



**ATLAS PUB Note**  
ATL-PHYS-PUB-2022-042  
24th August 2022



# DeXTer: Deep Sets based Neural Networks for Low- $p_T$ $X \rightarrow b\bar{b}$ Identification in ATLAS

The ATLAS Collaboration

Several flavor tagging algorithms exist in ATLAS to identify jets containing two  $b$ -hadrons. These *double- $b$  tagger* algorithms focus on high transverse-momentum jets, usually above 200 GeV. This work describes the development of a new double- $b$  tagger for jets with transverse momenta below 200 GeV. The algorithm relies on large-radius track-jets which can be reconstructed at low transverse momenta and implements a neural network architecture based on Deep Sets that uses displaced tracks, secondary vertices, and substructure information to identify the presence of multiple  $b$ -hadrons. A measurement of the efficiency of the algorithm is performed in  $t\bar{t}$  and  $Z + \text{jets}$  events using the collision data from the Large Hadron Collider at  $\sqrt{s} = 13$  TeV center-of-mass energy recorded with the ATLAS detector between 2015 and 2018, corresponding to an integrated luminosity of  $139 \text{ fb}^{-1}$ .

# 1 Introduction

Several Standard model (SM) and beyond the Standard model (BSM) processes produce final states with two or more collimated particle jets that are reconstructed as a single, merged jet in the detector. Once considered a challenging object, the excellent spatial resolution of modern trackers and calorimeters makes it possible to partially resolve different particle jets merged into a single reconstructed jet [1]. Resolving individual particle jets relies on substructure patterns formed when the jets are clustered together in a single reconstructed jet. In order to properly identify the substructure of merged jets, it is important to have information about all particles from the constituent particle jets.

Substructure identification algorithms are usually applied to jets clustered with large radius  $R$  parameters (large- $R$ ), and several algorithms have been developed by the ATLAS collaboration for  $R = 1.0$  jets [2]. Large-radius jets are used to capture the constituents of several particle jets, but can also include significant contributions from long-range phenomena or from simultaneous inelastic proton–proton interactions (pileup). These additional contributions have typically low transverse momentum ( $p_T$ ), but when integrated over a large jet area, they can have a non-negligible contribution to the jet energy scale. Several dedicated techniques have been developed to address these challenges [3], relying on the expected behavior of QCD showers to trim off regions of the jet with a small relative contribution to the jet  $p_T$ . The trimming of low  $p_T$  particles works better if the particles associated to the particle-jets are, themselves, of relatively large  $p_T$ . Thus, large- $R$  reconstructed jets are only defined and calibrated for  $p_T > 200$  GeV, rendering most substructure techniques inaccessible to analyses with final states with low- $p_T$ , merged jets. A particularly interesting case for the identification of merged reconstructed jets is when at least one of the particle jets comes from the hadronization of a heavy-flavored quark. In this case, the traditional identification of long-lived  $b$ - and  $c$ -hadrons can be used together with substructure techniques to resolve merged jets. This idea has been explored in algorithms for flavor tagging of reconstructed jets with multiple particle jets, like in the recent neural-network-based  $X \rightarrow b\bar{b}$  tagger [4].

The first half of this note describes the development of a new algorithm that identifies merged heavy-flavored particle jets inside a single reconstructed jets at low  $p_T < 200$  GeV. Instead of relying on large- $R$  calorimeter jets, the algorithm uses standard particle-flow jets [5] and an associated large- $R$  track jet that can be well-defined at low  $p_T$  values. The new algorithm is called **DEXTER** (**D**eep set **X**  $\rightarrow$   $b\bar{b}$  **T**agger). DEXTER uses information from reconstructed tracks and vertices in a Deep Set [6] Neural Network (NN) to identify merged heavy-flavored jets. The development of DEXTER is motivated by several BSM physics searches where the  $p_T$  of a  $b\bar{b}$  pair is constrained by the mass of an intermediate resonance. Examples of such a BSM scenario are exotic Higgs boson decays to a pair of new low-mass particles that subsequently decay to one or more  $b\bar{b}$  pairs [7]. This decay chain is present in models with an additional scalar  $H \rightarrow ss$  or pseudoscalar  $H \rightarrow aa$ , as well as in models with Higgs boson decays to dark photons  $H \rightarrow \gamma_D \gamma_D$  [8]. In some BSM models, new low-mass scalars and pseudoscalars can be produced in association with top-quark pairs [9]. In the SM, the production of low- $p_T$   $b\bar{b}$  pairs occurs mostly in gluon splitting  $g \rightarrow b\bar{b}$  processes. When produced in association with a top-quark pair, this process is an important background for  $t\bar{t}H$  measurements. DEXTER may be used to improve background rejection in  $t\bar{t}H$  analyses and help improve the direct measurement of the top-quark Yukawa coupling. The second half of the note describes the measurement of the DEXTER efficiency using data from Run 2 of the LHC. The efficiency is measured in  $t\bar{t}$  and  $Z + \text{jets}$  events, which provide large samples of merged and non-merged heavy-flavored jets.

## 2 ATLAS detector

The ATLAS experiment [10] at the LHC is a multipurpose particle detector with a forward–backward symmetric cylindrical geometry and a near  $4\pi$  coverage in solid angle<sup>1</sup>. It consists of an inner tracking detector (ID) surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field, electromagnetic and hadron calorimeters, and a muon spectrometer (MS) in a toroid magnet.

The ID is comprised of silicon pixel detector which covers the vertex region and typically provides four measurements per track. The innermost layer, known as the Insertable B-Layer (IBL), was added in 2014 and provides hits at small radius to improve the tracking performance. The silicon pixel detector is surrounded by a silicon microstrip tracker (SCT) that typically provides eight measurements. These silicon detectors are complemented by a transition radiation tracker (TRT), which enables radially extended track reconstruction up to the pseudorapidity  $|\eta| = 2.0$ . The TRT also provides electron identification information based on the fraction of hits above a higher energy-deposit threshold corresponding to transition radiation. The ID is immersed in a 2T axial magnetic field and provides charged-particle tracking in the pseudorapidity range  $|\eta| < 2.5$ .

Lead/liquid-argon (LAr) sampling calorimeters provide electromagnetic (EM) energy measurements with high granularity. A steel/scintillator-tile hadron calorimeter covers the central pseudorapidity range ( $|\eta| < 1.7$ ). The endcap and forward regions are instrumented with LAr calorimeters for both the EM and hadronic energy measurements up to  $|\eta| = 4.9$ . The MS surrounds the calorimeters and is based on three large superconducting air-core toroidal magnets with eight coils each. The field integral of the toroids ranges between 2.0 and 6.0 T m across most of the detector. The muon spectrometer includes a system of precision tracking chambers and fast detectors for triggering. A two-level trigger system is used to select events. The first-level trigger is implemented in hardware and uses a subset of the detector information to accept events at a rate below 100 kHz. This is followed by a software-based trigger that reduces the accepted event rate to 1 kHz on average depending on the data-taking conditions. An extensive software suite [11] is used in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

## 3 Object Reconstruction

Tracks are reconstructed in the ATLAS ID from clusters of energy deposits in the pixel and SCT detectors. Especially in dense environments inside jets, these clusters can be shared by several tracks [12]. Seeds are formed from triplets of clusters using a loose compatibility criterion to ensure high reconstruction efficiency. A Kalman filter is used to build tracks by extending seeds to additional clusters compatible with a track trajectory in the magnetic field and multiple scattering in the detector material [13]. Tracks are used to reconstruct vertices based on a pattern recognition algorithm and an adaptive vertex fitter with annealing [14]. The DEXTER algorithm uses tracks that satisfy a *loose* selection criteria, defined as  $p_T > 0.5$  GeV,  $|d_0| < 3.5$  mm, and  $|z_0 \sin \theta| < 5$  mm.

---

<sup>1</sup> ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the  $z$ -axis along the beam pipe. The  $x$ -axis points from the IP to the centre of the LHC ring, and the  $y$ -axis points upwards. Cylindrical coordinates  $(r, \phi)$  are used in the transverse plane. The pseudorapidity is defined in terms of the polar angle  $\theta$  as  $\eta = -\ln \tan(\theta/2)$ . Angular distance is measured in units of  $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ .

DEXTER uses standard ATLAS particle-flow (PFlow) jets reconstructed using the anti- $k_t$  algorithm with radius parameter  $R = 0.4$  [15]. The inputs to the clustering algorithm are determined with a PFlow algorithm that matches well-measured tracks to calorimeter topological clusters. Tracks consistent with the primary vertex of the event but not associated to reconstructed muons, and topological clusters not matched to tracks are clustered into jets.

The uncalibrated four-momentum of a jet is estimated as the sum of the four-momenta of the constituent tracks and topological clusters. The jet energy is corrected for pileup effects and simulations are used to calibrate the average jet energy response to the particle-jet level. Then, a global sequential correction method is used to improve the jet energy resolution by correcting the jet energy scale as a function of the tracks and topological cluster shapes without changing the average scale [16]. These set of corrections are important to account for variations coming from the flavor of the jet and the specific hadron composition coming from jet fragmentation. Finally, a residual *in situ* calibration is applied to data to correct the residual difference between data and MC simulations [17].

### 3.1 Track-jet Reconstruction

The radius parameter  $R = 0.4$  provides good containment for particles created in the fragmentation of a jet originating from the hadronization of a single parton. In order to capture the particles from the fragmentation or decay of heavy-flavored hadrons from multiple partons which may escape the  $R = 0.4$  jet, DEXTER associates an extended collection of tracks to a reconstructed jet by clustering all PFlow jets and ID tracks matched to the jets using the ghost-association method [18]. This method treats the tracks as four-vectors of infinitesimal magnitude during the jet reconstruction and assigns them to the jet with which it is clustered using an anti- $k_t$  algorithm with radius parameter  $R = 0.8$ . This process is called reclustering [19] and, here, its purpose is two-fold. First, it creates large- $R$  track-jets around each PFlow reconstructed jet. Second, it creates a technical definition of jet isolation. An isolated jet is defined here as a reclustered jet with a single PFlow jet constituent. The radius parameter  $R = 0.8$  is chosen as a compromise between capturing most of the tracks from heavy-flavored hadron decays and not including unnecessary uncorrelated tracks that cannot be removed with trimming procedures at this momentum range. The same radius parameter has been used successfully in a previous version of this work [20]. The  $R = 0.8$  track jets are used as basic input to the DEXTER algorithm.

### 3.2 Multiple Secondary Vertex Reconstruction

Secondary vertices (SV) in jets have been explored by many flavor tagging algorithms in ATLAS [21, 22]. These algorithms either reconstruct the cascade decay of detached  $b$ -hadron vertex and subsequent decay vertex to  $c$ -hadron, or only try to find the displaced vertex of a  $b$ -hadron.

To explore the multiple heavy-flavor decays with DEXTER, an algorithm to reconstruct multiple decay vertices was developed by combining two existing algorithms: the track-cluster-based low- $p_T$  vertex tagger (TC-LVT) [23], and the multiple secondary vertex finder algorithm (MSVF) [24].

The TC-LVT algorithm has been developed for soft  $b$ -hadron tagging and optimized to reconstruct low- $p_T$   $b$ -hadron decays. This work uses the clustering algorithm from TC-LVT to identify the collection of tracks that may have at least one displaced secondary vertex. The MSVF algorithm is used to identify multiple SVs in the track cluster. The algorithm builds all two-track proto-vertices consistent with displaced tracks that are not compatible with a hadronic material interaction, a photon conversion, or the decay

of a long-lived light-flavored hadrons. All displaced tracks reconstructed in the ID are used to build proto-vertices. Proto-vertices define relations between tracks. A single track can be part of more than one proto-vertex, and therefore be related to more than one other track. Each set of tracks that are mutually connected to each other form a secondary vertex. After secondary vertices are formed, tracks with small compatibility with the vertex are removed and the ambiguity caused by distant vertices sharing common tracks are resolved. Nearby vertices are also merged by the MSVF algorithm. Finally, reconstructed SVs are required to be  $\Delta R$ -matched to a reclustered  $R = 0.8$  track-jet.

### 3.3 Track Subjet Reconstruction

One of the main goals of substructure methods is to identify the flight axis of the several particle jets that are merged in a single reconstructed jet. Several algorithms have been developed for that purpose [25] with different performance and algorithmic complexity. A good compromise is obtained with the exclusive- $k_t$  clustering of the jet constituents [26]. The exclusive- $k_t$  algorithm is a sequential clustering algorithm that compares the relative  $k_t$ -distance  $\min(p_{T,i}, p_{T,j}) \times \Delta R_{ij}$  between pairs of components  $(i, j)$  and the so-called beam distance  $p_{T,i}$ . If the smallest value in the set is the beam distance, the component is removed. On the other hand, if the smallest value is the  $k_t$ -distance, then components  $i$  and  $j$  are clustered together in a pseudo-jet. The algorithm then iterates over the merged pseudo-jet and stops when a well-defined number of pseudo-jets remain, which are the subjets used to estimate the flight direction of the two particle-jets merged together.

DEXTER uses as inputs the direction of the subjets formed using the exclusive- $k_t$  algorithm on the tracks from the reclustered  $R = 0.8$  track jet, stopping when two pseudo-jets are left. In this note, this particular instance of the algorithm is referred to as  $\text{Ex}k_t^{(2)}$ .

### 3.4 Jet Selection

DEXTER is optimized for isolated jets obtained when reclustering calibrated PFlow jets with  $p_T > 20$  GeV and  $|\eta| < 2.5$ . The associated  $R = 0.8$  track jet is required to satisfy  $|\eta| < 2.0$  to account for the extended radius and the acceptance of the ID. Furthermore, each  $\text{Ex}k_t^{(2)}$  subjet is required to satisfy  $p_T > 5$  GeV where the track-subjet  $p_T$  is estimated from the sum of the tracks' four-momenta.

### 3.5 Leptons

Electron candidates are reconstructed from energy deposits (clusters) in the electromagnetic calorimeter associated to reconstructed tracks in the ID. Muon candidates are reconstructed from track segments in the various layers of the MS, and matched with tracks from the ID.

Leptons are required to satisfy  $p_T > 10$  GeV and  $|\eta| < 2.5$ . For electrons, the regions  $1.37 < |\eta| < 1.52$  and  $|\eta| > 2.47$  are excluded. Electrons are required to satisfy a tight quality requirement based on a likelihood calculated with shower shapes [27]. Muons are required to satisfy a medium quality requirements based on the number of hits on the MS and ID, as well as on the quality of the combined fitted track [28].

Selected leptons must satisfy a isolation requirement based on a multivariate algorithm that takes into account particles in the vicinity of leptons candidates as well as information about displaced tracks [29]. Electrons (muons) are further required to have  $|z_0 \sin \theta| < 0.5$  mm and  $|d_0/\sigma(d_0)| < 5(3)$ .

To reduce the background from muons from heavy-flavor decays inside jets, muons are required to be separated by  $\Delta R > 0.4$  from the nearest jet. If a muon with  $p_T > 4$  GeV is within  $\Delta R < 0.4$  of a jet, it is called a *soft-muon* if the jet has at least three associated tracks. If a jet has less than three tracks and a nearby muon, the jet is removed. Soft-muons are not used to identify events, but they are used to aid jet classification in data events.

## 4 Jet Truth Labeling and Simulation

Events from Monte Carlo (MC) simulation are used to develop and optimize the `DEXTER` algorithm. Different samples are used to select reconstructed jets with either zero, one or two heavy-flavor particle jets merged together. The true flavor of reconstructed jets is determined by the number of  $b$ - and  $c$ -hadrons with  $p_T > 5$  GeV ghost-associated [30] to the large- $R$  track-jet.  $b$ -hadrons ( $c$ -hadrons) in the decay chain of another  $b$ -hadron ( $c$ -hadron, respectively) are not ghost-associated to avoid double-counting. Jets are categorized as a  $B$ -labeled jet if at least two  $b$ -hadrons are ghost-associated to the jet. If exactly one  $b$ -hadron and no  $c$ -hadron are ghost-associated to the jet, it is categorized as a  $b$ -labeled jet. Jets with no ghost-associated  $b$ -hadron, but at least one ghost-associated  $c$ -hadron, are categorized as  $c$ -labeled jets. If no  $b$ - or  $c$ -hadron is ghost-associated, the reconstructed jet is categorized as a light-labeled jet. Only jets with  $B$ -,  $b$ -, and light-labeled are used in the training.

Multiple MC samples are used to obtain jets with different labels spanning a kinematic range.  $B$ -labeled jets are obtained from a suite of processes including  $VH$  with  $H \rightarrow aa \rightarrow (b\bar{b})(b\bar{b})$  and  $t\bar{t}a$  with  $a \rightarrow (b\bar{b})$  with different values of the mass of the resonance ( $m_a$ ). The  $VH$  samples are simulated at next-to-leading order (NLO) in perturbation theory with the `POWHEG HVJ` [31] process where the `MINLO` method is used to merge the 0- and 1-jet contributions. The gluon-initiated loop-induced  $ZH$  process is also generated with `POWHEG`, but at leading order (LO) in perturbation theory and without additional jets at matrix-element (ME) level. For both of these samples, the ME are interfaced with the parton distribution function (PDF) set `NNPDF3` [32] and with `PYTHIA8.186` [33] for showering, hadronization, and description of underlying events. The  $t\bar{t}a$  events are simulated with `MADGRAPH5_AMC@NLO 2.8.1` [34] at NLO in QCD perturbation theory. The PDF set `NNPDF2.3` [35] at LO is used and are also interfaced with `PYTHIA8`.

A  $Z + \text{jets}$  sample is used as source of  $B$ -labeled jets from  $g \rightarrow b\bar{b}$ , which are similar to those coming from  $a \rightarrow b\bar{b}$  but originating from a off-shell, color-octet, state. It is simulated with the `SHERPA 2.2.1` [36] generator using NLO ME for up to two partons, and LO ME for up to four partons calculated with the `Comix` [37] and `OPENLOOPS` [38–40] libraries. They are matched with the `SHERPA` parton shower [41] using the `MEPS@NLO` prescription [42–45] with the set of tuned parameters developed by the `SHERPA` authors. The `NNPDF3.0 NNLO` set of PDFs [32] is used and the samples are normalized to a next-to-next-to-leading-order (NNLO) prediction of the total cross-section.

Finally, a sample of top-quark pair ( $t\bar{t}$ ) events is used to select  $b$ -,  $c$ -, and light-labeled jets. The production of  $t\bar{t}$  events is modelled using the `PowhegBox` [46] v2 generator at NLO in QCD perturbation theory in the five flavor scheme. The  $h_{\text{damp}}$  parameter is set to  $1.5 \times m_{\text{top}}$  [47], and the functional form of the renormalization and factorization scale is set to  $\sqrt{m_{\text{top}}^2 + p_T^2}$ . Events are interfaced to `PYTHIA 8.186` [33] for the modelling of the parton shower, hadronization, and underlying events.

While developing and optimizing the `DEXTER` algorithm, samples of  $B$ -labeled and  $b$ -labeled jets obtained from the simulations described above are reweighted to the same  $p_T$  and  $\eta$  spectra of the light-labeled jets in

order to not use the different kinematics for flavor tagging. When performing the efficiency measurements in data, the simulated event samples are used to model SM processes and the expected flavor distribution of selected jets.

Additional samples are used to assess modeling uncertainties. An alternative  $t\bar{t}$  sample is simulated with the same configuration as the nominal one, but uses **HERWIG7** for shower and hadronization. The alternative  $Z + \text{jets}$  sample is simulated with **MADGRAPH5\_AMC@NLO 2.2.2** [34], using LO-accurate ME with up to four final-state partons. The ME calculation employed the NNPDF3.0NLO set of PDFs [32]. Events were interfaced to **PYTHIA 8.186** [33] for the modelling of the parton shower, hadronization, and underlying events. The overlap between ME and parton shower emissions was removed using the CKKW-L merging procedure [48, 49]. The A14 tune [50] of **PYTHIA 8** was used with the NNPDF2.3 PDF set [35]. The decays of bottom and charm hadrons were performed with **EVTGEN 1.2.0** [51].

Single-top t- and s-channel and  $tW$  production is modelled using the **POWHEGBox v2** [52–54] generator which provides ME at NLO in the strong coupling constant  $\alpha_s$ . For t-channel production, events are generated in the four flavor scheme with the NNPDF3.0 PDF set. For s-channel and  $tW$  production, events are generated in the five flavor scheme with the NNPDF3.0 PDF set. For  $tW$  production, the diagram removal scheme [55] was employed to handle the interference with  $t\bar{t}$  production [47]. Several rare processes were also considered but have negligible impact on the measurement, including  $t\bar{t}V$ ,  $t\bar{t}H$ ,  $t\bar{t}t\bar{t}$ ,  $W + \text{jets}$  and  $VV$  processes.

All simulated samples are processed using a **GEANT4** description of the ATLAS detector. The effect of pileup is modelled by overlaying the simulated hard-scattering event with inelastic proton–proton ( $pp$ ) events generated with **PYTHIA 8.186** [33] using the NNPDF2.3lo set of PDF [35] and the A3 tune [56].

## 5 DEXTER Algorithm

DEXTER is inspired by recent developments using the Deep Sets architecture to develop a permutation-invariant neural network (NN) for flavor tagging in ATLAS. The set-based impact-parameter  $b$ -tagging algorithm (DIPS) [57] uses track information as input for jet flavor tagging. This approach minimizes the empirical ordering typically introduced in the prior algorithm on tracks associated to jets. The same architecture has also been used for energy-flow and particle-flow event classification [58, 59].

The DEXTER algorithm extends this architecture to incorporate extra information from reconstructed secondary vertices and kinematical variables from the PFlow jet. The architecture comprises two separate feed-forward NNs which serve as the feature extractor for the tracks in the  $R = 0.8$  track jet and for the SVs. An additional global feed-forward NN combines the output of the feature-extraction NNs with the jet kinematics to learn correlations between them. The calibrated energy-momentum of the single PFlow jet in the isolated jet is used as estimate of the jet kinematics for the NN. A final layer with a softmax<sup>2</sup> activation function [60] is used to predict the probability for each flavor label. Figure 1 summarizes the structure of the complete algorithm.

The feature-extraction NNs contain 2 hidden layers with 100 neurons and an output layer with 128 features. The 256 features are used as input to the global NN which has 3 hidden layers. All NNs use the ELU activation function [61]. The global NN also receives the PFlow jet  $p_T$  and  $\eta$  as inputs.

<sup>2</sup> The softmax function is defined as  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\sigma(\mathbf{z})_i = e^{z_i} / \sum_{j=1}^n e^{z_j}$  for  $i = 1, \dots, n$  element of input vector  $\mathbf{z}$

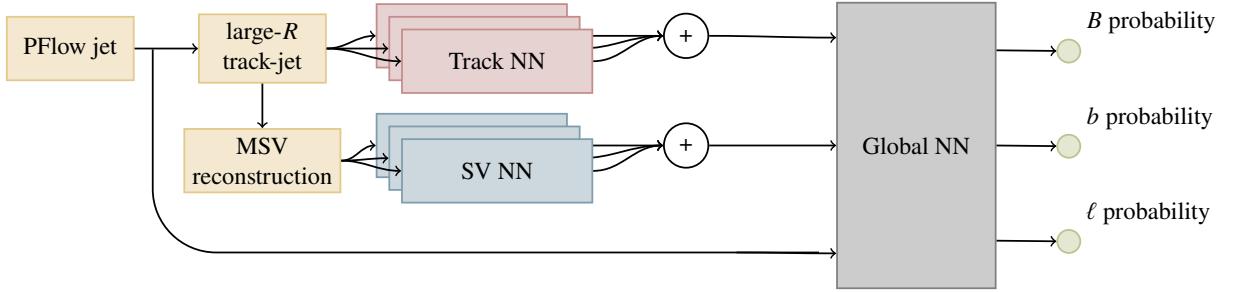


Figure 1: Architecture of the DEXTER algorithm. Two sets of feature-extracting NN are used for track and SV observables. A final global NN with a softmax output layer is used to interpret each output as a probability for each flavor.

## 5.1 Track Neural Network

The tracks clustered in the  $R = 0.8$  track-jet are ordered in decreasing value of transverse impact parameter significance  $S_{d_0}$  [62]. Properties of the 25 first tracks are used as input to the DEXTER algorithm. The track NN algorithm input variables are presented in Table 1. Variables that depend on both the track and the reconstructed jet, namely  $\Delta\eta(\text{track}, \text{jet})$  and  $\Delta\phi(\text{track}, \text{jet})$ , are calculated with respect to the axis of the  $\text{Ex}k_t^{(2)}$  track-subjet, since it is a better estimate of the original  $b$ -hadron flight direction than the PFlow jet. The sign of the impact parameter significance is not altered because the PFlow jet axis still represents the best estimate for reconstructed jets that do not come from merged particle jets.

## 5.2 Secondary Vertex Neural Network

The reconstructed secondary vertices matched to the  $R = 0.8$  track-jet are ordered in decreasing value of the transverse decay length significance  $S_{L_{xy}}$  and the properties of the 12 first vertices are used as input to the secondary vertex NN. The properties used are listed in Table 1. The definition of the variables is identical to the one used in other SV-based  $b$ -taggers [22], with the only differences being the choice of the  $\text{Ex}k_t^{(2)}$  jet axes as the reference for  $\Delta\eta(\text{track}, \text{jet})$  and  $\Delta\phi(\text{track}, \text{jet})$ . Decay lengths are always calculated with respect to the event PV:

$$L_{xy} = |(\vec{p}^{\text{SV}} - \vec{p}^{\text{PV}}) \times \hat{z}| \quad L_z = |(\vec{p}^{\text{SV}} - \vec{p}^{\text{PV}}) \cdot \hat{z}|.$$

The decay length significance values are calculated taking into account the covariance matrices of both primary and secondary vertices:

$$S_{L_{xy}} = (p^{\text{SV}} - p^{\text{PV}})_i [P_{xy}^{-1} (C^{\text{SV}} + C^{\text{PV}})^{-1} P_{xy}]_{ij} (p^{\text{SV}} - p^{\text{PV}})_j,$$

$$S_{L_z} = (p^{\text{SV}} - p^{\text{PV}})_i [P_z^{-1} (C^{\text{SV}} + C^{\text{PV}})^{-1} P_z]_{ij} (p^{\text{SV}} - p^{\text{PV}})_j.$$

where  $P_{xy}$  and  $P_z$  are projectors onto the transverse plane and longitudinal direction, respectively.

Even though an ordering is imposed to select the maximum number of input tracks and secondary vertices, the NN itself is permutation invariant.

Feature	Description
Jet	
$p_T$	Jet transverse momentum
$\eta$	Jet pseudorapidity
Track	
$\log p_T^{\text{frac}}(\text{track, PFlow jet})$	$\log p_T^{\text{track}}/p_T^{\text{PFlow jet}}$ of the $p_T$ fraction between track and PFlow jet $p_T$
$\Delta\eta$ (track, $\text{Ex}k_t^{(2)}$ track jet)	Pseudorapidity difference between track and $\text{Ex}k_t^{(2)}$ jet
$\Delta\phi$ (track, $\text{Ex}k_t^{(2)}$ track jet)	Angular difference between between track and $\text{Ex}k_t^{(2)}$ jet
$d_0$	Transverse impact parameter
$z_0 \sin \theta$	Longitudinal impact parameter
$S_{d_0}$	$d_0/\sigma_0$ : transverse IP significance
$S_{z_0 \sin \theta}$	$z_0 \sin \theta/\sigma_{z_0 \sin \theta}$ : longitudinal IP significance
PIX1 hits	Number of hits in the first pixel layer
IBL hits	Number of hits in the IBL
Shared IBL Hits	Number of shared hits in the IBL
Split IBL Hits	Number of split hits in the IBL
Shared pixel hits	Number of shared hits in the pixel layers
Split pixel hits	Number of split hits in the pixel layers
Shared SCT hits	Number of shared hits in the SCT
nPixHits	Number of hits in the pixel layers
nSCTHits	Number of hits in the SCT layers
Secondary Vertex	
$\log(m)$	Track mass of the secondary vertex
$\log p_T^{\text{frac}}(\text{vertex, PFlow jet})$	$\log p_T^{\text{SV}}/p_T^{\text{PFlow jet}}$ of the $p_T$ fraction between the secondary vertex and PFlow $p_T$ jet
$\Delta\eta$ (vertex, $\text{Ex}k_t^{(2)}$ track jet)	Pseudorapidity difference between the secondary vertex and the $\text{Ex}k_t^{(2)}$ jet
$\Delta\phi$ (vertex, $\text{Ex}k_t^{(2)}$ track jet)	Angular difference between between the secondary vertex and the $\text{Ex}k_t^{(2)}$ jet
$L_{xy}$	Transverse decay length relative to primary vertex
$L_z$	Longitudinal decay length relative to primary vertex
$S_{L_{xy}}$	Transverse decay length significance
$S_{L_z}$	Longitudinal decay length significance

Table 1: List of features used as input for the DEXTER algorithm. In the table,  $\text{Ex}k_t^{(2)}$  jet refers to track subjets reconstructed as described in Section 3.3.

### 5.3 Color Charge Adversarial Neural Network

The algorithm obtained with the NN depicted in Figure 1 would exhibit significant differences in the response for color singlet and octet  $b\bar{b}$  states. The difference stems from the larger track multiplicity observed in the shower from octet states and from the track distribution from the color interaction between two octet states [63]. This difference could be used to create optimized taggers for the singlet and octet cases. However, since the only calibration sample available comes from  $g \rightarrow b\bar{b}$ , a choice is made to minimize the differences by penalizing the feature extracting NNs using a domain-adversarial training of the NN (DANN) [64].

The  $B$ -labeled jet sample used for training is an ensemble mixture of  $a \rightarrow b\bar{b}$  and  $g \rightarrow b\bar{b}$  events. The adversarial NN receives the same inputs as the global NN and is built with a categorical loss  $L_A$  that discriminates between the two categories. The loss function  $L_D$  used to discriminate  $B$ -,  $b$ -, and light-labeled jets is built from the three output neurons on the global NN. While the backpropagation of the global NN is performed only with  $\partial L_D / \partial \theta_g$  and the backpropagation of the adversarial NN is performed only with  $\partial L_A / \partial \theta_a$ , the backpropagation of the feature-extraction NNs is done via the gradient:

$$\frac{\partial L_D}{\partial \theta_f} - \lambda \frac{\partial L_A}{\partial \theta_f},$$

where  $\theta_g, \theta_a, \theta_f$  are the weights for the global, adversarial, and feature-extracting NN, respectively. The optimized value  $\lambda = 10$  is chosen for the regulator when training the DEXTER algorithm in order to minimize the difference in the response between color singlets and octets.

DEXTER was trained with 3.4 million  $B$ -,  $b$ -, and light-labeled jets. From this dataset, 20% of jets are held out as a validation set. The architecture is implemented using Keras [65] with the TensorFlow backend [66]. An early stopping method that monitors the improvement of the validation set loss is used during the training. The Adam optimizer [67] is used with a learning rate of 0.001 and batch size of 512. Another independent set of 0.8 million jets is kept out for the performance studies shown in Section 6.

## 6 Algorithm Performance

As described in Section 5, the DEXTER algorithm is trained to classify reconstructed jets into three different categories based on the extracted input features. The three-class probabilities predicted by the model ( $p_B, p_b, p_l$ ) are combined in a tagging discriminant,  $D_B$ , defined as:

$$D_B = \ln \frac{p_B}{(1 - f_b)p_l + f_b p_b} \quad (1)$$

where  $f_b$  is a free parameter that balances between the rejection of light-labeled jets versus  $b$ -labeled jets. The value  $f_b = 0.4$  is used in this note based on the flavor composition of the  $t\bar{t}$  sample. Different applications may optimize  $f_b$  for background rejection. A clear separation between different categories can be observed in Figure 2.

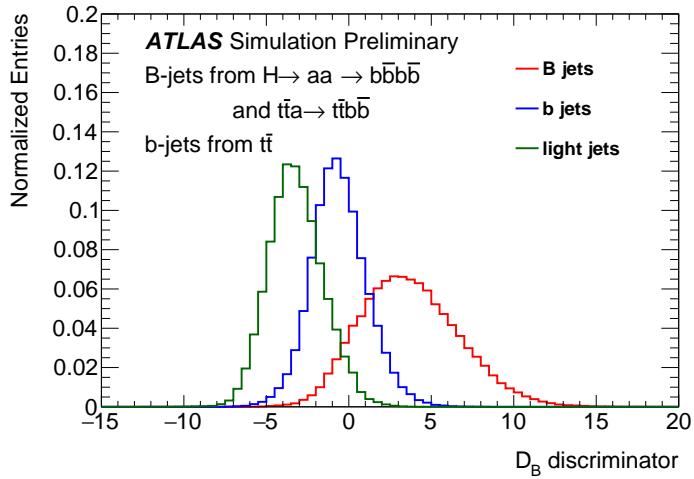


Figure 2: DEXTER discriminant distribution evaluated using  $B$ -labeled jets from  $H \rightarrow aa \rightarrow (b\bar{b})(b\bar{b})$  and  $t\bar{t}a$  with  $a \rightarrow (b\bar{b})$  and  $b$ -, and light-labeled jets from  $t\bar{t}$  samples with  $f_b = 0.4$ . The  $D_B$  score distributions are normalized to unity.

## 6.1 Performance Dependency on Resonance Color Charge

The exact response of the DEXTER algorithm depends on the adversarial penalty introduced to minimize the difference between the decay of color singlets and octets. Figure 3 compares the difference in response between the two types of decay to a  $b\bar{b}$  pair with and without the DANN. The value  $\lambda = 10$  was chosen to minimize the difference in the area under the receiver operating characteristic (ROC) curve for  $a \rightarrow b\bar{b}$  and  $g \rightarrow b\bar{b}$  events.

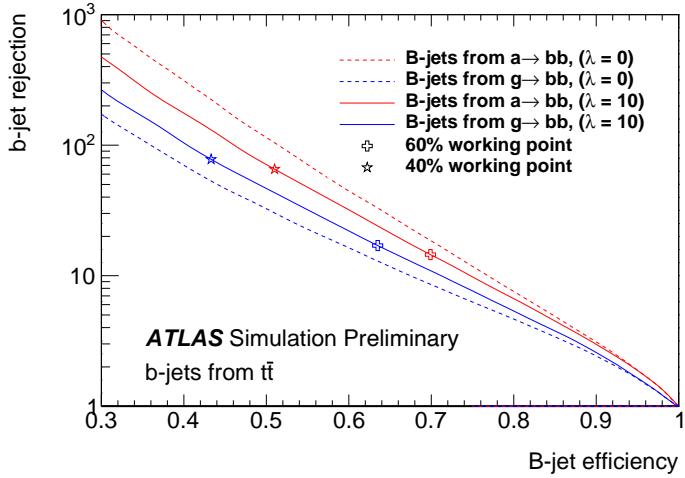


Figure 3: Comparison of the  $B$ -labeled jet tagging efficiency vs. the  $b$ -labeled jet rejection for  $B$ -jets originating from color singlets (red) and color octets (blue). The ROC curves are shown for architectures of DEXTER with (solid) and without (dashed) DANN. Two example working points are marked with labels corresponding to the approximate  $B$ -labeled jet efficiency for  $g \rightarrow b\bar{b}$ . The two working points are the boundaries of the intervals for which the efficiency are measured in data. The difference in efficiency for  $B$ -labeled jets between  $a \rightarrow b\bar{b}$  and  $g \rightarrow b\bar{b}$  is approximately 6 – 7% in the two working points.

## 6.2 Performance Dependency on Resonance Mass

Another desirable feature of the tagger is to be independent of the parent particle mass. Not only is it desirable for the tagger performance to be independent of unknown model parameters, but it is also fundamental when mass sidebands are used.

The PFlow jet mass is not a good proxy for the  $B$ -labeled jet particle-level mass since it doesn't provide good containment in the case of multiple particle-jets merged in the same reconstructed jet. Because of this, mass correlation is assessed using the true mass of the  $a$ -boson in decays  $a \rightarrow b\bar{b}$ .

Two strategies are adopted to minimize the mass correlation. First, the  $B$ -enriched sample used to train DEXTER is composed of an ensemble mix of  $H \rightarrow aa \rightarrow (b\bar{b})(b\bar{b})$  and  $t\bar{t}a, a \rightarrow b\bar{b}$  samples with different values of the  $a$ -boson mass. Second, the impact of each individual input variable with the mass correlation of the response is studied. When a variable is identified as the cause of a large difference in performance, it is either redefined to be less mass-sensitive, or removed when a suitable redefinition is not possible. The performance difference is examined by comparing the ROC curves across different  $m_a$  values in  $H \rightarrow aa \rightarrow (b\bar{b})(b\bar{b})$  decays, as shown in Figure 4, and  $t\bar{t}a, a \rightarrow b\bar{b}$  processes.

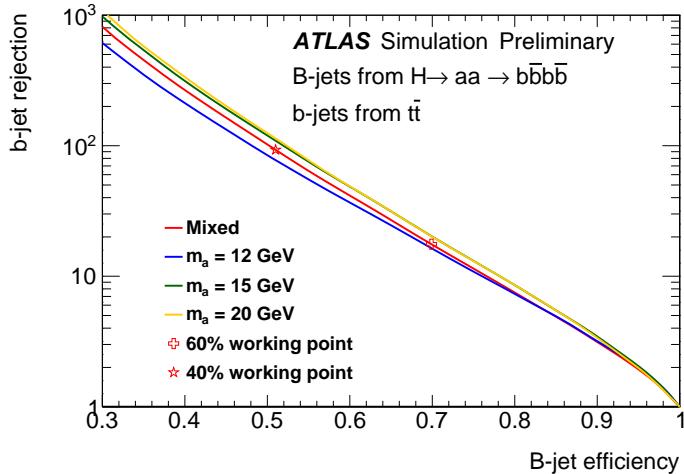


Figure 4: Performance for different  $a$ -boson masses in  $H \rightarrow aa \rightarrow (b\bar{b})(b\bar{b})$  events. The same working points defined in Figure 3 are marked. The  $B$ -labeled jet efficiency on the two reference working points differs by approximately 2 – 3% across the examined values of  $m_a$ .

## 7 Efficiency Measurements

The second half of this note describes the measurement of the algorithm efficiency with collider data. The efficiency measured with data can be used to correct the modeling in simulation and is essential for the use of DEXTER in physics analysis.

The measurement of the efficiency of  $X \rightarrow b\bar{b}$  taggers is challenging since it can be difficult to select merged  $B$ -flavored particle-jets in an unbiased way. A particularly clear decay topology that can be explored is  $Z(\rightarrow b\bar{b})\gamma$  [68]. However, this decay is only merged for  $p_T^Z \gtrsim 400$  GeV, outside the  $p_T$  range targeted by DEXTER. A previous version of this work [20] used  $g \rightarrow b\bar{b}$  processes in multijet events, but those events

require prescaled triggers, which limit the number of available events, and are challenging to simulate due to the large cross-section. This note described the calibration of jets tagged with the DEXTER algorithm using  $Z(\rightarrow \ell^+\ell^-) + g(\rightarrow b\bar{b})$  events as a source of  $B$ -labeled jets and  $t\bar{t}$  events as a source of  $b$ -labeled jets [69].

Efficiencies are measured in two tagging intervals. The first tagging interval is defined by  $D_B > 3.2$  and corresponds to efficiencies in the range 0 – 40%. The second tagging interval is defined as  $1.8 < D_B < 3.2$  and corresponds to average efficiencies in the range 40 – 60%. Jets with  $D_B < 1.8$  belong to the 60 – 100% tagging interval. The representative tagging efficiencies of each interval is determined using a  $g \rightarrow b\bar{b}$  sample and would be slightly higher if a  $a \rightarrow b\bar{b}$  sample were used, as shown in Figure 3. The 0 – 40% tagging interval has an average expected light-labeled jet efficiency of  $1/(2.4 \times 10^4)$ . For each tagging interval, the efficiency is measured in three exclusive  $p_T$  bins:  $20 < p_T^{\text{jet}} < 90 \text{ GeV}$ ,  $90 \leq p_T^{\text{jet}} < 140 \text{ GeV}$ , and  $140 \leq p_T^{\text{jet}} < 200 \text{ GeV}$ , for a total of 6 different measurements per jet label.

Figure 5 shows the efficiency observed in simulation for the 0 – 40% and 40 – 60% tagging intervals. Unless otherwise specified, all efficiencies for simulated samples are determined through truth matching of jets produced by PowHEGBox +PYTHIA8.

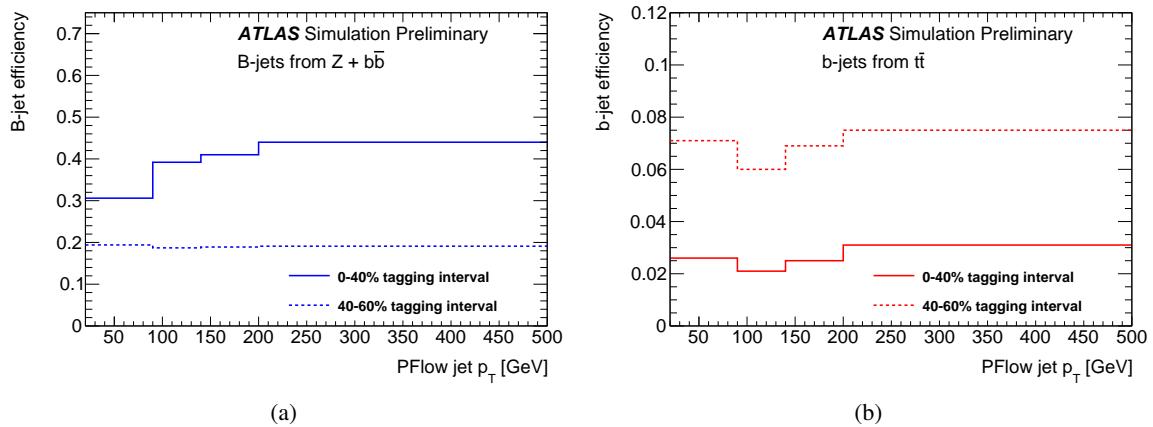


Figure 5: DEXTER efficiency in simulation for (a)  $B$ - and (b)  $b$ -labeled jets as a function of the jet  $p_T$ .

The efficiencies for  $B$ - and  $b$ -labeled jets in all  $p_T^{\text{jet}}$  ranges and tagging intervals are measured simultaneously. This allows to take into account the full correlation model in the propagation of uncertainties, which is not possible in other calibration strategies where tagging and mis-tagging efficiencies are measured separately [70].

This measurement uses data collected from the Large Hadron Collider at  $\sqrt{s} = 13 \text{ TeV}$  center-of-mass energy recorded with the ATLAS detector between 2015 and 2018, corresponding to an integrated luminosity of  $139 \text{ fb}^{-1}$  [71]. The data sample was selected with a set of single-electron [72] and single-muon triggers [73] that require the presence of at least one lepton with  $p_T > 26 \text{ GeV}$ . In 2015, the data were collected with lower instantaneous luminosity and events with electrons (muons) with  $p_T > 24$  (20) GeV were also recorded by the trigger system.

Simulated event samples are used to model SM processes and the expected flavor distribution of selected jets. The  $Z + \text{jets}$  and  $t\bar{t}$  samples are described in Section 4. Additional samples are used to assess modeling

uncertainties. An alternative  $t\bar{t}$  sample is simulated with the same configuration as the nominal one, but uses HERWIG7 for shower and hadronization.

The alternative  $Z + \text{jets}$  sample is simulated with MADGRAPH5\_AMC@NLO 2.2.2 [34], using LO-accurate ME with up to four final-state partons. The ME calculation employed the NNPDF3.0NLO set of PDFs [32]. Events were interfaced to PYTHIA 8.186 [33] for the modelling of the parton shower, hadronization, and underlying events. The overlap between ME and parton shower emissions was removed using the CKKW-L merging procedure [48, 49]. The A14 tune [50] of PYTHIA 8 was used with the NNPDF2.3 PDF set [35]. The decays of bottom and charm hadrons were performed with EvtGEN 1.2.0 [51].

Single-top t- and s-channel and  $tW$  production is modelled using the POWHEGBox v2 [52–54] generator which provides ME at NLO in the strong coupling constant  $\alpha_s$ . For t-channel production, events are generated in the four flavor scheme with the NNPDF3.0 PDF set. For s-channel and  $tW$  production, events are generated in the five flavor scheme with the NNPDF3.0 PDF set. For  $tW$  production, the diagram removal scheme [55] was employed to handle the interference with  $t\bar{t}$  production [47]. Events were interfaced to PYTHIA 8.230 for the modelling of the parton shower, hadronization, and underlying events.

Several rare processes were also considered but have negligible impact on the measurement, including  $t\bar{t}V$ ,  $t\bar{t}H$ ,  $t\bar{t}t\bar{t}$ ,  $W + \text{jets}$  and  $VV$  processes. Pileup interactions are modeled as described in Section 4. The distribution of simultaneous inelastic interactions in each individual simulated sample is reweighted to match the observed data.

## 8 Event Selection

This measurement uses data collected from the Large Hadron Collider at  $\sqrt{s} = 13$  TeV center-of-mass energy recorded with the ATLAS detector between 2015 and 2018, corresponding to an integrated luminosity of  $139 \text{ fb}^{-1}$  [71]. The data sample was selected with a set of single-electron [72] and single-muon triggers [73] that require the presence of at least one high- $p_T$  lepton. The trigger used between 2016 and 2018 required at least one electron or muon with  $p_T > 26$  GeV. In 2015, the data were collected with lower instantaneous luminosity and events with electrons (muons) with  $p_T > 24$  (20) GeV were also recorded.

Two regions are defined targeting different processes and jet flavor compositions. A top-region is defined by the presence of a high- $p_T$  electron and a high- $p_T$  muon produced in association with multiple jets to target top quark pair events. A  $Z$ -region is defined by the presence of a opposite-charge  $ee$  or  $\mu\mu$  pair with invariant mass around the  $Z$  pole mass. In both regions, the tight lepton identification suppresses any significant non-prompt background. Due to the use of single lepton triggers, the leading lepton is required to satisfy  $p_T > 27$  GeV in both top- and  $Z$ -region.

### 8.1 Top-region Event Selection and Channels

The definition of the top-region is inspired by the measurement of the  $b$ -jets identification efficiency with  $t\bar{t}$  events in ATLAS [69]. Candidate events are required to have exactly two leptons with opposite sign charge. To suppress the contamination from backgrounds with  $Z$ -bosons, events are required to have one electron and one muon. A selection on the mass of the two leptons  $m_{e\mu} > 50$  GeV is applied to reject events with low dilepton invariant mass which is not well modeled by the simulation. Events are also required to have

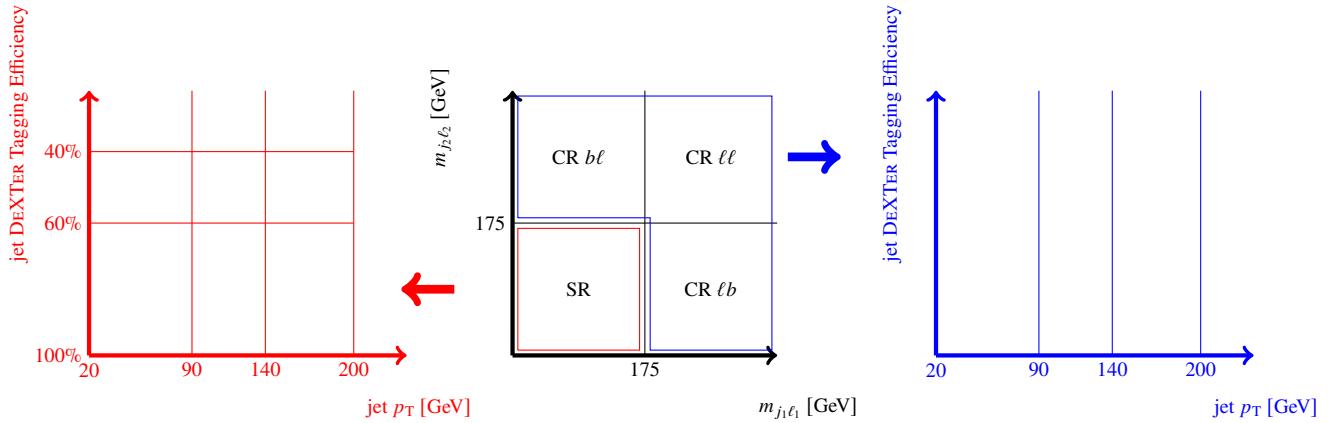


Figure 6: Diagram of the event selection and channel classification in the top-region.

exactly two reconstructed jets to suppress light-flavor jets from initial- or final-state radiation. Furthermore, a  $\Delta R(\text{jet, leptons}) > 0.8$  is applied to avoid leptons close to jets used to measure the efficiency.

A simple top-quark pair reconstruction is also adopted to achieve higher purity of two  $b$ -jets events from  $t\bar{t}$  decays. The electron and the muon are paired with the selected jets in the event using the following criterion:

$$\underset{\ell_1, \ell_2 \in \{e, \mu\}}{\operatorname{argmin}} (m_{j_1, \ell_1}^2 + m_{j_2, \ell_2}^2),$$

where  $j_1$  ( $j_2$ ) is the jet with highest (lowest)  $p_T$ . The invariant masses of the pairs formed are required to satisfy  $m_{j_1, \ell_1} > 20 \text{ GeV}$ , and  $m_{j_2, \ell_2} > 20 \text{ GeV}$  in order to avoid regions populated with low- $p_T$  jets.

The masses of the pairs  $(j_1, \ell_1)$  and  $(j_2, \ell_2)$  are used to define signal regions (SR) and control regions (CR):

- $m_{j_1, \ell_1} < 175 \text{ GeV}$  and  $m_{j_2, \ell_2} < 175 \text{ GeV}$  (SR  $bb$ ),
- $m_{j_1, \ell_1} < 175 \text{ GeV}$  and  $m_{j_2, \ell_2} > 175 \text{ GeV}$  (CR  $b\ell$ ),
- $m_{j_1, \ell_1} > 175 \text{ GeV}$  and  $m_{j_2, \ell_2} < 175 \text{ GeV}$  (CR  $\ell b$ ),
- $m_{j_1, \ell_1} > 175 \text{ GeV}$  and  $m_{j_2, \ell_2} > 175 \text{ GeV}$  (CR  $\ell\ell$ ).

Jets from events in the SR are further categorized depending on their  $p_T$  and on their DEXTER tagging interval so that the different efficiencies can be measured. Jets from events in the CR are further categorized only based on their  $p_T$ . These regions serve primarily to constrain the flavor composition of the selected jets by adding channels that, despite not being sensitive to the efficiency values, provide independent degrees of freedom. Figure 6 shows a diagrammatic representation of the event selection.

## 8.2 Z-region Event Selection and Channels

Events in the  $Z$ -region are required to have exactly two leptons, electrons or muons, with same flavor and opposite electric charge. The dilepton invariant mass is required to be in the  $Z$ -boson mass-pole region  $81 < m_{\ell\ell} < 101 \text{ GeV}$ . Events with exactly one *probe* jet, which is defined as the single jet selected in

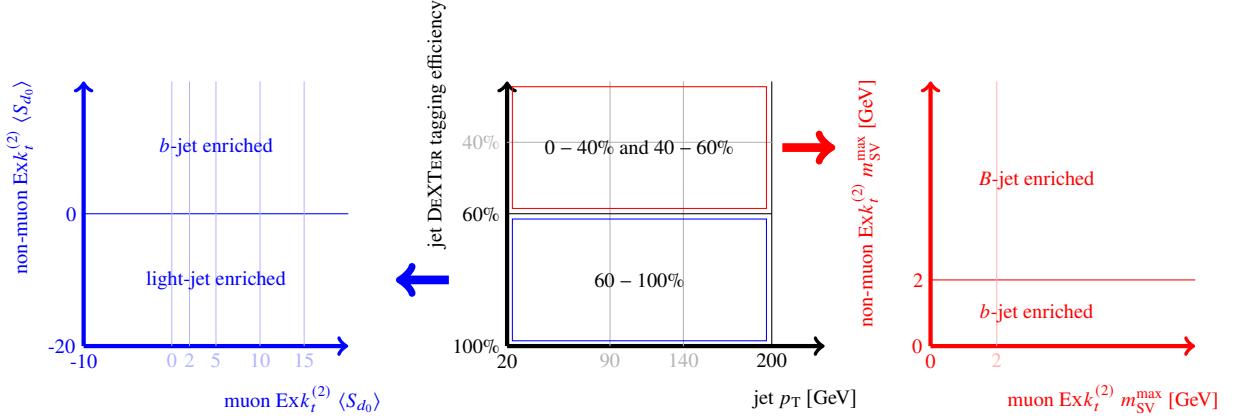


Figure 7: Diagram of the event selection and channel classification in the  $Z$ -region.

addition to the leptons and which is used to measure the efficiency of the DEXTER algorithm, are selected. A condition  $\Delta R(\text{jet}, \ell\ell) > 1.0$  is imposed on the angular separation between the probe jet and the dilepton pair to avoid mismodeling in the production of  $Z$  bosons with multiple jets. Another similar requirement  $\Delta R(\text{jet}, \text{lepton}) > 1.0$  is applied to avoid leptons too close to the probe that could impact the performance of the DEXTER algorithm.

In order to enrich the fraction of heavy-flavored jets in this region, the probe jet is required to have a soft-muon within  $\Delta R(\text{soft-muon}, \text{Ex}k_t^{(2)}) < 0.3$  of one of the two  $\text{Ex}k_t^{(2)}$  axes. The track-subjet closest to the soft-muon is called the *muon- $\text{Ex}k_t^{(2)}$*  track-subjet while the other is called the *non-muon- $\text{Ex}k_t^{(2)}$*  track-subjet.

Jets in the  $Z$ -region which fail the  $40 - 60\%$  tagging interval condition are further categorized in channels depending on the  $\langle S_{d0} \rangle$  of the muon- $\text{Ex}k_t^{(2)}$  and non-muon- $\text{Ex}k_t^{(2)}$  track-subjets. The observable  $\langle S_{d0} \rangle$  is defined as:

$$\langle S_{d0} \rangle = \frac{1}{3}(S_{d0}^1 + S_{d0}^2 + S_{d0}^3),$$

where  $S_{d0}^{1,2,3}$  are the signed transverse impact parameter significances of the three tracks ghost-associated to the track-subjet with highest value of  $S_{d0}$ .

The non-muon- $\text{Ex}k_t^{(2)}$  is binned in two regions  $\langle S_{d0} \rangle < 0$  and  $\langle S_{d0} \rangle \geq 0$  to create channels enriched in light- and  $b$ -labeled jet components. The number of  $B$ -labeled jets that fail the  $40 - 60\%$  tagging interval is small when compared to these two other components. The muon- $\text{Ex}k_t^{(2)}$  is binned in 6 different bins. A total of  $3 (p_T \text{ bins}) \times 12 (\langle S_{d0} \rangle \text{ bins})$  channels are used in the  $60 - 100\%$  tagging interval.

For jets in the  $0 - 40\%$  and  $40 - 60\%$  tagging intervals, an additional categorization is performed depending on the largest secondary-vertex mass  $m_{\text{SV}}^{\text{max}}$  within the non-muon- $\text{Ex}k_t^{(2)}$  track-subjet  $\Delta R(\text{SV}, \text{Ex}k_t^{(2)}) < 0.3$ . Two channels are defined by  $m_{\text{SV}}^{\text{max}} < 2 \text{ GeV}$  and  $m_{\text{SV}}^{\text{max}} \geq 2 \text{ GeV}$  which are enriched in  $b$ - and  $B$ -labeled jets, respectively. Jets without a reconstructed SV are assigned  $m_{\text{SV}}^{\text{max}} = 0 \text{ GeV}$ . An identical binning is made for the muon- $\text{Ex}k_t^{(2)}$  track-subjet. A total of  $3 (p_T \text{ bins}) \times 2 (\text{tagging interval}) \times 4 (m_{\text{SV}}^{\text{max}} \text{ bins})$  channels are used in the  $40 - 60\%$  and  $0 - 40\%$  tagging intervals. The total number of channels in the  $Z$ -region is 60. Figure 7 shows a diagram of the classification described.

## 9 Methodology

A binned Poisson likelihood is built based on the number of jets observed in the channels and regions described in Section 8. Each event from the top-region contributes with two jets while each event from the  $Z$ -region contributes with a single jet. The expected value for the Poisson probability distributions is built from all the simulated samples described in Section 4.

A series of multipliers are introduced:

**Flavor fraction ( $\mu$ )** Different multipliers are introduced for  $b$ - and light-flavored jets, separately for the top-region and for the  $Z$ -region, and for each  $p_T$ -bin. These multipliers provide a data-driven normalization of the corresponding flavor components in the phase-space used for the efficiency measurement.

**Efficiency scale factors (SF)** Scale factors represent the ratio between the efficiency in data to the one in simulation. Different multipliers are introduced for  $b$ - and  $B$ -flavored jets, separately for each  $p_T$  bin and for each tagging interval. The SF multipliers are defined so that the total number of predicted events is unchanged by efficiency scale factors.

Therefore, the expected value in each region  $r$  ( $r = \text{top or } Z$ ) and channel  $c$  can be written as:

$$\begin{aligned} \lambda_{rc} = & \mu_r^b \text{SF}_{0-40}^b n_{rc,0-40}^b + \mu_r^b \text{SF}_{40-60}^b n_{rc,40-60}^b + \mu_r^b \frac{(1 - \text{SF}_{0-40}^b \varepsilon_{0-40, \text{MC}}^b - \text{SF}_{40-60}^b \varepsilon_{40-60, \text{MC}}^b)}{(1 - \varepsilon_{0-40, \text{MC}}^b - \varepsilon_{40-60, \text{MC}}^b)} n_{rc,60-100}^b \\ & + \text{SF}_{0-40}^B n_{rc,0-40}^B + \text{SF}_{40-60}^B n_{rc,40-60}^B + \frac{(1 - \text{SF}_{0-40}^B \varepsilon_{0-40, \text{MC}}^B - \text{SF}_{40-60}^B \varepsilon_{40-60, \text{MC}}^B)}{(1 - \varepsilon_{0-40, \text{MC}}^B - \varepsilon_{40-60, \text{MC}}^B)} n_{rc,60-100}^B \\ & + \mu_r^{\text{light}} n_{rc}^{\text{light}} + \sum_{i \notin \{b, B, \text{light}\}} n_{rc}^i. \end{aligned} \quad (2)$$

All the parameters in Eq. 2 depend on the jet  $p_T$  bin, even though no additional index was introduced to simplify the notation. The MC yields  $n$  are corrected with dedicated MC-MC scale factors when simulated samples are generated with different programs. The tagging efficiency in simulation depends on the detailed description of the hadronization of heavy-flavored partons and of the resonant decays of hadrons inside jets [74]. The results in this note use the efficiency observed in PYTHIA 8 as reference as the scale factors.

Systematic uncertainties are introduced with constrained nuisance parameters. The nuisance parameters representing experimental and modeling uncertainties are accompanied by Gaussian constraint terms while those representing statistical variations arising from the limited number of simulated events are accompanied by Poisson constraint terms. The values of all nuisance parameters and flavor fractions  $\mu$  are determined by building a profile likelihood. The measured value and covariance for the several scale factors SF are determined by their maximum likelihood estimators.

## 10 Systematic Uncertainties

Several sources of uncertainties are considered in the efficiency measurement. They can be loosely divided into four categories: experimental uncertainties from object reconstruction and identification, uncertainties from simulation modeling, statistical uncertainties from Monte Carlo simulation, and extrapolation uncertainties due to the use of soft-muons for the selection of jets used in the efficiency measurement.

### 10.1 Experimental Uncertainties

For each object, a set of reconstruction uncertainties are included. Uncertainties associated with leptons arise from the trigger, reconstruction, identification, and isolation efficiencies [28, 75, 76], as well as the lepton momentum scale and resolution [27, 28]. Uncertainties related to jet reconstruction stem from the efficiency of pileup rejection by the jet vertex tagger [77], the jet energy scale [78] and resolution [79].

Since DEXTER uses detailed information from tracks in the feature-extracting NNs and to define the channels in the  $Z$ -region, tracking reconstruction uncertainties are included. Uncertainties related to the track selection efficiency and the number of fake tracks are considered. Finally, dedicated systematic uncertainties are considered for the tracking parameters including the transverse and longitudinal impact parameters, as well as the track sagitta. The uncertainty on the integrated luminosity for the full Run-2 dataset is 1.7% [71], as obtained using the LUCID-2 detector [80]. A variation in the pileup modeling based on different estimates of the total inelastic cross section is included as an uncertainty.

### 10.2 Modeling Uncertainties

Variations in the renormalization and factorization scales are used to estimate the uncertainty due to missing higher-order corrections. The combined PDF and  $\alpha_s$  uncertainties follow the PDF4LHC prescription [81]. Additional modeling systematic uncertainties are applied to specific samples and are described below.

**$Z + \text{jets}$  Modeling Uncertainties** Several additional modeling uncertainties on the  $Z + \text{jets}$  MC prediction are considered, related to the description of the  $p_T$  distribution, the modeling of associated heavy-flavor production and the choice of generator. The modeling of the  $Z$  boson  $p_T$  distribution is improved by reweighting the MC prediction to data using an inclusive selection. The full difference between the original MC prediction and the data is used as a systematic uncertainty. The uncertainty in the heavy-flavor components is evaluated by comparing the nominal  $Z + \text{jets}$  sample with alternatives with varied settings for the overlap between ME and parton shower emissions, and for the resummation scale [82].

A generator uncertainty is also considered by comparing the nominal SHERPA 2.2.1 sample to the MADGRAPH5\_AMC@NLO sample, as described in Section 4. The difference in the predictions is taken as a systematic uncertainty on the modeling of  $Z + \text{jets}$ .

**$t\bar{t}$  Modeling Uncertainties** For shower uncertainties, the settings of the nominal PowhegBox+Pythia8  $t\bar{t}$  sample are varied, resulting in different event weights; the uncertainty due to initial shower radiation is estimated by simultaneously changing renormalization and factorization scale in the ME and in the parton shower, while the uncertainty due to final state shower radiation is estimated by changing the associated scale in the simulation. For the uncertainties due to hadronization and NLO matching, the nominal PowhegBox+Pythia8  $t\bar{t}$  sample is compared to the PowhegBox+Herwig7  $t\bar{t}$  sample mentioned in Section 4.

**Single-top Modeling Uncertainties** Systematic uncertainties for the single top MC modeling follow the prescription in  $t\bar{t}$  modeling uncertainties. On top of these uncertainties, the interference between the single top and  $t\bar{t}$  production is considered by comparing the diagram removal (DR) and diagram subtraction (DS) schemes.

**Extrapolation Systematic Uncertainties** Systematic variations associated to the use of a soft-muon to select jets in the Z-region are estimated by comparing, for each tagging interval, the efficiency using SHERPA 2.2.1 and MADGRAPH5\_AMC@NLO. This additional uncertainty is added to account for any mismodeling in the efficiency of selecting jets with soft-muons and in the correlation of this efficiency with the DEXTER efficiency. The relative difference between the efficiency with and without soft-muon tagging in the two generators is used as an additional source of systematic uncertainty. This relative difference varies between 1 – 12% depending on the tagging interval and  $p_T$  range.

## 11 Results

The measured DEXTER efficiencies for  $B$ - and  $b$ -labeled jets in the two tagging intervals are shown in Figure 8 and Figure 9. The values are also summarized in Table 2. The statistical and total uncertainty are shown separately. The SFs for  $B$ -labeled jets in the 0 – 40% tagging interval are consistent with unity and uncertainties less than 20% showing the robustness of the choices made during the algorithm development. At low  $p_T$ , the uncertainty in this measurement is dominated by modeling uncertainties.

Figure 10 shows a comparison of the simulated and observed data in the top-region. Figure 11 shows a summary comparison in the Z-region. Figure 12 shows and example of  $\langle S_{d_0} \rangle$  and  $m_{SV}^{\max}$  distributions for two different bins of Figure 11. Tables 3 and 4 show a breakdown of the systematic uncertainties in grouped categories for  $B$ - and  $b$ -labeled jets, respectively.

Table 2: Ratio of the efficiency measured in data to MC for different jet labels,  $p_T$  ranges, and tagging intervals.

Jet label	Jet $p_T$ range (GeV)	Tagging Interval	Scale Factor
$B$ -label	[20, 90]	0 – 40%	$1.15 \pm 0.05$ (stat.) $\pm 0.26$ (syst.)
$B$ -label	[90, 140]	0 – 40%	$1.00 \pm 0.10$ (stat.) $\pm 0.14$ (syst.)
$B$ -label	[140, 200]	0 – 40%	$1.12 \pm 0.16$ (stat.) $\pm 0.16$ (syst.)
$B$ -label	[20, 90]	40 – 60%	$1.77 \pm 0.12$ (stat.) $\pm 0.46$ (syst.)
$B$ -label	[90, 140]	40 – 60%	$1.13 \pm 0.24$ (stat.) $\pm 0.31$ (syst.)
$B$ -label	[140, 200]	40 – 60%	$1.78 \pm 0.34$ (stat.) $\pm 0.34$ (syst.)
$b$ -label	[20, 90]	0 – 40%	$1.33 \pm 0.04$ (stat.) $\pm 0.18$ (syst.)
$b$ -label	[90, 140]	0 – 40%	$1.91 \pm 0.12$ (stat.) $\pm 0.13$ (syst.)
$b$ -label	[140, 200]	0 – 40%	$1.88 \pm 0.11$ (stat.) $\pm 0.08$ (syst.)
$b$ -label	[20, 90]	40 – 60%	$1.17 \pm 0.03$ (stat.) $\pm 0.13$ (syst.)
$b$ -label	[90, 140]	40 – 60%	$1.53 \pm 0.09$ (stat.) $\pm 0.13$ (syst.)
$b$ -label	[140, 200]	40 – 60%	$1.43 \pm 0.07$ (stat.) $\pm 0.09$ (syst.)

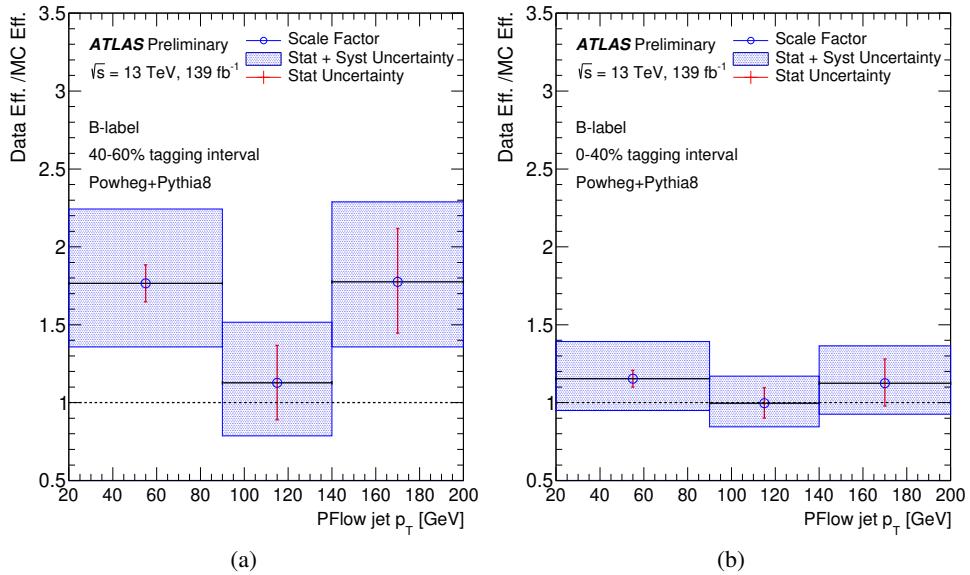


Figure 8: Data-to-MC scale factors obtained for  $B$ -labeled jets in the (a)  $40 - 60\%$  and (b)  $0 - 40\%$  tagging intervals. The blue error band includes systematic and statistical uncertainties. The red error bar represents the statistical uncertainty only.

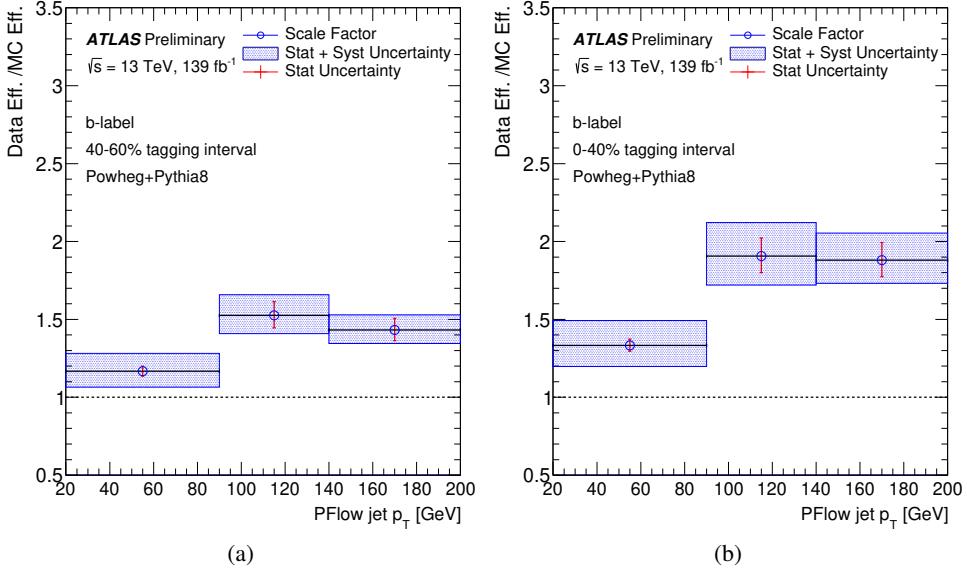


Figure 9: Data-to-MC scale factors obtained for  $b$ -labeled jets in the (a) 40 – 60% and (b) 0 – 40% tagging intervals. The blue error band includes systematic and statistical uncertainties. The red error bar represents the statistical uncertainty only.

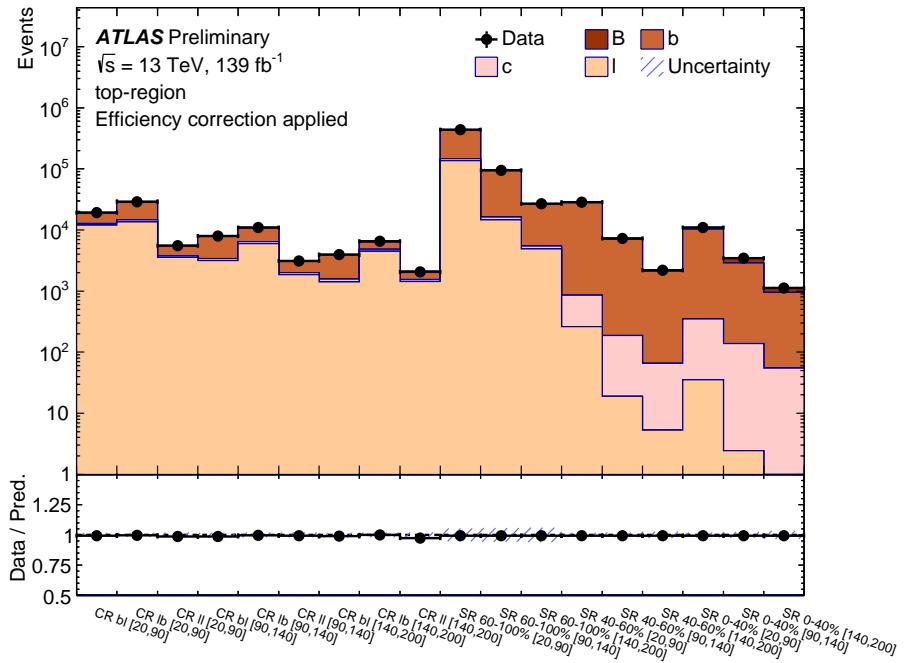


Figure 10: Comparison between data and prediction for the event yields in the top-region channels after the efficiency correction. The hatched band represents the systematic uncertainty in each channel.

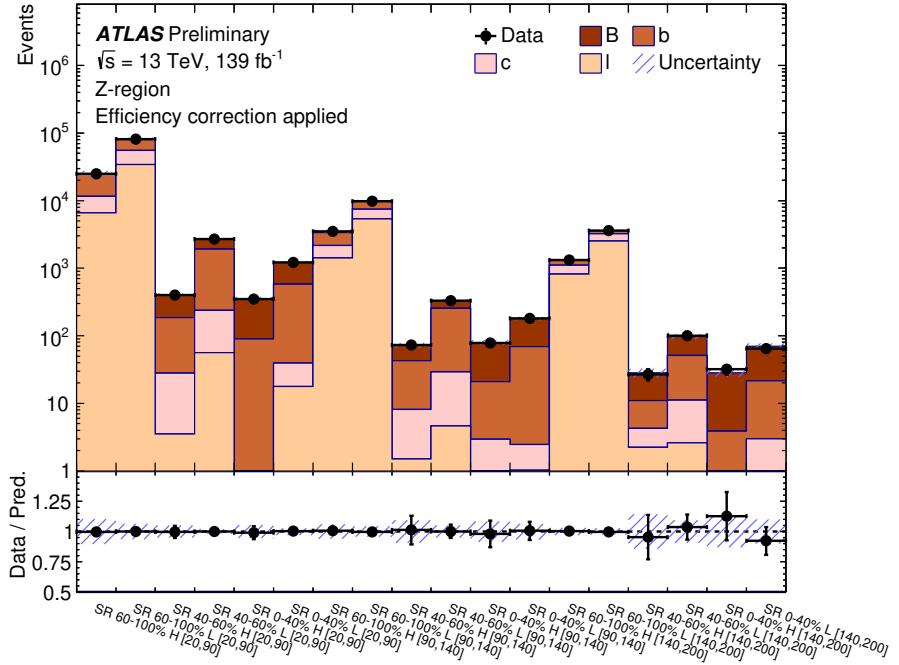


Figure 11: Comparison between data and prediction for the event yields in the Z-region channels after the efficiency correction. The labels  $H$  and  $L$  indicate the two different channels depending on the value of the non-muon- $Exk_t^{(2)}$  track-subjet  $m_{SV}^{\max}$  ( $H$ :  $m_{SV}^{\max} \geq 2$  GeV,  $L$ :  $m_{SV}^{\max} < 2$  GeV) or  $\langle S_{d0} \rangle$  ( $H$ :  $\langle S_{d0} \rangle \geq 0$ ,  $L$ :  $\langle S_{d0} \rangle < 0$ ). The hatched band represents the systematic uncertainty in each channel.

Table 3: Breakdown of the SF uncertainties for each summary category in different tagging intervals and  $p_T$  ranges for  $B$ -labeled jets. The statistical uncertainties are also included for comparison. The total systematic uncertainty is not equal to the sum in quadrature of the individual components due to correlations.

Source of uncertainties	0 – 40% tagging interval			40 – 60% tagging interval		
	$p_T^{\text{jet}}$ range (GeV)			$p_T^{\text{jet}}$ range (GeV)		
	[20, 90]	[90, 140]	[140, 200]	[20, 90]	[90, 140]	[140, 200]
Total uncertainty	0.26	0.17	0.23	0.48	0.39	0.48
Statistical	0.05	0.10	0.16	0.12	0.24	0.34
Systematic	0.26	0.14	0.16	0.46	0.31	0.34
Extrapolation	0.21	0.11	0.13	0.38	0.25	0.23
MC statistics	0.21	0.14	0.10	0.36	0.20	0.25
Jets	0.14	0.05	0.04	0.16	0.09	0.11
Leptons	0.02	0.01	0.01	0.05	0.02	0.01
Luminosity	0.01	0.02	0.02	0.06	0.03	0.02
Flavor fraction	0.19	0.10	0.12	0.30	0.20	0.26
Pileup	0.04	0.01	0.01	0.05	0.01	0.04
Tracking	0.16	0.04	0.05	0.19	0.11	0.09
Single top modeling	0.01	0.02	0.01	0.12	0.05	0.03
$t\bar{t}$ modeling	0.12	0.02	0.02	0.19	0.11	0.06
$Z + \text{jets}$ modeling	0.22	0.12	0.14	0.39	0.22	0.25

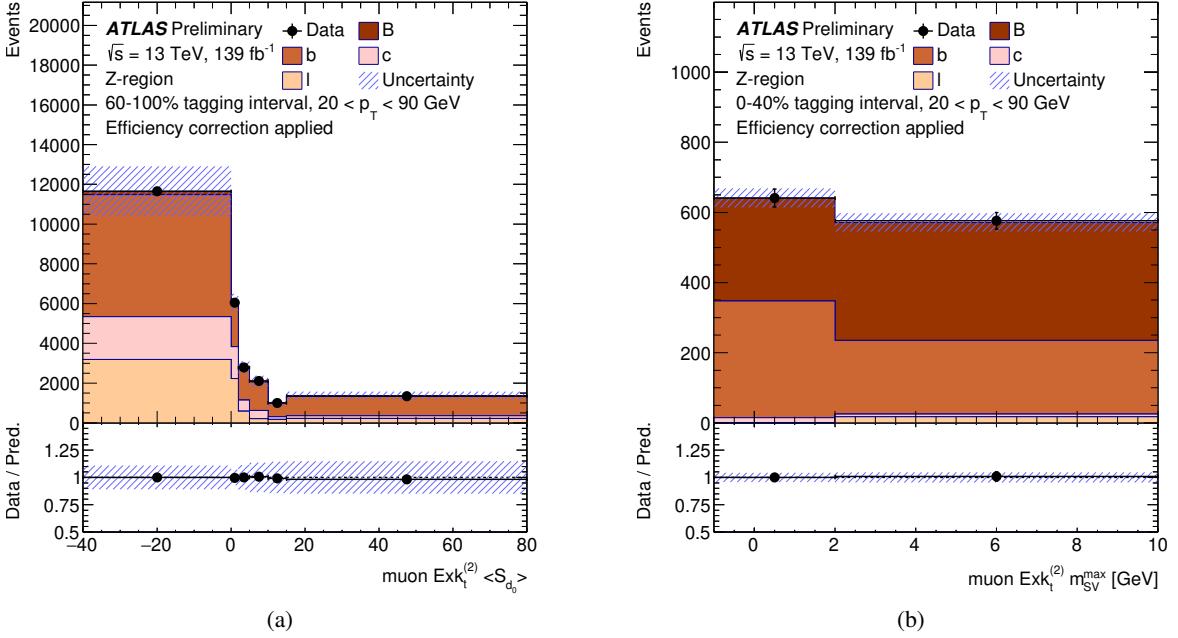


Figure 12: Comparison between data and prediction in the  $Z$ -region, after efficiency correction, for (a) the muon  $\text{Exk}_t^{(2)} \langle S_{d_0} \rangle$  distribution in the  $60 - 100\%$  tagging interval when requiring the non-muon  $\text{Exk}_t^{(2)} \langle S_{d_0} \rangle > 0$  and (b) the muon  $\text{Exk}_t^{(2)} m_{\text{SV}}^{\text{max}}$  distribution in the  $0 - 40\%$  tagging interval when requiring the non-muon  $\text{Exk}_t^{(2)} m_{\text{SV}}^{\text{max}} < 2 \text{ GeV}$ .

Table 4: Breakdown of the SF uncertainties for each summary category in different tagging interval and  $p_T$  ranges for  $b$ -labeled jets. The statistical uncertainties are also included for comparison. The total systematic uncertainty is not equal to the sum in quadrature of the individual components due to correlations.

Source of uncertainties	0 - 40% tagging interval			40 - 60% tagging interval		
	$p_T^{\text{jet}}$ range (GeV)			$p_T^{\text{jet}}$ range (GeV)		
	[20, 90]	[90, 140]	[140, 200]	[20, 90]	[90, 140]	[140, 200]
Total uncertainty	0.18	0.18	0.14	0.13	0.16	0.11
Statistical	0.04	0.12	0.11	0.03	0.09	0.07
Systematic	0.18	0.13	0.09	0.13	0.13	0.09
Extrapolation	0.18	0.12	0.07	0.13	0.12	0.07
MC statistics	0.13	0.04	0.02	0.10	0.04	0.02
Jets	0.07	0.04	0.03	0.06	0.04	0.03
Leptons	0.04	0.01	0.00	0.03	0.01	0.00
Luminosity	0.04	0.01	0.01	0.03	0.01	0.01
Flavor fraction	0.17	0.11	0.06	0.12	0.11	0.06
Pileup	0.05	0.02	0.00	0.04	0.02	0.00
Tracking	0.12	0.04	0.02	0.09	0.04	0.02
Single top modeling	0.14	0.03	0.04	0.09	0.03	0.04
$t\bar{t}$ modeling	0.17	0.07	0.03	0.13	0.07	0.03
$Z + \text{jets}$ modeling	0.18	0.12	0.07	0.13	0.12	0.07

## 12 Conclusion

This note presents the development of DEXTER, an algorithm that performs flavor tagging of low- $p_T$  jets arising from the hadronization of one or two heavy-flavor hadrons. The algorithm uses a neural network with two permutation-invariant deep set feature extractors, one for displaced tracks and one for secondary vertices. Displaced vertices are selected from a large- $R$  track jet reconstructed around PFlow jets. Secondary vertices are reconstructed with the TC-LVT clustering algorithm and the MSVF vertex finding algorithm.

The performance is studied in detail, ensuring a reduced dependency of the tagger performance with the parent particle color charge and mass. A measurement of the efficiency for two tagging intervals and three  $p_T^{\text{jet}}$  ranges using  $t\bar{t}$  and  $Z + \text{jets}$  events is reported for  $B$ - and  $b$ -labeled jets. These measurements make it possible to apply this algorithm in physics analyses, enabling and improving several new BSM searches and SM measurements.

# Appendix

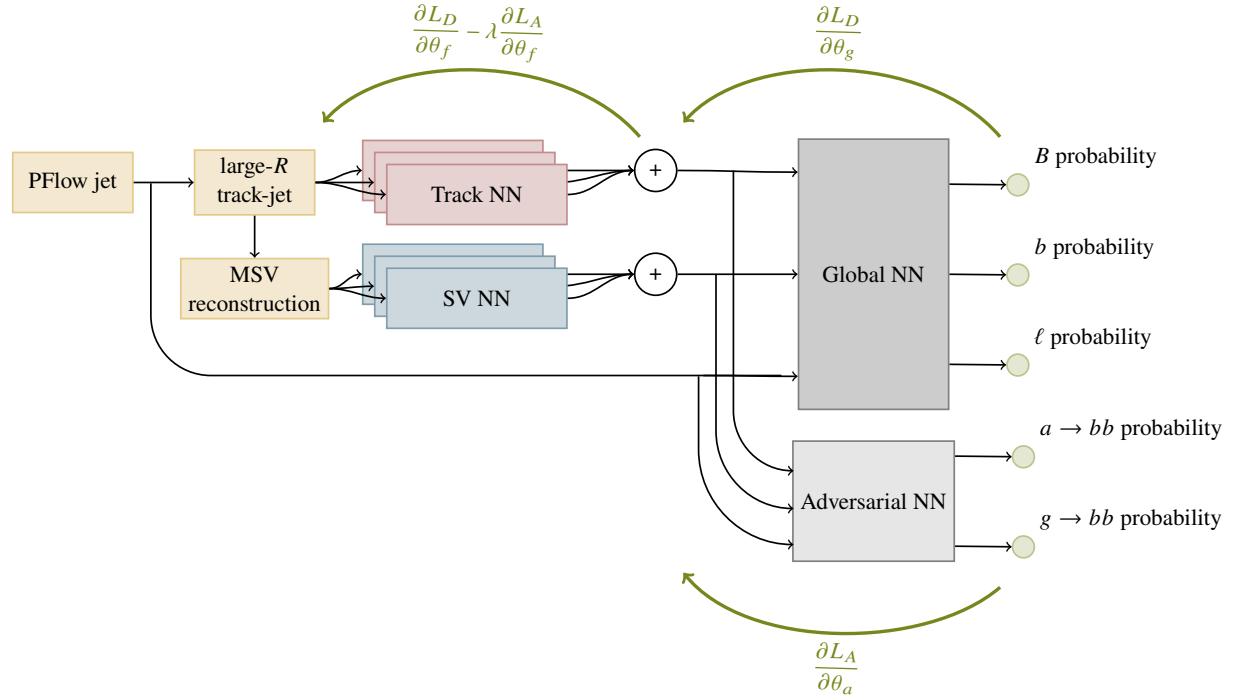


Figure 13: Architecture of the DEXTER algorithm with an adversarial NN. A DANN strategy is used to train the feature-extraction NNs. The adversarial NN discriminates  $a \rightarrow b\bar{b}$  and  $g \rightarrow b\bar{b}$  events included in the  $B$ -labeled jet ensemble mixture.

## References

- [1] R. Kogler et al., *Jet Substructure at the Large Hadron Collider: Experimental Review*, *Rev. Mod. Phys.* **91** (2019) 045003, arXiv: [1803.06991 \[hep-ex\]](https://arxiv.org/abs/1803.06991) (cit. on p. 2).
- [2] 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 \[hep-ex\]](https://arxiv.org/abs/2009.04986) (cit. on p. 2).
- [3] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*, *Eur. Phys. J. C* **79** (2019) 135, arXiv: [1807.09477 \[hep-ex\]](https://arxiv.org/abs/1807.09477) (cit. on p. 2).
- [4] 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).
- [5] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) 466, arXiv: [1703.10485 \[hep-ex\]](https://arxiv.org/abs/1703.10485) (cit. on p. 2).
- [6] M. Zaheer et al., *Deep Sets*, 2017, arXiv: [1703.06114 \[cs.LG\]](https://arxiv.org/abs/1703.06114) (cit. on p. 2).
- [7] D. Curtin et al., *Exotic decays of the 125 GeV Higgs boson*, *Phys. Rev. D* **90** (2014) 075004, arXiv: [1312.4992 \[hep-ph\]](https://arxiv.org/abs/1312.4992) (cit. on p. 2).
- [8] M. Cepeda, S. Gori, V. Martinez Outschoorn and J. Shelton, *Exotic Higgs Decays*, *Annu. Rev. Nucl. Part. Sci.* **72** (2022), arXiv: [2111.12751 \[hep-ph\]](https://arxiv.org/abs/2111.12751) (cit. on p. 2).
- [9] M. Casolino, T. Farooque, A. Juste, T. Liu and M. Spannowsky, *Probing a light CP-odd scalar in di-top-associated production at the LHC*, *Eur. Phys. J. C* **75** (2015) 498, arXiv: [1507.07004 \[hep-ph\]](https://arxiv.org/abs/1507.07004) (cit. on p. 2).
- [10] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003 (cit. on p. 3).
- [11] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187> (cit. on p. 3).
- [12] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, *JINST* **9** (2014) P09009, arXiv: [1406.7690 \[hep-ex\]](https://arxiv.org/abs/1406.7690) (cit. on p. 3).
- [13] 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 \[hep-ex\]](https://arxiv.org/abs/1704.07983) (cit. on p. 3).
- [14] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) 332, arXiv: [1611.10235 \[hep-ex\]](https://arxiv.org/abs/1611.10235) (cit. on p. 3).
- [15] M. Cacciari, G. P. Salam and G. Soyez, *The anti- $k_t$  jet clustering algorithm*, *JHEP* **04** (2008) 063, arXiv: [0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189) (cit. on p. 4).
- [16] ATLAS Collaboration, *Jet global sequential corrections with the ATLAS detector in proton–proton collisions at  $\sqrt{s} = 8$  TeV*, ATLAS-CONF-2015-002, 2015, URL: <https://cds.cern.ch/record/2001682> (cit. on p. 4).

[17] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector*, *Eur. Phys. J. C* **81** (2020) 689, arXiv: [2007.02645 \[hep-ex\]](#) (cit. on p. 4).

[18] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett. B* **659** (2008) 119, arXiv: [0707.1378 \[hep-ph\]](#) (cit. on p. 4).

[19] B. Nachman, P. Nef, A. Schwartzman, M. Swiatlowski and C. Wanotayaroj, *Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC*, *JHEP* **02** (2015) 075, arXiv: [1407.2922 \[hep-ph\]](#) (cit. on p. 4).

[20] ATLAS Collaboration, *Search for Higgs boson decays into two new low-mass spin-0 particles in the 4b channel with the ATLAS detector using pp collisions at  $\sqrt{s} = 13\text{ TeV}$* , *Phys. Rev. D* **102** (2020) 112006, arXiv: [2005.12236 \[hep-ex\]](#) (cit. on pp. 4, 12).

[21] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-PUB-2018-025, 2018, URL: <https://cds.cern.ch/record/2645405> (cit. on p. 4).

[22] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, URL: <https://cds.cern.ch/record/2270366> (cit. on pp. 4, 8).

[23] ATLAS Collaboration, *Soft b-hadron tagging for compressed SUSY scenarios*, ATLAS-CONF-2019-027, 2019, URL: <https://cds.cern.ch/record/2682131> (cit. on p. 4).

[24] V. Kostyukhin, *Secondary vertex based b-tagging*, tech. rep. ATL-PHYS-2003-033, revised version number 1 submitted on 2003-09-22 11:21:01: CERN, 2003, URL: <https://cds.cern.ch/record/685550> (cit. on p. 4).

[25] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **03** (2011) 015, arXiv: [1011.2268 \[hep-ph\]](#) (cit. on p. 5).

[26] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160, arXiv: [hep-ph/9305266](#) (cit. on p. 5).

[27] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, *JINST* **14** (2019) P12006, arXiv: [1908.00005 \[hep-ex\]](#) (cit. on pp. 5, 18).

[28] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13\text{ TeV}$* , *Eur. Phys. J. C* **76** (2016) 292, arXiv: [1603.05598 \[hep-ex\]](#) (cit. on pp. 5, 18).

[29] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, *Phys. Rev. D* **97** (2018) 072003, arXiv: [1712.08891 \[hep-ex\]](#) (cit. on p. 5).

[30] M. Cacciari, G. P. Salam and G. Soyez, *The catchment area of jets*, *JHEP* **2008** (2008) 005, ISSN: 1029-8479, URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/005> (cit. on p. 6).

[31] G. Luisoni, P. Nason, C. Oleari and F. Tramontano,  *$HW^\pm/HZ + 0$  and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, *JHEP* **10** (2013) 083, arXiv: [1306.2542 \[hep-ph\]](#) (cit. on p. 6).

[32] R. D. Ball et al., *Parton distributions for the LHC run II*, *JHEP* **04** (2015) 040, arXiv: [1410.8849 \[hep-ph\]](#) (cit. on pp. 6, 7, 14).

[33] 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 pp. 6, 7, 14).

[34] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, arXiv: [1405.0301 \[hep-ph\]](#) (cit. on pp. 6, 7, 14).

[35] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244, arXiv: [1207.1303 \[hep-ph\]](#) (cit. on pp. 6, 7, 14).

[36] E. Bothmann et al., *Event generation with Sherpa 2.2*, *SciPost Phys.* **7** (2019) 034, arXiv: [1905.09127 \[hep-ph\]](#) (cit. on p. 6).

[37] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039, arXiv: [0808.3674 \[hep-ph\]](#) (cit. on p. 6).

[38] F. Buccioni et al., *OpenLoops 2*, *Eur. Phys. J. C* **79** (2019) 866, arXiv: [1907.13071 \[hep-ph\]](#) (cit. on p. 6).

[39] F. Caccioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, *Phys. Rev. Lett.* **108** (2012) 111601, arXiv: [1111.5206 \[hep-ph\]](#) (cit. on p. 6).

[40] A. Denner, S. Dittmaier and L. Hofer, *COLLIER: A fortran-based complex one-loop library in extended regularizations*, *Comput. Phys. Commun.* **212** (2017) 220, arXiv: [1604.06792 \[hep-ph\]](#) (cit. on p. 6).

[41] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, *JHEP* **03** (2008) 038, arXiv: [0709.1027 \[hep-ph\]](#) (cit. on p. 6).

[42] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049, arXiv: [1111.1220 \[hep-ph\]](#) (cit. on p. 6).

[43] S. Höche, F. Krauss, M. Schönherr and F. Siegert, *QCD matrix elements + parton showers. The NLO case*, *JHEP* **04** (2013) 027, arXiv: [1207.5030 \[hep-ph\]](#) (cit. on p. 6).

[44] S. Catani, F. Krauss, B. R. Webber and R. Kuhn, *QCD Matrix Elements + Parton Showers*, *JHEP* **11** (2001) 063, arXiv: [hep-ph/0109231](#) (cit. on p. 6).

[45] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **05** (2009) 053, arXiv: [0903.1219 \[hep-ph\]](#) (cit. on p. 6).

[46] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126, arXiv: [0707.3088 \[hep-ph\]](#) (cit. on p. 6).

[47] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: <https://cds.cern.ch/record/2216168> (cit. on pp. 6, 7, 14).

[48] L. Lönnblad, *Correcting the Colour-Dipole Cascade Model with Fixed Order Matrix Elements*, *JHEP* **05** (2002) 046, arXiv: [hep-ph/0112284](#) (cit. on pp. 7, 14).

[49] L. Lönnblad and S. Prestel, *Matching tree-level matrix elements with interleaved showers*, *JHEP* **03** (2012) 019, arXiv: [1109.4829 \[hep-ph\]](#) (cit. on pp. 7, 14).

[50] 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 pp. 7, 14).

[51] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) 152 (cit. on pp. 7, 14).

[52] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, *JHEP* **09** (2012) 130, arXiv: [1207.5391 \[hep-ph\]](#) (cit. on pp. 7, 14).

[53] S. Alioli, P. Nason, C. Oleari and E. Re, *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, *JHEP* **09** (2009) 111, arXiv: [0907.4076 \[hep-ph\]](#) (cit. on pp. 7, 14), Erratum: *JHEP* **02** (2010) 011.

[54] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547, arXiv: [1009.2450 \[hep-ph\]](#) (cit. on pp. 7, 14).

[55] S. Frixione, E. Laenen, P. Motylinski, C. White and B. R. Webber, *Single-top hadroproduction in association with a W boson*, *JHEP* **07** (2008) 029, arXiv: [0805.3067 \[hep-ph\]](#) (cit. on pp. 7, 14).

[56] 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. 7).

[57] 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. 7).

[58] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121, arXiv: [1810.05165 \[hep-ph\]](#) (cit. on p. 7).

[59] ATLAS Collaboration, *Digluon Tagging using  $\sqrt{s} = 13$  TeV  $pp$  Collisions in the ATLAS Detector*, ATL-PHYS-PUB-2021-027, 2021, URL: <https://cds.cern.ch/record/2776780> (cit. on p. 7).

[60] J. S. Bridle, *Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition*, (1990), ed. by F. F. Soulié and J. Héault (cit. on p. 7).

[61] D.-A. Clevert, T. Unterthiner and S. Hochreiter, *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*, (2016), ed. by Y. Bengio and Y. LeCun, arXiv: [1511.07289 \[cs.LG\]](#) (cit. on p. 7).

[62] G. Aad et al., *Search for Higgs boson decays into two new low-mass spin-0 particles in the 4b channel with the ATLAS detector using  $pp$  collisions at  $\sqrt{s} = 13$  TeV*, (2020), arXiv: [2005.12236 \[hep-ex\]](#) (cit. on p. 8).

[63] ATLAS Collaboration, *Measurement of colour flow using jet-pull observables in  $t\bar{t}$  events with the ATLAS experiment at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **78** (2018) 847, arXiv: [1805.02935 \[hep-ex\]](#) (cit. on p. 10).

[64] Y. Ganin et al., *Domain-Adversarial Training of Neural Networks*, *JMLR* **17** (2016), arXiv: [1505.07818 \[stat.ML\]](#) (cit. on p. 10).

[65] F. Chollet et al., *Keras*, <https://keras.io>, 2015 (cit. on p. 10).

[66] Martín Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, URL: <https://www.tensorflow.org/> (cit. on p. 10).

[67] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 3rd International Conference for Learning Representations (2014), arXiv: [1412.6980 \[cs.LG\]](https://arxiv.org/abs/1412.6980) (cit. on p. 10).

[68] ATLAS Collaboration, *Identification of boosted Higgs bosons decaying into b-quark pairs with the ATLAS detector at 13 TeV*, *Eur. Phys. J. C* **79** (2019) 836, arXiv: [1906.11005 \[hep-ex\]](https://arxiv.org/abs/1906.11005) (cit. on p. 12).

[69] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with  $t\bar{t}$  events in pp collisions at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79** (2019) 970, arXiv: [1907.05120 \[hep-ex\]](https://arxiv.org/abs/1907.05120) (cit. on pp. 13, 14).

[70] 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 p. 13).

[71] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: <https://cds.cern.ch/record/2677054> (cit. on pp. 13, 14, 18).

[72] ATLAS Collaboration, *Performance of electron and photon triggers in ATLAS during LHC Run 2*, *Eur. Phys. J. C* **80** (2020) 47, arXiv: [1909.00761 \[hep-ex\]](https://arxiv.org/abs/1909.00761) (cit. on pp. 13, 14).

[73] ATLAS Collaboration, *Performance of the ATLAS muon triggers in Run 2*, *JINST* **15** (2020) P09015, arXiv: [2004.13447 \[hep-ex\]](https://arxiv.org/abs/2004.13447) (cit. on pp. 13, 14).

[74] ATLAS Collaboration, *Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration*, ATL-PHYS-PUB-2020-009, 2020, URL: <https://cds.cern.ch/record/2718610> (cit. on p. 17).

[75] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **79** (2019) 639, arXiv: [1902.04655 \[hep-ex\]](https://arxiv.org/abs/1902.04655) (cit. on p. 18).

[76] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at  $\sqrt{s} = 13$  TeV*, *Eur. Phys. J. C* **81** (2021) 578, arXiv: [2012.00578 \[hep-ex\]](https://arxiv.org/abs/2012.00578) (cit. on p. 18).

[77] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581, arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823) (cit. on p. 18).

[78] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, *Phys. Rev. D* **96** (2017) 072002, arXiv: [1703.09665 \[hep-ex\]](https://arxiv.org/abs/1703.09665) (cit. on p. 18).

[79] ATLAS Collaboration, *Determination of jet calibration and energy resolution in proton–proton collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, *Eur. Phys. J. C* **80** (2020) 1104, arXiv: [1910.04482 \[hep-ex\]](https://arxiv.org/abs/1910.04482) (cit. on p. 18).

- [80] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, **JINST** **13** (2018) P07017 (cit. on p. 18).
- [81] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, **J. Phys. G** **43** (2016) 023001, arXiv: [1510.03865 \[hep-ph\]](https://arxiv.org/abs/1510.03865) (cit. on p. 18).
- [82] F. Caccioli et al., *Precise Higgs-background predictions: merging NLO QCD and squared quark-loop corrections to four-lepton + 0,1 jet production*, **JHEP** **01** (2014) 046, arXiv: [1309.0500 \[hep-ph\]](https://arxiv.org/abs/1309.0500) (cit. on p. 18).