Flavour Tagging with Graph Neural Networks with the ATLAS experiment

Walteri Leinonen on behalf of the **ATLAS Collaboration**

ATLAS EXPERIMENT

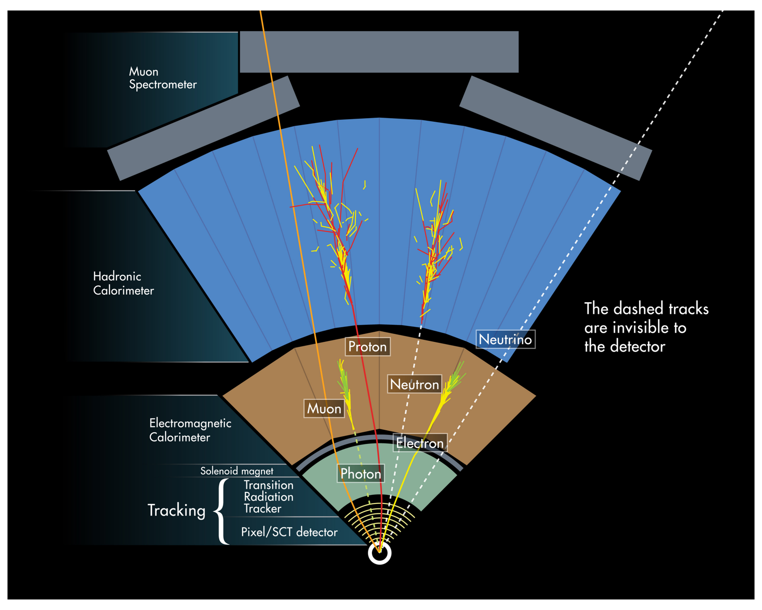
Nikhef

Radboud Universiteit

Introduction to ATLAS Physics

The ATLAS detector is a general-purpose barrel detector located at the LHC It studies high-energy fundamental particles such as the **Top quark**, and the **Higgs boson**. While these particles cannot be *directly detected* ($\tau < 10^{-20}\ s$), their decay products and fractions are predicted by the Standard Model [6].

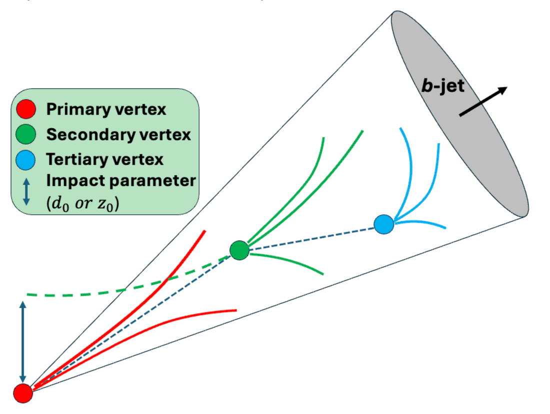
Source [6]	Mass	Lifetimes	Decay-mode	Branching Fraction
Top Quark	173.5 GeV	$\tau < 10^{-24}\ s$	$t \to W + b$	99.9%
Higgs Boson	125.09 GeV	$\tau < 10^{-22}\ s$	$h \to b\bar{b}$ / $h \to c\bar{c}$	58% / 20%



Physics objects are reconstructed by combining information from detector subsystems (**tracking**, **calorimeters**, and **muon chambers**). One important physics object that combines information from all subsystems is the **jet** – which represent a collimated spray of charged and un-charged hadrons produced in the pp-collision. Jets in ATLAS are reconstructed with a sequential **anti-kt algorithm** [3], where R=0.4 for "small-R jets" [4], and R=1.0 for "large-R jets" [5]. Internal to the jet, there are reconstructed **tracks** of charged particles – and at points where one or more tracks originate, one reconstructs **vertices**.

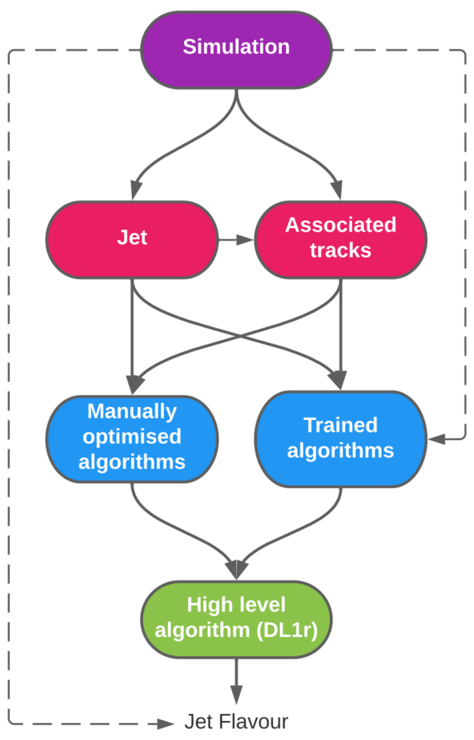
b-tagging in ATLAS

Due to the strong interaction, color-charged partons radiate and hadronize into the collimated spray we call a "jet". **Heavy-flavour partons** such as the b-quark and c-quark produce jets that can be distinguished from QCD "light-flavour jets" by their characteristic longer lifetimes, and decay-chains which can produce secondary (and tertiary) vertices in the jet. Identifying heavy-flavour jets with **high background rejection power** is key to many measurements – and a proper **calibration** of the technique is of paramount importance for results.



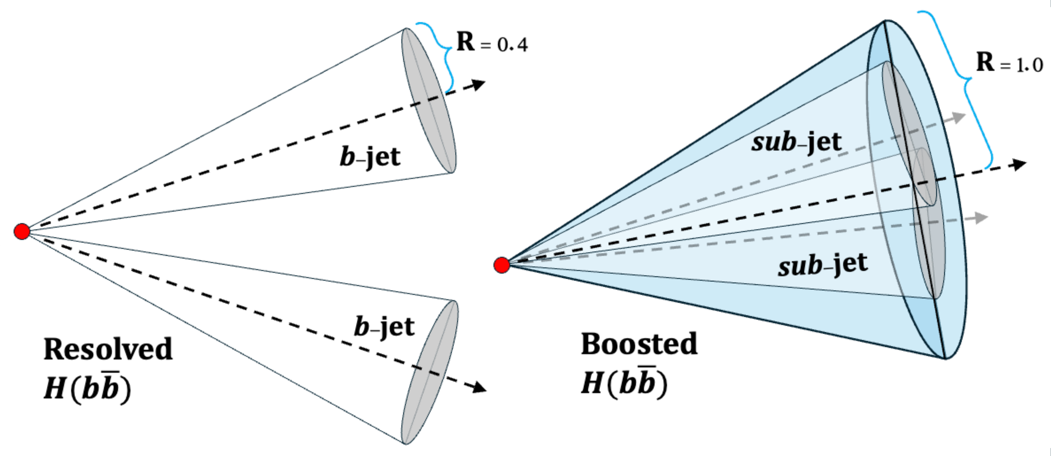
Flavour-tagging algorithms

Flavour-tagging has historically used a combination of "low-level" algorithms, which individually optimized to target a specific low-level observable (e.g. secondary-vertex reconstruction) and provide discrimination power from these physically interpretable observables alone. The outputs of these low-level taggers were then used as *input* into a "high-level" DNN known as DL1r [8]. This approach is labor intensive to train and to re-train for usage in the continually changing conditions at the LHC. A more **unified approach** is being pioneered, which uses only low-level input (jets, tracks) and **auxiliary tasks** (vertex reconstruction, track-origin) to create **physically interpretable observables**.



Boosted Object Tagging

Heavy particles are produced at the LHC with a continuum of momenta – should these particles decay into *hadrons* (e.g. b-quarks), these decay particles may be reconstructed as a **jet**. At higher momenta, *the jets will begin to overlap* (collimation) and the decay signature can no longer be distinguished as two separate objects.

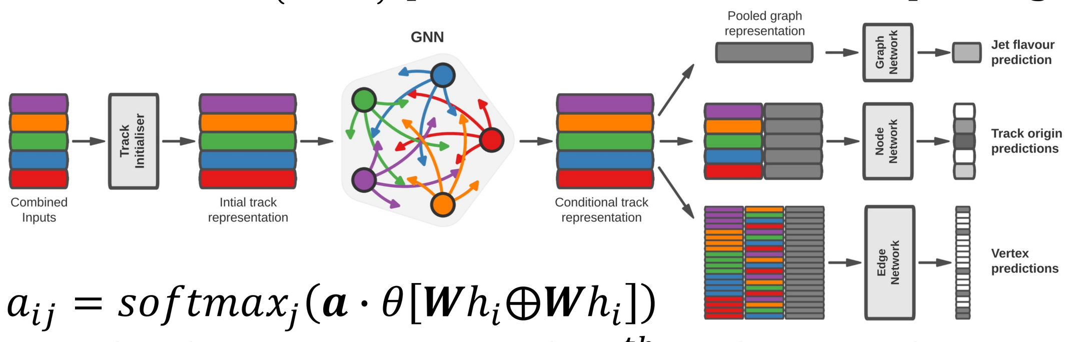


The decay products of Higgs bosons with a $p_T \gtrsim 250\ GeV$ will be **collimated**, and it is in this high-p_T "boosted" regime that sensitivity to BSM effects is highest. To reconstruct boosted $H(b\bar{b})$ and $H(c\bar{c})$, a large-R jet clustering is used. Like b-jet tagging, the main background is QCD multi-jet production. However, **boosted Top quarks** are a further background that could "fake" a boosted Higgs.

GN1 – a graph-attention b-tagger

Graph data-structures are well-suited to represent HEP data [9], which is generally unordered, feature-rich vectors that observe physical invariances. In flavour-tagging applications, each jet and track are presented by graph-level and node-level feature vectors, respectively. **Graph neural networks** (GNN) provide a framework for parsing such graph data-structures.

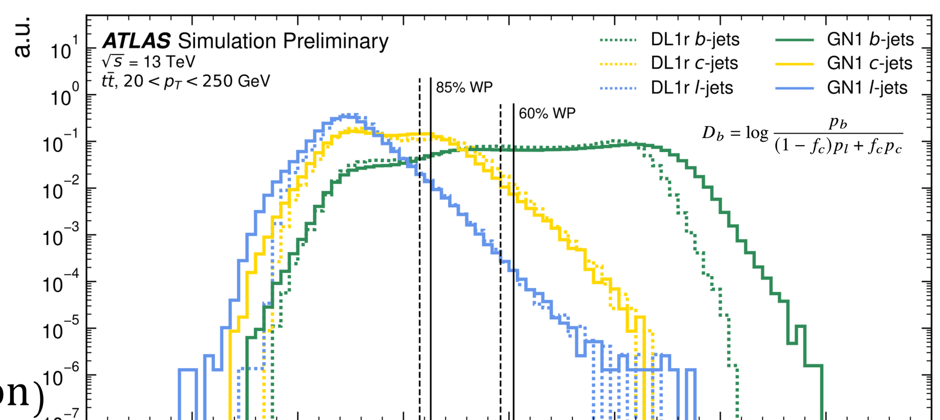
The ***GN1*** b-tagger is a GNN developed by ATLAS [1], which features the use of *self-attention* in its aggregation function.



$h_i \to h_i' = \sigma\left[\sum_{j \in N_i} a_{ij} \boldsymbol{W} h_j\right]$, where $a_{ij} = softmax_j(\boldsymbol{a} \cdot \theta[\boldsymbol{W}h_i \oplus \boldsymbol{W}h_i])$ are the **attention weights** given to each edge connecting to the i^{th} node, $\boldsymbol{W}$ and $\boldsymbol{\alpha}$ are fully-connected layers, and σ and θ are non-linear activation functions.

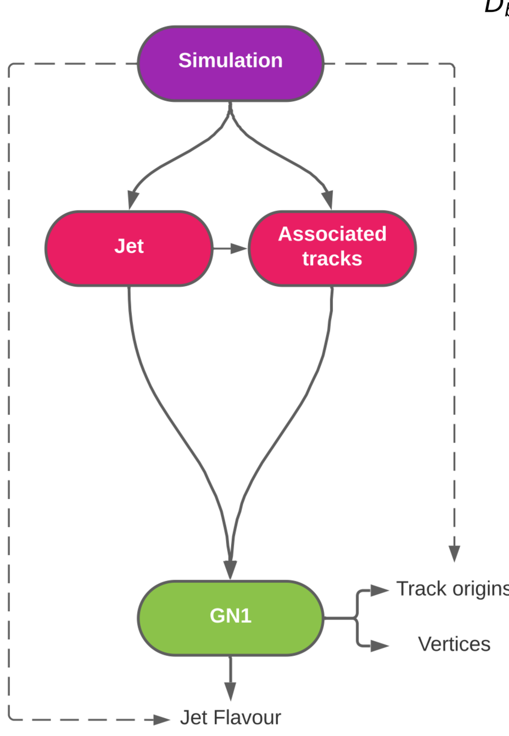
- 1 track initializer network
- 3 GNN blocks (as above)
- 1 final attention layer (weighted sum)
- 3 **independent fully-connected NNs**

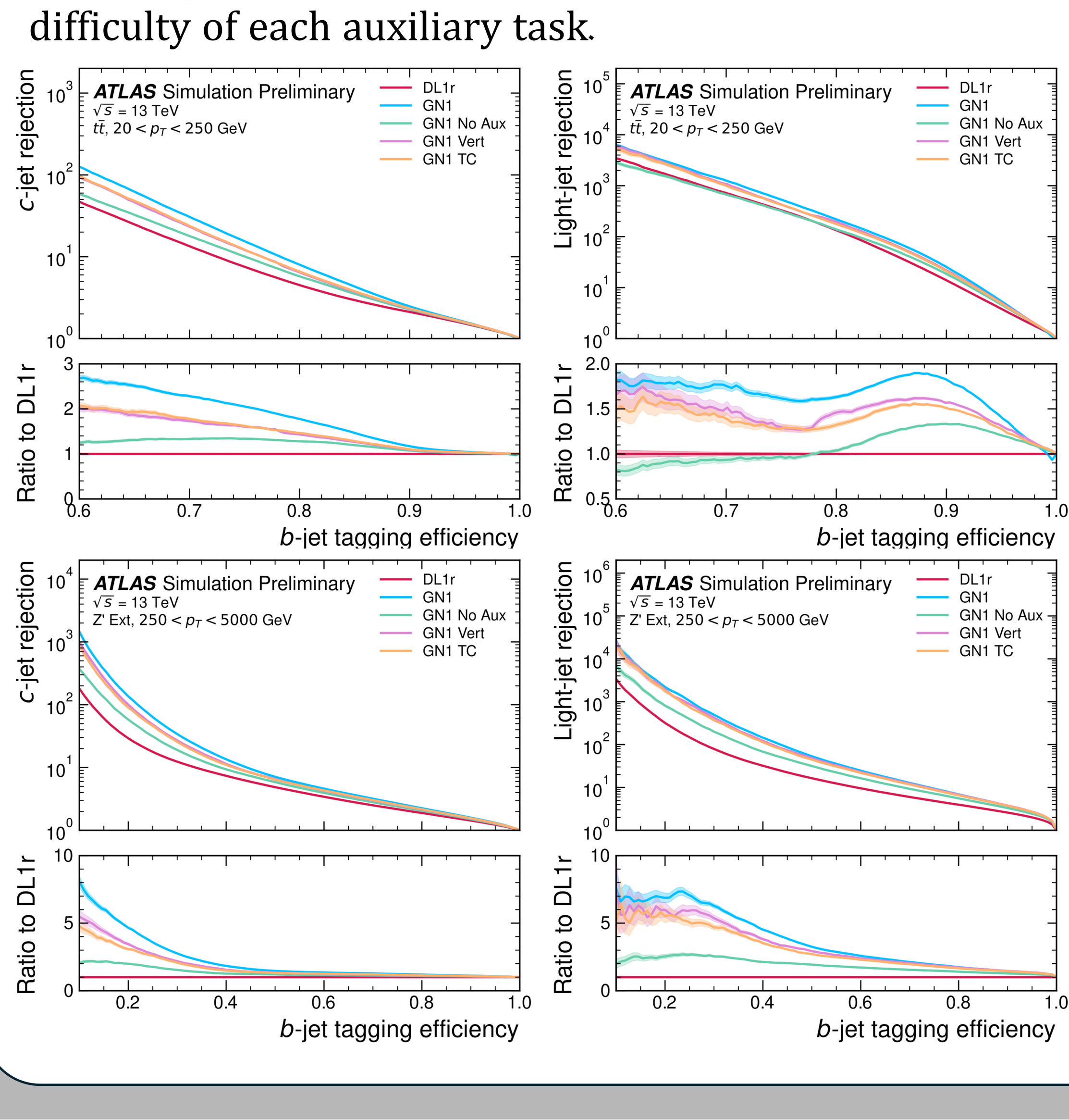
> Global classification (jet-origin tagging)
> Edge classification (vertex reconstruction)
> Node classification (track-origin tagging)



Auxiliary Training Tasks

Track origin classification and vertex reconstruction are **auxiliary training tasks** that are folded into the overall task of tagging b-jets (and c-jets) by additional terms in the loss function. The total loss function is given as a weighted sum: $\boldsymbol{L_{total} = L_{jet} + \alpha L_{vertex} + \beta L_{track}}$ where $\alpha = 1.5$ and $\beta = 0.5$ provide a weight on each task, signaling the relative difficulty of each auxiliary task.





L_{jet}: Categorical cross-entropy over all jet-flavours (b, c, light, τ)

L_{vertex}: Binary cross-entropy loss on track-pair compatibility, averaged over all track-pairs in *one batch*.

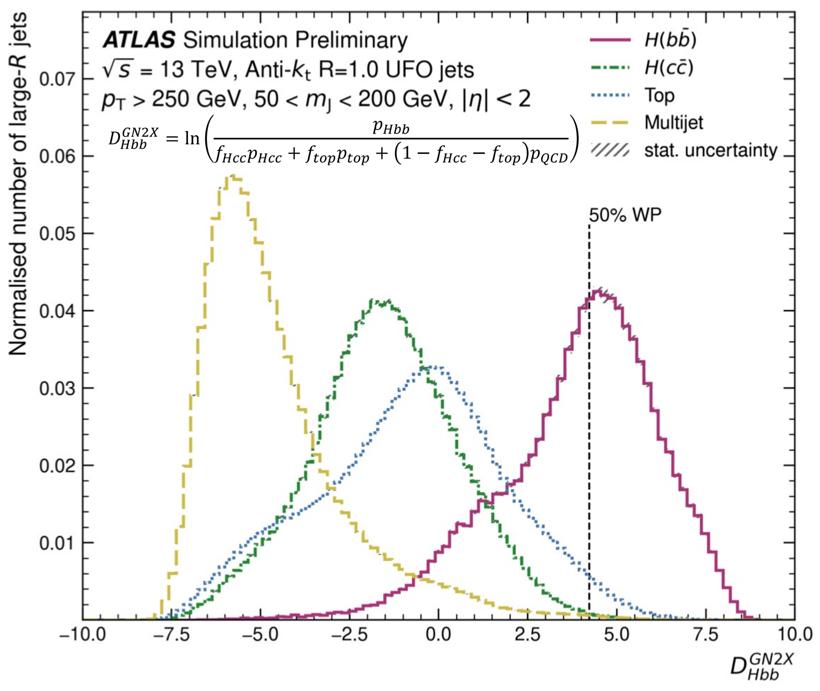
L_{track}: Categorical cross-entropy over track-origin prediction, averaged over all tracks in *one batch*.

GN2X – a Transformer $H(b\bar{b}/c\bar{c})$ tagger

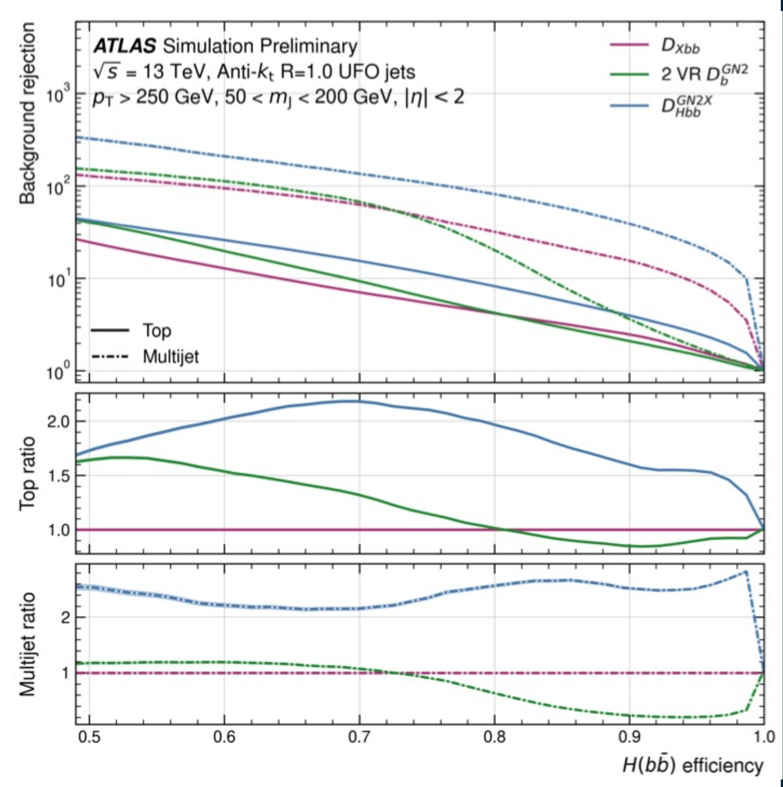
The ***GN2*** architecture is similar to ***GN1*** but replaces the GNN layer with a **Transformer encoder** which has four attention heads. Each encoder calculates the attention scores with a **scaled dot-product attention**

$$e_{ij} = \frac{(W_i h_i \cdot W_j h_j)}{\sqrt{s}} \to h_i' = softmax(e_{ij}) \cdot W_k h_j \ ,$$

closely resembling the original transformer architecture [7]. This network has already superseded the GN1 architecture and has been applied not only to b-tagging, but to **boosted Higgs tagging** in the GN2X.



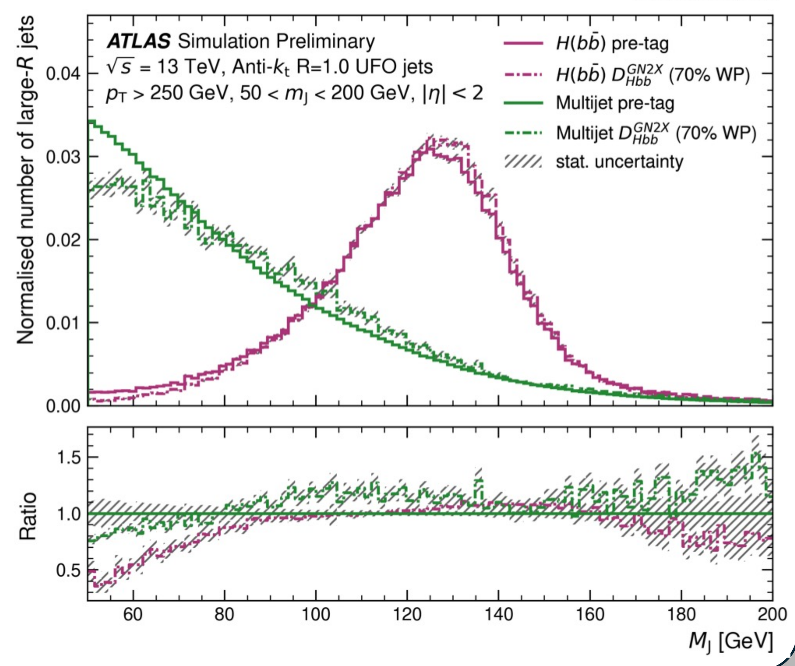
Prior to ***GN2X***, the state-of-the-art boosted "Xbb" tagger was a DNN large-R jet tagger that relied on the sub-jet DL1r b-tagging information [10,11] within the large-R jet.



Mass sculpting

One of the major challenges that ***GN2X*** (and the previous Xbb tagger) faces when taking analysis-level usage into consideration is the distribution-level **mass sculpting** effect on backgrounds such as QCD jets. If the mass of the large-R jet is used as an input, then one generally sees an *improved tagging efficiency* within the mass-range of your signal.

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).



[1] ATLAS Collaboration **Graph Neural Network Jet Flavour Tagging with the ATLAS Detector** (2022) URL: https://cds.cern.ch/record/2811135
[2] ATLAS Collaboration **Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS** (2023) URL: https://cds.cern.ch/record/2866601
[3] Matteo Cacciari *et al.* **The anti-kt jet clustering algorithm**. JHEP04(2008)063 URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063
[4] Aaboud, M., Aad, G., Abbott, B. *et al.* **Jet reconstruction and performance using particle flow with the ATLAS Detector**. *Eur. Phys. J. C* 77, 466 (2017)
[5] Aad, G., Abbott, B., Abbott, D.C. *et al.* **Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions**. *Eur. Phys. J. C* 81, 334 (2021)
[6] M. Tanabashi et al. (**Particle Data Group**), Phys. Rev. D 98, 030001 (2018) URL: https://journals.aps.org/prd/pdf/10.1103/PhysRevD.98.030001
[7] Vaswani, A et al. **Attention Is All You Need** arXiv:1706.03762 (2017)
[8] Aad, G., Abbott, B., Abbott, D.C. *et al.* **ATLAS flavour-tagging algorithms for the LHC Run 2 *pp* collision dataset**. *Eur. Phys. J. C* 83, 681 (2023)
[9] Jonathan Shlomi *et al.* **Graph neural networks in particle physics** (2021) *Mach. Learn.: Sci. Technol.* 2 021001
[10] ATLAS Collaboration., Aad, G., Abbott, B. *et al.* **Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV** . *Eur. Phys. J. C* 79, 836 (2019)
[11] ATLAS Collaboration **Efficiency corrections for a tagger for boosted $H \to b\bar{b}$ decays in *pp* collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector** (2021) URL: https://cds.cern.ch/record/2777811/files/ATL-PHYS-PUB-2021-035.pdf