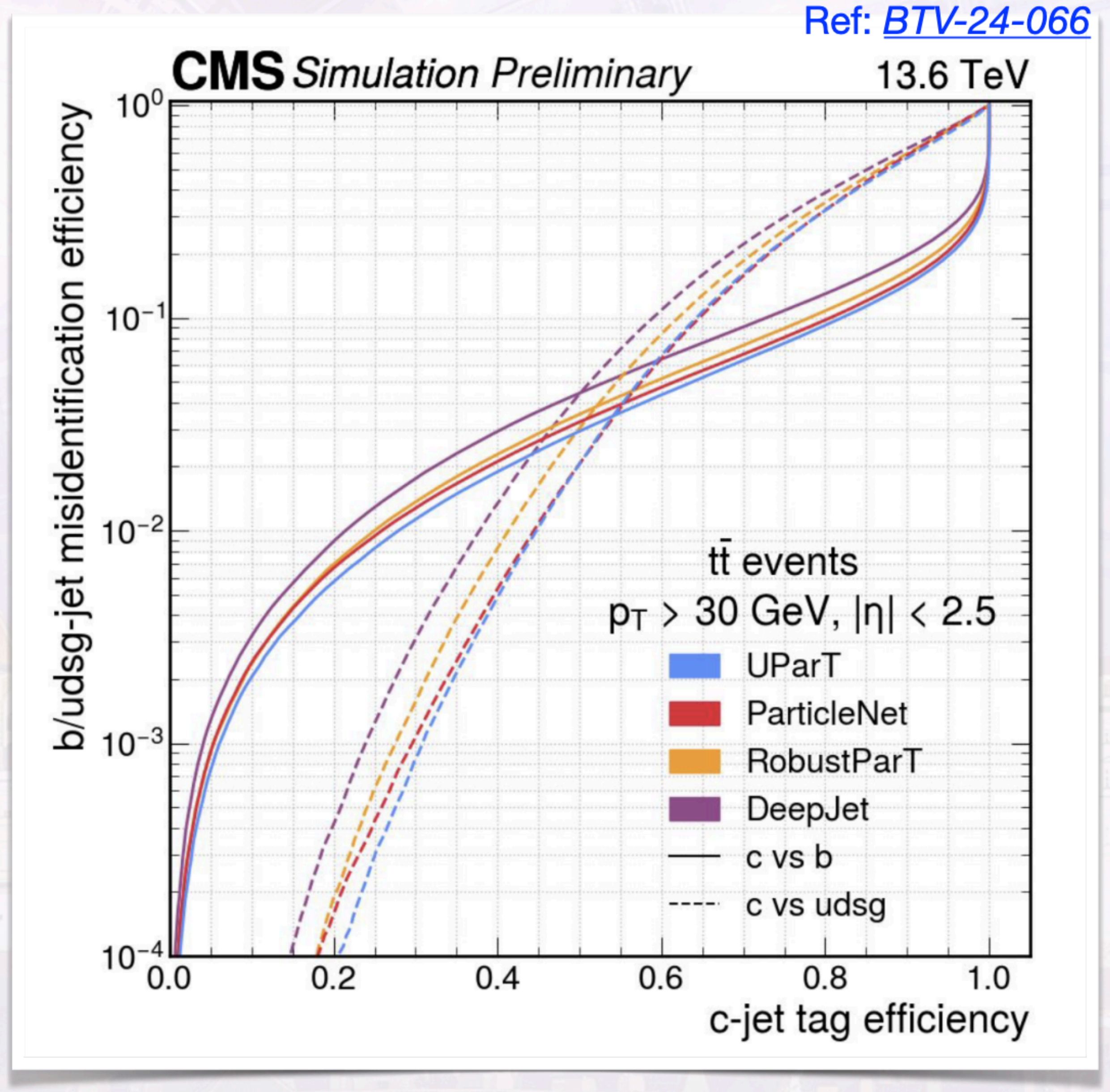


b-tagging



Significant improvement in b-tagging efficiency!

c-tagging



Improvement in c-tagging efficiency and c vs b discrimination

Extension of ParticleTransformer

Extended class: extending from b and c jet identification to s and hadronic tau (one per final state) identification

Extended regression: simultaneous flavor aware jet energy and resolution regression • Input variable distortion: • Reduce the observed differences prior to any calibration • Improve robustness of the classifier against injected mismodelings ➡Distortions of UParT: Preserving the Particle Cloud representation and the feature importance mapping