b-jet tagging in pp collisions using graph neural network with the ALICE experiment

Changhwan Choi[*1], Sanghoon Lim[1]
for the ALICE collaboration
1 Nuclear Physics Laboratory, Department of Physics, Pusan National University

• Introduction

Jets in heavy-ion collisions

Probe parton kinematics and fragmentation, interactions with the Quark-Gluon Plasma (QGP)

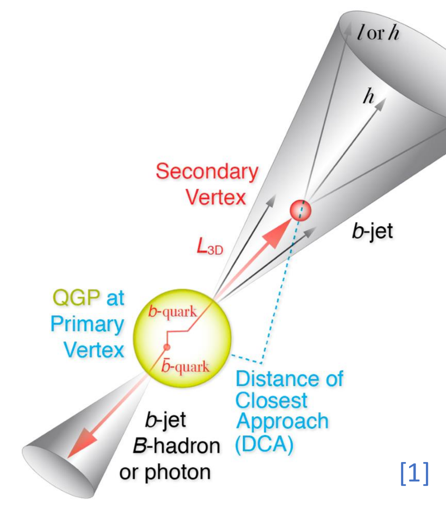
Beauty quarks in heavy-ion collisions

Due to their large mass,
1. Calculable production rates in pQCD
2. Predominant production from initial hard scatterings

b-jets in heavy-ion collisions

Kinematics of beauty quarks, probes of QGP

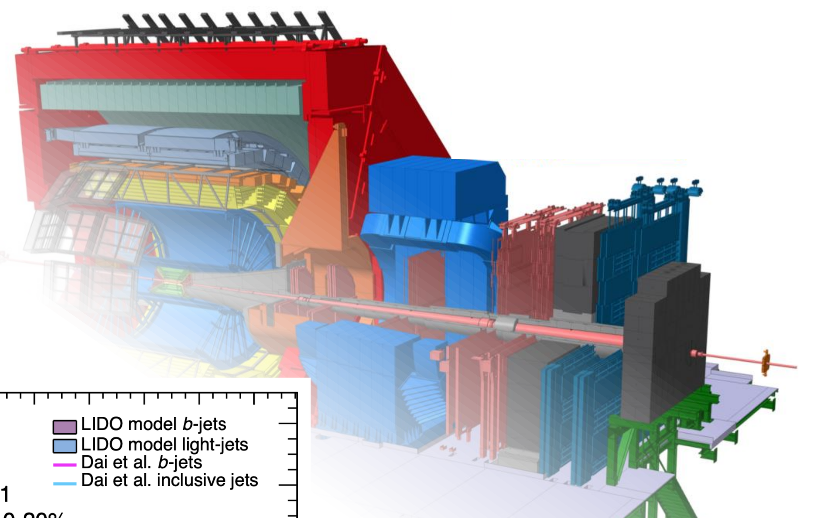
Especially in low-p_T region,
(comparable to beauty quark mass scale, 4.2 GeV/c^2)
the dominant partonic energy loss is due to collisions with QGP quasi-particles, rather than gluon radiation.
→ Excellent probes for studying QGP

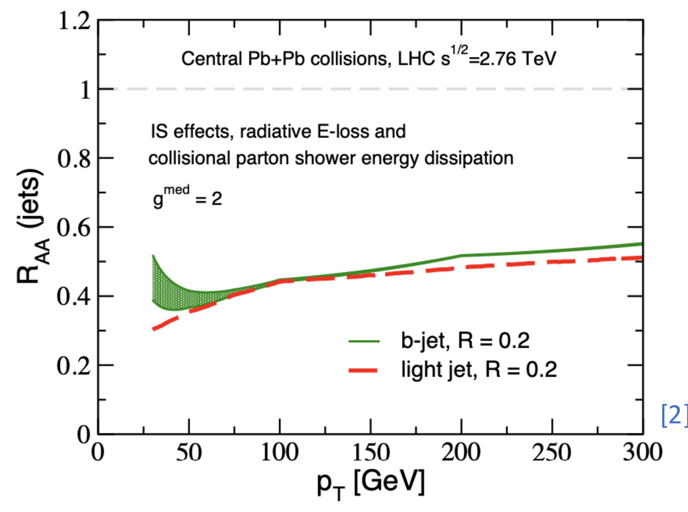


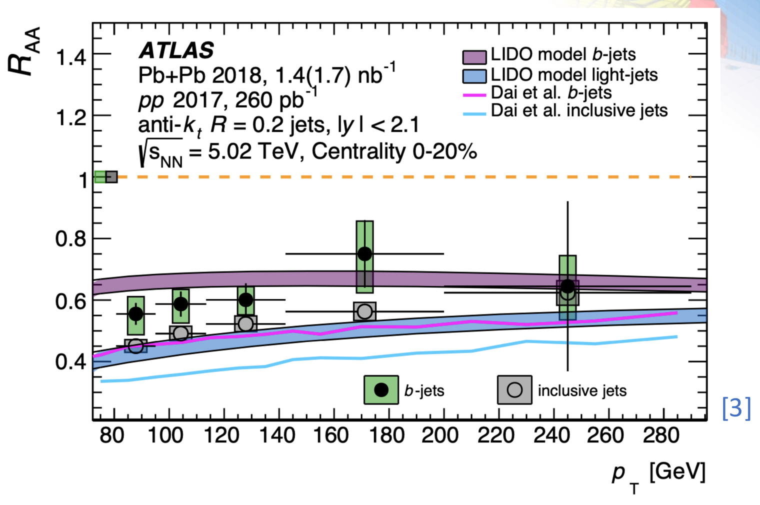
b-jet tagging using GNN in ALICE

Excellent capabilities of ALICE detector in low-p_T region
+ Superior b-jet tagging performance of GNN

→ Low-p_T b-jet measurement in heavy-ion collisions







• Dataset specification

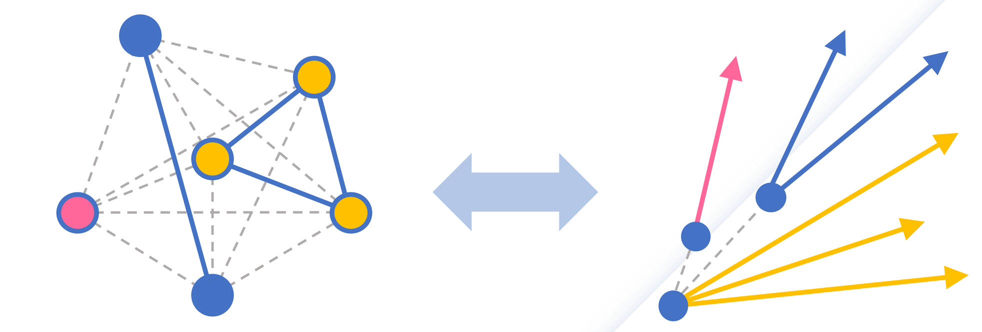
Monte Carlo sample	PYTHIA8 pp $\sqrt{s}$ = 5.02 TeV, $b\bar{b}$, $c\bar{c}$, jet-jet events
Detector	Run 2 configuration
Jet reconstruction	Charged-particle jets anti-k_T (R = 0.4)
Jet kinematic selection	$5 < p_T^{\text{ch jet}} < 100$ GeV/c, $\lvert\eta_{\text{ch jet}}\rvert < 0.5$, $N_{\text{tracks}} \geq 2$
Track kinematic selection	$p_T > 0.5$ GeV/c, $\lvert\eta\rvert < 0.9$
Dataset size	Training 6.4 M jets, validation 1.6 M jets, test 2.0 M jets (Same numbers of b-/c-/light jets)

Input features for b-jet tagging

Information of a jet & its constituent tracks

Jet: $p_T^{\text{ch jet}}$, $\varphi_{\text{ch jet}}$, $\eta_{\text{ch jet}}$
Tracks: p_T, $d\varphi$, $d\eta$, charge, impact parameter (IP), IP significance, χ^2/ndf, track reconstruction quality parameters

• Neural network

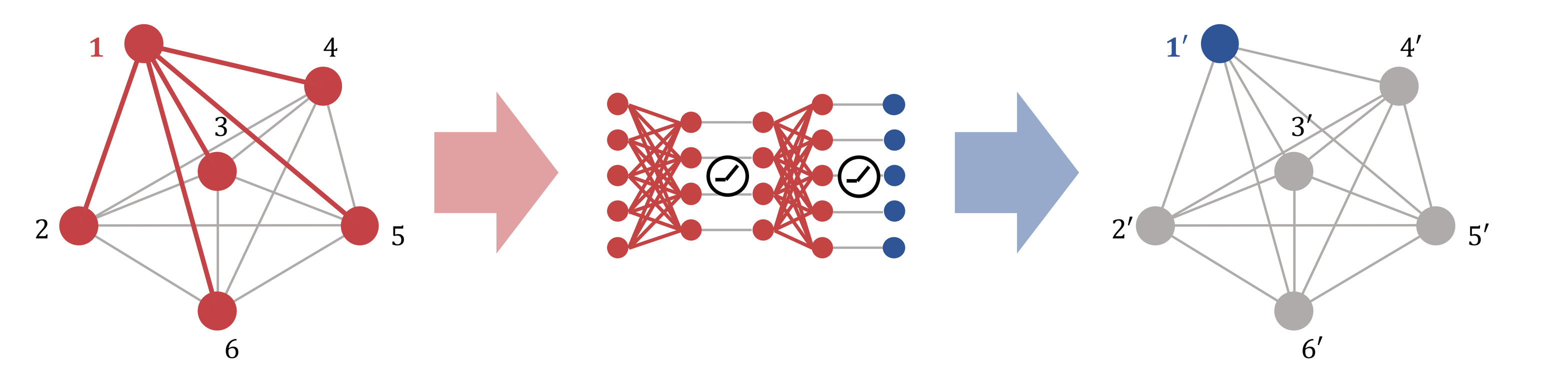


A graph is composed of nodes connected by edges.

A jet is composed of tracks connected by vertices.

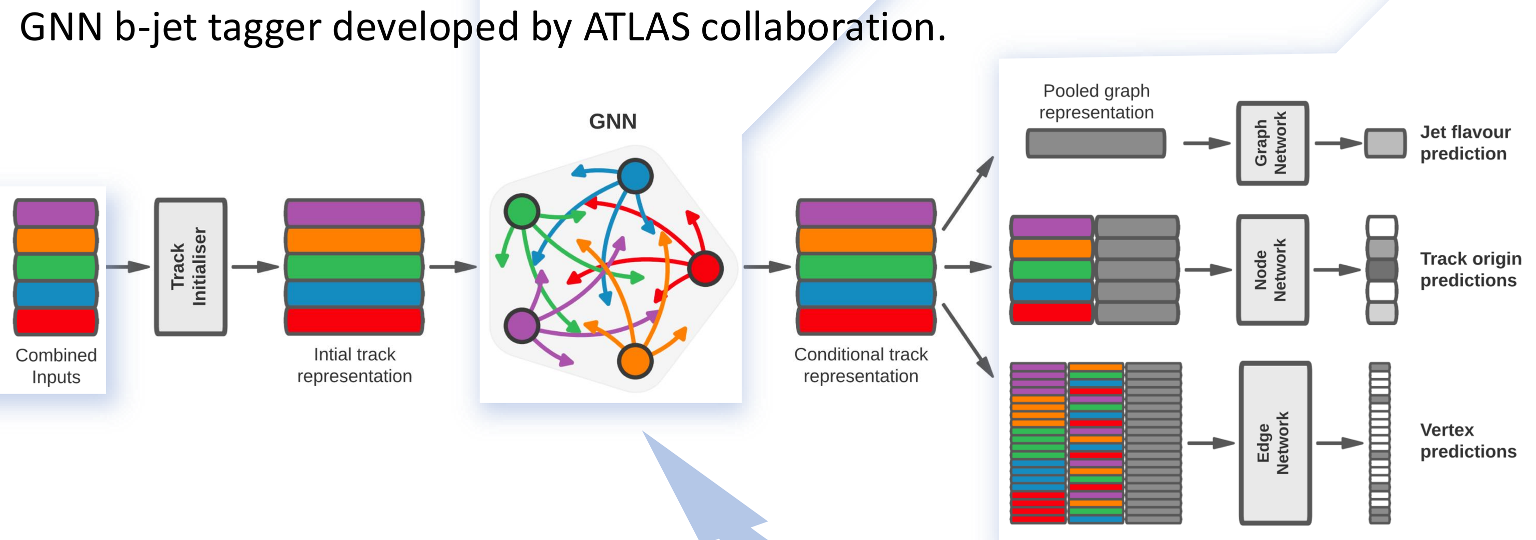
Graph Neural Network (GNN)

Captures significant features from the graph-shaped data by updating node representations by aggregating NN outputs of their adjacent nodes.



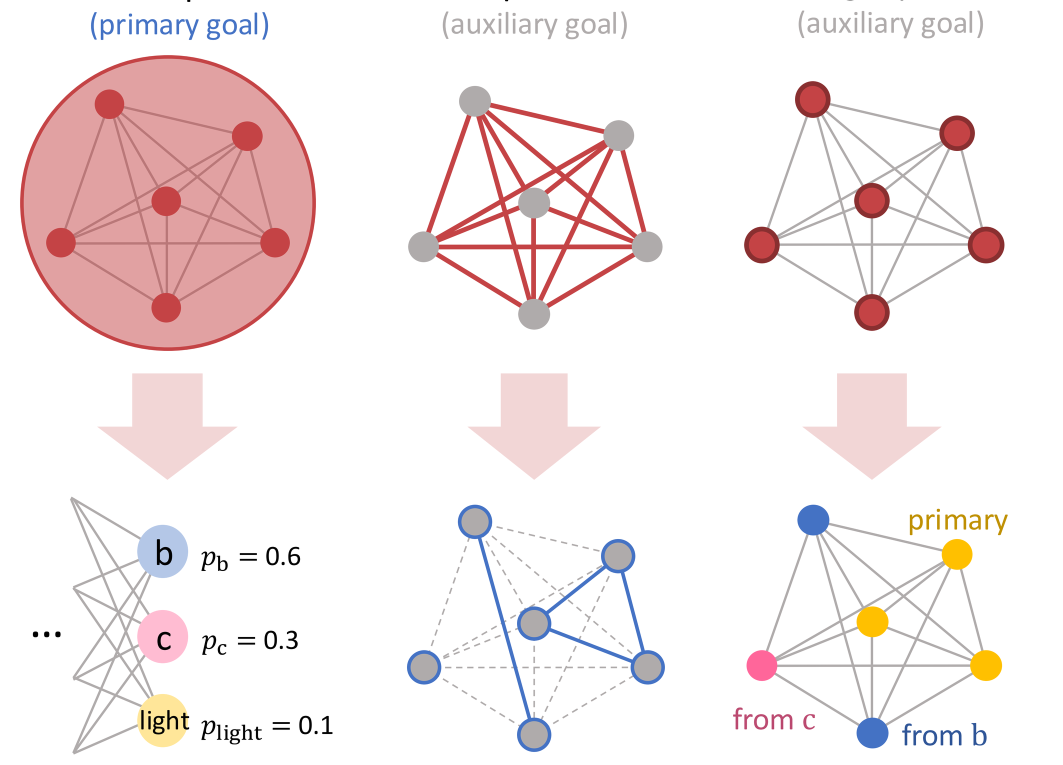
ATLAS GN1 [4]

GNN b-jet tagger developed by ATLAS collaboration.



Training goals

Jet flavour prediction (primary goal)
Vertex predictions (auxiliary goal)
Track origin predictions (auxiliary goal)



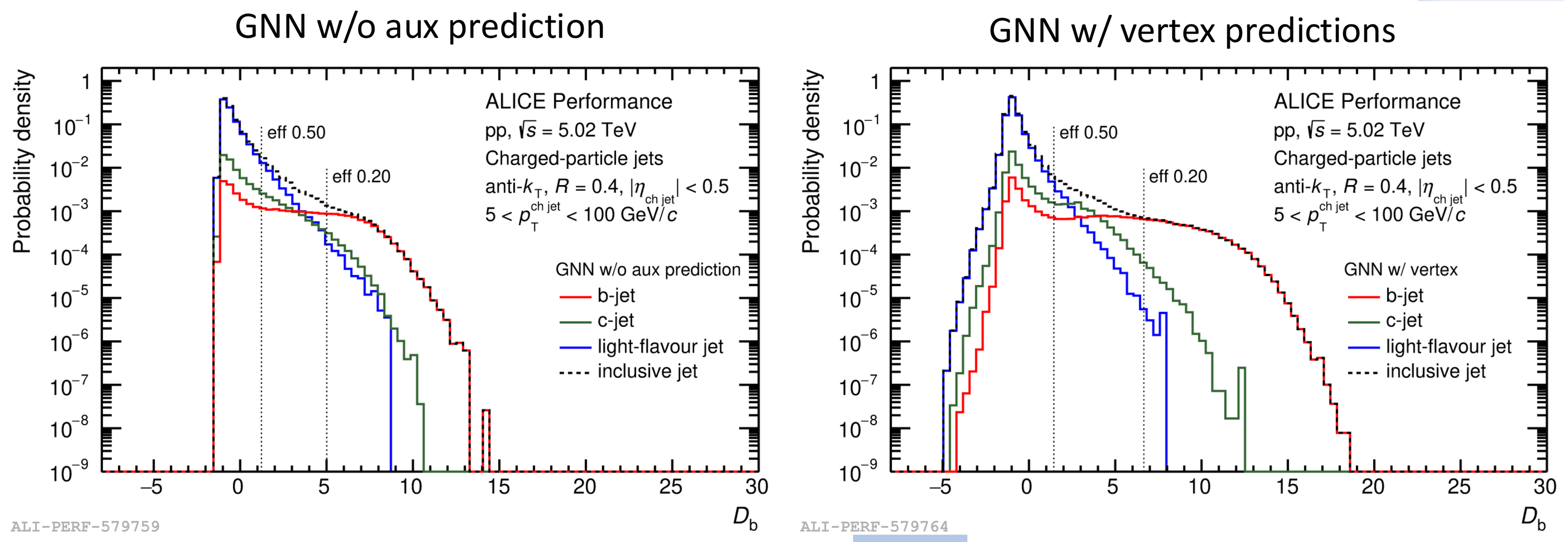
Which flavour is the jet?
Which tracks are originating from same vertex?
Where did the tracks come from?

Backpropagation

Loss functions of aux. goals can also **affect** on the training of GNN & eventually jet flavour tagging performance.

• b-jet tagging performance (MC)

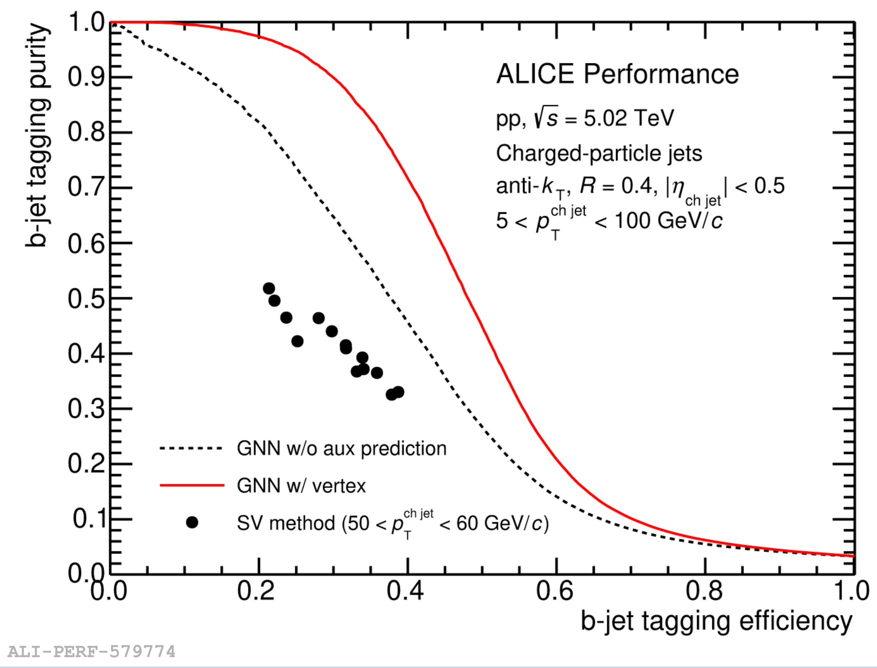
b-jet discriminant D_b



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

(f_c = 0.018, optimized parameter)

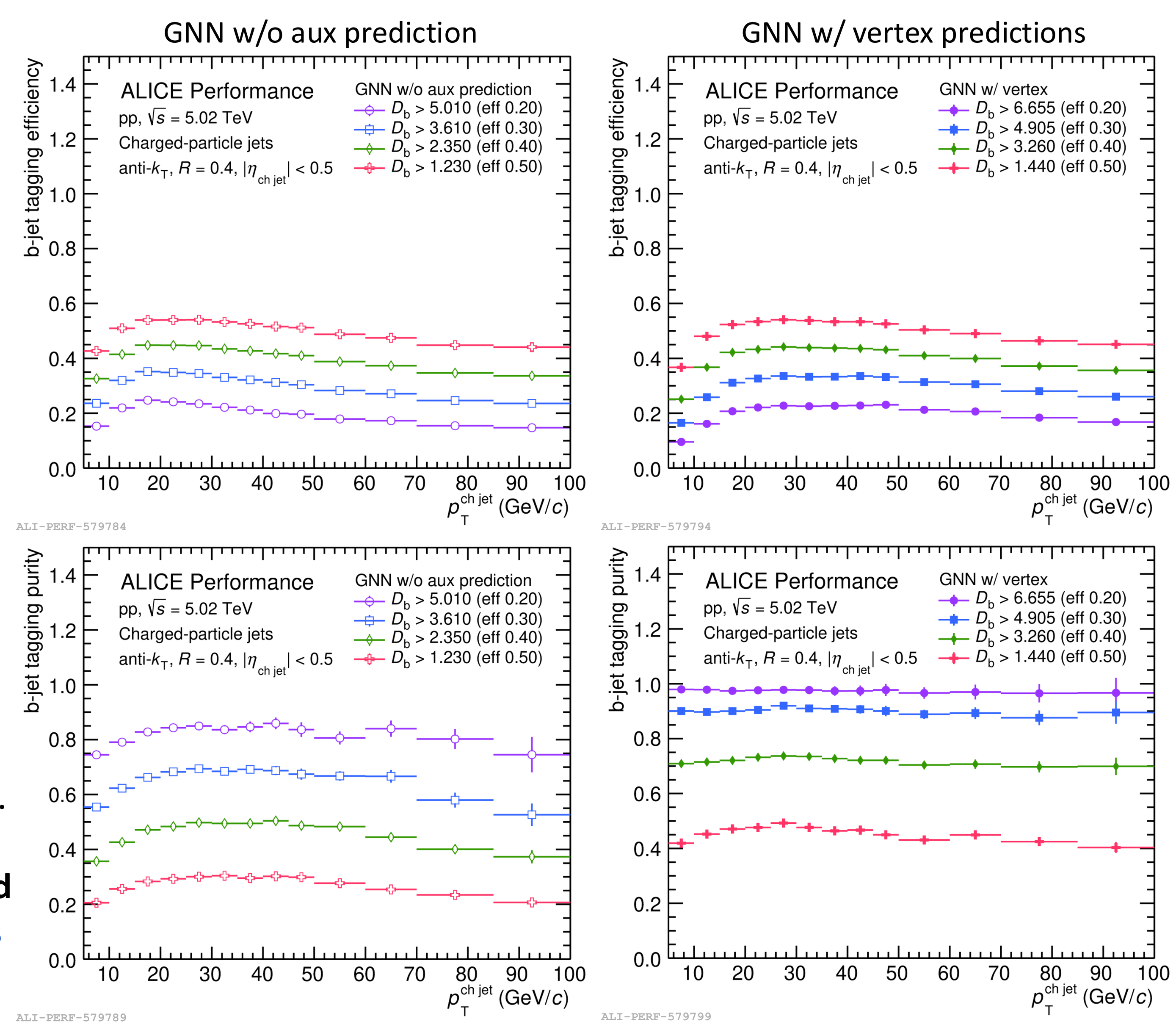
"if (D_b > threshold), then the jet is a b-jet candidate."



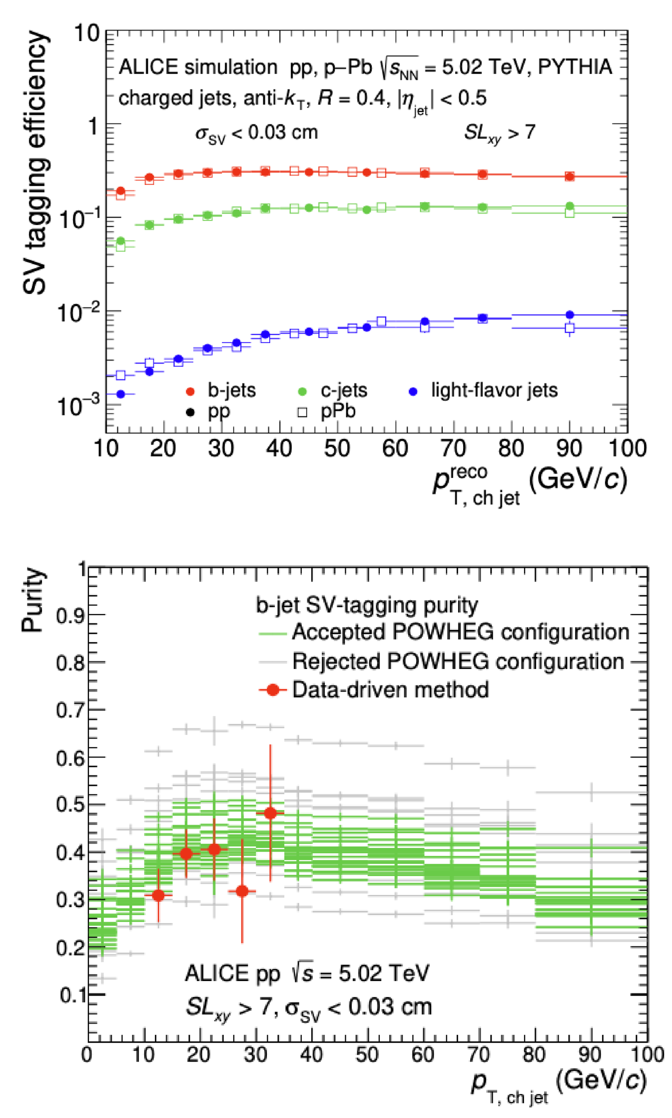
: GNN shows superior b-jet tagging performance compared to traditional secondary-vertex tagging (SV) method.

+ The performance is further **improved** when the (auxiliary) vertex predictions are considered.

b-jet tagging efficiency & purity



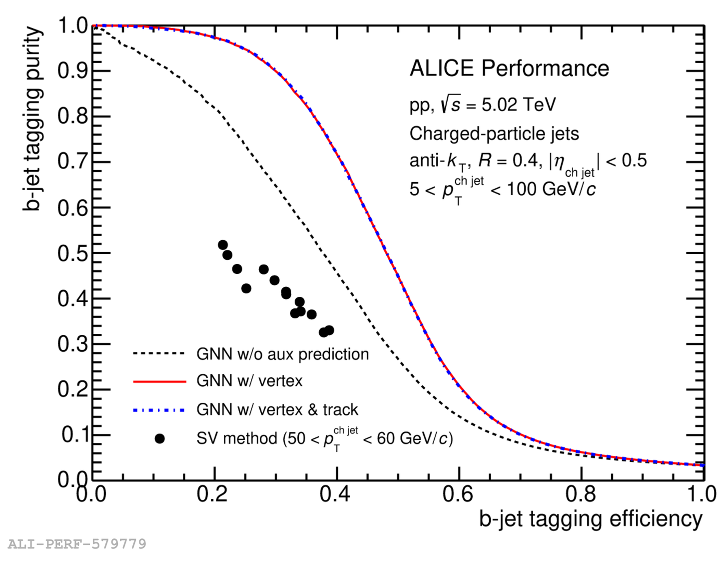
SV method [5]

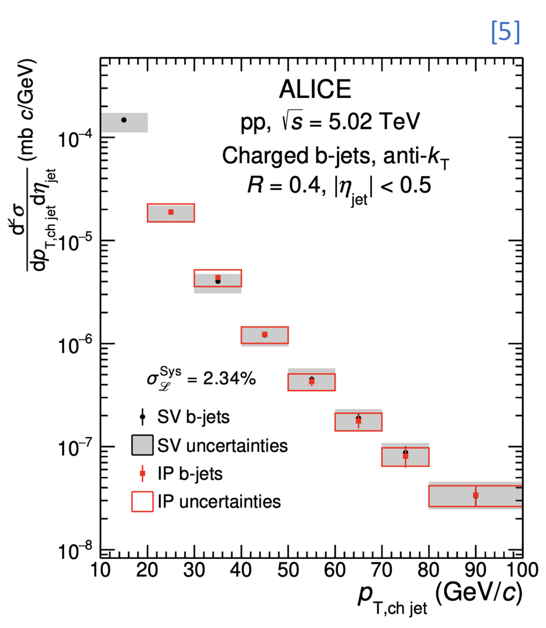


• Outlooks

- Further improvement in track origin predictions (left figure)
- Data analysis and evaluation of b-jet production cross section in pp collisions (right figure)
- ALICE Run 3 data

It is confirmed that GNN method shows superior performance than other tagging methods on ALICE Run 2 MC. We now plan to proceed with an analysis using Run 3, which have significantly better impact parameter and secondary vertex resolution and much larger data samples compared to Run 2.





- (Longer-term goal) b-jet tagging in heavy-ion collisions

• References

[1] The sPHENIX Collaboration, *sPH-HF-2017-001*
[2] J. Huang et al., *arXiv*:1306.0909
[3] The ATLAS Collaboration, *Eur.Phys.J.C* (2023) 83, 438
[4] The ATLAS Collaboration, *ATL-PHYS-PUB-2022-027*
[5] ALICE Collaboration, *JHEP* 01 (2022) 178