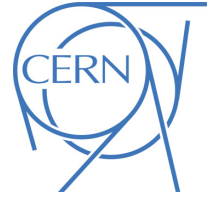




ATLAS PUB Note

ATL-PHYS-PUB-2023-021

1st August 2023



Transformer Neural Networks for Identifying Boosted Higgs Bosons decaying into $b\bar{b}$ and $c\bar{c}$ in ATLAS

The ATLAS Collaboration

Identifying boosted Higgs bosons decaying hadronically into a pair of b -quarks or c -quarks is an important capability which opens many opportunities to enrich the ATLAS physics programme at the Large Hadron Collider. In this work, a new algorithm using a transformer neural network architecture is introduced, GN2X, for identifying large-radius jets originating from boosted Higgs bosons decaying to $b\bar{b}$ and to $c\bar{c}$ pairs. GN2X directly uses information from charged particle tracks associated with the large-radius jet. For a working point exhibiting 50% $H(b\bar{b})$ efficiency, GN2X achieves a background rejection factor of 40 for jets from top-quark decays and 300 for multijet events. GN2X offers significant improvements in the rejection of background jets over the previous approach, which uses flavour tagging discriminants of individual track-based subjects in a feed-forward neural network architecture. Additional GN2X variants combining tracks, subjects and large-radius jet calorimeter constituents are also explored and found to achieve further gains in background rejection.

1 Introduction

Identifying the decays of high- p_T (boosted) Higgs bosons to pairs of bottom or charm quarks with high efficiency and background rejection is central to the physics programme of the ATLAS experiment [1] at the Large Hadron Collider [2]. A high-performance algorithm for tagging Higgs bosons decaying to $b\bar{b}$ can play a crucial role in improving the sensitivity of searches for new resonances in Beyond the Standard Model (SM) scenarios [3] or in the precise measurements of the Higgs boson p_T spectrum [4]. Additionally, direct searches for Higgs bosons decaying to charm quarks can benefit from techniques to identify the $c\bar{c}$ pair at higher momenta in a manner harmonised with the techniques used to identify the dominant $b\bar{b}$ decay [5–7].

In this work, a new algorithm, GN2X, is introduced, which is trained with the goal of classifying large-radius jets based on their origin, discriminating jets from boosted Higgs boson decaying into pairs of bottom quarks, $H(b\bar{b})$ -jets, and charm quarks, $H(c\bar{c})$ -jets, from those originating from background processes. The backgrounds considered are multijet processes and fully hadronic top-quark decays. GN2X benefits from recent advances in flavour tagging of small-radius jets using Graph Neural Networks (GNNs) and transformer architectures and auxiliary training objectives [8, 9].

Higgs boson candidates are reconstructed from a large-radius (large- R) jet, capable of capturing the majority of its decay products. In GN2X, charged particle tracks within the large- R jet are then used to identify the typical signatures of b - and c -hadron decays: displaced secondary (and possibly tertiary, in the case of b -hadrons) vertices and tracks with large impact parameters. Additional GN2X variants are also explored where tracks are combined with large- R jet calorimeter constituents and subjets. Subjets are defined with a variable-radius (VR) clustering algorithm [10, 11] to resolve and identify the individual b - or c -hadrons.

GN2X supersedes the previous state-of-the-art $H(b\bar{b})$ tagger [12, 13] which consists of a feed-forward network architecture combining flavour tagging discriminants (based on the DL1r tagger [14]) from up to three VR subjets associated to the large- R jet. It will be referred to as D_{Xbb} throughout this note and used as a benchmark for the GN2X performance.

This note is organised as follows: Section 2 details the simulation samples used in training and evaluation of GN2X. Section 3 describes the reconstruction of large- R jets, VR subjets and charged particle tracks. Section 4 introduces the algorithm and its architecture and input features. The GN2X performance for $H(b\bar{b})$ and $H(c\bar{c})$ tagging is studied in Section 5. Conclusions are presented in Section 6. The ATLAS detector is described in detail in Ref. [1].

2 Simulated Samples

The training and evaluation of the algorithms presented in this note are performed on a per-jet basis, using jets from simulated proton-proton collisions at a centre-of-mass energy of 13 TeV.

As signal, jets that originate from the decay of Higgs bosons to $b\bar{b}$ and to $c\bar{c}$ pairs produced in association with a Z boson (ZH) are considered. The goal is to train an algorithm that can discriminate between each jet class with minimal dependence on the p_T , η and mass distributions of the jets. Therefore, in order to ensure an appropriate sampling of the intended phase-space, the training sample corresponds to ZH production generated with a biased phase-space sampling that achieves an approximately flat distribution

of jet mass. In contrast, the evaluation sample consists of SM ZH production where the Higgs boson mass is set to 125 GeV and the jet p_T has a smoothly falling physical distribution.

Background jets arise from the decay of top quarks (with subsequent hadronic W boson decays) and from multijet production. The top quarks used in training are produced in the decay of a hypothetical Z' boson ($Z' \rightarrow t\bar{t}$) of mass 4 TeV. This sample is generated such that it has an approximately flat jet p_T distribution. For evaluation purposes, top quarks from SM top-pair production with a smoothly falling jet p_T distribution are considered. This SM $t\bar{t}$ sample is generated with filters on H_T – the scalar sum of the p_T of the objects in the event – in intervals of 600 GeV–1 TeV, 1 – 1.5 TeV and > 1.5 TeV. Finally, again in order to ensure sufficient population of jets at high momenta, the multijet process simulation is performed in slices of particle-level jet p_T [15] and in this case the same process is used to define the training and evaluation samples.

The versions of the MC event generators, tunes and parton distribution functions (PDF) used are detailed in Table 1 and Table 2, for training and evaluation samples respectively. In all cases, the decays of b - and c -hadrons are modelled by EVTGEN v1.6.0 [16], with the exception of the ZH evaluation samples where v1.2.0 is used.

Table 1: Signal and background processes for training with corresponding event generator versions, tunes and PDF sets.

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

Table 2: Signal and background processes for evaluation with corresponding event generator versions, tunes and PDF sets. $\ell = e, \mu$

Jet type	Process	Event generator and tune	PDF set
$H(b\bar{b})$	$q\bar{q}/g g \rightarrow ZH, Z \rightarrow \ell\bar{\ell}/\nu\bar{\nu}/q\bar{q}$	POWHEG v2 + PYTHIA 8.212 [20] with AZNLO [21]	NNPDF3.0 _{NLO}
$H(c\bar{c})$	$q\bar{q}/g g \rightarrow ZH, Z \rightarrow \ell\bar{\ell}/\nu\bar{\nu}/q\bar{q}$	POWHEG v2 + PYTHIA 8.212 with AZNLO	NNPDF3.0 _{NLO}
Top	$t\bar{t}$	POWHEG v2 + PYTHIA 8.230 with A14	NNPDF2.3 _{LO}
Multijet	Multijet	PYTHIA 8.235 with A14	NNPDF2.3 _{LO}

The samples are passed through the ATLAS full detector simulation [22] based on GEANT4 [23]. The effect of multiple interactions in the same and neighbouring bunch crossings (pile-up) was included by overlaying the simulated hard-scattering event with inelastic proton–proton events generated with PYTHIA 8.186 using the NNPDF2.3_{LO} PDF and the A3 set of tuned parameters [24].

The composite training sample undergoes a final kinematic resampling in jet p_T , η and mass to ensure identical distributions of these variables for each sample type. After applying the resampling and jet selections detailed in the following section, the resulting dataset contains 62 million jets, split between 15 million $H(b\bar{b})$, 15 million $H(c\bar{c})$, 10 million top and 22 million multijet jets.

3 Object Definitions

The inputs to the large- R jet algorithm used to reconstruct the Higgs boson candidate are Unified Flow Objects (UFOs) [25]. UFOs merge Particle-Flow Objects [26] and Track-Calo Clusters [27], which have neutral and charged components, to achieve optimal performance across a wide kinematic range. UFO large- R jets are built using the anti- k_t [28] algorithm with radius parameter $R = 1.0$ implemented in FASTJET [29]. Pile-up and underlying event contributions are removed via grooming with the Soft-Drop algorithm [30, 31] along with Constituent Subtraction [32] and SoftKiller [33]. For training, all jets with p_T in the range 200 GeV to 1.5 TeV, $|\eta| < 2.0$ and invariant mass in the window $50 \text{ GeV} < m_J < 300 \text{ GeV}$ are considered. For evaluation, jets are considered from $p_T > 250 \text{ GeV}$ to 1.5 TeV, $|\eta| < 2.0$ and $50 \text{ GeV} < m_J < 200 \text{ GeV}$.

Charged particles which traverse the inner detector register hits in the sub-detectors which are reconstructed as tracks. Tracks are ghost-associated [34] to the large- R jet and are required to satisfy the track selection summarised in Table 3 and further detailed in Ref. [35].

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	$> 500 \text{ MeV}$
$ d_0 $	$< 3.5 \text{ mm}$
$ z_0 \sin \theta $	$< 5 \text{ mm}$
Silicon hits	≥ 8
Shared silicon hits	< 2
Silicon holes	< 3
Pixel holes	< 2

In addition, inner detector tracks are used to form subjects using the variable-radius (VR) anti- k_t algorithm [37]. For VR jets, the radius R is a function of the jet p_T and a constant parameter ρ , scaling as ρ/p_T . Similar to Ref. [11], a value of $\rho = 30 \text{ GeV}$ is used. Minimum and maximum values of the radius parameter are also defined to avoid arbitrarily small or large radii with $R_{\min} = 0.02$ and $R_{\max} = 0.4$. The VR subjects are ghost-associated to the large- R jet and are required to have $p_T > 7 \text{ GeV}$.

All large- R jets in the signal samples are required to have a ghost-associated Higgs boson at generator level, using truth information from the generator’s event record. In the top quark background sample, large- R jets that are ghost-associated to any of the top quark hadronic decay products are considered. This means that large- R jets may fully or partially contain the top decay, including capturing only the hadronic W decay. No corresponding truth-labelling requirement is placed on the VR subjects.

4 Neural Network Architecture & Training

4.1 Model Inputs

The base GN2X model takes as input three large- R jet variables and 20 variables associated with each track fed to the network. The large- R jet inputs encompass the jet transverse momentum, signed pseudorapidity and mass. The full set of inputs are shown in Table 4. Up to 100 tracks associated with the jet are supplied to the network, sorted by a decreasing transverse impact parameter significance, denoted as $s(d_0)$ ¹. The ordering ensures that tracks that are more likely to have originated from displaced vertices are considered, but it has a negligible role in the algorithm since the average number of tracks in the jets will be less than 100 – the median number of associated tracks for $H(b\bar{b})$ jets is around 20.

Studies into architectures with heterogeneous inputs have also been performed. Two examples are presented in this note, GN2X + Subjets and GN2X + Flow. These models contain extra information in addition to the tracks fed into the model. GN2X + Subjets adds the kinematic and b -tagging information of the VR subjets, where the subjets are tagged using the GN2 tagger [9]. The GN2 tagger uses the same transformer architecture as GN2X but is trained to tag $R = 0.4$ jets containing a b - or c -hadron. GN2X + Flow uses UFO constituents which includes the use of charged and neutral calorimeter information. Charged constituents contain all the track inputs plus the extra calorimeter information, whereas for the neutral constituents only the calorimeter information is available. The extra inputs to each model are also documented in Table 4.

4.2 Architecture

The GN2X architecture is an evolution of the GN1 architecture [8]. GN1 makes use of a Graph Neural network, whereas GN2X (and GN2) utilises a Transformer network architecture [39]. The model takes as input the sequence of tracks in a jet, with jet- and track-level inputs outlined in Section 4.1. The jet and track inputs are concatenated and the combined jet-track sequence vectors are fed into a per-track initialiser network. This network uses a Deep Sets style [40] architecture but does not contain a reduction operation over the output track representations. For the GN2X + Subjets and GN2X + Flow variants, separate initialiser networks are defined to allow separate representations to be learned for each input type. The initialiser network for each input type consist of two dense layers projecting the input representations to an embedding dimension of 192.

The track representations are fed into a Transformer Encoder where the transformer architecture utilised follows that introduced in Ref. [41]. Multiple Layer Normalisation layers [42] are used to aid in providing stability during training along with residual connections. GN2X uses 6 encoder blocks with 4 attention heads. The heterogeneous models do not use separate transformer encoders for each input type due to the significant increase in the number of parameters to do so.

The output representation of each track is then combined to form a global representation of the jet to be used for classification. This global representation is formed by a weighted sum over the track representations, where the attention weights for the sum are learned during training.

¹ Impact parameter significances are defined as the IP divided by its corresponding uncertainty, $s(d_0) = d_0/\sigma(d_0)$ and $s(z_0) = z_0/\sigma(z_0)$. Track IP significances are lifetime signed according to the track's direction with respect to the jet axis and the primary vertex [38].

Table 4: Input features to the GN2X model. 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

In addition to the primary jet classification task, GN2X also has auxiliary training objectives – track origin classification and grouping tracks into vertices. Each of the tasks consist of three hidden layers containing 128, 64 and 32 neurons respectively. Further details on the auxiliary task objectives are outlined in Ref. [8]. In total GN2X contains approximately 1.5 million parameters.

4.3 Training

GN2X is trained using a training approach similar to GN1 [8]. GN2X trainings are done on a cluster with 4 NVIDIA A100 GPUs, taking approximately 1 hour to complete an epoch of 62 million jets. The model is optimised using the Adam optimiser with a batch size of 1,000 for 40 epochs. During training, the model’s primary and auxiliary tasks are weighted such that the losses were of similar magnitude, this approach is similar to that used in [8]. Moreover, the model is trained using a one-cycle learning rate policy, in which the learning rate starts at a very small value, 10^{-7} , and gradually increased over the first 4 epochs, after which it reached a maximum value of 0.005. Then, the learning rate is gradually decreased again and ends at 10^{-7} .

5 Results

The performance evaluation of the GN2X tagger includes an assessment of the efficiency for tagging both $H(b\bar{b})$ and $H(c\bar{c})$. The $H(b\bar{b})$ -tagging efficiency is defined as the number of correctly identified $H(b\bar{b})$ jets divided by the total number of $H(b\bar{b})$ jets, while treating $H(c\bar{c})$, top, and multijet (often also denoted QCD) processes as backgrounds. The $H(c\bar{c})$ efficiency is also defined in the same manner, with $H(b\bar{b})$ as a background. The network generates probability scores that indicate the likelihood of a given jet being identified as $H(c\bar{c})$, $H(b\bar{b})$, top, or multijet. When assessing the $H(b\bar{b})$ -tagging efficiency, these probability scores are combined into a discriminant, $D_{\text{Hbb}}^{\text{GN2X}}$, defined as:

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

where f_{Hcc} and f_{top} are two free parameters that determine the relative weights of p_{Hcc} and p_{top} respectively to p_{QCD} , controlling the trade-off among $H(c\bar{c})$, top and multijet rejections. For the following performance studies, $f_{\text{Hcc}} = 0.02$ and $f_{\text{top}} = 0.25$ were used, which were obtained after following an optimisation procedure to maximise the rejection for a given efficiency.

Figure 1 shows the normalised distribution of the discriminant score for $H(b\bar{b})$, $H(c\bar{c})$, top, and multijet jets. The double peak structure observed in the top jet distribution is a result of the difference between large- R jets the fully or only partially contain the top decay, where jets that don’t contain the b -hadron from the top decay appear much more similar to the multijet background. Jets are considered tagged if they have a score higher than a given threshold value of the $D_{\text{Hbb}}^{\text{GN2X}}$ discriminant, usually defined to correspond to a specific desired signal efficiency. The D_{Xbb} model evaluated on the same dataset serves as a baseline. It should be noted, however, that it has not been retrained with the samples used for training GN2X.

An additional baseline considered is the case where two individual VR subjets are b - or c -tagged based on their corresponding GN2 discriminants (noted as 2 VR D_b^{GN2} or 2 VR D_c^{GN2} in the plots). In this case, only the two leading VR subjets associated to the large- R jet are considered, and the minimum GN2 value

between the two is used to define the signal efficiency. This approach does not utilise any correlations between subjects and corresponds to tagging them independently.

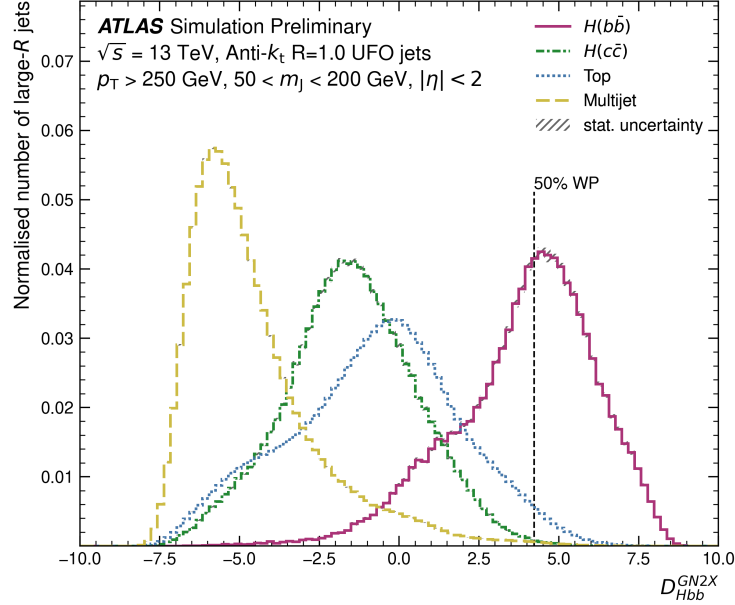


Figure 1: GN2X discriminant, D_{Hbb} with $f_{top} = 0.25$ and $f_{Hcc} = 0.02$ for the four jet classes. Working points are derived by imposing a requirement on the discriminant distribution of the $H(b\bar{b})$ jets to give the desired signal efficiency. The distribution is shown for the SM evaluation samples.

5.1 $H(b\bar{b})$ performance

The $H(b\bar{b})$ -tagging performance of the tagger can be quantified by its power to reject $H(c\bar{c})$, top, and multijet backgrounds for a given $H(b\bar{b})$ -tagging efficiency.

Figure 2 shows the background rejection as a function of the $H(b\bar{b})$ signal efficiency in the signal efficiency range used for most physics analyses. The rejection is evaluated at a set number of points and can change rapidly at high efficiencies, which causes the observed sharp behaviour. GN2X demonstrates a significant improvement in both the top and multijet background rejection, significantly outperforming both baselines across the relevant signal efficiency range. At a 50% $H(b\bar{b})$ signal efficiency GN2X provides a factor of 1.6 increase in the top jet rejection and a factor of 2.5 increase in the multijet rejection. GN2X also outperforms the 2-tag VR subjects baseline across all efficiencies.

The 2-tag VR subjects baseline only outperforms D_{Xbb} at low $H(b\bar{b})$ efficiencies, which shows the benefits of using the correlations between subjects. Furthermore, D_{Xbb} uses up to three VR subjects which improves its efficiency for cases where the b -hadrons are not contained in the leading two subjects. In this comparison D_{Xbb} was trained with DL1r discriminants [43] whereas the 2-tag VR subjects baseline benefits from the state-of-the-art GN2 tagger which has superior c -jet rejection.

The top and multijet rejection factors are also shown as a function of the jet p_T in Figure 3 at a 50% $H(b\bar{b})$ efficiency where it can be seen that GN2X outperforms the baseline taggers across the entire p_T spectrum. The top jet rejection is improved by a factor of 1.4 over the D_{Xbb} tagger at 250 GeV and this increases to a factor of 2.2 increase in rejection at 1.5 TeV. The multijet rejection is improved by more

than a factor of two across the entire p_T spectrum. In Figure 4, the $H(b\bar{b})$ signal efficiency against jet p_T is shown for the same 50% working point. It shows a remarkably stable efficiency as a function of the jet p_T , in contrast to what is observed with the baseline taggers. While the D_{Xbb} efficiency drops from the expected 50% efficiency at 250 GeV to approximately 35% at 1.5 TeV, GN2X is able to retain an approximately constant efficiency over the entire p_T range.

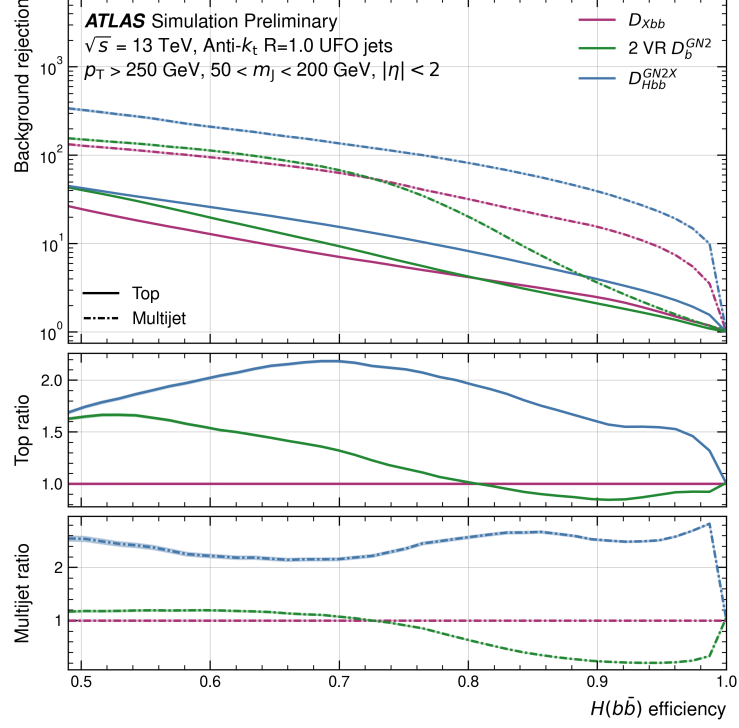


Figure 2: Top and multijet rejections as a function of the $H(b\bar{b})$ efficiency for jets with $p_T > 250$ GeV and mass $50 \text{ GeV} < m_J < 200 \text{ GeV}$. Performance of the GN2X algorithm is compared to the D_{Xbb} and VR subjects baselines. Statistical uncertainty bands (calculated with a binomial model) are denoted. The distribution is shown for the SM evaluation samples.

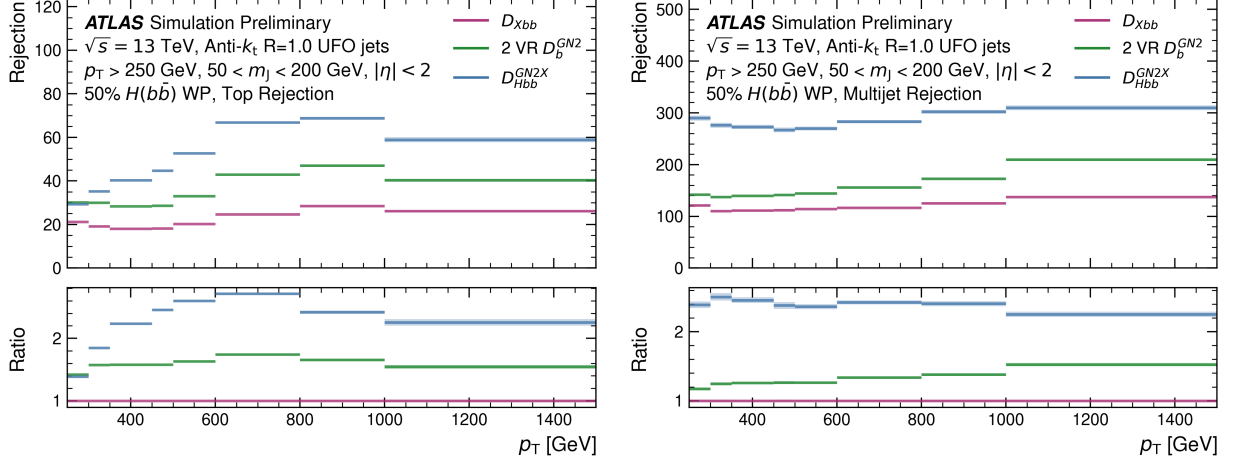


Figure 3: The top (left) and multijet (right) rejection against the jet p_T for a 50% $H(b\bar{b})$ efficiency. Statistical uncertainty bands (calculated with a binomial model) are denoted. The distribution is shown for the SM evaluation samples.

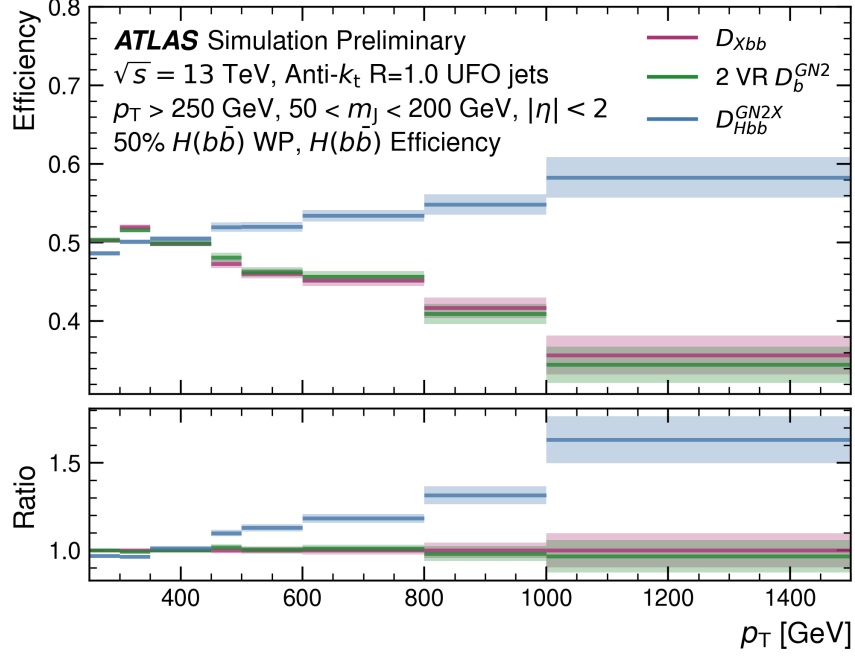


Figure 4: The $H(b\bar{b})$ signal efficiency as a function of the jet p_T at a 50% working point. Statistical uncertainty bands (calculated with a binomial model) are denoted. The distribution is shown for the SM evaluation samples.

As described in Section 4 preliminary studies using heterogeneous input types have also been conducted. Figure 5 shows the $H(b\bar{b})$ tagging performance of the GN2X + Subjets and GN2X + Flow architectures compared to the baseline GN2X model. The GN2X + Subjets model provides an additional 100% increase in the top rejection relative to GN2X by including complementary information on the large- R jet substructure, though it results in a reduction in the multijet rejection performance. The GN2X + Flow model however takes advantage of information on neutral jet components missing from other versions and provides more than a 50% improvement in the multijet rejection.

5.1.1 Mass sculpting

To ensure that the signal and background mass shapes remain distinct after applying $H(b\bar{b})$ tagging in an analysis, it is important that the tagger scores are not correlated with the large- R jet mass. If there is such a correlation, it may cause the mass distribution of the background to resemble that of the signal after tagging. This phenomenon is referred to as mass sculpting and it is undesirable when backgrounds are estimated with data-driven approaches that rely on smoothly-falling distributions, as is often the case for the multijet process. This is important for searches for Beyond the SM phenomena (e.g. [44, 45]) and also in the calibration of any $H(b\bar{b})$ tagging algorithm that relies on identifying $Z \rightarrow b\bar{b}$ decays in Z +jets or $Z + \gamma$ events [13]. The top process is not considered here given that the jet mass is naturally correlated with flavour and can usually be estimated via simulation.

Figure 6 displays the normalised large- R jet mass distributions pre- and post-tagging with GN2X, for the $H(b\bar{b})$ signal and the multijet background. A requirement on the discriminant corresponding to a 70% $H(b\bar{b})$ efficiency has been used.

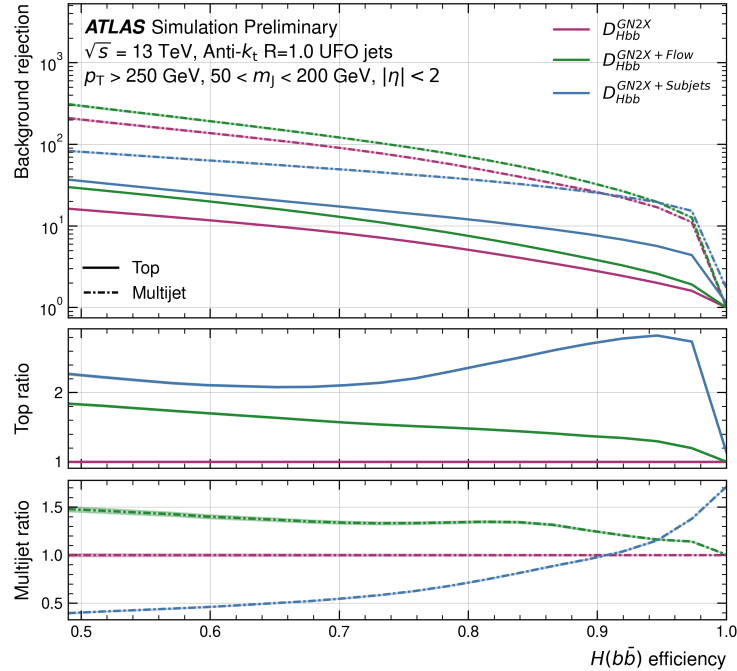


Figure 5: The top and multijet background rejection as a function of the $H(b\bar{b})$ tagging efficiency for the two heterogeneous input type architectures compared to the baseline GN2X model. Statistical uncertainty bands (calculated with a binomial model) are denoted. The distribution is shown for the SM evaluation samples.

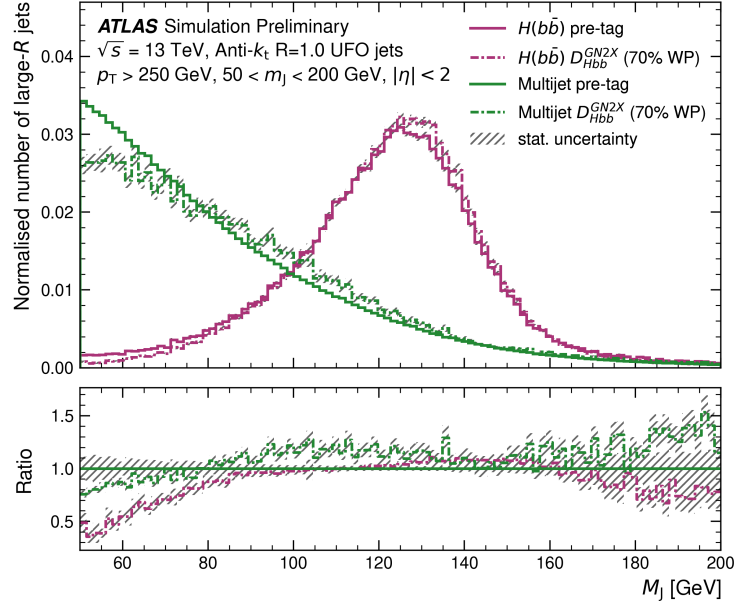


Figure 6: Large- R jet mass distributions for $H(b\bar{b})$ and multijet samples, before and after applying a 70% $H(b\bar{b})$ efficiency D_{Hbb}^{GN2X} cut. The distribution is shown for the SM evaluation samples.

Close to the peak of the $H(b\bar{b})$ mass distribution there is little change in the shape of the multijet distribution, as desired. However, at both high and low mass values changes in shape of the multijet distribution begin to become more significant. This is expected due to the low training statistics for jets in this kinematic regime. Extending the training distribution to higher masses is expected to reduce sculpting in the high mass region. These conclusions also hold for stricter requirements on the discriminant. Sculpting of the signal distribution is not considered a concern for most physics use cases, but does provide useful information on the population of $H(b\bar{b})$ jets that are not tagged.

5.2 $H(c\bar{c})$ performance

Since GN2X outputs the likelihoods of the jets being identified as $H(b\bar{b})$, $H(c\bar{c})$, top and multijet, it can also be used for $H(c\bar{c})$ tagging. The likelihoods can be combined similarly as to D_{Hbb}^{GN2X} in Section 5 to obtain a discriminant, D_{Hcc}^{GN2X} , defined as:

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

where, f_{Hbb} is a free parameter, similar to f_{Hcc} , that determines the relative rejection weight of p_{Hbb} to the rest of the backgrounds. For the following performance studies, $f_{Hbb} = 0.3$ and $f_{\text{top}} = 0.25$ were used, which were obtained via an optimisation procedure identical to the one used for D_{Hbb}^{GN2X} .

Figure 7 shows the background rejection as a function of the $H(c\bar{c})$ signal efficiency in the range relevant for most physics analyses. As the previous D_{Xbb} tagger was not designed for $H(c\bar{c})$ identification, only the 2-tag VR track-jet baseline can be compared to. The minimum discriminant value of the leading two VR track-jets must pass a c -tagging requirement. For $H(c\bar{c})$ identification, in addition to the top and

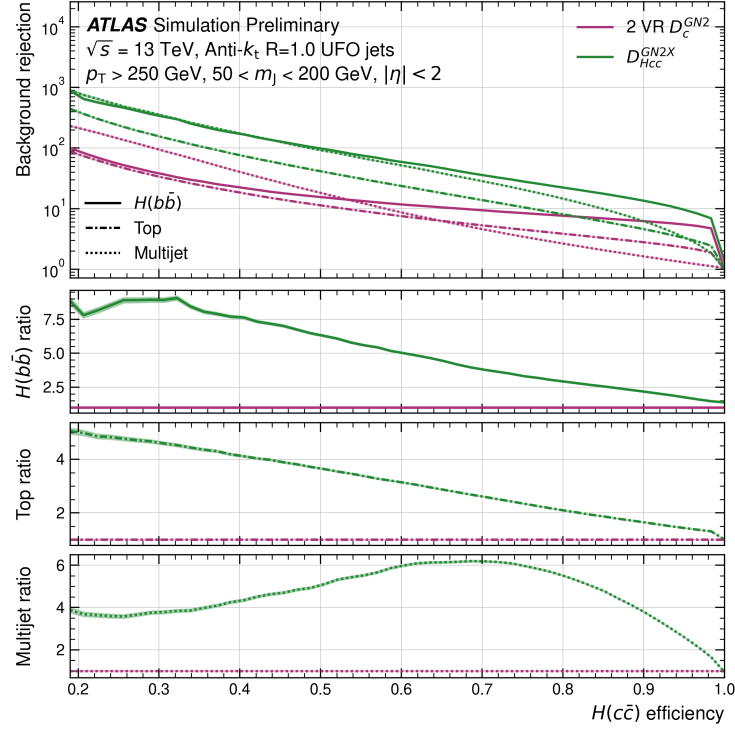


Figure 7: $H(b\bar{b})$, top and multijet rejections as a function of the $H(c\bar{c})$ efficiency for jets with $p_T > 250$ GeV and mass $50 \text{ GeV} < m_J < 200 \text{ GeV}$. The performance of the GN2X algorithm is compared to the 2-tag VR track-jet baseline. Statistical uncertainty bands (calculated with a binomial model) are denoted. The distribution is shown for the SM evaluation samples.

multijet background, the background from $H(b\bar{b})$ events is non-negligible so the $H(b\bar{b})$ rejection is also shown. At a 50% signal efficiency GN2X provides significant increases in background rejection, a factor of 3 improvement for top jet rejection, a factor of 5 improvement of the multijet rejection and a factor 6 improvement in the $H(b\bar{b})$ rejection.

6 Conclusion

This note presents a novel boosted double b - and double c -tagging algorithm, GN2X, building upon a graph neural network and transformer architecture. GN2X improves upon the current ATLAS Xbb tagger by improving the top and multijet background rejections by a factor of 1.6 and 2.5 respectively at a 50% signal efficiency when evaluated on jets with $p_T > 200$ GeV and a mass range of $50 \text{ GeV} < m_J < 200 \text{ GeV}$. GN2X can also operate as a boosted $H(c\bar{c})$ tagger, the first of its kind in ATLAS. As an $H(c\bar{c})$ tagger it provides significant background rejection against not only the top and multijet backgrounds but also rejects 99% of $H(b\bar{b})$ jets at a 50% $H(c\bar{c})$ efficiency. Preliminary studies using heterogeneous inputs have also been conducted using extra subjet and UFO constituents information, the constituent model providing a further 50% increase in multijet rejection at a 50% $H(b\bar{b})$ signal efficiency and the subjet model further doubles the top rejection. GN2X has been successfully implemented into the ATLAS software [46] and is ready for further studies to verify its performance on real collision data.

Appendix

A Multijet Rejection

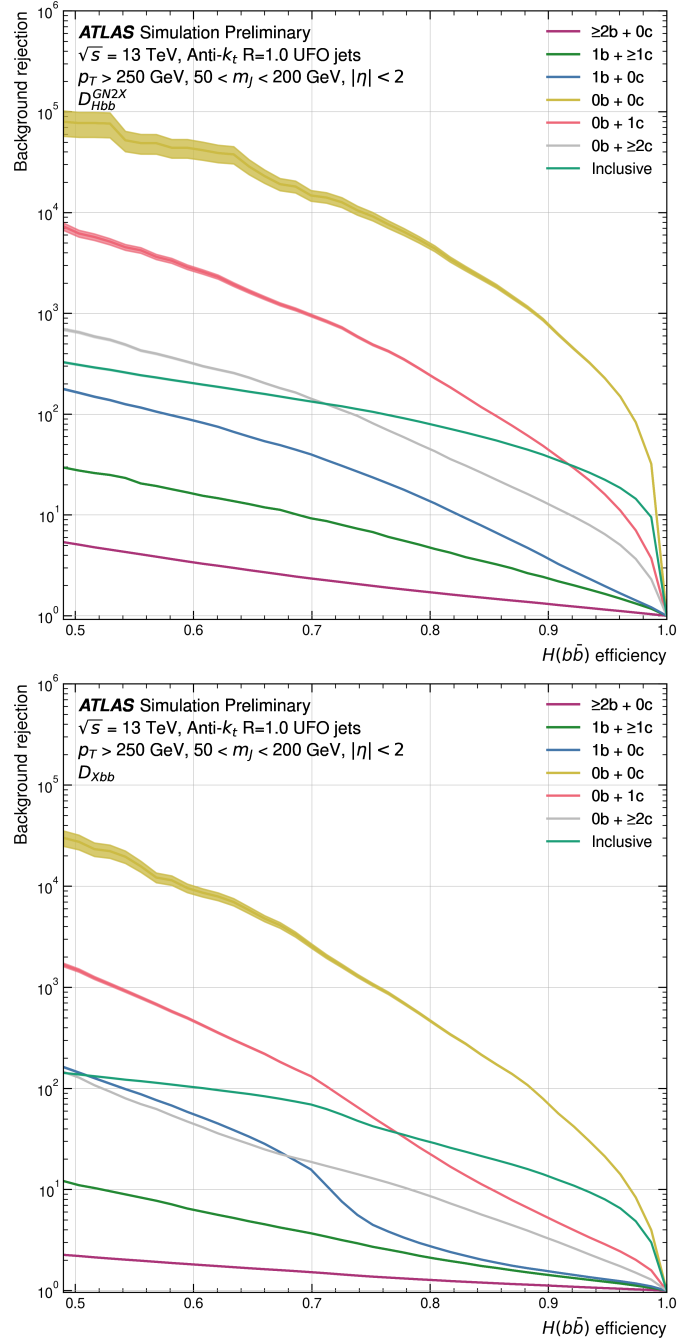


Figure 8: Multijet rejection split into the flavour components determined by the flavour of the VR subjects for GN2X and D_{Xbb} . Statistical uncertainty bands (calculated with a binomial model) are denoted. D_{Xbb} has not been retrained for this comparison. The distribution is shown for the SM evaluation samples.

References

- [1] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, [JINST 3 \(2008\) S08003](#) (cit. on p. 2).
- [2] L. Evans and P. Bryant, *LHC Machine*, [JINST 3 \(2008\) S08001](#) (cit. on p. 2).
- [3] ATLAS Collaboration, *Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, (2023), arXiv: [2306.03637 \[hep-ex\]](#) (cit. on p. 2).
- [4] ATLAS Collaboration, *Constraints on Higgs boson production with large transverse momentum using $H \rightarrow b\bar{b}$ decays in the ATLAS detector*, [Phys. Rev. D 105 \(2021\) 092003](#), arXiv: [2111.08340 \[hep-ex\]](#) (cit. on p. 2).
- [5] ATLAS Collaboration, *Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector*, [Eur. Phys. J. C 82 \(2022\) 717](#), arXiv: [2201.11428 \[hep-ex\]](#) (cit. on p. 2).
- [6] CMS Collaboration, *Search for boosted Higgs boson decay to a charm quark–antiquark pair in proton–proton collisions at $\sqrt{s} = 13$ TeV*, (2022), arXiv: [2211.14181 \[hep-ex\]](#) (cit. on p. 2).
- [7] CMS Collaboration, *Search for Higgs boson decay to a charm quark–antiquark pair in proton–proton collisions at $\sqrt{s} = 13$ TeV*, (2022), arXiv: [2205.05550 \[hep-ex\]](#) (cit. on p. 2).
- [8] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-027, 2022, URL: <https://cds.cern.ch/record/2811135> (cit. on pp. 2, 5, 7).
- [9] ATLAS Collaboration, *Jet Flavour Tagging With GNI and DL1d. Generator dependence, Run 2 and Run 3 data agreement studies*, [ATL-PLOT-FTAG-2023-01](#), 2023 (cit. on pp. 2, 5).
- [10] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R* , [JHEP 06 \(2009\) 059](#), arXiv: [0903.0392 \[hep-ph\]](#) (cit. on p. 2).
- [11] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017, URL: <https://cds.cern.ch/record/2268678> (cit. on pp. 2, 4).
- [12] ATLAS Collaboration, *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*, ATL-PHYS-PUB-2020-019, 2020, URL: <https://cds.cern.ch/record/2724739> (cit. on p. 2).
- [13] ATLAS Collaboration, *Efficiency corrections for a tagger for boosted $H \rightarrow b\bar{b}$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATL-PHYS-PUB-2021-035, 2021, URL: <https://cds.cern.ch/record/2777811> (cit. on pp. 2, 11).
- [14] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: <https://cds.cern.ch/record/2273281> (cit. on p. 2).
- [15] ATLAS Collaboration, *Multijet simulation for 13 TeV ATLAS Analyses*, ATL-PHYS-PUB-2019-017, 2019, URL: <https://cds.cern.ch/record/2672252> (cit. on p. 3).
- [16] D. J. Lange, *The EvtGen particle decay simulation package*, [Nucl. Instrum. Meth. A 462 \(2001\) 152](#) (cit. on p. 3).

- [17] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, [*Comput. Phys. Commun.* **178** \(2008\) 852](#), arXiv: [0710.3820 \[hep-ph\]](#) (cit. on p. 3).
- [18] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419> (cit. on p. 3).
- [19] NNPDF Collaboration, R. D. Ball et al., *Parton distributions with LHC data*, [*Nucl. Phys. B* **867** \(2013\) 244](#), arXiv: [1207.1303 \[hep-ph\]](#) (cit. on p. 3).
- [20] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, [*JHEP* **11** \(2007\) 070](#), arXiv: [0709.2092 \[hep-ph\]](#) (cit. on p. 3).
- [21] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, [*JHEP* **09** \(2014\) 145](#), arXiv: [1406.3660 \[hep-ex\]](#) (cit. on p. 3).
- [22] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, [*Eur. Phys. J. C* **70** \(2010\) 823](#), arXiv: [1005.4568 \[physics.ins-det\]](#) (cit. on p. 3).
- [23] S. Agostinelli et al., *GEANT4 – a simulation toolkit*, [*Nucl. Instrum. Meth. A* **506** \(2003\) 250](#) (cit. on p. 3).
- [24] ATLAS Collaboration, *The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model*, ATL-PHYS-PUB-2016-017, 2016, URL: <https://cds.cern.ch/record/2206965> (cit. on p. 3).
- [25] ATLAS Collaboration, *Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions*, [*Eur. Phys. J. C* **81** \(2020\) 334](#), arXiv: [2009.04986 \[hep-ex\]](#) (cit. on p. 4).
- [26] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, [*Eur. Phys. J. C* **77** \(2017\) 466](#), arXiv: [1703.10485 \[hep-ex\]](#) (cit. on p. 4).
- [27] ATLAS Collaboration, *Improving jet substructure performance in ATLAS using Track-CaloClusters*, ATL-PHYS-PUB-2017-015, 2017, URL: <https://cds.cern.ch/record/2275636> (cit. on p. 4).
- [28] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, [*JHEP* **04** \(2008\) 063](#), arXiv: [0802.1189 \[hep-ph\]](#) (cit. on p. 4).
- [29] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, [*Eur. Phys. J. C* **72** \(2012\) 1896](#), arXiv: [1111.6097 \[hep-ph\]](#) (cit. on p. 4).
- [30] A. J. Larkoski, S. Marzani, G. Soyez and J. Thaler, *Soft Drop*, [*JHEP* **05** \(2014\) 146](#), arXiv: [1402.2657 \[hep-ph\]](#) (cit. on p. 4).
- [31] M. Dasgupta, A. Fregoso, S. Marzani and G. P. Salam, *Towards an understanding of jet substructure*, [*JHEP* **09** \(2013\) 029](#), arXiv: [1307.0007 \[hep-ph\]](#) (cit. on p. 4).
- [32] P. Berta, M. Spousta, D. W. Miller and R. Leitner, *Particle-level pileup subtraction for jets and jet shapes*, [*JHEP* **06** \(2014\) 092](#), arXiv: [1403.3108 \[hep-ex\]](#) (cit. on p. 4).
- [33] M. Cacciari, G. P. Salam and G. Soyez, *SoftKiller, a particle-level pileup removal method*, [*Eur. Phys. J. C* **75** \(2015\) 59](#), arXiv: [1407.0408 \[hep-ph\]](#) (cit. on p. 4).

- [34] M. Cacciari, G. P. Salam and G. Soyez, *The catchment area of jets*, **JHEP** **2008** (2008) 005 (cit. on p. 4).
- [35] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014, 2020, URL: <https://cds.cern.ch/record/2718948> (cit. on p. 4).
- [36] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, **Eur. Phys. J. C** **77** (2017) 673, arXiv: [1704.07983](https://arxiv.org/abs/1704.07983) [[hep-ex](#)] (cit. on p. 4).
- [37] D. Krohn, J. Thaler and L.-T. Wang, *Jets with variable R* , **JHEP** **09** (2009) 059 (cit. on p. 4).
- [38] ATLAS Collaboration, *Performance of b -jet identification in the ATLAS experiment*, **JINST** **11** (2016) P04008, arXiv: [1512.01094](https://arxiv.org/abs/1512.01094) [[hep-ex](#)] (cit. on p. 5).
- [39] A. Vaswani et al., *Attention is all you need*, Advances in neural information processing systems **30** (2017) (cit. on p. 5).
- [40] M. Zaheer et al., *Deep Sets*, 2018, arXiv: [1703.06114](https://arxiv.org/abs/1703.06114) [[cs.LG](#)] (cit. on p. 5).
- [41] S. Shleifer, J. Weston and M. Ott, *Normformer: Improved transformer pretraining with extra normalization*, arXiv preprint arXiv:2110.09456 (2021) (cit. on p. 5).
- [42] J. L. Ba, J. R. Kiros and G. E. Hinton, *Layer Normalization*, 2016, arXiv: [1607.06450](https://arxiv.org/abs/1607.06450) [[stat.ML](#)] (cit. on p. 5).
- [43] ATLAS Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, (2022), arXiv: [2211.16345](https://arxiv.org/abs/2211.16345) [[physics.data-an](#)] (cit. on p. 8).
- [44] ATLAS Collaboration, *Search for heavy particles in the b -tagged dijet mass distribution with additional b -tagged jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, **Phys. Rev. D** **105** (2021) 012001, arXiv: [2108.07586](https://arxiv.org/abs/2108.07586) [[hep-ex](#)] (cit. on p. 11).
- [45] ATLAS Collaboration, *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2021-035, 2021, URL: <https://cds.cern.ch/record/2777861> (cit. on p. 11).
- [46] A. Collaboration, *Athena*, version 21.0.127, 2021, URL: <https://doi.org/10.5281/zenodo.4772550> (cit. on p. 13).